Biologie structurale de complexes protéine:ARN

Thibaut Crépin

UNIVERSITÉ Grenoble Alpes

2017

Mémoire de

Thibaut Crépin

Pour obtenir l'

HABILITATION A DIRIGER DES RECHERCHES

de l'

ECOLE DOCTORALE DE CHIMIE ET SCIENCE DU VIVANT

COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Biologie structurale de complexes protéine:ARN

soutenu publiquement le 07 Juin 2017,
devant le jury composé de :

Dr. Etienne Decroly (Marseille)	*Rapporteur*
Dr. Roland Marquet (Strasbourg)	*Rapporteur*
Prof. Dino Moras (Strasbourg)	*Rapporteur*
Dr. Emmanuelle Schmitt (Palaiseau)	*Examinateur*
Prof. Winfried Weissenhorn (Grenoble)	*Examinateur*

ibs Institut de Biologie Structurale

Institut de Biologie Structurale - UMR 5075

Résumé

Ce manuscrit résume mes années passées dans la recherche, depuis ma thèse au laboratoire de biochimie de l'École Polytechnique jusqu'au poste de chargé de recherche CNRS que j'occupe maintenant à l'Institut de Biologie Structurale. Durant ces années, mes travaux ont visé à détailler par différentes techniques de la biologie structurale, des constituants du vivant ayant vocation à interagir entre eux en vue de l'accomplissement d'un processus biologique. Une partie de ces travaux a consisté à étudier les mécanismes de la spécificité du fonctionnement des aminoacyl-ARNt synthétases, et ainsi apporter des réponses quant aux bases moléculaires de la fidélité du processus traductionnel. Mon recrutement par le CNRS s'est accompagné d'une mobilité thématique, me faisant découvrir un autre monde, celui des virus et de leurs complexes réplicatifs. J'ai notamment pris part à la grande aventure visant à détailler les constituants du complexe réplicatif du virus de la grippe, aventure qui a abouti à la résolution de la structure du complexe polymérasique grippal. Loin de considérer cette dernière comme une finalité, j'ambitionne maintenant de comprendre son assemblage, d'étendre ces travaux à d'autres des complexes réplicatifs mais tout en intégrant la dynamique cellulaire associée.

Summary

This manuscript presents my research activities, from my Ph.D. in the laboratory of biochemistry at the École Polytechnique to my current CNRS position at the Institut de Biologie Structurale. During these years, through use of various modern structural biology techniques, I have sought to understand how interactions of biological partners underpin different processes. My early work focused on the mechanisms of specificity of different aminoacyl-tRNA synthetases and helped to explain the molecular basis of their fidelity during protein translation. After joining the CNRS, I discovered another research field: the world of viruses and their replication machines. In particular, I have contributed to the great adventure of the influenza virus replication complex, an adventure that led to the resolution of the structure of the full polymerase trimer. Far from being the end of the project, I am now trying to understand influenza polymerase assembly and its cellular dynamic, and am extending this work towards other viral replication machines.

SOMMAIRE

CURRICULUM VITAE

Thibaut CREPIN
Né le 31 décembre 1974 à Lesquin (Nord)
Nationalité française

Adresse professionnelle
IBS - UMR5075 CEA-CNRS-UGA
71 avenue des Martyrs
38042 Grenoble Cedex 9
☎ 04 76 20 94 39
E-mail : thibaut.crepin@ibs.fr

RECHERCHE

depuis 2016	**CR1** - *IBS*, Grenoble (groupe de M. Jamin ; équipe de R. Ruigrok) Études structurales et fonctionnelles de machineries réplicatives virales et de leurs interactions avec des partenaires cellulaires
2007 - 2015	**CR2 & CR1** - *UVHCI*, Grenoble (groupe de R. Ruigrok & M. Jamin) Études structurales et fonctionnelles d'ARN-polymérases virales et de leurs interactions avec des partenaires cellulaires
2003 - 2007	**Post-doctorant** - *EMBL*, Grenoble (groupe de S. Cusack) Études structurales de différentes aminoacyl-ARNt synthétases
1999 - 2002	**Doctorant** - *Ecole Polytechnique*, Palaiseau (équipe d'Y. Mechulam) Études structurales et fonctionnelles de méthionyl-ARNt synthétases

FORMATION

2002	**Thèse de doctorat** - Biologie *École Polytechnique*, Palaiseau
2001	**Diplôme d'ingénieur** - Sciences de la vie *Conservatoire National des Arts et Métiers*, Paris
1992 - 1998	**Cursus universitaire** - Biochimie *Université des Sciences et Technologies de Lille*, Lille

PRODUCTION SCIENTIFIQUE

- 41 publications parues ; h-index = 17 (ISI) ou 20 (Google Scholar)
- 2 chapitres d'ouvrage
- 48 structures déposées à la PDB
- 1 brevet déposé

AUTRES ACTIVITES

- Coordinateur du projet ANR RNAP-IAV (ANR-14-CE09-0017) depuis 2014 ; 459 k€ alloués
- Encadrement de 3 post-doctorants, 3 doctorants et 17 stagiaires (tous niveaux de BTS au M2)
- Participation au cours de Virologie Fondamentale de l'Institut Pasteur depuis 2012
- Participation à l'enseignement théorique de l'Université Joseph Fourier depuis 2004
- Participation à l'enseignement expérimental de l'Ecole Polytechnique de 1999 à 2002
- Membre du comité d'organisation des Journées Francophones de Virologie (JFV) depuis 2014
- Membre du comité d'évaluation de projets « biologie-santé » (CP5) du synchrotron SOLEIL depuis 2012
- Responsable administratif du BAG EMBL-UVHCI pour l'ESRF depuis 2011
- Local contact sur les lignes de lumière dédiées à la biologie de l'ESRF de 2005 à 2014

I - Financements

Depuis 2013, un projet de recherche collaboratif relatif à l'assemblage du complexe réplicatif du virus de la grippe était soumis à financement auprès de l'ANR. Il a été financé en 2014 (Interactions protéine-protéine et protéine-ARN au sein du complexe réplicatif du virus de la grippe de type A, acronyme : RNAP-IAV, ANR-14-CE09-0017). Ce projet regroupe deux partenaires, l'IBS (précédemment l'UVHCI) à Grenoble (partenaire 1 : Thibaut Crépin) et le VIM de l'INRA à Jouy-en-Josas (partenaire 2 : Anny Slama-Schwok). Le montant de l'aide totale allouée par l'ANR pour sa réalisation est de 459 924 €, repartis entre les deux partenaires, à hauteur de 249 924 € pour l'IBS et 209 999 € pour le VIM. Ce projet vise à détailler les bases moléculaires de l'assemblage de l'ensemble du complexe réplicatif du virus de la grippe ainsi que de ses interactions avec des partenaires cellulaires.

II - Encadrement

Dans cette partie, sont regroupées toutes les personnes que j'ai été amené à superviser. Sont présentés les personnels dont j'ai été directement (co)responsable, puis les doctorants auxquels je n'ai apporté qu'une expertise, puis ceux pour lesquels j'ai pris part à leur jury de thèse.

a) Post-doctorant

Nom Prénom	Année(s)	Co-responsable	Article	revue
Sylvie Chenavas	2009-2012	Rob Ruigrok	1	4
Denis Bouvier	2008-2009	Rob Ruigrok	1	1
Andrès Palencia	2007-2009	Stephen Cusack	1	-

b) Doctorants co-encadrés

Nom Prénom	Soutenance	Co-directeur	Article	revue
Alice Labaronne	2017	Rob Ruigrok & Guy Schoehn	4 (2 en révision)	1
Christopher Swale	13/11/2015	Rob Ruigrok	3 (1 en révision)	2
Alexandre Monod	06/11/2014	Guy Schoehn	2	4

c) Stagiaires encadrés

Année	Durée	Nom Prénom	Niveau	Rapport
2017	6 mois	Amélie Donchet	M2	oui
	3 semaines	Mickael Gloaguen	L3	non
	14 semaines	Justine Sensini	BTS	oui
2016	7 semaines	Sébastien Cammarata	BTS (2^{ième} année)	oui
	7 semaines	Amélie Donchet	M1	oui

	6 mois	Myriam Miloudi	M2	oui
	7 semaines	Amélie Vessière	BTS (1ère année)	oui
2015	7 semaines	Célia Boyat	BTS (2ième année)	oui
	7 semaines	Sébastien Cammarata	BTS (1ère année)	oui
	7 semaines	Elisa Tournebize	L3	non
2014	7 semaines	Célia Boyat	BTS (1ère année)	oui
	4 semaines	Kenneth Mermet	L3	oui
	4 semaines	Blandine Vallat	M1	non
2013-2014	11 mois	Laura Tengo	Licence Pro (alternance)	oui
2013	7 semaines	Frédéric Achard	BTS (1ère année)	oui
	7 semaines	Laura Tengo	BTS (2ième année)	oui
2012	6 semaines	Marie-Claire Conde	L3	non
	7 semaines	Katinka Ruigrok	M1	oui
	6 mois	Christopher Swale	M2	oui
	7 semaines	Laura Tengo	BTS (1ère année)	oui
2011	4 semaines	Paul Alouani	L3	non
	6 mois	Alexandre Monod	M2	oui
	7 semaines	Edouard Zuurdeg	BTS (2ième année)	oui
2010	6 mois	Manuel Blanc	M2	oui
	7 semaines	Alexandre Monod	M1	oui
	7 semaines	Christopher Swale	L3	non
2009	7 semaines	Christopher Swale	BTS (2ième année)	oui
2008	7 semaines	Christopher Swale	BTS (1ère année)	oui

d) Doctorants supportés

Nom Prénom	**Directeur(s)**	**Soutenance**	**Article conjoint**
Catherine Fallecker	Emmanuel Drouet	09/01/2014	1 en préparation
Guillaume Communie	Martin Blackledge & Rob W Ruigrok	24/10/2013	1
Ivan Ivanov	Rob W Ruigrok & Marc Jamin	02/12/2011	1
Alexandre Dias	Florence Baudin	08/12/2009	3

e) Membre de Jury de thèse

Nom Prénom	Soutenance	Directeur	Titre
Marie Gerber	27/09/2016	Catherine Isel & Roland Marquet	Recherche de signaux d'empaquetage spécifiques du génome des virus influenza A
Louis-Marie Bloyet	18/12/2015	Denis Gerlier	Formation et régulation du complexe polymérase du virus de la rougeole
Julien Sourimant	20/05/2015	Jean-Francois Eléouët	Caractérisation structurale et fonctionnelle de polymérase du virus respiratoire syncytial

III - Enseignement

Cette partie résume toutes les activités d'enseignement que j'ai effectué.

- **Intervenant** - Master 2 « cours de virologie fondamentale » (responsables N. Tordo & P.E. Ceccaldi) des Universités Diderot (Paris 7); Descartes (Paris 5); Pierre &Marie Curie (Paris 6)
  Titre: Les *complexes de réplication des virus à ARN négatifs* (2 h)
  Intervention en français **depuis 2012**

- **Intervenant** - Module « biologie structurale » (responsable W. Weissenhorn) du Master 2 « biochimie et biologie structurale » (responsable M. Jamin) de l'Université Joseph Fourier de Grenoble
  Titre: *la biologie structurale appliquée aux complexes réplicatifs des virus à ARN négatifs* (2 h)
  Intervention en français **de 2011 à 2016**

- **Intervenant** - Module « acides nucléiques » (responsable R. Ruigrok) du Master 2 « biochimie et biologie structurale » (responsable M. Jamin) de l'Université Joseph Fourier de Grenoble
  Titre: *les méthodes de production d'ARN appliquées à la biologie structurale* (6 h)
  Intervention en français **depuis 2004 à 2016**

- **Responsable de module** - Module « acides nucléiques » (40h) du Master 2 « biochimie et biologie structurale » (responsable M. Jamin) de l'Université Joseph Fourier de Grenoble
  En remplacement de R. Ruigrok pendant l'**année universitaire 2013-2014**

- **Intervenant** - HKU-Pasteur Virology Course « influenza and other emerging respiratory viral infections (responsables N. Tordo & M. Peiris) au centre de recherche Pasteur de Hong-Kong
  Titre: *structure - function of the influenza viral particle components* (1 h 30)
  Intervention en anglais en **2010**

- **Tuteur** - EMBO practical course « structural characterization of macromolecular complexes » (responsable C. Petosa) organisé à l'EMBL-Grenoble
  Titre: *Protein/Nucleic Acid Interactions* (20 h/cours)
  Travaux pratiques en anglais pour les cours organisés en **2006**, **2008** et **2010**

- **Chargé d'enseignement -** Travaux expérimentaux du Département de Biologie (responsable S. Blanquet & Y. Mechulam) de l'Ecole Polytechnique (Palaiseau)
  Travaux pratiques en français pour les cours organisés de **1999 à 2001**

IV - Participation à l'organisation de conférences

- **Membre du comité d'organisation** - Journées Francophone de Virologie (JFV) depuis **2014**
- **Membre du comité d'organisation** - aaRS2008; International Conference on Aminoacyl-tRNA Synthetases : From basic mechanisms to systems biology. *Veyrier du lac*,France. **7 - 11 septembre 2008**

V - Participation à des conférences

1 - Swale C, Monod A, Tengo L, Garzoni F, Bourhis JM, Cusack S, Schoehn G, Berger I, Ruigrok RW & **Crépin T**. (intervention orale de C Swale)
L'importine-β cellulaire RanBP5 interagit avec le sous-complexe PA-PB1 du virus de la grippe
XVIII^{ème} journées francophones de virologie. Paris, France. 24-25 mars 2016.

2 - Fournier G, Tengo L, Magnus J, Chiang C, Munier S, Tomoiu A, Demeret C, Vidalain PO, Jacob Y, Ruigrok RW, **Crépin T** & Naffakh N. (intervention orale de N Naffakh)
Functional and structural analysis of the interaction between influenza virus polymerase and the spliceosomal RED-SMU1
NSV 2015. Sienna, Italie. 14-19 juin 2015.

3 - **Crépin T.** (intervention orale)
Structure de la machinerie réplicative du virus de la grippe
XVII^{ème} journées francophones de virologie. Paris, France. 9-10 avril 2015.

4 - **Crépin T**, Chenavas S, Slama-Schwok A, Delmas B, & Ruigrok RW. (intervention orale)
La forme monomérique de la nucléoprotéine du virus de la grippe
XV^{ème} journées francophones de virologie. Paris, France. 18-19 avril 2013.

5 - **Crépin T**, Cusack S, Hart DJ & Ruigrok RW. (intervention orale)
Structure based drug-design on influenza virus RNA polymerase
Colloque scientifique grippe (AH1N1). Veyrier du Lac, France. 18-19 Juin 2010.

6 - Dias A, Bouvier D,**Crépin T**, McCarthy A, Hart DJ, Baudin F, Cusack S & Ruigrok RW.
The cap-snaching endonuclease of influenza virus polymerase resides in the PA subunit
5th Orthomyxovirus research conference. Freiburg, Allemagne. 09-12 septembre 2009.

7 - **Crépin T**, Dias A, Swale C, Cusack S & Ruigrok R (intervention orale)
Mutational analysis of the endonuclease active site of the influenza virus polymerase PA subunit
4th FLUPOL annual meeting. London, Angleterre. 1-2 septembre 2009.

8 - **Crépin T**, Dias A, Bouvier D, McCarthy A, Baudin F, Cusack S & Ruigrok RW. (intervention orale)
Structure and function of the influenza virus endonuclease
6th International Virus Assembly Symposium. Crête, Grèce. 03-07 Mai 2009.

9 - **Crépin T**, Vu M, Palencia A, Alley MRK, Martinis S & Cusack S. (intervention orale)
High resolution structure of the *Escherichia coli* LeuRS:tRNA^{Leu}:AN2679 complex
aaRS2008 - international conference on aminoacyl-tRNA synthetases: from basic mechanisms to system biology. Veyrier du Lac, France. 7-11 septembre 2008.

10 - **Crépin T**, Vu M, Alley MR, Martinis S & Cusack S. (intervention orale)
Structural and functional properties of *Escherichia coli* LeuRS:tRNA^{Leu} complex
22nd tRNA workshop. Uppsala, Suède. 1-6 novembre 2007.

11 - **Crépin T** & Cusack S.
tRNA^{Pro} recognition by *Rhodopseudomonas palustris* prolyl-tRNA synthetase
2006 international conference on aminoacyl-tRNA synthetases: from the genetic code to human diseases & medicine. San Diego, Californie, USA. 1-6 octobre 2006.

12 - Cusack S & **Crépin T.** (intervention orale de S. Cusack)

Structural basis for activation and aminoacylation by *Rhodopseudomonas palustris* prolyl-tRNA synthetase
2006 international conference on aminoacyl-tRNA synthetases: from the genetic code to human diseases & medicine. San Diego, Californie, USA. 1-6 octobre 2006.

13 - Sukuru SC, **Crépin T**, Morris JC, Grøtli M, Danel F, Page MG, Cusack S, Hartlein M, Kron M & Kuhn LA.
The contributions of conformational flexibility to *Brugia malayi* asparaginyl-tRNA synthetase specificity for inhibitors identified by structure-based screening
231st american chemical society national meeting, Atlanta, Géorgie, USA. 26-30 mars 2006.

14 - **Crépin T** & Cusack S. (intervention orale)
Structural basis for substrate discrimination by *Rhodopseudomonas palustris* prolyl-tRNA synthetase
21rst tRNA workshop. Bangalore, Inde. 2-7 décembre 2005.

15 - **Crépin T**, Fridd S, Legrand P & Cusack S.
Rhodopseudomonas palustris prolyl-tRNA synthetase crystal structure
2004 international conference on aminoacyl-tRNA synthetases : ancient molecules for future biology and medicine. Séoul, Corée. 4-9 juillet 2004.

16 - **Crépin T**, Kron M, Kuhn L, Grøtli M, Haertlein M, Danel F & Cusack S.
Crystallographic studies of *Brugia malayi* asparaginyl-tRNA synthetase in complex with substrate analogues and inhibitors
2004 international conference on aminoacyl-tRNA synthetases : ancient molecules for future biology and medicine. Séoul, Corée. 4-9 juillet 2004.

17 - Yaremchuk A, **Crepin T**, Tukalo M & Cusack S. (intervention orale de S. Cusack)
Editing by class I and class II aminoacyl-tRNA synthetases
2004 international conference on aminoacyl-tRNA synthetases : ancient molecules for future biology and medicine. Séoul, Corée. 4-9 juillet 2004.

18 - **Crépin T**, Kron M, Kuhn L, Grøtli M, Hartlein M & Cusack S.
Design of macrofilaricidal aminoacyl-tRNA synthetase inhibitors
20th tRNA workshop. Banz, Allemagne. 2-7 octobre 2003.

19 - Mechulam Y, **Crépin T**, Schmitt E, Sampson PB, Vaughan MD, Honek JF & Blanquet S. (intervention orale de Y. Mechulam)
Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in *Escherichia coli* methionyl-tRNA synthetase
20th tRNA workshop. Banz, Allemagne. 2-7 octobre 2003.

20 - **Crépin T**, Schmitt E, Blanquet S & Mechulam Y. (intervention orale)
Structure et fonction du domaine C-terminal de la méthionyl-ARNt synthétase
4ième rencontre SifrARN. Nancy, France. 14-17 octobre 2002.

21 - Mechulam Y, **Crépin T**, Schmitt E & Blanquet S.
3D-structure of the C-terminal domain of *Pyrococcus abyssi* methionyl-tRNA synthetase
19th tRNA workshop. Shanghai, Chine. 6-11 avril 2002.

22 - **Crépin T**, Schmitt E, Blanquet S & Mechulam Y.
3D-structure of the C-terminal domain of *Pyrococcus abyssi MetRS*
Aminoacyl-tRNA synthetases in Biology, Medicine & Evolution. Pacific grove, Californie, USA. 13-18 janvier 2002.

VI - Animation et management de la recherche

a) Gestion d'équipe

En octobre 2013, suite à un problème de santé, Rob W Ruigrok (Professeur, UGA) a été totalement absent du laboratoire pendant une année, avant de reprendre son activité (partiellement pendant 1 an puis

totalement depuis Septembre 2015). J'ai donc assuré la responsabilité et la gestion de l'équipe (composée alors de 3 doctorants et de moi-même) de façon à assurer la pérennité de l'activité de l'équipe mais aussi assurer aux étudiants la continuité de leur formation. Pendant cette période, j'ai pu obtenir un financement ANR et un doctorant a soutenu de thèse.

b) Membre du PRC5 - SOLEIL

Depuis 2012, je suis membre du comité de programme n°5 (CP5) du synchrotron SOLEIL. Ce comité est chargé, deux fois par an, d'évaluer les projets relatifs à la biologie et la santé pour l'attribution des temps de faisceau sur les différentes lignes de lumière dédiées à la biologie du synchrotron français SOLEIL.

c) Responsable administratif de BAG - ESRF

Depuis 2011, je suis responsable administratif du « BAG » EMBL-UVHCI. Mon rôle consiste à gérer les demandes de temps de faisceau pour mon unité sur les différents synchrotrons qu'elle sollicite ainsi que les rapports correspondants. Mon ancienne unité (UVHCI) disposait majoritairement d'un accès l'ESRF de Grenoble mais a sollicité également d'autres grands instruments comme ce fut le cas en 2010-2011, lors de la phase d'upgrade de l'ESRF. Au quotidien, je gérais également la répartition des temps de faisceau alloués dans le cadre du BAG. Malgré mon rattachement à l'IBS au 1er janvier 2016, j'ai exprimé ma volonté de continuer cette activité dans le cadre d'un partenariat avec l'EMBL de Grenoble.

d) Local contact - ESRF

Lors de mon travail post-doctoral, l'expérience que j'ai acquise en matière d'utilisation des lignes de lumière, m'avait valu d'être sollicité par le groupe de cristallographie macromoléculaire (Manager des lignes de lumière : David Flot) de l'ESRF. J'assurais ainsi, en qualité de « local contact », l'accueil et le suivi des utilisateurs extérieurs sur les lignes de lumière dédiées à la biologie structurale et je leur apportais si besoin, mon expertise. Le volume horaire de cette activité correspondait à un jour de travail par mois. En intégrant le CNRS, j'ai souhaité conserver cette activité, et ce pour différentes raisons. Elle me permet dans un premier temps de me tenir au courant des améliorations apportées régulièrement sur les lignes de lumière et donc de conserver mon savoir-faire et mes compétences. Cette activité me permet deuxièmement d'avoir un accès privilégié sur les lignes de lumière dédiées à la biologie de l'ESRF, raison non négligeable me permettant une grande réactivité pour les projets prioritaires auxquels je suis amené à participer. Enfin, la troisième raison qui m'a poussée à conserver cette activité est qu'elle me permet d'être en contact constant avec la communauté internationale des cristallographes et ainsi d'avoir des échanges facilités.

e) Membre de jury de concours pour le recrutement d'un agent en situation de handicap

En 2014, j'ai pris part au jury de concours CNRS pour le recrutement d'un agent en situation de handicap (concours 14INSB11), en qualité de représentants du laboratoire. Ce concours a abouti au recrutement d'un ingénieur de recherche (Emilie Stermann) dans l'équipe de Pascal Fender (DR2, CNRS). En 2015, j'ai pris part au jury de titularisation d'Emilie Stermann.

Je me souviens…

… Je me souviens de ce matin d'automne de l'an 2000… Juste arrivé au labo, attendant que le divin nectar finisse de passer mais déjà en compagnie de la bino… Dans la quiétude de la chambre thermostatée, le moment d'observation des gouttes de la veille, mélange d'angoisse du vide et d'excitation de l'inconnu. La même question que pour le chat de Schrödinger : je regarde et j'ai une réponse ou alors je ne regarde pas et dans ce cas, reste l'espoir mais surtout le doute (sauf que pour le chat, au bout d'un moment, à l'odeur, on peut arriver à savoir !!!). Résonne aussi dans ma tête la voie pleine de sarcasmes de Lionel: « 'mach'ra pas !!! »… Et si il avait raison ? une fois de plus… …

… Je veux savoir. Je me lance. Première boite, première goutte. Rien. Déjà, l'espoir en prend un coup ! La seconde, idem

¬ *« Bon, ben allé, c'est parti ! ».*

Les gouttes s'enchainent alors, imprégnant la pupille comme des comètes fugaces…

¬ *« Qu'est-ce que c'est que ça ? … oh putain !…OH PUTAIN DE MERDE !!! »*

Des cristaux ! Partout. Entremêlés par endroits, isolés dans d'autres. Je relève la tête. Je tremble. Je recolle les paupières aux oculaires pour vérifier que je ne rêve pas… ils sont toujours la !...

¬ *« Les chefs ! … faut que je demande aux chefs »*

Je sors en trombe de la pièce, ma subtilité habituelle. Je n'irais pas jusqu'à parler de secousses telluriques mais si il fallait éviter les vibrations et les chocs, c'était trop tard ! Je les trouve assis l'un en face de l'autre autour de la table basse de la cafet, les yeux encore embrumés, café fumant à la main.

¬ *« Bonjour, faut que je vous montre quelque-chose »*

Tandis que je me sers un café, retentit derrière moi

¬ *« Yves. Emmanuelle. Mon-sieur Crépin. Mes hommages ! »*

SB est d'humeur joyeuse ce matin… « Je vais le laisser avec son anticoopérativité !!! » me dis-je. Le café avalé, je les entraine dans la caverne. Là, le verdict tombe :

¬ *« cool, des cristaux ! »*

vite suivit de :

¬ *« ouais, mais tant qu'on a pas diffracté, on peut pas dire ! »*

Et la, après un moment fugace d'extrême jouissance et d'autosatisfaction, ressurgissent les doutes et incertitudes parsemant le long chemin de croix du cristallographe : « est-ce que mes cristaux ne vont pas s'abimer ? - Faut congeler. - Faut un cryo. - C'est quoi le meilleur cryo ? 15, 20, 25 % de glycérol ? Ethylène glycol ? PEG400 ?- on peut aussi les monter en capillaire. Et si ça diffracte pas, on fait quoi ? »… et ça, ce n'était que le début…

Depuis, les robots de cristallisation ont écourté les séances de « cryothérapie » forcées et les systèmes de visionnage automatiques ont remplacé les lampes binoculaires. Cette scène, navrante de banalité pour beaucoup, constitue un tournant majeur dans ma vie de scientifique. J'ai toujours été curieux et j'ai toujours éprouvé un besoin de comprendre les choses, d'obtenir des réponses. J'ai parfaitement conscience qu'à l'issue de ma vie, je ne les aurai pas toutes, loin de là. Je serai un monstre d'arrogance si je le pensais, mais en aurai-je quelques-unes au moins ? J'espère ! Ce métier que j'ai choisi, me permet de m'épanouir en assouvissant ce besoin d'avoir une explication (« mais est-ce vraiment l'Explication ? ») de mieux comprendre les choses et de me rapprocher du fameux « Pourquoi », celui qui précède tous les autres. Pour écrire ce manuscrit, je me suis ainsi posé la question : « mais que fait-il que j'en sois arrivé là ? ». Cette scène m'a semblé être cette sorte de passage initiatique, matérialisation par l'obtention du cristal de cette possibilité de se rapprocher de la vérité. Cependant, on peut se demander si la biologie structurale et la cristallographie en particulier, sont la meilleure des solutions pour la recherche de la vérité, biologiquement parlant. Sans conteste, sa finalité tend à l'être, mais au prix d'une dualité biochimico-cristallographique empreinte de masochisme et de schizophrénie. Je me surprends des fois à imaginer certains de mes pairs errant sans répit comme des âmes en peine, grimés tel Gollum à la recherche de son prréé́é-cieux. Que veulent-ils *in fine*, eux ? Je n'en sais rien et ne veux pas savoir si c'est un besoin de reconnaissance, de gloire, d'argent, de « Nature », de compenser quelque-chose ... Cela leur appartient. Leur Vérité. Moi, je veux comprendre ou tout du moins, essayer. Comprendre comment ça marche… Et pourquoi ? Un manque d'ambition de ma part ?, un coté terre-à-terre ? Non, je ne le pense pas. J'ai conscience de la finalité de mon travail, de ce que pourquoi j'ai été recruté, même si maintenant j'ai de plus en plus le sentiment que de n'avoir qu'à gérer la compréhension.

J'ai voulu cet avant-propos comme une digression introductive la moins scientifique possible mais aussi une ébauche introspective de mon Moi professionnel. En le mettant noir sur blanc, mon cheminement peut s'apparenter à celui d'un biochimiste mercenaire de la biologie structurale, tant les sujets abordés peuvent paraitre disparates et variés. Il faut simplement y voir une progression continue couplée à une notion d'échelle différente dans l'étude de constituants du vivant ayant vocation à interagir entre eux pour

l'accomplissement d'un processus biologique. Mon recrutement au CNRS m'a fait basculer du domaine de la traduction à celui des virus, deux mondes ayant comme point commun la molécule d'ARN.

Bases moléculaires de la fidélité de la traduction

La similarité entre les acides aminés représente pour toute cellule, un challenge de taille pour assurer la fidélité de son processus traductionnel. Les bases d'un « contrôle-qualité » permettant une discrimination efficace entre les acides aminés ont été lancées par Linus Pauling dans les années 50s (Pauling, 1957) et depuis, ce mécanisme n'a cessé d'être décortiqué. La traduction de l'ARNm s'effectue en deux étapes critiques. Les aminoacyl-ARNt synthétases (aaRSs) décryptent le code génétique en attachant les acides aminés sur les ARNt correspondant. Ces aminoacyl-ARNt sont ensuite pris en charge par le ribosome puis positionnés sur l'ARNm permettant *in fine* la synthèse protéique. La réaction d'aminoacylation de l'ARNt s'effectue elle aussi, en deux étapes: (i) l'activation de l'acide aminé par une molécule d'ATP donnant l'aminoacyl adenylate puis (ii) son transfert sur l'extrémité CCA-3' de l'ARNt correspondant. La fidélité traductionnelle repose en grande partie sur la précision de la réaction enzymatique catalysée par les aaRSs. Un phénomène de correction d'erreur, aussi appelé édition, peut se produire en amont (édition pré-transfert) ou en aval (édition post-transfert) du transfert de l'acide aminé sur l'ARNt. Les premières preuves d'une édition post-transfert ont été observées grâce à l'IleRS et la PheRS, pour lesquelles il a été montré que ces deux enzymes étaient capables d'hydrolyser très efficacement, respectivement le Val-ARNt^{Ile} et l'Ile-ARNt^{Phe} (Eldred & Schimmel, 1972; Yarus, 1972). L'édition pré-transfert a quant à elle, été mise en évidence avec l'IleRS (Baldwin & Berg, 1966) et confirmée par la suite par « fast-kinetic » (Fersht, 1977). Depuis, ces mécanismes d'édition ont été mis en évidence pour la moitié des aaRSs. Pour expliquer la rareté des ARNt mal-aminoacylés, un modèle dit à « double tamis » (*double sieve* en anglais) a été proposé (Fersht & Kaethner, 1976). Dans ce modèle, un premier « tamisage » grossier serait réalisé par le site actif de l'enzyme qui lierait sur l'ARNt, le « bon » acide aminé ainsi que ceux proches stériquement, tout en excluant les acides aminés plus « volumineux ». Un second site présent sur l'enzyme, appelé site d'édition, réaliserait ensuite un second criblage, hydrolysant uniquement les ARNt mal-chargés grâce à une reconnaissance beaucoup plus fine de l'acide aminé. De nombreuses structures cristallines sont venues renforcer ce modèle, sans pour autant apporter d'explication claire à ce mécanisme. En revanche, les structures de certaines aaRSs ont compliqué la donne, montrant que le site d'édition, porté par le domaine d'édition pouvait être distant du site actif de l'enzyme de plus de 35 - 40 Å (Fukai *et al.*, 2000; Nureki *et al.*, 1998; Sankaranarayanan *et al.*, 1999; Silvian *et al.*, 1999; Yaremchuk *et al.*, 2000a).

Les aaRS forment une famille d'enzymes isofonctionnelles indispensables à la vie cellulaire. Elles représentent l'archétype de la protéine modulaire, au sein desquelles différents domaines individualisés ayant une fonction qui leur est propre, fonctionnent de manière synergique en vue de l'accomplissement d'un processus biologique. Cette modularité est la source de nombreuses spéculations évolutionnistes et a

notamment servi de base à la partition des aaRS en deux classes distinctes, suivant que leur domaine catalytique soit agencé autour d'un repliement de type « Rossmann-fold » ou d'un feuillet de brins antiparallèlles (Eriani *et al.*, 1990). De manière intéressante, il est apparu qu'un premier processus de discrimination des acides aminés aurait pour origine le type de repliement à partir duquel ont dérivé les domaines catalytiques des aaRS. Mes premiers travaux ont visé à détailler par des approches structurales, les mécanismes de la spécificité du fonctionnement de certaines aaRSs. Différents systèmes enzymatiques ont été étudiés au cours de ces travaux, la MetRS et la LeuRS ainsi que l'AsnRS et la ProRS (Figure 1).

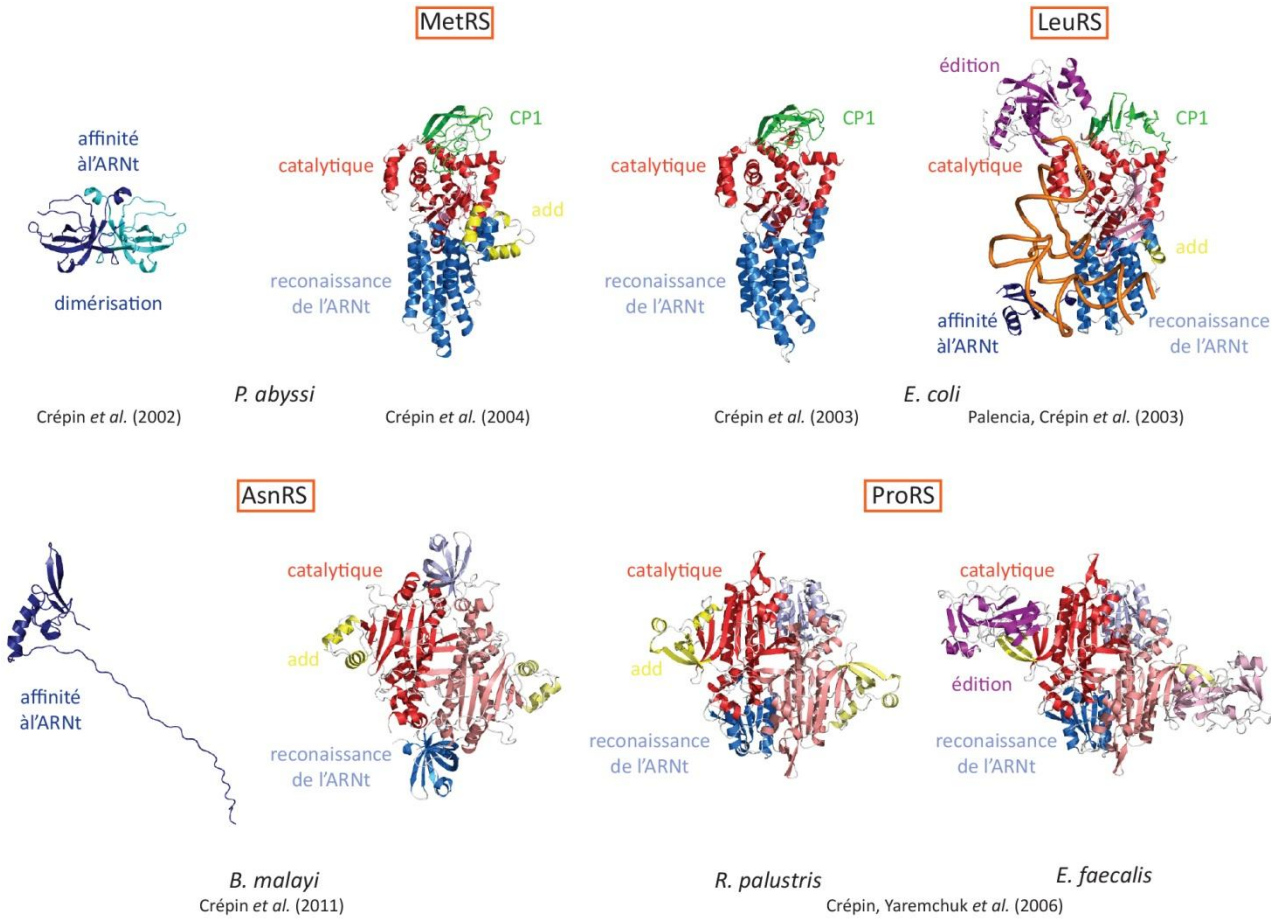


Figure 1: Représentation macromoléculaire d'après leurs domaines des différentes aminoacyl-ARNt synthétases ayant fait l'objet de publications. Les enzymes dites de « classe 1 » sont dessinées sur la partie haute alors que celles dites de « classe 2 » sur la partie basse. Le même code couleur est employé pour toutes les enzymes, à savoir les domaines catalytiques en rouge, les domaines impliqués dans la reconnaissance et l'affinité de l'ARNt en bleu, les domaines de connexion (CP1 domain) des aaRS de classe 1 en vert, les domaines impliqués dans l'édition post-transfert en violet et les domaines additionnels en jaune. Les aaRS de classe 2 étant dimériques, la distinction des deux monomères est faite par la nuance dans les couleurs utilisées. Il en va de même pour le domaine Cterminal de la MetRS (en haut à gauche).

Outre leur aspect macromoléculaire, j'ai cherché par mes travaux, à mettre en lumière les réarrangements moléculaires de la discrimination des différents substrats et ainsi pouvoir détailler les

différentes étapes de la réaction d'aminoacylation. L'utilisation de molécules mimant des états de transition réactionnels s'est avérée être un puissant outil, à la condition d'avoir dans le même temps des cristaux ayant un fort pouvoir de diffraction. Pour illustrer mes propos, je prends l'exemple des travaux réalisés sur la MetRS sous la supervision des Dr Emmanuelle Schmitt et Yves Mechulam (Crépin *et al.*, 2002; Crépin *et al.*, 2004; Crépin *et al.*, 2003). Cette enzyme a longtemps été étudiée par le laboratoire de biochimie de l'École Polytechnique et l'est toujours. Durant mon doctorat, le laboratoire avait établi une collaboration avec le Pr. JJ Honek (Université de Waterloo, Canada). Ce dernier a fourni des composants dérivés de la L-méthionine, utilisées alors dans des expériences de trempage dans le but de collecter des données de diffraction à haute résolution (1,7 <résolution < 1,93 Å). Grâce à ces molécules, j'ai pu décortiquer les réarrangements atomiques du site actif de l'enzyme, notamment celui de la tyrosine-15, consécutifs à l'enchainement des étapes de la réaction d'aminoacylation (Figure 2). En effet, la fixation de la L-méthionine dans le site actif de la MetRS s'accompagne d'un basculement la chaine latérale de la Tyr-15 et de celle du Trp-253. Ces deux résidus semblent essentiels à la discrimination de la méthionine. Par contre, une augmentation de l'encombrement stérique au niveau de sa fonction carboxyle, telle qu'envisagée lors de son transfert sur l'extrémité 3'OH de l'ARNt^{Met}, semble s'accompagner d'un mouvement vibratoire de la chaine latérale de la Tyr-15. La Figure 2 montre notamment l'évolution de la densité électronique associée à la Tyr-15 au fur et à mesure de l'augmentation de la « gêne » stérique au niveau de la fonction carboxyle de la méthionine grâce à l'utilisation des composés MetI et MetP mimant des intermédiaires réactionnels. Ces résultats montrent ainsi l'enchainement des réarrangements atomiques du site actif de la MetRS lors la réaction de méthionylation de l'ARNt^{Met} (Crépin *et al.*, 2003).

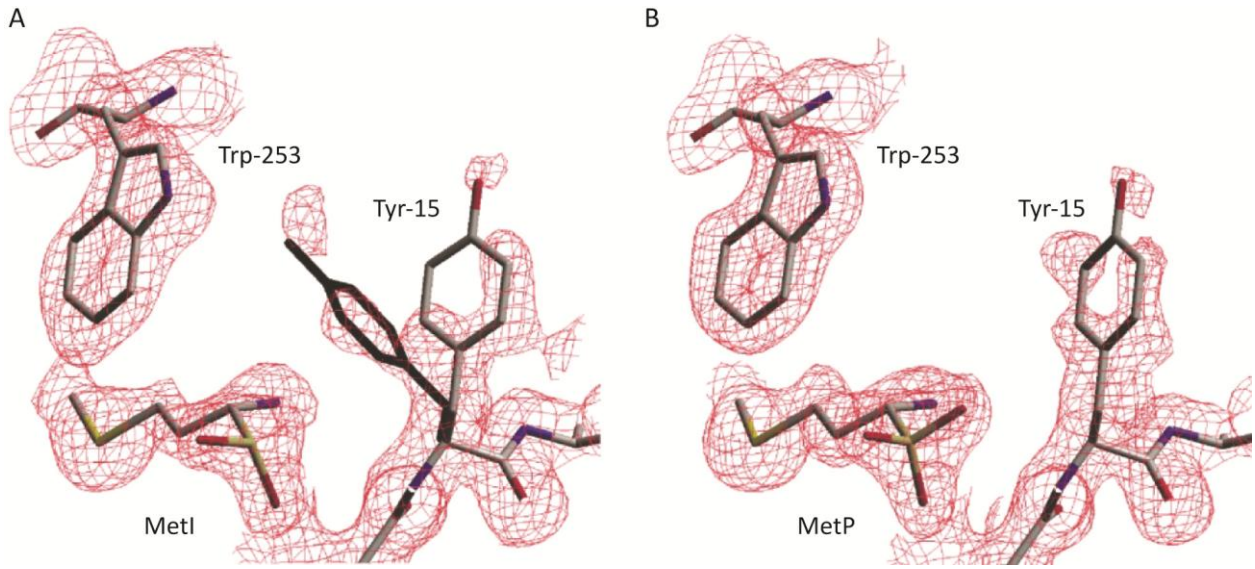


Figure 2: Effet de l'augmentation de l'encombrement stérique au niveau de la fonction carboxyle de la L-méthionine sur les acides aminés du site actif de la MetRS. La figure montre la densité électronique (*2Fo-Fc*) de la tyrosine-15, celle du tryptophane-253 et celle du composé MetI (A) ou celle du MetP (B) à la fin du processus d'affinement. Figure tirée de (Crépin *et al.*, 2003).

J'ai poursuivi cette approche dans le cadre de mes travaux post-doctoraux, avec la volonté d'une progression dans la notion d'échelle des cibles étudiées, en intégrant notamment la composante ARN dans le système. La réussite ne fut pas toujours là. Dans le cadre d'un projet portant sur *Brugia malayi*, une espèce de nématodes associée à la filariose lymphatique, j'ai été amené à tenter de cristalliser le complexe AsnRS:ARNt^{Asn}. Confronté au dynamisme du système, la stratégie a été réorientée vers la partition de la protéine en vue de l'étude structurale du corps de l'AsnRS et de sa spécificité par cristallographie, et celle du domaine Nterminal spécifique de l'enzyme eucaryote par RMN (Crépin *et al.*, 2011). Dans le cadre de deux autres systèmes, la ProRS de *Rhodopseudomonas palustris* et la LeuRS d'*Escherichia coli*, l'obtention de cristaux des complexes protéine:ARN m'a en revanche permis, outre la possibilité d'appréhender la réaction d'aminoacylation dans sa globalité, de détailler précisément les mécanismes de l'édition pré- et post-transfert. Une partie des résultats obtenus avec la ProRS de *R. palustris* n'ayant pas été encore publiée à ce jour, cette partie est plus légèrement détaillée ci-dessous.

Les analyses de séquences montrent que, pour une espèce d'aaRS donnée, son architecture est globalement stable au sein d'un même règne. La ProRS fait exception. En effet, la comparaison de séquences de ProRSs bactériennes montre une grande variabilité dans l'architecture de l'enzyme. Ainsi autour d'un cœur constitué d'un domaine catalytique (feuillet de brins β antiparallèles, caractéristique des aaRSs de classe 2) et d'un domaine de reconnaissance de l'anticodon, des domaines additionnels sont présents. Ils sont insérés au sein de la séquence primaire de l'enzyme ou situés aux extrémités N- et/ou Cterminales (Figure 3). Cette différence d'architecture se traduirait par une différence de spécificité, certaines ProRSs ayant une plus forte propension à l'édition pré-transfert alors que d'autres, une plus forte disposition pour l'édition post-transfert (Beuning & Musier-Forsyth, 2001; Kamtekar *et al.*, 2003).

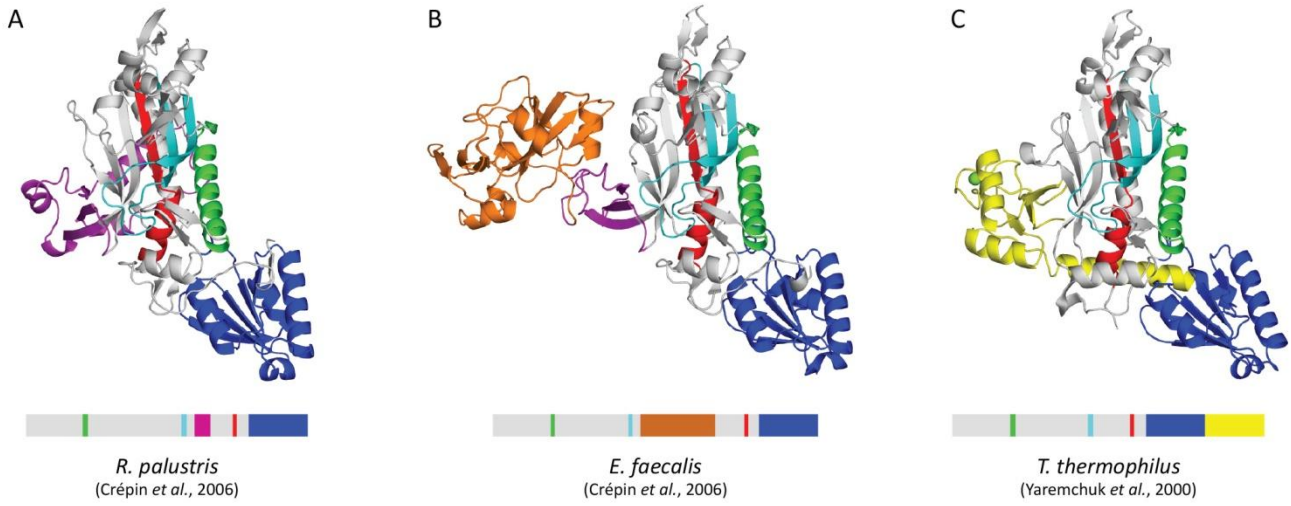


Figure 3: Organisation modulaire de la ProRS. La figure montre les structures cristallographiques de trois ProRS bactériennes, respectivement (A) *Rhodopseudomonas palustris*, (B) *Enterococcus faecalis* et (C) *Thermus thermophilus* ainsi que leur architecture. Les domaines catalytiques sont représentés en gris avec les 3 motifs conservés (motif-1, -2 et -3), respectivement en vert, cyan et rouge, les domaines de liaison à l'anticodon en bleu et les différents domaines additionnels, respectivement en violet, orange et jaune. Figure tirée de (Crépin *et al.*, 2006).

Au commencement de ce travail sur la ProRS, seule les structures de ProRS identiques à celle de *Thermus thermophilus* (Figure 3.C) étaient connues (Kamtekar *et al.*, 2003; Yaremchuk *et al.*, 2000b; Yaremchuk *et al.*, 2000c). Celles-ci se caractérisent par un domaine catalytique (gris sur la Figure 3) et un domaine de reconnaissance de l'anticodon (bleu) possédant à son extrémité Cterminale, un domaine additionnel (jaune), les deux étant reliés par une hélice-α. Cette hélice-α positionne le domaine additionnel à proximité du domaine catalytique. La délétion de ce domaine Cterminal additionnel affecte l'activité d'aminoacylation de la ProRS sans impacter son affinité pour l'ARNt^{Pro} (Ambrogelly *et al.*, 2005).

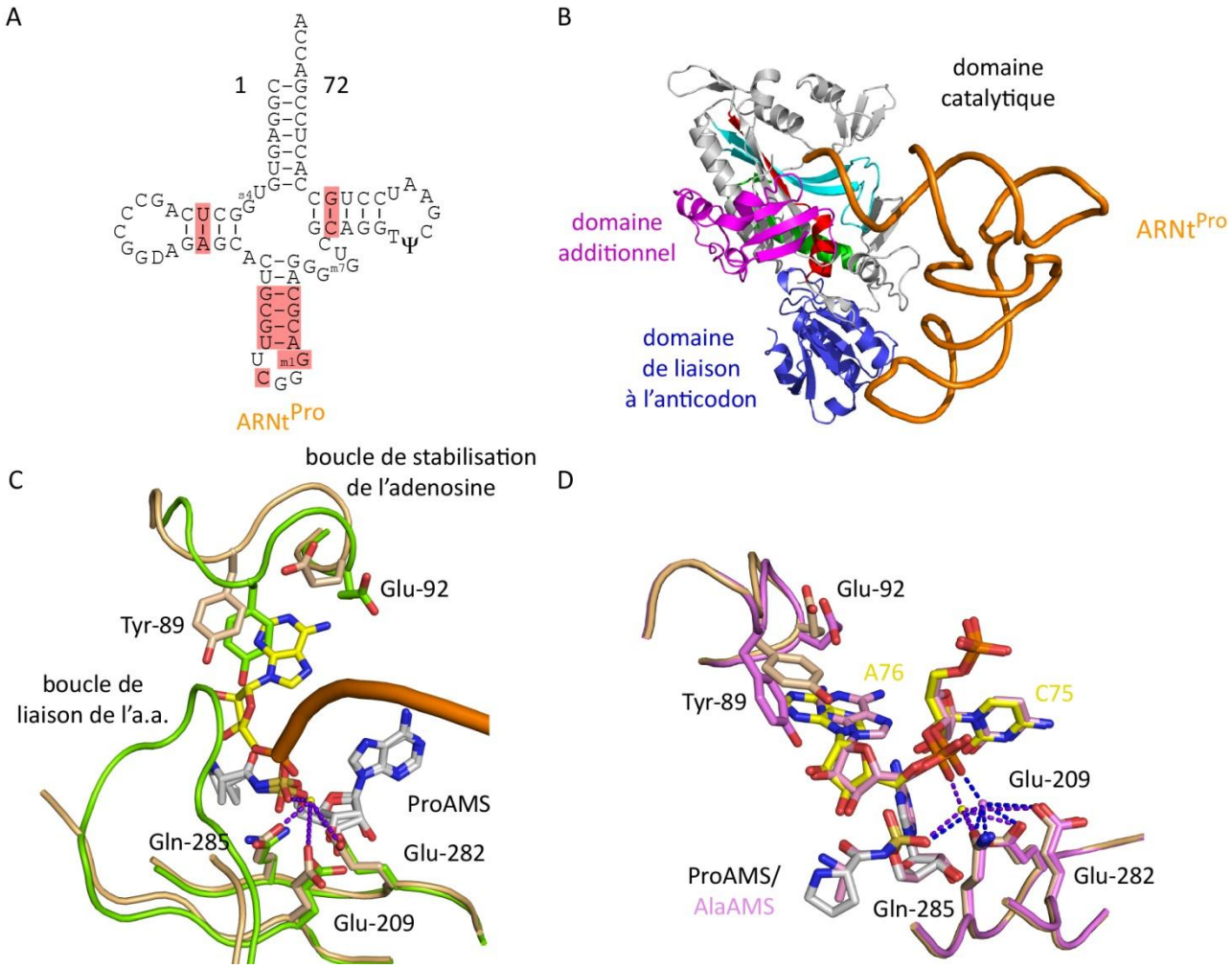


Figure 4: Spécificité de la ProRS de *R. palustris*. (A) Détail de la séquence de l'ARNt^{Pro} de *R. palustris*. Les bases soulignées en rouge correspondent aux bases non-conservées chez les 3 ARNt^{Pro} iso-accepteurs trouvés dans cet organisme. (B) Vue globale de la structure du complexe ProRS:ProAMS:ARNt^{Pro}. Le domaine catalytique est représenté en gris avec les 3 motifs conservés (motif-1, -2 et -3), respectivement en vert, cyan et rouge, le domaine de liaison à l'anticodon en bleu et le domaine additionnel en violet. (C) Effet de l'interaction de l'ARNt^{Pro} dans le site catalytique. Le panneau montre la superposition du site actif de l'enzyme, dans lequel est accommodé le ProAMS (molécule grise), en absence (vert) et présence (beige) de l'ARNt^{Pro}. (D) Spécificité de la réaction de prolylation de l'ARNt^{Pro}. Le panneau montre le site actif en présence du « bon » acide aminé (i.e. ProAMS, protéine en beige) ou d'un « mauvais » acide aminé (i.e. AlaAMS, protéine en violet) pouvant être accommodé en son sein.

Table 1 : Statistiques des collectes des données de diffraction et d'affinement des structures du projet sur la ProRS de *R. palustris*.

		ProRSRp:tRNA1^{Pro}				ProRSRp:tRNA3^{Pro}
		ProAMS	CysAMS	AlaAMS	Proline:AMPPcP	ProAMS
Data collection						
X-ray source		ID14-EH1	ID14-EH1	ID14-EH1	ID14-EH1	ID23-EH1
Wavelength (Å)		0.934	0.934	0.934	0.934	1.0063
Space group		C2	C2	C2	C2	$P2_12_12_1$
Cell dimensions	*a* (Å)	74.00	74.21	73.71	74.45	79.18
	b (Å)	126.07	126.19	125.91	126.50	128.75
	c (Å)	76.43	76.53	76.39	76.54	312.28
	β (º)	93.47	93.87	93.56	93.33	90.0
Resolution range[(a)] (Å)		40-1.7 (1.8-1.7)	40-1.8 (1.9-1.8)	40-2.0 (2.11-2.0)	40-2.0 (2.06-1.98)	40-2.7 (2.82-2.7)
Completeness[(a)] (%)		97.1 (94.3)	95.9 (82.6)	97.9 (86.6)	96.1 (78.5)	87.5 (74.7)
Rsym (I) [(a) (b)] (%)		8.0 (41.7)	5.4 (24.7)	6.2 (33.1)	6.0 (32.4)	7.4 (38.7)
I/σI[(a)]		7.4 (2.3)	15.9 (4.9)	23.3 (4.9)	9.3 (2.5)	11.9 (3.0)
Total reflections[(a)]		157604 (23698)	257677 (29650)	330811 (32366)	98538 (7962)	265985 (19065)
Unique reflections[(a)]		74394 (11329)	62265 (7983)	45898 (6015)	47231 (4311)	77631 (7978)
Multiplicity[(a)]		2.1 (2.1)	4.1 (3.7)	7.2 (5.4)	2.1 (1.85)	3.4 (2.4)
Refinement statistics						
Complex/AU		1	1	1	1	4
R-factor (%)		19.9	18.8	18.4	18.1	19.3
R_{free} (%)		23.4	22.5	22.5	22.4	27.8
R.m.s deviations from ideal values						
Bond length (Å)		0.013	0.012	0.015	0.018	0.031
Bond angle (°)		1.773	1.732	1.987	2.031	3.596
Mean *B*-factor (Å^2)						
Protein		19.4	22.7	23.9	24.0	49.3
RNA		24.7	28.1	29.1	29.5	53.2
ligand		10.5	14.1	14.1	17.1	30.8
Ramachandran plot (%)						
Favored		92.1	91.3	92.1	91.8	79.4
Additional		7.9	8.7	7.9	8.2	20.2

[(a)] Values in parentheses are for the highest-resolution shell.

[(b)] *R*sym (*I*) = $[\sum_{hkl}\sum_i|<I_{hkl}> - I_{hkl,i}|]/[\sum_{hkl}\sum_i|I_{hkl}|]$, where *i* is the number of reflection *hkl*.

Pour détailler cet état de fait, j'ai démarré une étude structurale de la ProRS de *R. palustris*, une bactérie pourpre non sulfureuse de la famille des *Bradyrhizobiaceae*. L'analyse de sa séquence montre une architecture minimale, un domaine de reconnaissance de l'anticodon et un domaine catalytique possédant en son sein, une petite insertion de 80 résidus. L'enzyme a été cristallisée, sa structure résolue à 2,0 Å de résolution (Figure 3.A). De manière intéressante, il apparait que l'insertion dans la séquence primaire de la ProRS de *R. palustris* se matérialise par la présence d'un domaine spatialement superposable au domaine Cterminal des ProRS identiques à celle de *T. thermophilus*, domaine dont la ProRS de *R. palustris* est dépourvue. De plus cette insertion semble servir de socle à l'encrage du domaine d'édition mis en évidence dans le même temps pour la ProRS d'*Enterococcus faecalis* (Figure 3.B). L'utilisation des substrats naturels (L-proline, ATP) et non-conventionnel (L-cystéine) et d'analogues structuraux (ProAMS et CysAMS, analogues non-hydrolysables du prolyl et cystéinyl adénylate) dans des expériences de co-cristallisation a permis de décortiquer les bases atomiques de leur reconnaissance en absence de l'ARNt^{Pro} (Crépin *et al.*, 2006). Les résultats obtenus ont mis en lumière les réarrangements atomiques consécutifs à la liaison des substrats, notamment le repliement d'une boucle dans le site catalytique consécutif à l'interaction de la L-proline (boucle verte sur la Figure 4.C) qui couvre/protège alors l'aminoacyl adénylate et semble obstruer le site actif. Par contre, ces travaux ont montré qu'aucune différence notable n'est observable dans le site catalytique lorsque celui-ci accommode le ProAMS (analogue non hydrolysable du prolyl adénylate) ou le CysAMS (analogue non hydrolysable du cystéinyl adénylate), semblant ainsi suggérer une absence totale de spécificité. Pour compléter cette étude, j'ai entrepris l'étude du complexe avec les différents ARNt^{Pro} isoaccepteurs de cet organisme. Des cristaux ont été obtenus et les structures correspondantes résolues (Figure 4 et Tableau 1). De même que pour l'enzyme seule, l'utilisation du substrat naturel (L-proline) et d'analogues structuraux (ProAMS, CysAMS, AlaAMS et AMPPcP) dans des expériences de co-cristallisation a permis de décortiquer les bases moléculaires de la spécificité pré- et post-transfert. Les résolutions des structures ProRS-ARNt^{Pro} obtenues sont respectivement 1,70 Å avec le ProAMS, 1,8 Å avec le CysAMS, 2,0 Å avec l'AlaAMS et 1,98 Å pour le mélange L-proline + AMPPcP. La comparaison des structures des complexes ProRS:ProAMS en absence et en présence de l'ARNt^{Pro} (Figure 4.C) montre que l'arrivée de l'extrémité CCA-3' de l'ARNt entraine l'ouverture du site catalytique. En effet, l'adénosine terminale (A76) est stabilisée par la Tyr-89, dont le changement de conformation entraine l'ouverture de la boucle de liaison de l'acide aminé. Par contre, la fonction amine de l'aminoacyl adénylate présent dans le site catalytique, apparait conditionner la position de l'A76 et sa stabilisation lors du déroulement de la réaction de transfert (Figure 4.D). Ainsi la fonction amine du cystéinyl et celle de l'alanyl adénylate semblent rejeter l'A76 qui, étant alors moins bien stabilisée dans le site catalytique, serait non susceptible pour le transfert. Ayant une demi-vie (présence) dans le site catalytique ouvert plus importante que le prolyl adénylate, le cystéinyl et l'alanyl adénylate auraient une plus grande susceptibilité à subir un *turn-over* expliquant probablement le phénomène d'édition pré-transfert.

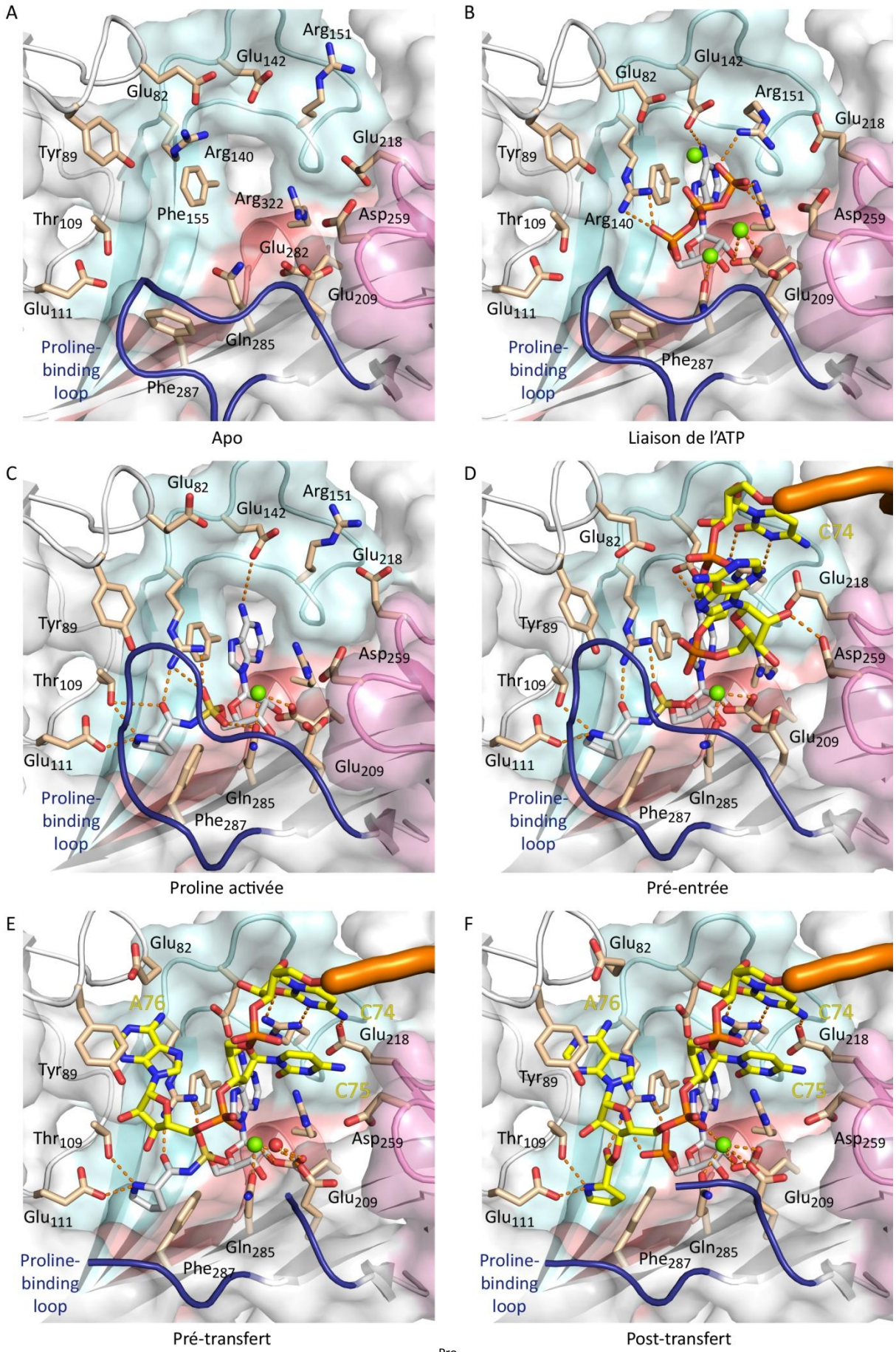


Figure 5: Bases atomiques de la prolylation de l'ARNt^{Pro} par la ProRS de *R. palustris*.

Par contre, lors de ce travail, de façon surprenante, la co-cristallisation du complexe ProRS-ARNt^{Pro} en présence d'un mélange L-proline + AMPPcP (analogue structural de l'ATP, supposé non hydrolysable !) a permis d'obtenir la structure d'un complexe stable entre la ProRS et le Pro-ARNt^{Pro}, le produit final de la réaction de prolylation de l'ARNt^{Pro}. Il semble que cette conformation découle de l'utilisation de l'AMPPcP, la cristallogène simultanée du même complexe mais en présence d'ATP, n'ayant jamais abouti. A noter également que dans les mêmes conditions, il n'a jamais été possible d'obtenir de cristaux en présence d'un mélange L-cystéine + AMPPcP ou L-alanine + AMPPcP. Ce travail permet donc de disposer de la structure cristallographique à haute résolution de l'ensemble des étapes de la réaction d'aminoacylation de la ProRS de *R. palustris*, depuis la reconnaissance des différents substrats de cette réaction enzymatique complexe, jusqu'à l'obtention du produit final de la réaction, stabilisé dans le site catalytique de l'enzyme (Figure 5).

Ayant ainsi travaillé sur l'aspect pré-transfert de l'édition, j'ai voulu travaillé sur son aspect post-transfert. Bien que les précédents travaux m'aient amenés à travailler avec une ProRS possédant un vrai domaine d'édition (i.e. la ProRS d'*E. faecalis*, Figure 3.B), je n'ai pu poursuivre sur ce modèle. J'ai alors été amené à travailler avec autre système enzymatique, la LeuRS d'*E. coli*. Un second objectif du travail sur la LeuRS d'*E. coli* était, en collaboration avec une entreprise américaine (Anacor Pharmaceuticals, Inc actuellement partie intégrante du groupe Pfizer) de l'intégrer dans un processus de conception de molécules à visée thérapeutique ciblant le domaine d'édition (Figure 6.A). Cette entreprise avait développé toute une série de composés basés sur la chimie du bore dont l'un d'entre eux (AN2690) avait prouvé son efficacité dans le traitement de l'onychomycose, l'infection fongique des ongles. J'ai participé au travail initial visant à montrer que l'atome de bore présent dans le AN2690 réagissait avec les fonctions 2'- et 3'-OH de l'Adénosine-76 de l'ARNt, formant une structure spiroborate tétraédrique (Figure 6.B) stabilisée dans le site d'édition de la LeuRS, notamment par les résidus strictement conservés décrits comme étant largement impliqués dans la réaction d'édition post-transfert (Crépin & Cusack, 2007; Rock *et al.*, 2007).

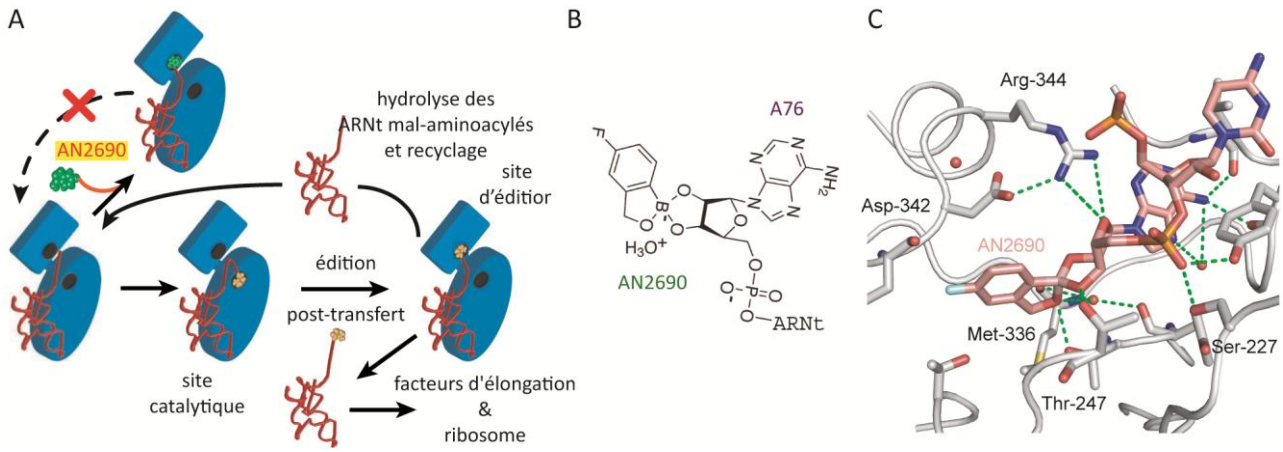


Figure 6: Cibler le domaine d'édition des aaRS pour concevoir des médicaments. (A) Schéma du mode d'action du AN2690. (B) Formule de la structure spiroborate tétraédrique résultant de la réaction de l'atome de bore avec les fonctions 2'- et 3'-OH de l'Adénosine-76 de l'ARNt^{Leu}. (C) détail de la structure du complexe avec le AN2690. La partie ARN est représentée en rose. La protéine est en gris. Figure adaptée de (Crépin & Cusack, 2007; Rock *et al.*, 2007).

La LeuRS est une enzyme modulaire dont l'architecture est globalement plus conservée dans son agencement que la ProRS pour les différents règnes. Les structures préalablement obtenues (*T. thermophilus* et *Pyrococcus horikoshii*) montraient un site d'édition et un site catalytique distant de 40 Å (Fukunaga & Yokoyama, 2005b; Yaremchuk *et al.*, 2000a). En revanche, même si leur position dans la séquence primaire est identique, les domaines d'édition de LeuRS présentent une grande hétérogénéité dans leurs propres et/ou dans leurs repliements. Les structures des LeuRS de *T. thermophilus* et *P. horikoshii* avaient également été résolues en complexe avec l'ARNt^{Leu}, montrant chacune des conformations réactionnelles bien différentes, à savoir respectivement la conformation d'édition (l'extrémité CCA-3' dirigée vers le site d'édition) et la conformation d'aminoacylation (l'extrémité CCA-3' dirigée vers le site catalytique) (Fukunaga & Yokoyama, 2005a). Dans ce contexte, j'ai amorcé l'étude structurale du complexe entre la LeuRS d'*E. coli* et l'ARNt^{Leu,} et résolu la « *structure-zéro* » de ce projet à 2,7 Å de résolution. La structure montre le complexe protéine-ARN avec l'ARNt^{Leu} en configuration d'édition, c'est-à-dire avec son extrémité CCA-3' entrant dans le site d'édition (Figure 7.B). La structure est largement similaire à celle obtenue pour le complexe LeuRS-ARNt^{Leu} de *T. thermophilus*, que ce soit concernant l'homologie de structure du domaine d'édition et aussi donc dans le placement de la molécule d'ARNt. En revanche, le grand intérêt de cette structure est qu'elle concerne une bactérie « commune » par rapport aux deux autres systèmes déjà développés et apparentée à de nombreux agents pathogènes. Elle constitue par conséquent un outil stratégique pour une société pharmaceutique dans la conception de médicaments potentiels. Ainsi, une première molécule développée par Anacor a été utilisée dans des expériences de co-cristallisation en présence de deux partenaires bactériens. La structure du complexe a été résolue, montrant le composé cyclisé par l'atome de bore avec les fonctions 2'- et 3'-OH de l'Adénosine-76, le tout stabilisé dans le site d'édition de la LeuRS. L'utilisation du composé a de plus permis d'améliorer la résolution de la structure du complexe, passant de 2,7 à 2,0 Å de résolution (Figure 6.C). Ayant été validées, l'utilisation des molécules développées par Anacor a été systématisée, permettant leur optimisation grâce à la structure, avec un total d'une vingtaine de structures de complexes obtenue en ce qui me concerne. Ces structures faisant l'objet d'une collaboration avec une entreprise privée, seules certaines d'entre elles ont pu être déposées et utilisées pour publication (Hernandez *et al.*, 2013).

Au cours de ce travail, j'ai été amené à former un chercheur post-doctorant en vue de la pérennité du projet et conjointement, en utilisant le LeuAMS (un analogue structural du leucyl adenylate), la structure du complexe LeuRS:ARNt^{Leu} en configuration d'aminoacylation a été résolue à 2,4 Å de résolution. Elle montre cette fois, l'extrémité CCA-3' de l'ARNt^{Leu} entrant dans le site catalytique de l'enzyme (Figure 7.A). Dans cette structure, le ribose de l'A76 est positionné près du LeuAMS, prêt à être chargé et de matière intéressante, la Tyr-43 (équivalente de la Tyr-15) montre les mêmes réarrangements que ceux observés pour la MetRS d'*E. coli* neuf ans plus tôt (Crépin *et al.*, 2003). Dans leur globalité, ces travaux sur la LeuRS sont les premiers à montrer la translocation de l'extrémité CCA-3' de l'ARNt du site d'édition au site

d'aminoacylation et *vice versa*, sur un même système enzymatique. Ils montrent que la translocation implique des mouvements corrélés des différents domaines flexibles de l'enzyme, dont notamment celui du domaine portant le motif KMSKS essentiel à la réaction d'aminoacylation (Hountondji *et al.*, 1985; Hountondji *et al.*, 1986). De façon inattendue, le domaine d'édition stabilise l'ARNt^{Leu} lorsque l'extrémité CCA-3' de celui-ci se trouve dans le site catalytique pour y être aminoacylé (Figure 7.C). Ces résultats montrent l'importance de la dynamique moléculaire et la synergie intrinsèque entre les macromolécules pour assurer la fidélité du processus traductionnel.

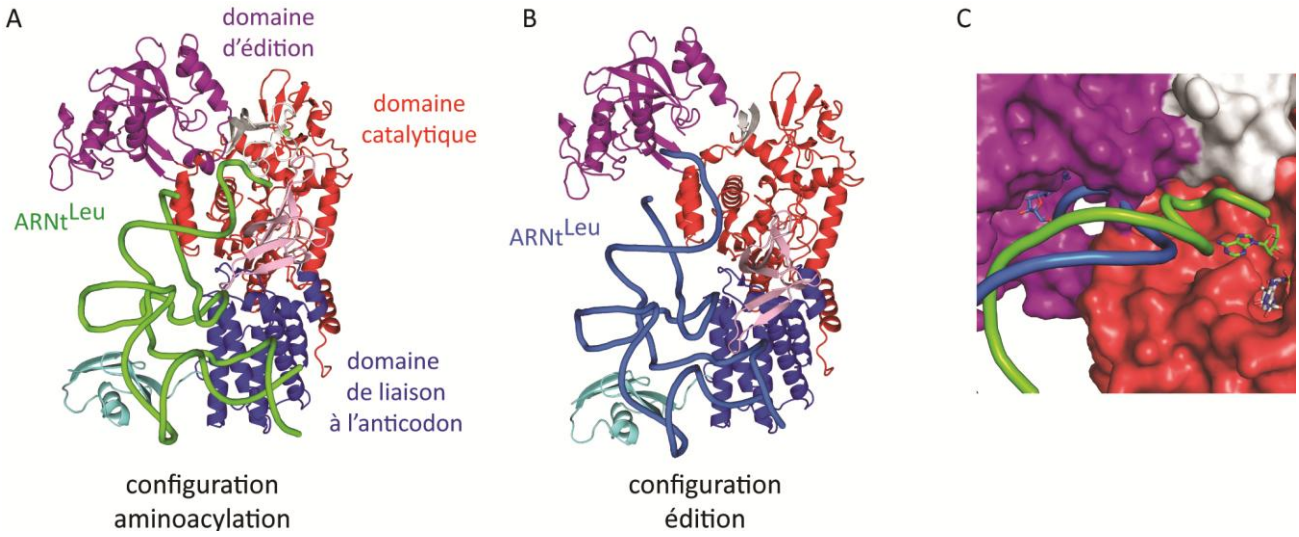


Figure 7: Bases moléculaires de l'édition post-transfert. En utilisant le même d'ARNt^{Leu}, la structure du complexe LeuRS-ARNt^{Leu} d'*Escherichia coli* a pu être déterminée dans deux conformations différentes, respectivement (A) dans une conformation d'aminoacylation et (B) une conformation d'édition. Les deux parties utilisent le même code couleur, respectivement blanc pour le domaine catalytique, 2 nuances de bleu pour les domaines d'interaction avec l'anticodon, rose pour le domaine spécifique à la Leurs d'*E. coli*, en violet le domaine de liaison au zinc et orange pour le domaine d'édition. La partie (C) correspond à un agrandissement de la superposition des deux structures pour lesquelles les protéines sont représentées en surface. Le code couleur est le même que pour les parties (A) et (B). L'ARNt^{Leu} en configuration d'aminoacylation est en vert. Il est représenté en bleu lorsqu'il est en configuration d'édition. Figure adaptée de (Palencia *et al.*, 2012).

Mes travaux sur les aaRSs ont permis, grâce à l'obtention de nombreuses structures cristallographiques à haute résolution, de couvrir l'ensemble du processus d'aminoacylation, depuis la reconnaissance spécifique des substrats jusqu'aux bases moléculaires de l'édition (pré- et post-transfert) et d'apporter ainsi des données importantes. Mes derniers travaux (pour le moment j'espère !) dans le vaste domaine qu'est la traduction, ont ciblé le facteur d'élongation eucaryote eEF1A2 (Crépin *et al.*, 2014; Yaremchuk *et al.*, 2012). L'ensemble de ces travaux m'ont conforté dans mon choix de savoir maitriser l'utilisation de nombreuses macromolécules pour la compréhension des processus biologiques complexes. Mais outre leur aspect très

fondamental, mes travaux sur les aaRSs peuvent être appliqués à la conception rationnelle de molécules à visée thérapeutique.

Dynamique cellulaire de la réplication virale

Mon recrutement par le CNRS en 2007 s'est accompagné d'une mobilité thématique, me faisant ainsi passer des protéines impliquées dans la traduction aux complexes macromoléculaires engagés dans la réplication virale. Un virus est un agent infectieux nécessitant un hôte, dont il utilise le métabolisme et ses constituants pour se répliquer. Un virus existe sous une forme extra-cellulaire (entité matérielle indépendante appelée virion) ou intra-cellulaire. Sous cette dernière, le virus est un élément génétique pouvant se répliquer indépendamment du matériel génétique cellulaire, mais non indépendamment de la cellule hôte. Le matériel génétique (génome viral) peut-être composé d'ADN ou d'ARN, circulaire ou linéaire, bicaténaire ou monocaténaire. La classification dite de « *Baltimore* », est basée sur le type d'acide nucléique formant le génome des virus et son mode d'expression (Baltimore, 1971). Le groupe V de cette classification correspond aux virus à ARN simple brin, de polarité négative, regroupant de nombreux pathogènes animaux/humains. Dans ce groupe, on distingue les virus dont le génome est segmenté (ARNss ; virus de la grippe, virus Thogoto, virus de la fièvre hémorragique de Crimée-Congo, …) et ceux dont le génome est non segmenté (ARNns : virus de la rage, virus de la rougeole, virus Ebola, …). Leur génome doit impérativement être recouvert de nucléoprotéines pour être « réplicable ». Le complexe entre l'ARN viral (ARNv) et la nucléoprotéine (NP pour les ARNss et N pour les ARNns) est appelé nucléocapside. Du fait de sa polarité, l'ARNv est complémentaire de l'ARNm. L'étape de synthèse des ARNm viraux à partir de l'ARNv s'appelle la transcription. L'étape d'amplification de l'ARNv, au cours de laquelle il y a synthèse d'un ARN de polarité positive (ARNc) servant de matrice à la synthèse de l'ARNv, est appelée réplication. Ces deux étapes sont effectuées par une ARN polymérase ARN-dépendante codée par le génome viral. Encore une fois, il existe une différence entre les virus à ARNss et ceux à ARNns. Dans le cas des virus à ARNss, l'ARN polymérase reconnait et se fixe spécifiquement sur les extrémités 5' et 3' formant alors une ribonucléoprotéine (RNP). L'ARN polymérase virale va alors directement transcrire et répliquer l'ARNv sur lequel elle est accrochée. Dans le cas des virus à ARNns, l'enzyme ne se fixe pas sur la matrice et a besoin d'un cofacteur, la phosphoprotéine (P), pour effectuer les étapes de transcription et de réplication. Le projet pour lequel j'ai été recruté, visait à détailler le fonctionnement de la machinerie réplicative du virus de la grippe. Cette partie de mes travaux va être détaillée ci-dessous. Cependant, en parallèle de ces travaux, j'ai également participé à l'étude des domaines d'oligomérisation de la P de deux virus à ARNns, respectivement les virus de la rage et de la rougeole (Communie *et al.*, 2013; Ivanov *et al.*, 2010). La figure 8 résume l'ensemble des travaux réalisés.

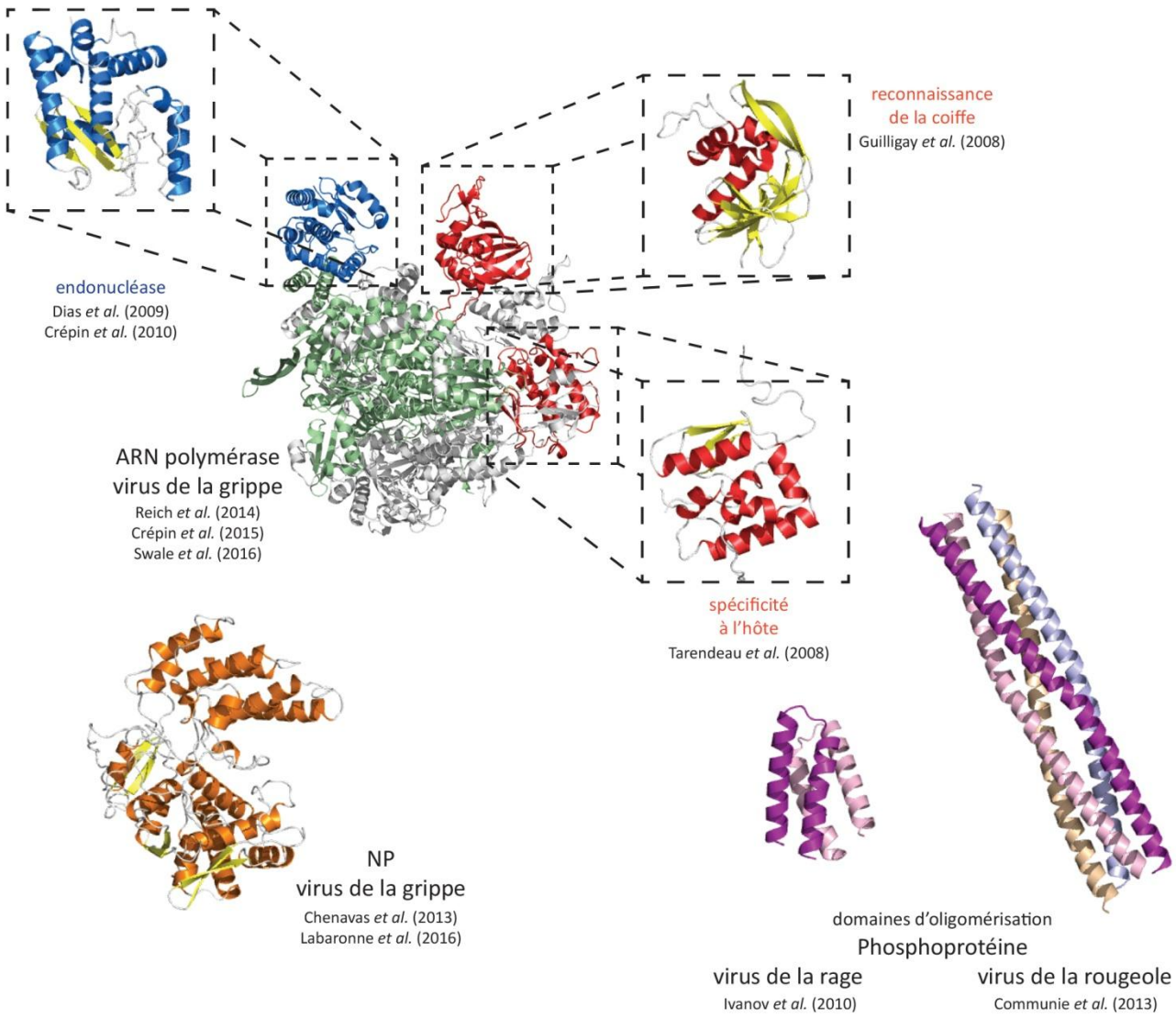


Figure 8: Structures cristollographiques des différentes protéines des machineries réplicatives virales étudiées.

La machinerie réplicative du virus de la grippe

Le virus de la grippe est un des agents pathogènes majeurs actuellement en circulation dans le monde et du fait de son lourd passif, il fait l'objet d'une attention toute particulière de la part de l'OMS. Le virus de la grippe est un virus enveloppé à ARNss. Son génome code au total pour une dizaine de protéines qui vont *in fine*, former les nouvelles particules virales. Il existe différents types de virus grippaux. Les virus de type A ont un large tropisme d'hôtes (oiseaux, mammifères dont l'homme) alors que ceux de type B sont presque exclusivement humains. Le génome de ces deux types est formé de huit ribonucléoprotéines (ARNv couvert de NP avec une ARN polymérase interagissant avec les extrémités 5' et 3'). Les particules virales de type C et D, infectant respectivement l'homme et les ruminants, contiennent quant à elles sept RNPs. Les RNPs constituent les entités fonctionnelles autonomes des virus grippaux. L'une de leurs particularités est que la réplication et la transcription (Figure 9) se déroulent dans le noyau de la cellule infectée. Cette particularité implique par conséquent un important « va-et-vient » de matériel biologique viral entre le cytoplasme et le

noyau: (i) import nucléaire des RNPs consécutif à l'entrée du virus, (ii) export des ARNm viraux après leur transcription, (iii) import des protéines virales néosynthétisées pour former les nouvelles RNPs et (iv) leur export pour reconstituer les nouveaux virions. Une seconde particularité des RNPs est que l'ARN polymérase correspond un assemblage hétérotrimérique (PA/P3, PB1 et PB2) formant un complexe macromoléculaire d'environ 240 kDa. La dernière spécificité concerne le fonctionnement de ce complexe et notamment l'initiation de la phase transcriptionnelle (synthèse d'ARNm). Dans toute cellule eucaryote, le recrutement de la particule ribosomale nécessite un signal moléculaire particulier à l'extrémité 5' de l'ARNm (Asano *et al.*, 2000; Banerjee, 1980). Ce signal, appelé coiffe, correspond à un 7-méthyl-guanosine (^{7m}G) lié par une liaison 5'-5' triphosphate au premier nucléotide du messager. A noter que différentes structures de coiffe existent, suivant que des méthylations additionnelles soient opérées sur les riboses des nucléotides N1 et/ou N2 de l'ARN. Alors que les ARN polymérases des virus à ARNns possèdent toutes les activités enzymatiques assurant une synthèse endogène de la coiffe sur les ARNm qu'elles synthétisent, celles de nombreux virus à ARNss sont totalement dépourvues de cette activité. Les ARNm sont alors synthétisés par un mécanisme particulier appelé vol de coiffe (Plotch *et al.*, 1981) (Figure 9.B) propre aux virus grippaux, aux *Bunyaviridae* (virus de La Crosse, virus Crimée-Congo, …) ou encore aux *Arenaviridae* (virus Lassa) (Jin & Elliott, 1993). Dans le cas du virus de la grippe, l'ARN polymérase virale reconnaît spécifiquement la coiffe située à l'extrémité 5' des pré-ARNm cellulaires puis la clive grâce à une activité endonucléase, générant alors une amorce de 10-13 nucléotides coiffée qui va servir à initier la transcription des messagers viraux. Ainsi coiffés, les messagers viraux sont reconnus comme des ARNm endogènes par la machinerie cellulaire, exportés du noyau et traduits en protéines.

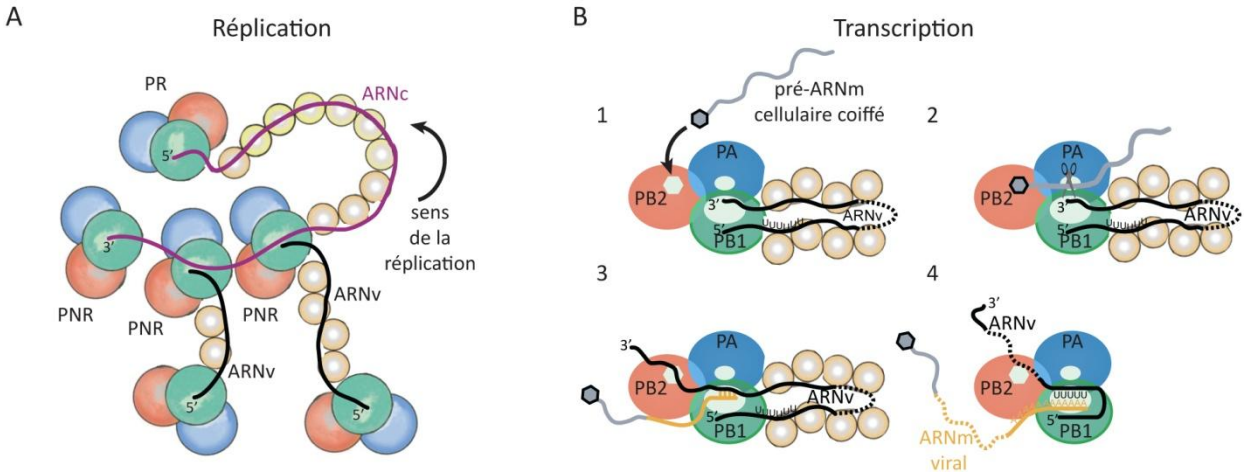


Figure 9: Principes généraux du fonctionnement de l'ARN polymérase du virus de la grippe. (A) La réplication en trans. Lors d'une première phase, l'ARNc (violet) généré par duplication de l'ARNv est encapsidé par une ARN polymérase dite « résidente » (PR) et des protomères de NP. Son extrémité 3' serait accessible à des complexes hétérotimériques « non-résidents » (PNR) produisant les ARNv (noir) qui, à leur tour encapsidés, formeraient les RNP amplifiées. (B) La transcription coiffe-dépendante. (1) Le pré-ARNm cellulaire (gris) est reconnu par le site de liaison à la coiffe porté par PB2 puis (2) coupé 10-13 nucléotides après la coiffe par le domaine endonucléase porté par PA. (3) Le fragment d'ARN cellulaire coiffé sert alors d'amorces pour initier la transcription des ARNm viraux qui, (4) sont polyadénylés par bégaiement de l'enzyme. Figure adaptée de (Chenavas *et al.*, 2013b).

Du fait de leur complexité de séquence, de repliement et/ou d'assemblage, les ARN polymérases des virus à ARN négatif ont pendant longtemps semblé inaccessibles pour la biologie structurale. Ce n'est que très récemment, grâce aux avancées technologiques, que ce verrou a sauté (Chang *et al.*, 2015; Gerlach *et al.*, 2015; Liang *et al.*, 2015; Pflug *et al.*, 2014; Reich *et al.*, 2014). J'ai grandement participé au projet ciblant l'ARN polymérase grippale au cours de deux contrats européens (FluPol et FLUPHARM) du fait de mon appartenance au groupe du Pr. Rob Ruigrok (UVHCI-UMI3265 puis IBS-UMR5075 depuis le 1er janvier 2016). La stratégie employée a consisté, après de nombreux essais pour produire le complexe hétérotrimérique que constitue l'ARN polymérase du virus de la grippe dans son intégralité, à le décortiquer en domaines isolés, solubles et fonctionnels. Cette stratégie a payé dans un premier temps. Différents domaines ont été caractérisés, chronologiquement pour PB2 (i) le domaine Cterminal ($PB2_{Cter}$) impliqué dans son transport au noyau (Tarendeau *et al.*, 2007), (ii) le domaine de liaison à la coiffe ($PB2_{Cap}$) (Guilligay *et al.*, 2008) et (iii) son domaine de spécificité à l'hôte ($PB2_{627}$) (Tarendeau et al, 2008), pour PA (i) son grand domaine Cterminal (PA_{Cter}) assurant l'interaction avec PB1 (He et al, 2008; Obayashi et al, 2008) et son domaine Nterminal (PA_{Nter}) portant l'activité endonucléase (Dias et al, 2009a; Yuan et al, 2009). Pour PB1, la sous-unité portant l'activité de synthèse d'ARN, seule une courte zone de contact avec PB2 a pu être mise en évidence (Sugiyama et al, 2009). Ce découpage a été réalisé par différents laboratoires à travers le monde. Pour ma part, j'ai été impliqué dans la caractérisation des domaines $PB2_{Cap}$, $PB2_{627}$ et PA_{Nter} (Crépin et al, 2010; Dias et al, 2009a; Guilligay et al, 2008; Tarendeau et al, 2008). A titre d'exemple pour ce travail, l'obtention de la structure cristallographique du domaine PA_{Nter} a permis de mettre en évidence une similarité structurale du domaine avec le site catalytique de certaines enzymes de restriction (EcoRV par exemple). La combinaison de l'approche structurale et du travail fonctionnel a notamment mis en évidence une dépendance de l'activité nucléase du domaine vis-à-vis du manganèse, dépendance qui a été décortiquée par la suite (Figure 10.A, 10.B et 10.C) (Crépin *et al.*, 2010; Dias *et al.*, 2009a). Les ions manganèse sont principalement présents dans les mitochondries et dans le noyau des cellules (Maynard & Cotzias, 1955; Sakurai *et al.*, 1985). Cependant la distribution des ions métalliques dans les cellules infectées par le virus de la grippe était non documentée et donc aucun lien ne pouvait être établi entre l'activité nucléase manganèse-dépendante de PA_{Nter} et les cellules cibles du virus. Pour étudier cet aspect, j'ai proposé un projet sur une ligne de microscopie rayons-X pour examiner la distribution cellulaire des ions métalliques durant un cycle d'infection par un virus grippal. L'expérience réalisée sur la ligne ID21 de l'ESRF, a permis de montrer que dans des cellules de référence utilisées pour les travaux en infectiologie avec les virus grippaux (cellules épithéliales alvéolaires humaines A549), les ions manganèse étaient principalement localisés dans le noyau (Figure 10.D) avec une modification de leur distribution au cours du cycle infectieux. Ces expériences permettent ainsi de faire un lien entre l'aspect moléculaire de la réplication virale et son aspect cellulaire mais elles doivent cependant encore être compléter par une analyse quantitative, avant leur prochaine

publication. A noter aussi que parallèlement au travail sur le virus de la grippe, la même approche de découpage a été réalisée sur les sous-unités de l'ARN polymérase du virus Thogoto, un virus de la même famille que le virus de la grippe (i.e famille des *Orthomyxoviridae*), infectant les tiques et potentiellement transmissible à l'homme. Ces travaux structure/fonction ont montré que même si les domaines étaient structuralement conservés, leurs activités associées au mécanisme de vol de coiffe étaient par contre inexistantes, laissant suggérer un mécanisme d'initiation de la transcription différent pour ce virus, qui reste encore à étudier (Guilligay *et al.*, 2014).

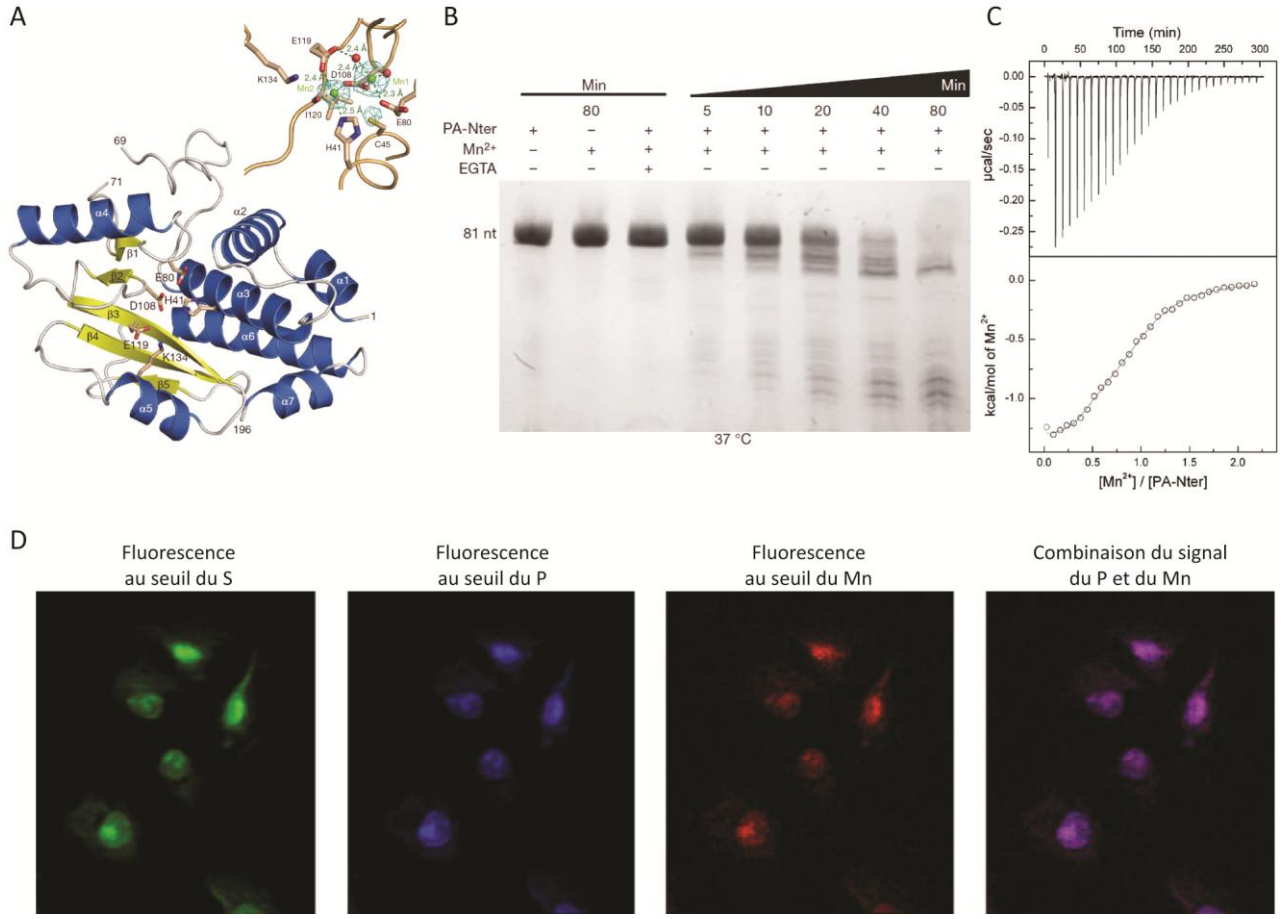


Figure 10: Caractérisation du domaine endonucléase isolé PA_{Nter}. (A) Structure du domaine avec les hélices-α représentées en bleu et les brins-β en jaune. L'encart supérieur montre la densité électronique des atomes de manganèse (sphères vertes) dans le site actif du domaine endonucléase. La carte de densité électronique *fo-fc* à partir de données de diffraction anomales collectées au seuil du manganèse est contourée à 3σ. (B) Activité nucléase de PA_{Nter}. (C) Titration par microcalorimétrie de l'affinité de PA_{Nter} pour le manganèse. (D) Cartographie du signal de fluorescence des rayons-X de différents éléments de cellules épithéliales alvéolaires humaines (A549). Figure adaptée de (Crépin *et al.*, 2010; Dias *et al.*, 2009a; Dias *et al.*, 2009b).

La stratégie consistant à décortiquer l'ARN polymérase en domaines solubles et fonctionnels a parfaitement fonctionné pour les domaines périphériques et aux vues de la complexité de l'enchevêtrement des sous-unités observée grâce aux récentes structures du complexe dans sa totalité, cela est largement explicable. En particulier, l'obtention d'informations structurale sur PB1, la sous-unité portant l'activité ARN-

polymérasique, n'a pu être possible que dans le cadre des complexes (Pflug *et al.*, 2014; Reich *et al.*, 2014; Swale *et al.*, 2016). Les cinq années écoulées entre la dernière structure de domaine isolé et la première structure du complexe dans sa totalité, correspond à la mise au point d'une stratégie d'expression innovante, stratégie à laquelle j'ai grandement participé avec deux de mes étudiants. Elle consistait à exprimer le complexe sous la forme d'une protéine de fusion dans laquelle les sous-unités sont séparées les unes des autres par des sites reconnus par une protéase, qui est conjointement exprimée par le système (cellules d'insecte). Ainsi produites en quantités exactement stœchiométriques, les sous-unités du complexe sont sensées s'assembler et former un complexe fonctionnel (Crépin *et al.*, 2015; Reich *et al.*, 2014; Swale *et al.*, 2016). L'ARN polymérase du virus de la grippe de type A humain s'est montré très difficile à produire et notamment PB2 qui est apparu comme le « facteur » limitant dans ce cas. Disposant néanmoins d'un large panel de constructions différentes de l'ARN polymérase du virus de la grippe de type A humain, leur caractérisation structurale et fonctionnelle a été réalisée, notamment l'interaction avec l'ARNv. Ainsi l'hétérodimère PA-PB1 est capable de reconnaitre spécifiquement l'extrémité 5' de l'ARNv avec une très forte affinité (k_d de 0.2 nM) alors que la reconnaissance de l'extrémité 3' requiert la présence de la sous-unité PB2. Une construction correspondant à l'hétérodimère PA-PB1$_{1\text{-}686}$ a cristallisé. Les cristaux se sont révélés très délicats et malgré de nombreuses tentatives, la diffraction n'a jamais excédé 8 Å de résolution (Swale *et al.*, 2016). Le projet sur l'étude structurale de l'ARN polymérase a été mené *conjointement* avec le groupe du Dr Stephen Cusack (EMBL). La stratégie d'expression appliquée au complexe enzymatique du virus de la grippe de type B humain et celui de chauve-souris de type A a été nettement plus prolifique, aboutissant *in fine* aux structures cristallographiques précédemment mentionnées (Pflug *et al.*, 2014; Reich *et al.*, 2014). La comparaison de ces deux structures, notamment dans le placement unique du domaine endonucléase et des différentes conformations que peut adopter le domaine de liaison à la coiffe, permet d'imaginer la dynamique qui semble régir le mécanisme de vol de coiffe (Reich *et al.*, 2014).

Le second partenaire viral impliqué dans cet assemblage macromoléculaire complexe qu'est la RNP, est la nucléoprotéine (NP). Son rôle principal consiste à couvrir et protéger l'ARNv, donnant à la RNP une supra-structure particulière en « collier de perles torsadé ». Son rôle n'est pas qu'architecturale. NP est essentielle pour transcrire et répliquer l'ARNv. Lors du cycle viral, elle jouerait également le rôle d'interrupteur, permettant à l'ARN polymérase de basculer du mode transcriptionnel au mode réplicatif (Medcalf *et al.*, 1999; Portela & Digard, 2002). La première structure de NP grippale a été résolue en 2006 (Ye *et al.*, 2006). Celle-ci correspondait à la structure d'une forme trimérique sans ARN dans laquelle la boucle d'oligomérisation d'un protomère vient interagir dans un sillon d'un second et ainsi de suite pour former le trimère. Depuis, d'autres structures de NP grippales ont été obtenues (Figure 11.A), toutes en absence d'ARN (Ng *et al.*, 2012; Ng *et al.*, 2008). Les structures résolues ne répondent pas à la question de l'interaction de NP avec l'ARN.

Avec une post-doctorante, j'ai travaillé à la caractérisation du mutant strictement monomérique NP-R416A du virus de la grippe de type A. Il a été cristallisé et sa structure résolue à 2,7 Å (Figure 11.B). Cette structure montre dans ce cas, que la boucle d'oligomérisation est repliée sur son propre corps. A partir de cela, nous nous sommes attelés à purifier une forme monomérique sauvage stable de NP, chose qui n'avait pas été faite par le passé. Ayant établi les conditions nous avons essayé de reconstituer des formes oligomériques stables du complexe NP:ARN. Nous avons ainsi montré que seul le monomère permet d'obtenir des formes complexes en anneaux et/ou en particules de type RNP alors que si l'on purifie NP sous une forme tri/tétramérique, ces supra-structures NP-ARN ne peuvent pas être obtenue (Chenavas *et al.*, 2013a). Ces travaux ont depuis été repris par une étudiante de thèse. Une caractérisation poussée de l'interaction protéine:ARN impliquant les NP des virus de type A et B a été réalisée (Labaronne *et al.*, 2016). Ella a permis de faire le lien entre l'interaction NP:ARN et son processus d'auto-oligomérisation (Figure 11.C). A partir des conditions établies, des expériences de cryo-microscopie électroniques vont être menées prochainement avec l'objectif d'obtenir la structure du complexe NP:ARN et ainsi progresser dans la compréhension de l'assemblage de la machinerie réplicative du virus de la grippe. A noter que la suite du manuscrit comprend différents manuscrits qui viennent d'être soumis pour publication, dont deux portant sur NP. L'un, associé au transport nucléaire, concerne l'étude structure/fonction du complexe entre NP du virus de la grippe de type B et l'importine-α. Le second décrit la structure de NP du virus de la grippe de type D, un nouveau virus grippal, affectant les bovins et les ovins. Les résultats obtenus montrent également le rôle joué par la partie non repliée Cterminale, dans l'affinité protéine:ARN.

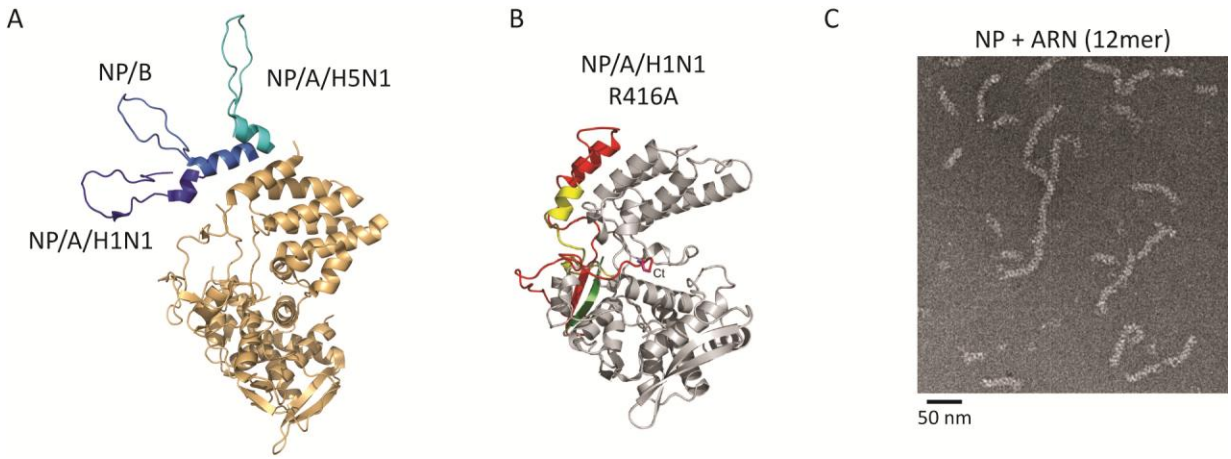


Figure 11: dynamique moléculaire de la nucléoprotéine du virus de la grippe. (A) Superposition des structures de NP préalablement résolues, montrant les différentes conformations (différentes nuances de bleu) que peut adopter la boucle d'oligomérisation par rapport au corps de la protéine (orange), lors de la formation du complexe NP:NP. (B) Structure du mutant monomérique NP-R416A montrant la boucle d'oligomérisation (rouge et jaune) collée sur le corps (gris) de NP. (D) Cliché de microscopie électronique (coloration négative) montrant les structures supramoléculaires que peut adopter le complexe NP:ARN reconstitué à partir des 2 protagonistes. Figure adaptée de (Chenavas *et al.*, 2013a; Chenavas *et al.*, 2013b; Labaronne *et al.*, 2016).

Le transport nucléaire et assemblage du complexe réplicatif

Comme mentionné précédemment, le virus de la grippe présente la particularité d'avoir une phase transcriptionnelle/réplicative nucléaire. Cela implique qu'à l'instar de toutes les protéines de l'hôte ayant une fonction nucléaire, les protéines virales impliquées dans ce(s) processus, une fois synthétisées dans le cytoplasme, doivent être importées dans le noyau. De manière générale, ces mécanismes d'import/export nécessitent des signaux spécifiques (signal de localisation nucléaire, NLS ou signal d'export nucléaire NES). Ces signaux, présents sur les protéines considérées, permettent l'interaction spécifique avec les complexes de transport nucléaire puis avec le pore nucléaire (Gorlich *et al.*, 1996; Gorlich & Mattaj, 1996; Kalderon *et al.*, 1984a; Kalderon *et al.*, 1984b; Kosugi *et al.*, 2009a; Kosugi *et al.*, 2009b; Pemberton & Paschal, 2005; Robbins *et al.*, 1991). Dans le cas des protéines formant les RNPs du virus de la grippe, il a été montré que NP et la sous-unité PB2 utilisent la voie des importines-α (Melen *et al.*, 2003; Naito *et al.*, 2007; Portela & Digard, 2002; Tarendeau *et al.*, 2007; Weber *et al.*, 1998) alors que l'import des 2 autres sous-unités de l'ARN polymérase virale se ferait conjointement sous forme d'un hétérodimère PA-PB1, médié spécifiquement par un interaction avec le facteur d'import nucléaire RanBP5 (Deng *et al.*, 2006; Deng *et al.*, 2005; Fodor & Smith, 2004; Hutchinson *et al.*, 2011) (Figure 12). Contrôler ces interactions hôte-pathogène constitue aussi une piste pour bloquer la prolifération virale.

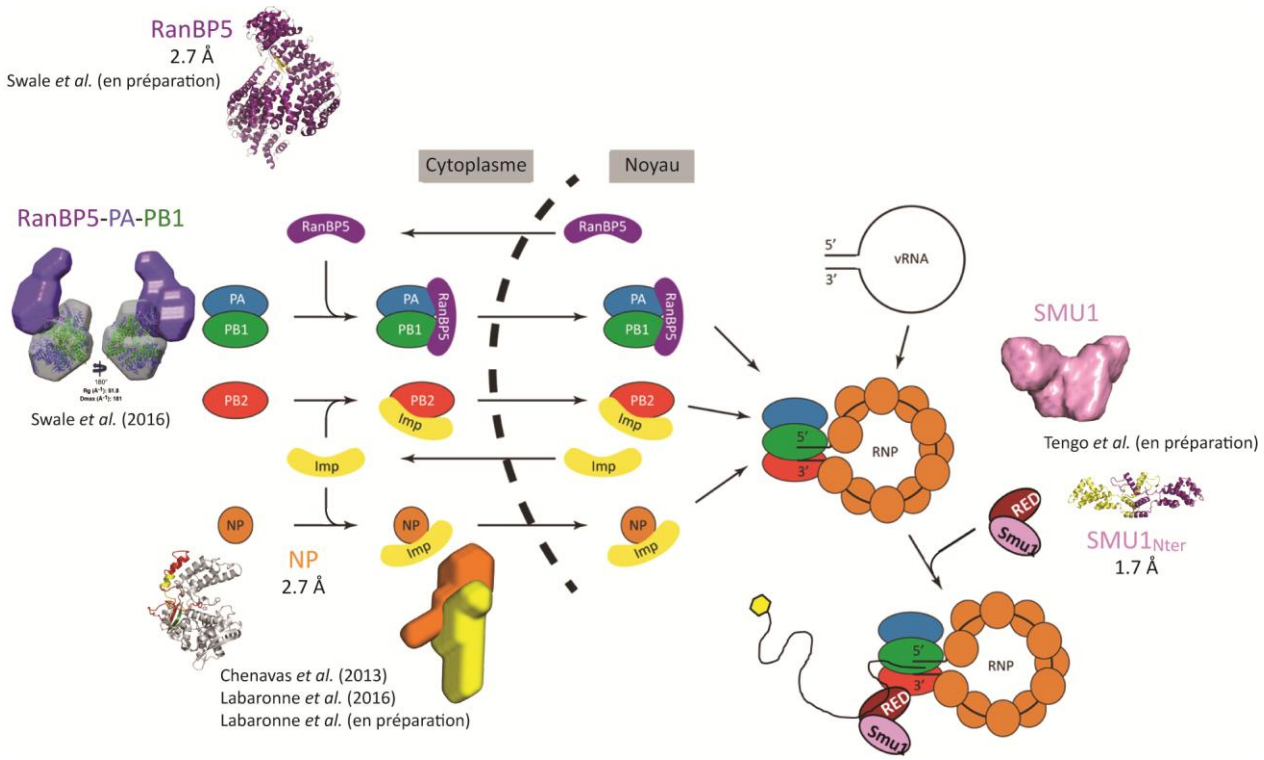


Figure 12: Devenir des partenaires néosynthétisés de la RNP virale. Cette figure schématise les connaissances sur l'import nucléaire des constituants de la RNP. Les données structurales obtenues récemment y ont été ajoutées.

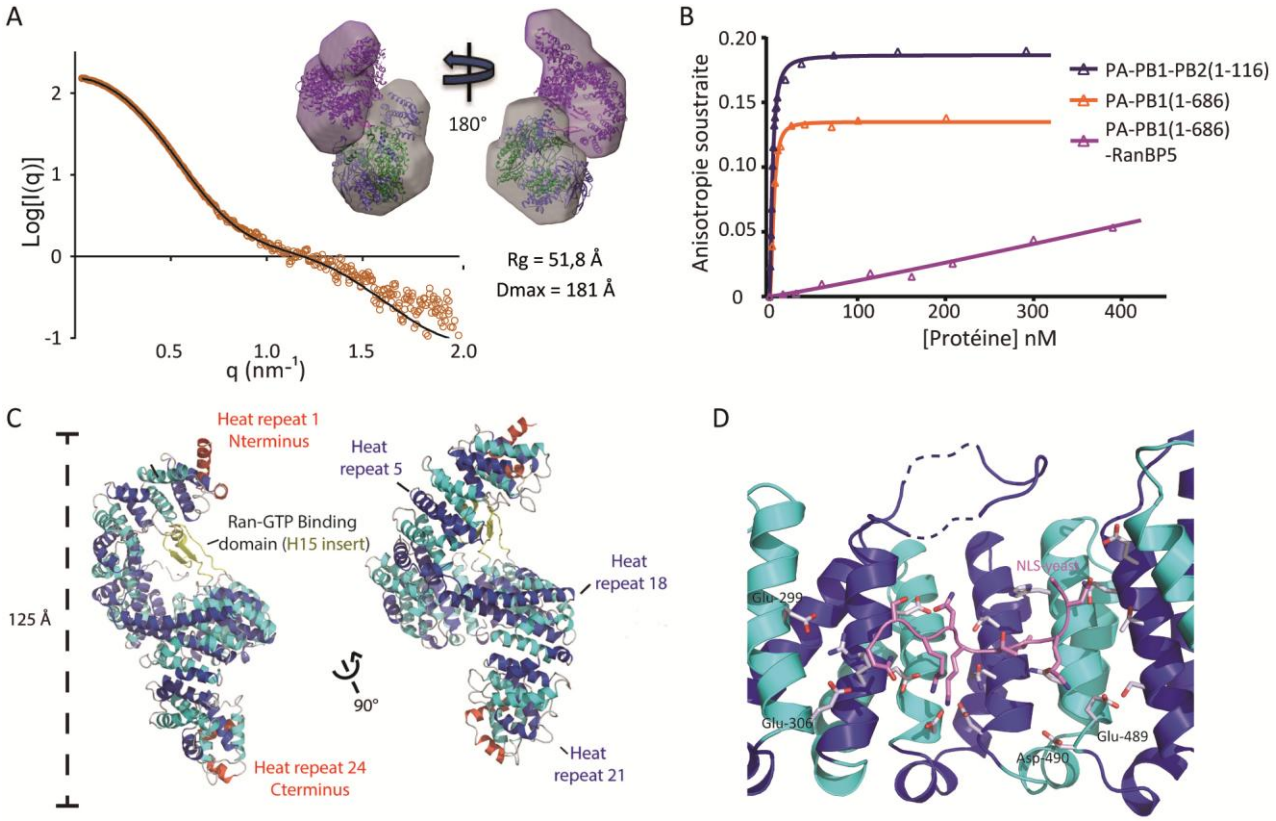


Figure 13: Le facteur d'import nucléaire RanBP5. (A) Courbe de diffusion des rayons X aux petits angles (SAXS) mesurée pour l'hétérotrimère PA-PB1-RanBP5à partir de laquelle une enveloppe moyenne a été obtenue et dans laquelle le programme MONSA (Svergun, 1999) a positionné les coordonnées des modèles cristallographiques de PA-PB1 (bleu et vert) et de Kap121 (violet), homologue de RanBP5 chez la levure. (B) Courbes d'anisotropie de fluorescence obtenues lors de l'interaction de différents complexes avec une molécule d'ARN avec mimant l'ARNv-5' du virus de la grippe. (C) Structure cristallographique de RanBP5 obtenue à 2,5 Å de résolution. (D) Site potentiel d'interaction avec la NLS (rose) identifié en superposant la structure de Kap121 obtenue avec une NLS (Kobayashi & Matsuura, 2013).

L'interaction entre la sous-unité PB2 et l'importine-α5 a déjà été caractérisée par le passé (Tarendeau *et al.*, 2007). La structure cristallographique montre que la séquence NLS portée par le domaine Cterminal de PB2 présente différentes conformations lorsqu'elle interagit ou non, avec la protéine d'import cellulaire. Comme mentionné précédemment, une étude comparable a été effectuée concernant l'interaction entre NP et les importines-α, étude réalisée par RMN (collaboration avec le groupe du Dr Martin Blackledge ; IBS) et par SAXS. Elle fait l'objet d'un manuscrit qui vient d'être soumis. Dans le cas de l'implication du facteur d'import nucléaire RanBP5 dans le processus d'assemblage de la machinerie réplicative virale, avec un de mes étudiants de thèse, j'ai démarré la caractérisation de la karyophérine humaine (Swale *et al.*, 2016). En effet, lors de la mise au point de la stratégie pour la production de l'hétérotrimère formant l'ARN polymérase du virus de la grippe, la séquence codante pour RanBP5 a été ajoutée à la polyprotéine. Il s'est

avéré que seule la co-expression du facteur d'import RanBP5 et de l'hétérodimère PA-PB1 permettait de reconstituer un complexe macromoléculaire homogène et stable. Il a été caractérisé structuralement par SAXS (Figure 13.A). De plus, nous avons montré que l'interaction avec RanBP5 limite la capacité de PA-PB1 à reconnaitre spécifiquement l'ARNv (Figure 13.B). Des cristaux de RanBP5 ont également été obtenus et la structure a pu être résolue à 2,5 Å de résolution (Figure 13.C). L'analyse de la structure et sa comparaison avec celle de son homologue chez la levure (Kobayashi & Matsuura, 2013) a permis d'identifier le site potentiel d'interaction avec la NLS (Figure 13.D), dont la caractérisation est en cours en collaboration avec le groupe du Dr Bernard Delmas (INRA, Jouy-en-Josas). Ce travail constitue un objectif important du projet RNAP-IAV (ANR-14-CE09-0017) pour lequel j'ai obtenu un financement auprès de l'ANR en 2014.

La maturation des messagers viraux

Les RNPs du virus de la grippe répliquent le génome viral et le transcrit en ARNm dans le noyau de la cellule infectée. La synthèse des messagers viraux s'effectue en étroite collaboration avec la machinerie transcriptionnelle de l'hôte. Seules NS2 (protéine non structurale 2) et M2 (protéine transmembranaire formant le canal à protons), deux protéines essentielles à la multiplication virale, sont produites à la suite d'un processus d'épissage des ARNm viraux, à partir de ceux codant respectivement pour NS1 (protéine non structurale 1) et M1 (protéine de matrice). Très peu de choses sont connues sur cet épissage. L'équipe du Dr Nadia Naffakh (Institut Pasteur, Paris) a récemment identifié RED et SMU1, deux facteurs du splicéosome humain, semblant contrôler l'expression de NS2 (Fournier *et al.*, 2014). Dans des cellules infectées par le virus de la grippe, l'ARN polymérase virale recruterait un complexe formé par RED et SMU1, par des interactions avec les sous-unités PB2 et PB1 (Figure 14.A). Dans des cellules dépourvues de RED et/ou SMU1, l'épissage de l'ARNm viral NS1 est affecté, conduisant à une diminution de la production de l'ARNm épissé NS2, une quantité de protéine NS2 réduite, une altération du transport vers le cytoplasme des RNPs nouvellement synthétisées, et une forte réduction de la production de virions infectieux (Fournier *et al.*, 2014). Ainsi, les facteurs d'épissage RED et SMU1 sembleraient agir conjointement pour réguler l'expression de gènes clés du virus de la grippe. J'ai donc établi une collaboration avec l'équipe du Dr Nadia Naffakh afin d'apporter une vision structurale à ce processus d'épissage, avec pour objectif de documenter cette interaction hôte-pathogène impliquant la machinerie réplicative du virus de la grippe. Ce travail pourrait également ouvrir une nouvelle voie pour inhiber la prolifération virale.

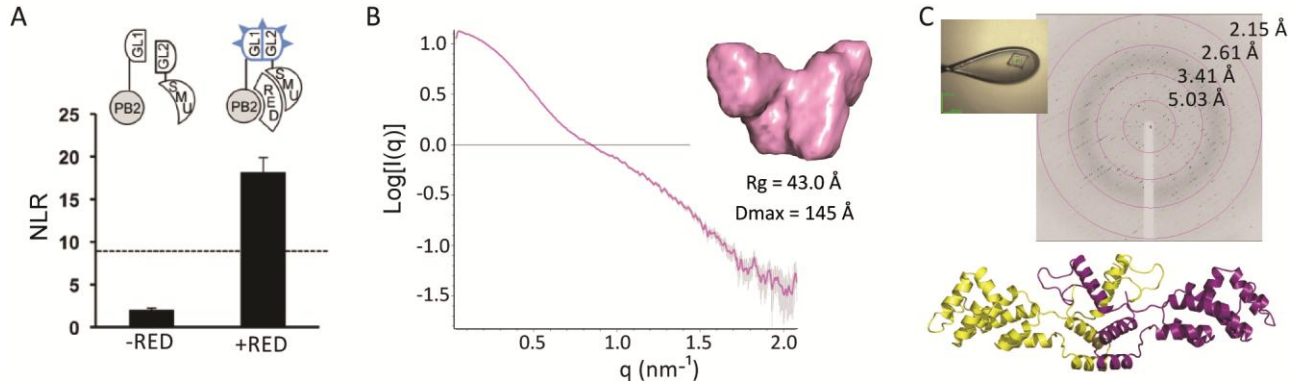


Figure 14: Etude du complexe cellulaire d'épissage RED-SMU1. (A) Détection du complexe ternaire PB2-RED-SMU1 par criblage double hybride en levure (Fournier *et al.*, 2014). (B) Courbe de SAXS mesurée pour SMU1 passé sur une colonne de chromatographie d'exclusion de taille. L'interprétation de ces données donne les valeurs de Rg et Dmax à partir desquelles une enveloppe moyenne a été calculée. (C) Cristallisation du domaine Nterminal de SMU1 dont la structure a pu être résolue à 1,9 Å de résolution (R_{fac} = 17,8 %, R_{free} = 20,7 %).

Le projet a été amorcé, avec un certain nombre de constructions réalisées. Les formes entières de SMU1 et RED ne peuvent être purifiées qu'une fois produites en cellules d'insecte. SMU1 forme alors un dimère stable en solution. Par contre, RED n'est soluble qu'en présence de SMU1, avec une stœchiométrie RED-SMU1 de 1:2. Des tentatives de cristallisation ont été menées. Les échantillons ont été également caractérisés par SAXS (Figure 4.B), permettant d'apporter des renseignements sur ces deux protéines, dans leur globalité. Pour compenser le manque d'information structurale à haute résolution, l'identification de domaines isolés a été entreprise. Elle a montré que le domaine Nterminal de SMU1 ($SMU1_{Nter}$) pouvait être produit cette fois, en bactérie et qu'il interagissait avec le domaine Nterminal de RED (RED_{Nter}). $SMU1_{Nter}$ a été cristallisé et sa structure résolue à 1,9 Å de résolution (Figure 4.C). Des cristaux du complexe RED_{Nter}-$SMU1_{Nter}$ ont été obtenus. Un jeu de diffraction des rayons X a pu être collecté à 3,0 Å et la résolution de la structure est en cours.

Ce projet fait actuellement l'objet d'une demande de financement à l'ANR (coordinatrice, Nadia Naffakh). Ce projet collaboratif vise à documenter du point structural et fonctionnel l'interaction entre le complexe cellulaire RED-SMU1 et l'ARN polymérase du virus de la grippe, la reconnaissance spécifique de l'ARNm viral NS2 et le mécanisme d'épissage. Ce projet comprend également une partie visant à concevoir des molécules à visée thérapeutique ciblant ce processus.

Références

Ambrogelly A, Kamtekar S, Stathopoulos C, Kennedy D, Soll D (2005) Asymmetric behavior of archaeal prolyl-tRNA synthetase. *FEBS Lett* **579:** 6017-6022

Asano K, Clayton J, Shalev A, Hinnebusch AG (2000) A multifactor complex of eukaryotic initiation factors, eIF1, eIF2, eIF3, eIF5, and initiator $tRNA^{Met}$ is an important translation initiation intermediate *in vivo*. *Genes Dev* **14:** 2534-2546

Baldwin AN, Berg P (1966) Transfer ribonucleic acid-induced hydrolysis of valyladenylate bound to isoleucyl ribonucleic acid synthetase. *J Biol Chem* **241:** 839-845

Baltimore D (1971) Expression of animal virus genomes. *Bacteriological reviews* **35:** 235-241

Banerjee AK (1980) 5'-terminal cap structure in eucaryotic messenger ribonucleic acids. *Microbiol Rev* **44:** 175-205

Beuning PJ, Musier-Forsyth K (2001) Species-specific differences in amino acid editing by class II prolyl-tRNA synthetase. *J Biol Chem* **276:** 30779-30785

Chang S, Sun D, Liang H, Wang J, Li J, Guo L, Wang X, Guan C, Boruah BM, Yuan L, Feng F, Yang M, Wang L, Wang Y, Wojdyla J, Li L, Wang J, Wang M, Cheng G, Wang HW, Liu Y (2015) Cryo-EM structure of influenza virus RNA polymerase complex at 4.3 A resolution. *Mol Cell* **57:** 925-935

Chenavas S, Estrozi LF, Slama-Schwok A, Delmas B, Di Primo C, Baudin F, Li X, Crépin T, Ruigrok RW (2013a) Monomeric nucleoprotein of influenza A virus. *PLoS Pathog* **9:** e1003275

Chenavas S, Monod A, Ruigrok RW, Crépin T (2013b) Le complexe de transcription et de réplication des virus influenza, une cible pour de nouvelles molécules antivirales. *Virologie* **17:** 6-16

Communie G, Crépin T, Maurin D, Jensen MR, Blackledge M, Ruigrok RW (2013) Structure of the tetramerization domain of measles virus phosphoprotein. *J Virol* **87:** 7166-7169

Crépin T, Cusack S (2007) [Trap the product to specifically inhibit the enzyme]. *Med Sci (Paris)* **23:** 1087-1089

Crépin T, Dias A, Palencia A, Swale C, Cusack S, Ruigrok RW (2010) Mutational and metal binding analysis of the endonuclease domain of the influenza virus polymerase PA subunit. *J Virol* **84:** 9096-9104

Crépin T, Peterson F, Haertlein M, Jensen D, Wang C, Cusack S, Kron M (2011) A hybrid structural model of the complete *Brugia malayi* cytoplasmic asparaginyl-tRNA synthetase. *J Mol Biol* **405:** 1056-1069

Crépin T, Schmitt E, Blanquet S, Mechulam Y (2002) Structure and function of the C-terminal domain of methionyl-tRNA synthetase. *Biochemistry-Us* **41:** 13003-13011

Crépin T, Schmitt E, Blanquet S, Mechulam Y (2004) Three-dimensional structure of methionyl-tRNA synthetase from *Pyrococcus abyssi*. *Biochemistry-Us* **43:** 2635-2644

Crépin T, Schmitt E, Mechulam Y, Sampson PB, Vaughan MD, Honek JF, Blanquet S (2003) Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in *Escherichia coli* methionyl-tRNA synthetase. *J Mol Biol* **332:** 59-72

Crépin T, Shalak VF, Yaremchuk AD, Vlasenko DO, McCarthy A, Negrutskii BS, Tukalo MA, El'skaya AV (2014) Mammalian translation elongation factor eEF1A2: X-ray structure and new features of GDP/GTP exchange mechanism in higher eukaryotes. *Nucleic Acids Res* **42:** 12939-12948

Crépin T, Swale C, Monod A, Garzoni F, Chaillet M, Berger I (2015) Polyproteins in structural biology. *Curr Opin Struct Biol* **32:** 139-146

Crépin T, Yaremchuk A, Tukalo M, Cusack S (2006) Structures of two bacterial prolyl-tRNA synthetases with and without a cis-editing domain. *Structure* **14:** 1511-1525

Deng T, Engelhardt OG, Thomas B, Akoulitchev AV, Brownlee GG, Fodor E (2006) Role of ran binding protein 5 in nuclear import and assembly of the influenza virus RNA polymerase complex. *J Virol* **80:** 11911-11919

Deng T, Sharps J, Fodor E, Brownlee GG (2005) In vitro assembly of PB2 with a PB1-PA dimer supports a new model of assembly of influenza A virus polymerase subunits into a functional trimeric complex. *J Virol* **79:** 8669-8674

Dias A, Bouvier D, Crépin T, McCarthy AA, Hart DJ, Baudin F, Cusack S, Ruigrok RW (2009a) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458:** 914-918

Dias A, Bouvier D, Cusack S, Ruigrok RW, Crépin T (2009b) [New strategy for targeting influenza virus replication]. *Med Sci (Paris)* **25:** 352-354

Eldred EW, Schimmel PR (1972) Rapid deacylation by isoleucyl transfer ribonucleic acid synthetase of isoleucine-specific transfer ribonucleic acid aminoacylated with valine. *J Biol Chem* **247:** 2961-2964

Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347:** 203-206

Fersht AR (1977) Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthetase. *Biochemistry-Us* **16:** 1025-1030

Fersht AR, Kaethner MM (1976) Enzyme hyperspecificity. Rejection of threonine by the valyl-tRNA synthetase by misacylation and hydrolytic editing. *Biochemistry-Us* **15:** 3342-3346

Fodor E, Smith M (2004) The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex. *J Virol* **78:** 9144-9153

Fournier G, Chiang C, Munier S, Tomoiu A, Demeret C, Vidalain PO, Jacob Y, Naffakh N (2014) Recruitment of RED-SMU1 complex by Influenza A Virus RNA polymerase to control Viral mRNA splicing. *PLoS Pathog* **10:** e1004164

Fukai S, Nureki O, Sekine S, Shimada A, Tao J, Vassylyev DG, Yokoyama S (2000) Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of $tRNA^{Val}$ and valyl-tRNA synthetase. *Cell* **103:** 793-803

Fukunaga R, Yokoyama S (2005a) Aminoacylation complex structures of leucyl-tRNA synthetase and $tRNA^{Leu}$ reveal two modes of discriminator-base recognition. *Nat Struct Mol Biol* **12:** 915-922

Fukunaga R, Yokoyama S (2005b) Crystal structure of leucyl-tRNA synthetase from the archaeon *Pyrococcus horikoshii* reveals a novel editing domain orientation. *J Mol Biol* **346:** 57-71

Gerlach P, Malet H, Cusack S, Reguera J (2015) Structural Insights into Bunyavirus Replication and Its Regulation by the vRNA Promoter. *Cell* **161:** 1267-1279

Gorlich D, Henklein P, Laskey RA, Hartmann E (1996) A 41 amino acid motif in importin-alpha confers binding to importin-beta and hence transit into the nucleus. *EMBO J* **15:** 1810-1817

Gorlich D, Mattaj IW (1996) Nucleocytoplasmic transport. *Science* **271:** 1513-1518

Guilligay D, Kadlec J, Crépin T, Lunardi T, Bouvier D, Kochs G, Ruigrok RW, Cusack S (2014) Comparative structural and functional analysis of orthomyxovirus polymerase cap-snatching domains. *PLoS One* **9:** e84973

Guilligay D, Tarendeau F, Resa-Infante P, Coloma R, Crépin T, Sehr P, Lewis J, Ruigrok RW, Ortin J, Hart DJ, Cusack S (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol* **15:** 500-506

He X, Zhou J, Bartlam M, Zhang R, Ma J, Lou Z, Li X, Li J, Joachimiak A, Zeng Z, Ge R, Rao Z, Liu Y (2008) Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature* **454:** 1123-1126

Hernandez V, Crépin T, Palencia A, Cusack S, Akama T, Baker SJ, Bu W, Feng L, Freund YR, Liu L, Meewan M, Mohan M, Mao W, Rock FL, Sexton H, Sheoran A, Zhang Y, Zhang YK, Zhou Y, Nieman JA, Anugula MR, Keramane el M, Savariraj K, Reddy DS, Sharma R, Subedi R, Singh R, O'Leary A, Simon NL, De Marsh PL, Mushtaq S, Warner M, Livermore DM, Alley MR, Plattner JJ (2013) Discovery of a novel class of boron-based antibacterials with activity against gram-negative bacteria. *Antimicrob Agents Chemother* **57:** 1394-1403

Hountondji C, Blanquet S, Lederer F (1985) Methionyl-tRNA synthetase from *Escherichia coli*: primary structure at the binding site for the 3'-end of $tRNA_f^{Met}$. *Biochemistry-Us* **24:** 1175-1180

Hountondji C, Lederer F, Dessen P, Blanquet S (1986) *Escherichia coli* tyrosyl- and methionyl-tRNA synthetases display sequence similarity at the binding site for the 3'-end of tRNA. *Biochemistry-Us* **25:** 16-21

Hutchinson EC, Orr OE, Man Liu S, Engelhardt OG, Fodor E (2011) Characterization of the interaction between the influenza A virus polymerase subunit PB1 and the host nuclear import factor Ran-binding protein 5. *J Gen Virol* **92:** 1859-1869

Ivanov I, Crépin T, Jamin M, Ruigrok RW (2010) Structure of the dimerization domain of the rabies virus phosphoprotein. *J Virol* **84:** 3707-3710

Jin H, Elliott RM (1993) Non-viral sequences at the 5' ends of Dugbe nairovirus S mRNAs. *J Gen Virol* **74 (Pt 10):** 2293-2297

Kalderon D, Richardson WD, Markham AF, Smith AE (1984a) Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature* **311:** 33-38

Kalderon D, Roberts BL, Richardson WD, Smith AE (1984b) A short amino acid sequence able to specify nuclear location. *Cell* **39:** 499-509

Kamtekar S, Kennedy WD, Wang J, Stathopoulos C, Soll D, Steitz TA (2003) The structural basis of cysteine aminoacylation of $tRNA^{Pro}$ by prolyl-tRNA synthetases. *Proc Natl Acad Sci U S A* **100:** 1673-1678

Kobayashi J, Matsuura Y (2013) Structural basis for cell-cycle-dependent nuclear import mediated by the karyopherin Kap121p. *J Mol Biol* **425:** 1852-1868

Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, Yanagawa H (2009a) Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J Biol Chem* **284:** 478-485

Kosugi S, Hasebe M, Tomita M, Yanagawa H (2009b) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A* **106:** 10171-10176

Labaronne A, Swale C, Monod A, Schoehn G, Crépin T, Ruigrok RW (2016) Binding of RNA by the nucleoproteins of influenza viruses A and B. *Viruses* **8:** In press

Liang B, Li Z, Jenni S, Rahmeh AA, Morin BM, Grant T, Grigorieff N, Harrison SC, Whelan SP (2015) Structure of the L protein of vesicular stomatitis virus from electron cryomicroscopy. *Cell* **162:** 314-327

Maynard LS, Cotzias GC (1955) The partition of manganese among organs and intracellular organelles of the rat. *J Biol Chem* **214:** 489-495

Medcalf L, Poole E, Elton D, Digard P (1999) Temperature-sensitive lesions in two influenza A viruses defective for replicative transcription disrupt RNA binding by the nucleoprotein. *J Virol* **73:** 7349-7356

Melen K, Fagerlund R, Franke J, Kohler M, Kinnunen L, Julkunen I (2003) Importin alpha nuclear localization signal binding sites for STAT1, STAT2, and influenza A virus nucleoprotein. *J Biol Chem* **278:** 28193-28200

Naito T, Momose F, Kawaguchi A, Nagata K (2007) Involvement of Hsp90 in assembly and nuclear import of influenza virus RNA polymerase subunits. *J Virol* **81:** 1339-1349

Ng AK, Lam MK, Zhang H, Liu J, Au SW, Chan PK, Wang J, Shaw PC (2012) Structural basis for RNA binding and homo-oligomer formation by influenza B virus nucleoprotein. *J Virol* **86:** 6758-6767

Ng AK, Zhang H, Tan K, Li Z, Liu JH, Chan PK, Li SM, Chan WY, Au SW, Joachimiak A, Walz T, Wang JH, Shaw PC (2008) Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *Faseb J* **22:** 3638-3647

Nureki O, Vassylyev DG, Tateno M, Shimada A, Nakama T, Fukai S, Konno M, Hendrickson TL, Schimmel P, Yokoyama S (1998) Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science* **280:** 578-582

Obayashi E, Yoshida H, Kawai F, Shibayama N, Kawaguchi A, Nagata K, Tame JR, Park SY (2008) The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* **454:** 1127-1131

Palencia A, Crépin T, Vu MT, Lincecum TL, Jr., Martinis SA, Cusack S (2012) Structural dynamics of the aminoacylation and proofreading functional cycle of bacterial leucyl-tRNA synthetase. *Nat Struct Mol Biol* **19:** 677-684

Pauling L (1957) The probability of errors in the process of synthesis of protein molecules. In *Festschrift fuer Pr. Dr. Arthur Stoll*, Birkhauser B (ed), pp 597–602.

Pemberton LF, Paschal BM (2005) Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* **6:** 187-198

Pflug A, Guilligay D, Reich S, Cusack S (2014) Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516:** 355-360

Plotch SJ, Bouloy M, Ulmanen I, Krug RM (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23:** 847-858

Portela A, Digard P (2002) The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication. *J Gen Virol* **83:** 723-734

Reich S, Guilligay D, Pflug A, Malet H, Berger I, Crépin T, Hart D, Lunardi T, Nanao M, Ruigrok RW, Cusack S (2014) Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516:** 361-366

Robbins J, Dilworth SM, Laskey RA, Dingwall C (1991) Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64:** 615-623

Rock FL, Mao W, Yaremchuk A, Tukalo M, Crépin T, Zhou H, Zhang YK, Hernandez V, Akama T, Baker SJ, Plattner JJ, Shapiro L, Martinis SA, Benkovic SJ, Cusack S, Alley MR (2007) An antifungal agent inhibits an aminoacyl-tRNA synthetase by trapping tRNA in the editing site. *Science* **316:** 1759-1761

Sakurai H, Nishida M, Yoshimura T, Takada J, Koyama M (1985) Partition of divalent and total manganese in organs and subcellular organelles of $MnCl_2$-treated rats studied by ESR and neutron activation analysis. *Biochim Biophys Acta* **841:** 208-214

Sankaranarayanan R, Dock-Bregeon AC, Romby P, Caillet J, Springer M, Rees B, Ehresmann C, Ehresmann B, Moras D (1999) The structure of threonyl-tRNA synthetase-$tRNA^{Thr}$ complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* **97:** 371-381

Silvian LF, Wang J, Steitz TA (1999) Insights into editing from an ile-tRNA synthetase structure with $tRNA^{ile}$ and mupirocin. *Science* **285:** 1074-1077

Sugiyama K, Obayashi E, Kawaguchi A, Suzuki Y, Tame JR, Nagata K, Park SY (2009) Structural insight into the essential PB1-PB2 subunit contact of the influenza virus RNA polymerase. *Embo J* **28:** 1803-1811

Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* **76:** 2879–2886

Swale C, Monod A, Tengo L, Labaronne A, Garzoni F, Bourhis JM, Cusack S, Schoehn G, Berger I, Ruigrok RW, Crépin T (2016) Structural characterization of recombinant IAV polymerase reveals a stable complex between viral PA-PB1 heterodimer and host RanBP5. *Scientific reports* **6:** 24727

Tarendeau F, Boudet J, Guilligay D, Mas P, Bougault C, Boulo S, Baudin F, Ruigrok RWH, Daigle N, Ellenberg J, Cusack S, Simorre J-P, Hart DJ (2007) Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat Struct Mol Biol* **14:** 229-233

Tarendeau F, Crépin T, Guilligay D, Ruigrok RW, Cusack S, Hart DJ (2008) Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog* **4:** e1000136

Weber F, Kochs G, Gruber S, Haller O (1998) A classical bipartite nuclear localization signal on Thogoto and influenza A virus nucleoproteins. *Virology* **250:** 9-18

Yaremchuk A, Cusack S, Gudzera O, Grotli M, Tukalo M (2000a) Crystallization and preliminary crystallographic analysis of *Thermus thermophilus* leucyl-tRNA synthetase and its complexes with leucine and a non-hydrolysable leucyl-adenylate analogue. *Acta Crystallogr D Biol Crystallogr* **56:** 667-669

Yaremchuk A, Cusack S, Tukalo M (2000b) Crystal structure of a eukaryote/archaeon-like protyl-tRNA synthetase and its complex with $tRNA^{Pro}$(CGG). *Embo J* **19:** 4745-4758

Yaremchuk A, Cusack S, Tukalo M (2000c) Crystallization and preliminary X-ray diffraction analysis of *Thermus thermophilus* prolyl-tRNA synthetase. *Acta Crystallogr D Biol Crystallogr* **56:** 195-196

Yaremchuk A, Shalak VF, Novosylna OV, Negrutskii BS, Crépin T, El'skaya AV, Tukalo M (2012) Purification, crystallization and preliminary X-ray crystallographic analysis of mammalian translation elongation factor eEF1A2. *Acta crystallographica Section F, Structural biology and crystallization communications* **68:** 295-297

Yarus M (1972) Phenylalanyl-tRNA synthetase and isoleucyl-$tRNA^{Phe}$: a possible verification mechanism for aminoacyl-tRNA. *Proc Natl Acad Sci U S A* **69:** 1915-1919

Ye Q, Krug RM, Tao YJ (2006) The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444:** 1078-1082

Yuan P, Bartlam M, Lou Z, Chen S, Zhou J, He X, Lv Z, Ge R, Li X, Deng T, Fodor E, Rao Z, Liu Y (2009) Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature* **458:** 909-913

Publications

XX. Labaronne A, Milles S, Ringkjøbing Jensen M, Blackledge M, Bourhis JM, Ruigrok RW and **Crépin T**. Complex between nucleoprotein of influenza B and importin-α. *En préparation.*

XX. Labaronne A, Miloudi M, Tengo L, Mas C, Schoehn G, Ruigrok RW and **Crépin T**. The RNA binding of influenza D virus nucleoprotein is regulated by NP_{Tail}. *soumis à J Virol.*

41. Labaronne A, Swale C, Monod A, Schoehn G, **Crépin T*** and Ruigrok RW (2016) Binding of RNA by the nucleoproteins of influenza viruses A and B. *Viruses*, 8, 247; doi:10.3390/v8090247.
*Crépin T est correspondant

40. Swale C, Monod A, Tengo L, Laboronne A, Garzoni F, Bourhis JM, Cusack S, Schoehn G, Berger I, Ruigrok RW and **Crépin T** (2016) Structural characterization of recombinant IAV polymerase reveals a stable complex between viral PA-PB1 heterodimer and host RanBP5. *Sci Rep*, 6:24727. doi: 10.1038/srep24727.

39. Dégut C, Monod A, Brachet F, **Crépin T**, Tisné C (2016) *In vitro*/*in vivo* production of tRNA for X-Ray studies. *Methods Mol Biol.*, 1320:37-57.

38. **Crépin T***, Swale C, Monod A, Garzoni F, Chaillet M, Berger I* (2015) Polyproteins in structural biology. *Curr Opin Struct Biol.*, 32:139-46.
*Crépin T et Berger I sont correspondants

37. Monod A, Swale C, Tarus B, Tissot A, Delmas B, Ruigrok RW, **Crépin T** & Slama-Schwok A (2015) Learning from structure-based drug design and new antivirals targeting the ribonucleoprotein complex for the treatment of influenza. *Expert Opin Drug Discov*, 10:345-71.

36. Reich S, Guilligay D, Pflug A, Malet H, Berger I, **Crépin T**, Hart D, Lunardi T, Nanao M, Ruigrok RW, Cusack S (2014) Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature*, 516:361-6.

35. **Crépin T**, Shalak VF, Yaremchuk AD, Vlasenko DO, McCarthy A, Negrutskii BS, Tukalo MA, El'skaya AV (2014) Mammalian translation elongation factor eEF1A2: X-ray structure and new features of GDP/GTP exchange mechanism in higher eukaryotes. *Nucleic Acids Res*, 42:12939-48.

34. Chiu W, **Crépin T** & Ruigrok RW (2014) Editorial overview: Virus structure and function. *Curr Opin Virol*, 5.

33. Klumpp K and Crépin T (2014) Capsid proteins of enveloped viruses as antiviral drug targets. Curr Opin Virol, 5:63-71.

32. Guilligay D, Kadlec J, **Crépin T**, Lunardi T, Bouvier D, Kochs G, Ruigrok RW, Cusack S (2014) Comparative structural and functional analysis of orthomyxovirus polymerase cap-snatching domains. *PLoS One*, 9:e84973.

31. Chenavas S, **Crépin T**, Delmas B, Ruigrok RW and Slama-Schwok A (2013) Influenza virus nucleoprotein: structure, RNA binding, oligomerization and antiviral drug target. *Future Microbiol.*: 10.2217/fmb.13.128.

30. Communie G, **Crépin T**, Maurin D, Ringkjøbing Jensen M, Blackledge M and Ruigrok RW (2013) Structure of the tetramerization domain of measles virus phosphoprotein. *J Virol*, 87:7166-7169.

29. Chenavas S, Estrozi LF, Slama-Schwok A, Delmas B, Di Primo C, Baudin F, Li X, **Crépin T** and Ruigrok RW. (2013) Monomeric nucleoprotein of influenza A virus. *PLoS Pathog*, 9:e1003275.

28. Hernandez V, **Crépin T**, Palencia A, Cusack S, Akama T et al. (2013) Discovery of a novel class of boron-based antibacterials with activity against gram-negative bacteria. *Antimicrob Agents Chemother*, 57:1394-1403.

27. Chenavas S, Monod A, Ruigrok RW and **Crépin T** (2013) Le complexe de transcription et de réplication des virus influenza, une cible pour de nouvelles molécules antivirales. *Virologie*, 17:6-16.

26. Palencia A[±], **Crépin T**[±], Vu MT, Lincecum TM, Martinis SA and Cusack S (2012) Structural dynamics of the aminoacylation and proof-reading functional cycle of bacterial leucyl-tRNA synthetase. *Nat Struct Mol Biol*, 19:677-684.
[±]Palencia A et Crépin sont co-premiers auteurs

25. Yaremchuk A, Shalak VF, Novosylna OV, Negrutskii BS, **Crépin T**, El'skaya AV and Tukalo M (2012) Purification, crystallization and preliminary X-ray crystallographic analysis of mammalian translation elongation factor eEF1A2. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. F68:295-297.

24. Ruigrok RW, **Crépin T** and Kolakofsky D (2011) Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr Opin Microbiol*. 14:504-510.

23. **Crépin T**, Peterson F, Haertlein M, Jensen D, Wang C, Cusack S and Kron M (2011) A hybrid structural model of the complete Brugia malayi cytoplasmic asparaginyl-tRNA synthetase. *J Mol Biol*, 405:1056-1069.

22. Chenavas S, Guilligay D, Cusack S, Hart DJ, Ruigrok RW and **Crépin T** (2010) Avancées dans l'étude structurale et la compréhension moléculaire du mécanisme de transcription du virus de la grippe. *Virologie*, 14:435-445.

21. Guilligay D, Chenavas S, Cusack S, Ruigrok RW and **Crépin T** (2010) L'ARN polymérase du virus de la grippe : de la recherche fondamentale à l'élaboration de nouveaux antiviraux. *Biofutur*, 313:48-52.

20. **Crépin T**, Dias A, Palencia A, Swale C, Cusack S and Ruigrok RW (2010) Mutational and metal binding analysis of the endonuclease domain of the influenza virus polymerase PA subunit. *J Virol*, 84:9096-9104.

19. Ivanov I, **Crépin T**, Jamin M and Ruigrok RW (2010) Structure of the dimerization domain of the rabies virus phosphoprotein. *J Virol*, 84:3707-3710.

18. Ruigrok RW, **Crépin T**, Hart DJ and Cusack S (2010) Towards an atomic resolution understanding of the influenza virus replication machinery. *Curr Opin Struct Biol*, 20:104-113.

17. Ruigrok RW and **Crépin T** (2010) Commentary: Nucleoproteins of Negative Strand RNA Viruses; RNA Binding, Oligomerisation and Binding to Polymerase Co-Factor. *Viruses*, 2:27-32.

16. **Crépin T**, Hart DJ, Cusack S and Ruigrok RW (2009) L'étude de l'ARN polymérase du virus de la grippe, un projet pour définir des cibles thérapeutiques nouvelles. *Regard sur la biochimie*, décembre, 7-9.

15. Johansson S, Nilsson E, Qian W, Guilligay D, **Crépin T**, Cusack S, Arnberg N and Elofsson M (2009) Design, synthesis, and evaluation of N-acyl modified sialic acids as inhibitors of adenoviruses causing epidemic keratoconjunctivitis. *J Med Chem*, 52:3666-3678.

14. Dias A[±], Bouvier D[±], **Crépin T**[±], McCarthy AA, Hart DJ, Baudin F, Cusack S and Ruigrok RW (2009) The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*, 458:914-918.
[±]Dias A, Bouvier D et Crepin T sont co-premiers auteurs

13. Dias A, Bouvier D, Cusack S, Ruigrok RW and **Crépin T** (2009) New strategy for targeting influenza virus replication. *Med Sci* (Paris), 25:352-354.

12. Boesecke P, Bois JM, **Crépin T**, Hunte C, Kahn R, Kao WC, Nauton L, Winther AM, Moller J, Nissen P, Nury H, Olesen C, Pebay-Peyroula E, Vicat J and Stuhrmann H (2009) A first low-resolution difference Fourier map of phosphorus in a membrane protein from near-edge anomalous diffraction. *J Synchrotron Radiat*, 16:658-665.

11. Tarendeau F, **Crépin T**, Guilligay D, Ruigrok RW, Cusack S and Hart DJ (2008) Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog*, 4:e1000136.

10. Guilligay D, Tarendeau F, Resa-Infante P, Coloma R, **Crépin T**, Sehr P, Lewis J, Ruigrok RW, Ortin J, Hart DJ and Cusack S (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol*, 15:500-506.

9. Rock FL, Mao W, Yaremchuk A, Tukalo M, **Crépin T**, Zhou H, Zhang YK, Hernandez V, Akama T, Baker SJ, Plattner JJ, Shapiro L, Martinis SA, Benkovic SJ, Cusack S and Alley MR (2007) An antifungal agent inhibits an aminoacyl-tRNA synthetase by trapping tRNA in the editing site. *Science*, 316:1759-1761.

8. **Crépin T** and Cusack S (2007) Trap the product to specifically inhibit the enzyme. *Med Sci* (Paris), 23:1087-1089.

7. Sukuru SC, **Crépin T**, Milev Y, Marsh LC, Hill JB, Anderson RJ, Morris JC, Rohatgi A, O'Mahony G, Grotli M, Danel F, Page MG, Hartlein M, Cusack S, Kron MA and Kuhn LA (2006) Discovering new classes of Brugia malayi asparaginyl-tRNA synthetase inhibitors and relating specificity to conformational change. *J Comput Aided Mol Des*, 20:159-178.

6. **Crépin T**[±], Yaremchuk A[±], Tukalo M and Cusack S (2006) Structures of two bacterial prolyl-tRNA synthetases with and without a cis-editing domain. *Structure,* 14:1511-1525.
[±]Crepin T et Yaremchuk A sont co-premiers auteurs

5. **Crépin T**, Schmitt E, Blanquet S and Mechulam Y (2004) Three-dimensional structure of methionyl-tRNA synthetase from Pyrococcus abyssi. *Biochemistry*, 43:2635-2644.

4. Blanquet S, **Crépin T**, Mechulam Y and Schmitt E (2004) Class 1 aminoacyl-tRNA synthetases: Methionyl-tRNA synthetases. In Ibba M, Francklyn C and Cusack S (eds.), The aminoacyl-tRNA synthetases. Landes Biosciences.

3. **Crépin T**, Schmitt E, Mechulam Y, Sampson PB, Vaughan MD, Honek JF and Blanquet S (2003) Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in Escherichia coli methionyl-tRNA synthetase. *J Mol Biol*, 332:59-72.

2. **Crépin T**, Schmitt E, Blanquet S and Mechulam Y (2002) Structure and function of the C-terminal domain of methionyl-tRNA synthetase. *Biochemistry*, 41:13003-13011.

1. Chihib NE, **Crépin T**, Delattre G and Tholozan JL (1999) Involvement of cell envelope in nisin resistance of Pectinatus frisingensis, a Gram-negative, strictly anaerobic beer-spoilage bacterium naturally sensitive to nisin. *FEMS Microbiol Lett*, 177:167-175.

Coordonnées déposées à la PDB

code	Organism	Protein	Ligand	Resolution	R_{value}/R_{free} (%)
1MKH	*P. abyssi*	C12K	-	2.01 Å	27.5 / 28.9
1PFU	*E. coli*	MetRS	MetI	1.91 Å	19.5 / 22.4
1PFV	*E. coli*	MetRS	DFM	1.70 Å	18.6 / 20.3
1PFW	*E. coli*	MetRS	TFM	1.78 Å	18.4 / 20.9
1PFY	*E. coli*	MetRS	MetSA	1.93 Å	18.8 / 22.4
1PG0	*E. coli*	MetRS	Metol-AMP	1.9 Å	22.3 / 25.0
1PG2	*E. coli*	MetRS	Met + adenosine	1.75 Å	20.1 / 22.6
1P7P	*E. coli*	MetRS	MetP	1.80 Å	19.5 / 22.5
1RQG	*P. abyssi*	MetRS	-	2.90 Å	24.4 / 29.6
2I4L	*R. palustris*	ProRS	-	2.00 Å	19.2 / 22.1
2I4M	*R. palustris*	ProRS	ProAMS	2.80 Å	20.9 / 26.9
2I4N	*R. palustris*	ProRS	CysAMS	2.85 Å	21.4 / 27.7
2I4O	*R. palustris*	ProRS	ATP	2.40 Å	20.1 / 26.1
2J3L	*E. faecalis*	ProRS	ProAMS	2.30 Å	20.8 / 26.8
2J3M	*E. faecalis*	ProRS	prolinol + ATP	2.30 Å	19.8 / 26.0
2KQR	*B. malayi*	AsnRS	-	NMR model	
2V0C	*T. thermophilus*	LeuRS	AMP + AN2690	1.85 Å	18.1 / 20.2
2V0G	*T. thermophilus*	LeuRS	$tRNA^{Leu}$ + AN2690	3.50 Å	24.9 / 31.4
2VQZ	Influenza virus	PB2-cap	m_7GTP	2.30 Å	18.9 / 23.5
2VY6	Influenza virus	PB2-Cter	-	1.95 Å	20.4 / 25.8
2VY7	Influenza virus	PB2-627	-	1.53 Å	14.6 / 18.5
2VY8	Influenza virus	PB2-627	-	1.20 Å	19.1 / 21.4
2WGT	Adenovirus 37	Fiber-head	17a	1.80 Å	18.5 / 22.6
2WGU	Adenovirus 37	Fiber-head	17g	1.80 Å	18.5 / 22.3
2W69	Influenza virus	PA-Nter	Mn^{2+}	2.05 Å	22.0 / 26.8
2XGT	*B. malayi*	AsnRS	AsnAMS	1.90 Å	20.4 / 25.0
2XTI	*B. malayi*	AsnRS	ATP + LβNOH	2.40 Å	21.6 / 29.8
3L32	Rabies virus	P-dim	-	1.50 Å	19.9 / 23.5
3ZDO	Measles virus	P-tetra	-	2.07 Å	21.3 / 28.4
3ZDP	Influenza virus	NP-R416A	-	2.69 Å	20.0 / 20.5
3ZJT	*E. coli*	LeuRS	$tRNA^{Leu}$ + AN3017	2.20 Å	20.9 / 25.1
3ZJU	*E. coli*	LeuRS	$tRNA^{Leu}$ + AN3016	2.40 Å	21.0 / 26.1

3ZJV	*E. coli*	LeuRS	$tRNA^{Leu}$ + AN3213	2.31 Å	20.3 / 24.7
4AQ7	*E. coli*	LeuRS	$tRNA^{Leu}$ + LeuAMS	2.50 Å	19.1 / 25.0
4ARC	*E. coli*	LeuRS	$tRNA^{Leu}$ + Leucine	2.00 Å	21.2 / 25.7
4ARI	*E. coli*	LeuRS	$tRNA^{Leu}$ + AN2679	2.08 Å	20.1 / 24.4
4AS1	*E. coli*	LeuRS	$tRNA^{Leu}$ + AN2679	2.02 Å	21.0 / 24.9
4C0S	*O. cuniculus*	eEF1A2	GDP	2.7 Å	20.5 / 25.6
4CGS	Dhori virus	PA-Nter	-	1.3 Å	17.5 / 22.0
4CGX	Thogoto virus	PA-Nter	-	2.7 Å	20.4 / 28.0
4CHC	Thogoto virus	PA-Nter	-	2.7 Å	23.1 / 27.9
4CHD	Thogoto virus	PB2-627	-	2.4 Å	20.9 / 23.9
4CHE	Thogoto virus	PB2-cap	-	1.8 Å	17.3 / 19.2
4CHF	Thogoto virus	PB2-cap	-	3.0 Å	20.0 / 21.6
4WCY	*M. musculus*	AZ130-Fab	-	2.0 Å	17.2 / 22.4
4WRT	Influenza B	Polymerase	vRNA	2.7 Å	17.5 / 21.1
4WSA	Influenza B	Polymerase	vRNA	3.4 Å	23.0 / 26.5
5N2U	Influenza D	NP	-	2.35 Å	20.4 / 26.0

Publications choisies

Structure and Function of the C-Terminal Domain of Methionyl-tRNA Synthetase†

Thibaut Crepin, Emmanuelle Schmitt, Sylvain Blanquet, and Yves Mechulam*

Laboratoire de Biochimie, Unité Mixte de Recherche 7654, CNRS-Ecole Polytechnique, F-91128 Palaiseau Cedex, France

Received June 24, 2002; Revised Manuscript Received August 30, 2002

ABSTRACT: The minimal polypeptide supporting full methionyl-tRNA synthetase (MetRS) activity is composed of four domains: a catalytic Rossmann fold, a connective peptide, a KMSKS domain, and a C-terminal α helix bundle domain. The minimal MetRS behaves as a monomer. In several species, MetRS is a homodimer because of a C-terminal domain appended to the core polypeptide. Upon truncation of this C-terminal domain, subunits dissociate irreversibly. Here, the C-terminal domain of dimeric MetRS from *Pyrococcus abyssi* was isolated and studied. It displays nonspecific tRNA-binding properties and has a crystalline structure closely resembling that of Trbp111, a dimeric tRNA-binding protein found in many bacteria and archaea. The obtained 3D model was used to direct mutations against dimerization of *Escherichia coli* MetRS. Comparison of the resulting mutants to native and C-truncated MetRS shows that the presence of the appended C-domain improves $tRNA^{Met}$ binding affinity. However, dimer formation is required to evidence the gain in affinity.

Methionyl-tRNA synthetase (MetRS) is a ubiquitous enzyme responsible for the specific aminoacylation of methionine transfer RNAs. Depending on the species from which it is isolated, MetRS behaves either as a monomer or a dimer. The reason for these two molecular states for a same enzyme activity is intriguing. Indeed, it was early demonstrated that dimeric *Escherichia coli* MetRS could be converted by C-terminal truncation into a minimal core of nearly 550 residues keeping activity for both the activation of methionine and the aminoacylation of $tRNA^{Met}$ (*1*). The truncated polypeptide can no more dimerize. Such shortened units derived from *E. coli* and *Thermus thermophilus* MetRS have been extensively studied, and in particular, the 3D structures of both enzymes were determined (*2*, *3*). The enzyme is made of four domains: a catalytic Rossmann fold, a connective peptide inserted between the two halves of the Rossmann fold, a KMSKS domain characteristic of class 1 aaRS (*4*−*6*), and a C-terminal α helix bundle domain, responsible for the recognition of the CAU anticodon of methionine tRNAs. These four domains form the structural core of MetRS and are systematically encountered in all MetRS sequenced so far (Figure 1). Thirty-four percent of known MetRS consist solely of this main unit and behave as monomers. The other MetRS (66%) contain domains appended to this minimal core and either form homodimers (48%) or remain long monomers (18%; Figure 1).

Dimer formation of MetRS is always related to the presence, at the C-side of the core enzyme, of an appended domain. Interestingly, some organisms express paralogs of this dimerization domain (Figure 1). Such paralogs are, for example, Trbp111[1] (tRNA-binding protein) of *Aquifex aeolicus* (*7*), the cold-shock protein CsaA from *B. subtilis* (*8*), yeast Arc1p (aminoacyl-tRNA cofactor protein, *9*), and the mammalian cytokine endothelial monocyte activating poly-

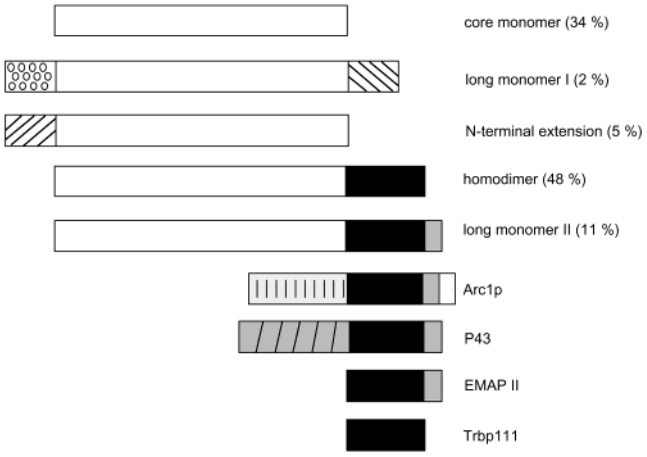


FIGURE 1: Structural organization of MetRS proteins and alignment of their C-terminal extensions with paralogous proteins. The open boxes represent the core structural unit of MetRS, whereas the dark boxes correspond to the C-terminal dimerization domain of MetRS. The light gray box schematizes the added domain in Arc1p, p43, EMAP II, and long monomer 1, which prevents dimerization. Other unrelated domains have specific patterns. The numbers in parentheses indicate the percentage of related oligomeric states according to a sample of 65 examined MetRS sequences. Bacterial MetRS belong either to the core monomeric or to the homodimeric type. In archaea, euryarchaeotal enzymes are monomeric, whereas crenarchaeotal enzymes are homodimeric. In eukaryotes, fungal MetRS present an N-terminal extension, whereas in viridiplantae and metazoa, these enzymes are long monomers. The 3D structure of the core monomer of MetRS is known in the cases of *E. coli* (*2*) and *T. thermophilus* (*3*). The 3D structures of human EMAP II (*14*, *15*) and *A. aeolicus* Trbp111 (*12*) have also been determined.

† Coordinates have been deposited at the Protein Data Bank with accession number 1MKH.

* Corresponding author. Tel: +33 1 69334885. Fax: +33 1 69333013. E-mail: yves@botrytis.polytechnique.fr.

[1] Abbreviations: SDS−PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; EDTA, ethylenediamine tetraacetic acid; EMAP II, endothelial monocyte activating polypeptide II; Arc1p, aminoacyl-tRNA cofactor protein; Trbp, tRNA-binding protein.

10.1021/bi026343m CCC: $22.00
Published on Web 10/05/2002

peptide II (EMAP II, *10*). Both Arc1p and EMAP II interfere with the function of aminoacyl-tRNA synthetases in a trans-acting manner. Arc1p binds the N-terminal extension of monomeric MetRS of yeast, thereby increasing the affinity of the synthetase for tRNA (*9*). In mammals, a high molecular weight complex assembles nine aaRS with non-synthetase components. Cleavage of p43, one of these nonsynthetase components, by an apoptotic protease produces EMAP II (*11*).

The 3D structures of *A. aeolicus* Trbp111 (*12*), of *T. thermophilus* CsaA (*13*), and of human EMAP II (*14*, *15*) were determined. In all cases, an OB-fold forms the core of the structure. However, Trbp111 dimerizes, while EMAP II is monomeric. The monomeric state of EMAP II can be explained by a C-terminal extension (Figure 1), which masks the dimerization region (*15*). Interestingly, an EMAP II-like structure is found at the C-termini of long monomeric II MetRS (Figure 1), thereby explaining that these proteins cannot dimerize.

Biochemical characterizations show general tRNA-binding properties for both Trbp111 and EMAP II (*7*, *16*). Specificity of the binding appears rather broad, with recognition of the L shape of tRNA. It has been suggested that Trbp111 has the activity of a tRNA chaperone. Upon gene fusion, this activity would have been incorporated in MetRS, as well as in some other aminoacyl-tRNA synthetases such as PheRS and human TyrRS (*7*).

Here, we studied the C-terminal domain of dimeric MetRS from the archaeon *Pyrococcus abyssi*. Its three-dimensional structure was determined. This structure was used to direct mutations against the dimerization of *E. coli* MetRS. Study of the mutants indicates that the appended C-terminal domain is a cis-acting element providing the enzyme with a better tRNA affinity. However, to gain the increase in tRNA affinity, subunits of the enzyme must associate.

RESULTS AND DISCUSSION

Production and Characterization of the Dimerization Domain of MetRS. The *metS* gene encoding MetRS from the hyperthermophilic archaeon *Pyrococcus abyssi* was PCR-amplified and cloned between the *Nco*I and *Sac*II sites of pET15blpa. In the resulting construction, the region encoding the His tag is removed. When expressed in *E. coli*, this vector leads to the production of two thermostable proteins. The heaviest protein has a molecular weight of 84 KDa, corresponding to that of *P. abyssi* MetRS (722 residues), whereas the lightest one has an apparent molecular weight of 12 KDa and is preponderant (Figure 2). To characterize this polypeptide, N-terminal sequencing was performed after excision of the protein from the gel. The sequence MYVKFD was obtained. It corresponds to an internal sequence of *P. abyssi* MetRS, starting at residue 616. Examination of the *metS* gene in this region shows that the ATG codon at position 616 is preceded 14 bases upstream by a putative AGGAGG Shine-Dalgarno sequence (*17*) (Figure 2). The polypeptide predicted to be produced from this ATG codon has a molecular weight of 12.4 KDa, in good agreement with the one observed upon SDS−PAGE analysis. This strongly indicates that the observed protein is produced from the ATG codon internal to *metS* and recognized as translation start signal by the *E. coli* ribosome. The protein was therefore called MetRS−C12K. At this stage, the possibility that MetRS−C12K is expressed in *Pyrococcus abyssi* remains open. Indeed, it should be underlined that the Shine-Dalgarno-like sequence AGGAGG is also complementary to a sequence at the 3′ end of the *P. abyssi* 16S rRNA (*18*). It might therefore represent a translation start signal in this species (*19*). To undertake the characterization of MetRS−C12K, the protein was purified from overproducing cells. Mass spectrometry measurements were performed, and confirmed the identity between the purified protein (12 041.6 ± 2.8 Da) and the one predicted from the gene sequence (12 041.2 Da). Moreover, as shown by molecular sieving, MetRS−C12K (30 μM) behaves as a dimer in solution.

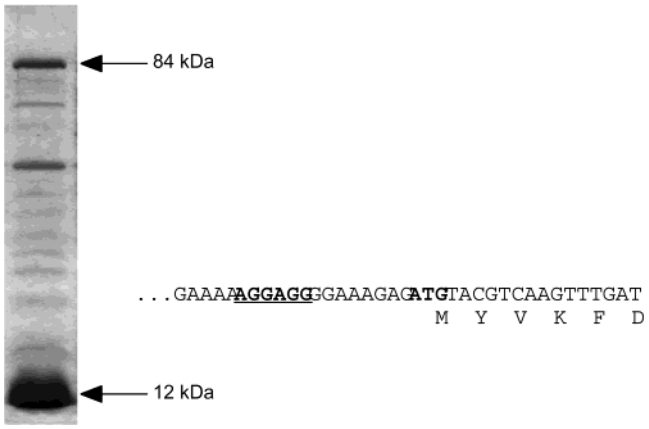


FIGURE 2: SDS−PAGE analysis of a heated extract of overproducing *E. coli* cells carrying the pET15bMet-pa plasmid. The bands corresponding to full-length *P. abyssi* MetRS (84 kDA) and to the isolated C-terminal domain (12 kDa) are flagged by arrows. The two other visible bands are also present in a nonoverproducing control extract and therefore correspond to thermostable *E. coli* proteins. In the left part of the figure, the sequence around the internal ATG codon corresponding to the start of MetRS−C12K is shown. The Shine−Dalgarno sequence is underlined.

As described in the Introduction section, MetRS−C12K belongs to a family of tRNA-binding proteins, including *A. aeolicus* Trbp111 (*7*), a part of the C-terminal domain of *S. cerevisiae* Arc1p (*9*), and the N-domain of the EMAP II cytokine (*10*, *16*). Trbp111 and Arc1p bind tRNAs with dissociation constants in the 10 nM range (*7*, *9*), whereas EMAP II binds tRNAs with an apparent K_d of 20 μM (*16*). tRNA-binding properties of MetRS−C12K were studied by the gel-shift method. Assuming a stoichiometry of one tRNA molecule per dimer, experiments using purified *E. coli* $tRNA_f^{Met}$ or $tRNA_1^{Val}$ (0.5 μM) showed that MetRS−C12K can bind either of the two tRNAs with an equilibrium constant smaller than 0.5 μM (data not shown).

Crystal Structure of MetRS−C12K. Hexagonal crystals of MetRS−C12K were obtained by using 20% PEG8000 as precipitating agent in the presence of 50 mM potassium phosphate. Crystals belong to space group $P3_221$, with cell parameters $a = b = 38.3$ Å, $c = 161.6$ Å. Cryoprotection of the crystals could be achieved by soaking in a solution containing 10% ethylene glycol, 20% PEG8000, and 50 mM KH_2PO_4. A complete dataset to 2 Å of resolution was recorded at 100 K at the ESRF ID14-EH1 beamline. The structure was solved by molecular replacement by using residues 10−95 in the human EMAP II structure (*14*) as a starting model. The structure, containing one monomer per asymmetric unit, was refined to 2 Å resolution with *R* and

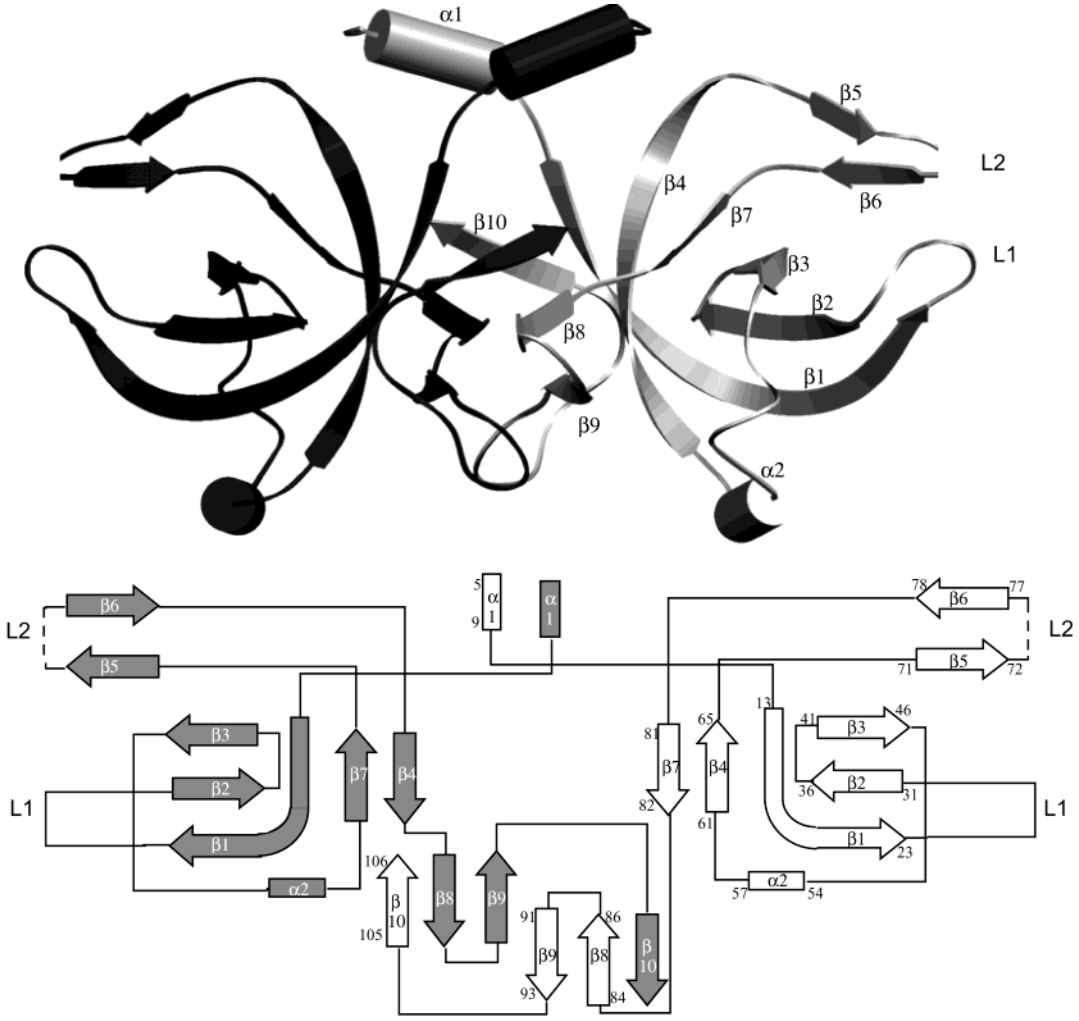


FIGURE 3: Structure of the dimeric MetRS−C12K domain. Top: representation of the 3D structure of the domain drawn with Setor (*46*). One monomer is colored in dark gray and the other monomer in light gray. Bottom: scheme of the topology showing secondary structure elements.

free-*R* values of 27.5% and 28.9%, respectively. The overall structure shows a *B* value of 69 Å^2, reflected at the level of the value of 49 Å^2 deduced from the Wilson plot of the diffraction data. This high *B* value is mainly due to two mobile loops (L1 and L2 in Figure 3), corresponding to residues 25−28 and 72−75, which have inaccurately determined positions (no density is observed for residue 74). Taking into account the small size of the protein, the presence of the two poorly ordered loops made it difficult to refine the structure to lower *R* values.

In the crystal, the structure of MetRS−C12K is homodimeric, with the dimer axis corresponding to a crystallographic axis. As shown in Figure 3, the structure of the subunit is mainly composed of β strands. Only two short α helices are found, α1 at the N-terminus of the protein and α2 between β3 and β4. The structure defined by strands β1, β2, β3, β4, and β7 is reminiscent of that of an OB-fold (*20*), capped by the α2 helix. α1, β8, β9, and β10 ensure the dimer interface. Finally, it is remarkable that the interface of MetRS−C12K forms two symmetrical β barrel-like structures (Figure 4). Therefore, on the whole, the MetRS−C12K dimer appears to result from the association of four β-barrels. The structure of MetRS−C12K is highly similar to those of Trbp111 from *A. aeolicus* (*12*) and CsaA from *T. thermophilus* (*13*). The rms deviation between MetRS−C12K and CsaA is 1.2 Å for 101 pairs of C α atoms. The model of the MetRS−C12K dimer can be superimposed on that of EMAP II (*14*, *15*) with an rmsd of 1.4 Å for 99 compared C alpha atoms. This superimposition involves C alpha atoms from both MetRS−C12K monomers, since the C-terminal extension of EMAP II aligns with the second monomer of MetRS−C12K. Such a superimposition reflects the previous observation that the C-terminal domain of monomeric EMAP II is related by degenerate 2-fold symmetry to the OB-fold domain, in such a way that the domain interface in EMAP II mimics the dimer interface of Trbp111-like proteins (*15*).

In the case of *A. aeolicus* Trbp111, binding of tRNA was studied by engineering site-directed mutants (*12*). The most important residues map in two loops. The first loop (L1) lies between β1 and β2, while the second loop (L2) connects β5 to β6. In MetRS−C12K, as in Trbp111, these two loops are face-to-face and form a clamp protruding out of the barrel (Figure 3). Most of the residues found important for tRNA binding in the case of Trbp111 are conserved in MetRS−C12K. These residues are K30 (L1), N66 and S78 (L2), and M81 (N-terminus of β7).

Design of a Long Monomeric MetRS. To investigate the role of the C-terminal domain in the context of MetRS, we wished to design mutations impairing the dimerization process. Our goal was to construct long monomers of *E. coli*

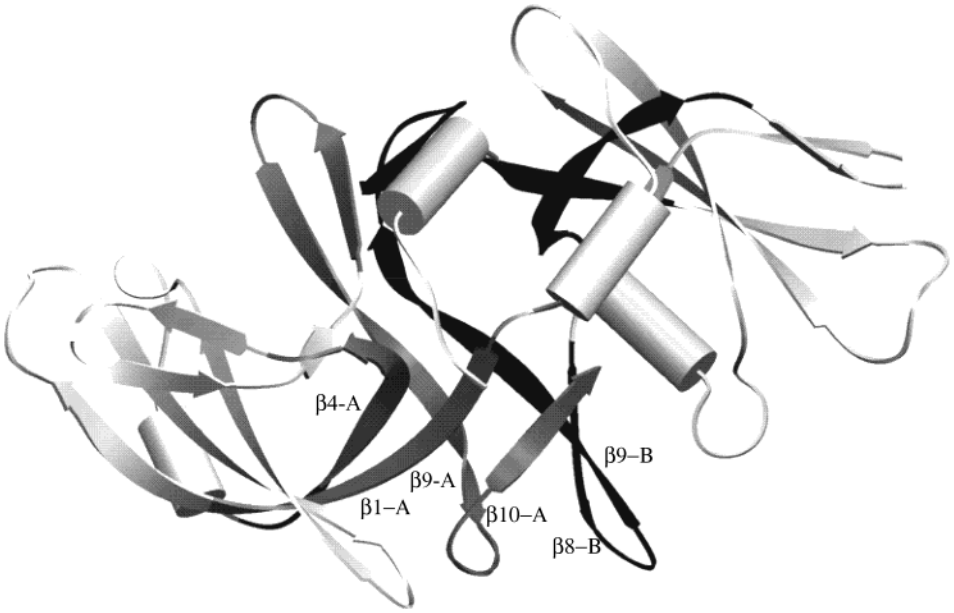


FIGURE 4: Dimer interface of MetRS−C12K. Top: schematic view of the MetRS−C12K dimer showing the internal β barrel-like structure made upon dimerization. The secondary structure elements involved in the formation of this barrel are shown in dark gray for one monomer and in black for the other monomer. Only the bottom barrel is labeled; the symmetrical one is less visible, at the top of the view. For the sake of clarity, the β9−A and β9-B strands were prolonged by 4 residues at their C-termini.

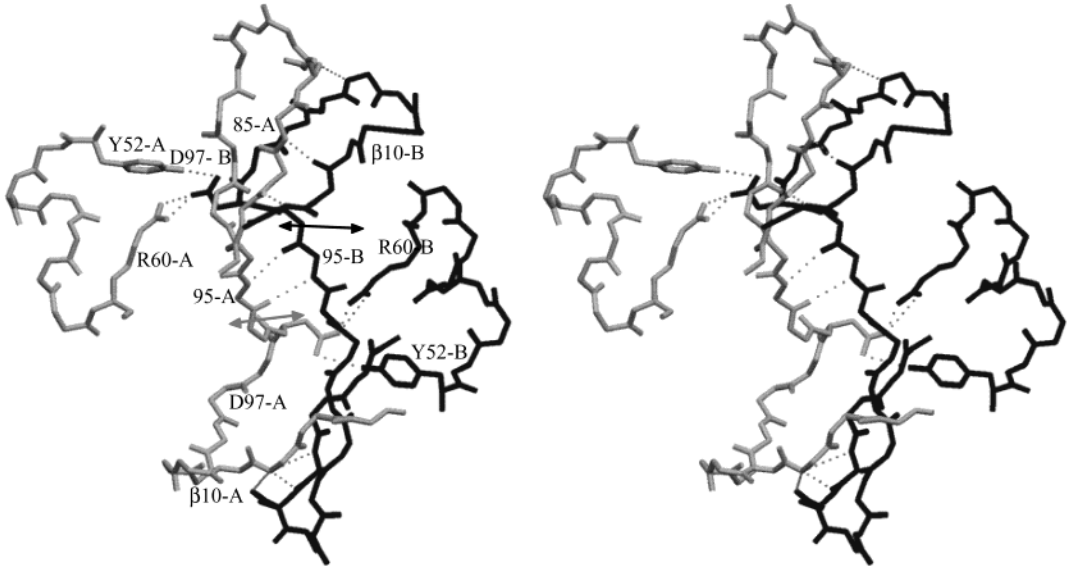


FIGURE 5: Stereo representation of the dimer interface modified by site-directed mutagenesis. The side chain interactions involving the conserved residues Tyr52 and Arg60 of one monomer and Asp96 of the other monomer are shown. The β strands interactions between β10 and β8 are shown with dashed lines. The arrows indicate the position of the cleavage leading to the M665 variant. Only the relevant side chains are drawn.

MetRS to investigate the roles in enzyme activity of the C-terminal domain by itself on the one hand and of the dimerization of MetRS on the other hand. In MetRS−C12K, the dimer interface mainly results from antiparallel β-strand interactions. In addition, the two α1 helices are packed together, and the N-termini are in contact through hydrogen bonds involving the carbonyl group of M1 of one monomer and the side chain of N67 of the other monomer. β strand interactions involve the β10 C-terminal strand of one monomer, which contacts the β8 strand of the other monomer. In addition, the two symmetrical extended peptides after β9 (residues 95−97) also make main chain atom contacts (Figure 5). A notable exception is D97, whose side chain establishes electrostatic contacts with the side chains of Y52 and R60 of the other monomer (Figure 5). Y52 is strictly conserved in all dimeric Trbp111-like domains, while R60 and D97 only show conservative replacements at these positions (K or E, respectively). Interestingly, none of these residues is found in the monomeric EMAP II-like domains.

Because the dimer interaction primarily involves main chain atom contacts, mutations that cause dissociation of the dimer may be difficult to obtain. However, taking into account the interface contacts, we decided to construct two *E. coli* MetRS variants: the first one corresponds to the mutation into A of D666, the residue corresponding to D97 in MetRS−C12K (D666A mutant). In the second one, the β10 strand and the extended peptide at its N-terminus were removed by inserting a stop codon at position 666 (M665

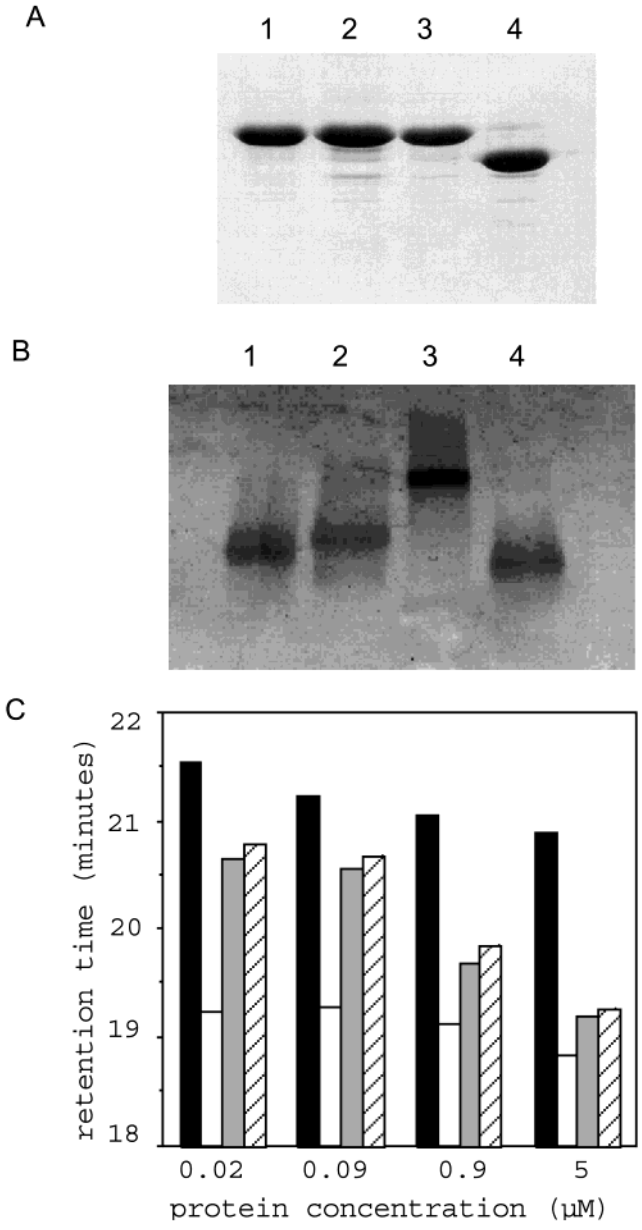


FIGURE 6: Panel A: SDS−PAGE analysis of MetRS variants. Lane 1, wild-type MetRS dimer; lane 2, M665; lane3, D666A; lane 4, M547 truncated monomer. Panel B: native PAGE analysis of MetRS variants. Electrophoresis of the purified proteins (6 μM) was performed on a 4−15% polyacrylamide gradient gel. Lane 1, M665; lane 2, D666A; lane 3, wild-type MetRS dimer; lane 4, M547 truncated monomer. Panel C: chromatographic analysis of MetRS variants. MetRS variants were injected at different concentrations onto a TSK4000SWXL molecular sieving column (TOSOHAAS 7.5 × 3000 mm) operated at 0.5 mL/min in 20 mM Tris. pH 7.6, 7 mM $MgCl_2$, 150 mM KCl. and 10 mM 2-mercaptoethanol. The retention times are indicated: filled bars M547, open bars M676, gray bars D666A, hatched bars M665. The indicated concentrations are those in the eluted peaks. Similar results were obtained in 20 mM Tris, pH 7.6, 10 mM $MgCl_2$.

mutant). The two variants could be overproduced as stable proteins in *E. coli* cells. They were purified to homogeneity, as judged by SDS−PAGE analysis (Figure 6). The thermal stabilities of the mutant enzymes were kept intact, as compared to that of the wild-type dimeric enzyme (half-life time of 16 min at 51 °C). The oligomeric states of D666A and M665 were studied by using native PAGE analysis and molecular sieving methods. As shown in Figure 6, the two mutant enzymes clearly behave as monomers on a native gel, although a faint band corresponding to the dimer is detectable in the case of D666A. Interestingly, when 10mM $MgCl_2$ is added to the gel buffer, the migration pattern of the D666A enzyme shifts toward that of the dimeric form, whereas M665 shows multiple bands indicative of an equilibrium between the monomeric and the dimeric forms (Figure 7C). Taking into account this salt-dependence, it appeared necessary to define the oligomeric state of the two mutants under the conditions routinely used for the biochemical characterization of MetRS. To achieve this, the retentions of the wild-type and mutant proteins on a HPLC molecular sieving column were compared (Figure 6). Two conditions were chosen: the first one corresponds to that used in the tRNA aminoacylation assay, i.e., 150 mM KCl plus 7 mM $MgCl_2$ (*21*), and the second one is that used in the fluorescence-monitored tRNA-binding assay (10 mM $MgCl_2$) (*22*). The results showed that, in either buffer, the mutants are monomeric, provided that their concentration is kept lower than 0.1 μM. Therefore, the two mutants must be monomeric under the conditions of the aminoacylation assay, where the catalyst is added at a concentration on the order of 0.3 nM. Although accurate dimer dissociation constants cannot be deduced from the chromatographic data, they could be estimated on the order of 1 μM for the two mutant proteins. As noted above, using PAGE analysis, the truncated M655 mutant appeared less prone to dimerization than the point mutant D666A. The whole data indicate that a mixture of monomers and dimers should form under the conditions of the tRNA-binding assay, when the enzyme concentration is in the micromolar range.

Biochemical Characterization of the Mutant Enzymes. The kinetic parameters of the two mutants were measured and compared to those of the dimeric M676 and minimal monomeric M547 enzymes (Table 2). Regarding the aminoacylation reaction, both mutant enzymes, monomeric in the assay, showed an increase in the Km of $tRNA_f^{Met}$. However, as compared to that with wild-type M676, the increase did not exceed a factor of 2. To facilitate comparisons, the k_{cat} constants in Table 2 were systematically calculated for one protomer. Their values are slightly increased in the cases of one protomer of D666A or of one protomer of M665, as compared to one protomer of M676. The gains in k_{cat} at saturating tRNA remain small, yet significant. Surprisingly, for the two mutants, the rate of the ATP−PPi exchange reaction was reduced by a factor of 2, whereas the k_{cat} values are identical for M676 and M547 (*1* and Table 1). This might indicate that, in the presence of the C-terminal domain, dimerization influences the conformation of the active site. This effect is not reversed upon saturation of the tRNA-binding site, since the ATP−PPi exchange rates were not significantly changed in the presence of 5 μM $tRNA_f^{Met}$ (data not shown).

In a second set of experiments, the role of the C-terminal domain in tRNA binding was studied. Saturation of MetRS by $tRNA_f^{Met}$ causes an important quenching of the intrinsic enzyme fluorescence (*23*). This quenching is by 60% in the case of the M547 monomer and by 37% in the case of the M676 dimer. This difference can be explained by the anticooperative binding of a single tRNA molecule per dimer under the used conditions. Moreover, the dissociation constant of $tRNA_f^{Met}$ from the strong site in the M676 dimer

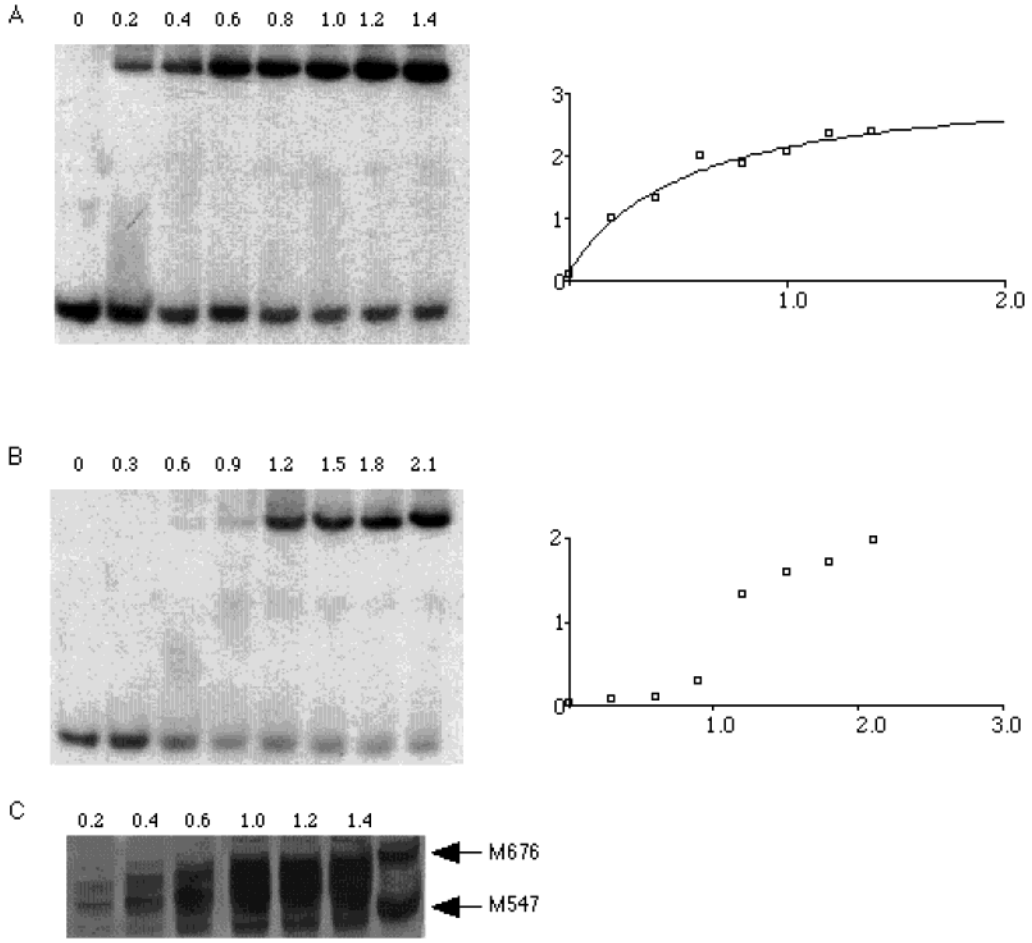


FIGURE 7: Binding of $tRNA_f^{Met}$ to MetRS variants. Panel A: gel shift assay of M676 associated to radioactively labeled $tRNA_f^{Met}$. The enzyme was incubated at increasing concentrations (0.2−1.4 μM) with 15 nM 5′-^{32}P-$tRNA_f^{Met}$ and electrophoresed on a native acrylamide gel. The curve on the right corresponds to a plot of the intensity of the shifted band as a function of enzyme concentration. It is fitted according to a simple binding equilibrium. Panel B: same experiment as that in panel A, with the M665 mutant. Panel C: same experiment as that in panel B, with the M665 mutant, but silver stained. This staining allows visualization of the proteins and shows the equilibrium between the monomeric and dimeric forms of M665 under the exact conditions of the gel shift assay. Migrations of M676 and M547 are indicated by arrows.

Table 1: Native Data Used in Structure Determination of MetRS−C12K[a]

data set	native
X-ray source	ID14eh1
unique reflections	9286
resolution (Å)	2
completeness	97.3 (75.9)
redundancy	3.0 (1.9)
R_{sym} (I) (%)[b]	5.2 (20.8)
R/free-R (%)	27.5/28.9
rms deviatiations bonds (Å)	0.0086
angles (deg)	1.53
mean B values (Å^2) protein atoms	68.6
water	59.5

[a] Each data set was collected at 100 K with a single crystal. The values in parentheses correspond to the highest shell of resolution (2.06−2.01 Å). Note that the relatively low completeness is due to the 2.02−2.01 Å shell; completeness is 92.1% in the 2.07−2.02 Å shell. [b] R_{sym} (I) = $[\sum_{hkl}\sum_i|\langle I_{hkl}\rangle - I_{hkl,i}|]/[\sum_{hkl}\sum_i|I_{hkl}|]$, where i is the number of reflection hkl.

Table 2: Michaelis−Menten Parameters of the Aminoacylation Reaction and of the [32P]PPi-ATP Isotopic Exchange Catalyzed by Ec-MetRS Variants[a]

	aminoacylation of $tRNA_f^{Met}$		[32P]PPi-ATP isotopic exchange		tRNA binding
	K_m (μM)	k_{cat} (s^{-1})	K_m (μM)	k_{cat} (s^{-1})	K_d (μM)
M676	4.4 ± 0.5	2.8	20 ± 3	42	0.20 ± 0.05
M547	5.8 ± 0.7	3.6	22 ± 3	47	2.4 ± 0.4
D666A	6.4 ± 0.7	3.1	29 ± 4	23	0.40 ± 0.05
M665	8.7 ± 1.0	4.4	34 ± 4	28	0.80 ± 0.2

[a] The values were determined by using homogeneous enzyme in standard assays as described in Materials and Methods. Parameters and standard errors were derived from iterative nonlinear least-squares fits of the Michaelis−Menten equation to the measured values. k_{cat} values were calculated for one enzyme protomer. In the case of ATP−PPi isotopic exchange, K_m is that of methionine measured at 2 mM ATP−Mg^{2+}. K_d values were measured by spectrofluorometric titrations of the enzyme, assuming one tRNA-binding site per M547 monomer and one tRNA-binding site per dimer in the other cases.

(0.2 μM) is 12-fold lower than that from the single binding site on the M547 truncated monomer (K_d value of 2.4 μM; *23* and Table 2). Affinity co-electrophoresis experiments showed a similar enhancement of the affinity of a $tRNA_f^{Met}$ acceptor microhelix to the M676 as compared to that of the M547 MetRS (*24*). In the cases of the M665 and D666A mutants, the quenchings upon tRNA saturation were by 40% and 37%, respectively. Such amplitudes rather argue in favor of a half of the site-binding behavior of the mutants in the titration assay. Nevertheless, dissociation constants of the two mutant enzyme:tRNA complexes were increased as

compared to that with the wild-type dimer (Table 2), showing that perturbation of the dimer interface had affected the strength of the tRNA binding.

Finally, formations of the enzyme:tRNA complexes were also studied by using a gel-shift assay for M676 and for the most affected mutant M665. In these experiments enzyme concentrations were varied in the range 0.2−2 μM. For the M676:tRNA complex, a K_d value of 0.6 ± 0.2 μM could be deduced (Figure 7). This value agrees well with that obtained using affinity co-electrophoresis (*24*). With the M665 variant, the tRNA-binding curve displayed a sigmoidal shape, with a sharp transition at an enzyme concentration of 1.2 μM. This may indicate that tRNA preferentially binds the dimeric form of the enzyme. The oligomeric state of the mutant enzyme in the gel shift assay was assessed by silver staining of the proteins. As can be observed in panel C of Figure 7, the migration profile shows a multiple band pattern indicative of an equilibrium between the monomeric and the dimeric forms. Upon increasing the enzyme concentration, the amount of the dimeric form increases. Since the autoradiography shows a single shifted band, one can conclude that the enzyme:tRNA complex preferentially forms with the dimeric protein. The sharp transition of the curve could therefore be explained by similar equilibrium constants for dimer formation from M665 protomers and for tRNA binding by dimers of this mutant.

CONCLUDING REMARKS

The data presented in this work establish that the C-terminal domain appended to dimeric MetRS provides the enzyme with additional tRNA-binding affinity, through general tRNA-binding properties. Therefore, this domain is indeed a structural and functional homologue of the Trbp111 protein.

E. coli MetRS mutants still containing the OB-folded tRNA-binding domain but affected in their capacity to dimerize were designed. Study of these mutants strongly indicates that the enhancement of tRNA-binding affinity conferred by the C-terminal domain is only expressed when the enzyme is in the dimeric state. The requirement for a dimeric form of the generic Trbp module to obtain full tRNA-binding affinity agrees with the low tRNA-binding affinity exhibited by monomeric EMAP II protein (K_d of 20 μM, *16*). Moreover, the requirement for a dimeric state is in keeping with the previous proposal that the tRNA-binding cleft in Trbp111 is formed by both the L1-L2 loops and the dimer interface (*12*). In such a model, only one tRNA site can be occupied at a time, in agreement with the experimentally measured stoichiometry of one tRNA per Trbp111 dimer (*7*). This view also explains the anticooperative behavior observed for *E. coli* dimeric MetRS upon binding two tRNA molecules (*23*, *25*, *26*). Indeed, it was early demonstrated that the enzyme binds a first tRNA molecule with strong affinity, and a second one with a much lower affinity. It can be imagined that only the first bound tRNA benefits from an access to the C-terminal domain. In this context, the question whether the strongly bound tRNA cross bridges the MetRS subunits through other contacts remains open. Binding of the tRNA molecule might occur with the enzyme on one side and MetRS−C12K on the opposite side, as proposed by Nomanbhoy et al (*27*). The size of the linker peptide (20 residues) between the MetRS core and dimerization domain is sufficient to allow such a complex.

It has been previously proposed that the anticooperative binding of two tRNAs to dimeric MetRS might serve to couple the working conditions of the enzyme to the concentration of free $tRNA^{Met}$ (*25*). Indeed, at high tRNA concentration, occupation of the second weak binding site would stimulate tRNA dissociation from the strong binding site. However, under the conditions of the tRNA aminoacylation assay, the advantage in catalytic turnover of native MetRS as compared to the monomeric minimal core enzyme does not exceed a factor of 1.5 (*21*, Table 1). With the *Bacillus stearothermophilus* MetRS, the kinetic consequences of the anticooperative behavior are more apparent, with the occurrence of two sets of K_m and V_m values upon varying the tRNA substrate (*28*).

The case of *S. cerevisiae* MetRS is particular, since this monomeric enzyme associates with the Arc1p protein. This association confers on MetRS a much higher affinity for tRNA. However, it is notable that Arc1p, which contains an EMAP II-like C-terminal domain (Figure 1), is monomeric. In addition to this domain, Arc1p possesses an N-terminal domain important for MetRS binding and a central domain also involved in tRNA binding (*29*). It can be imagined that the central domain compensates for the absence of dimerization of this tRNA-binding protein.

Actually, nonessential appended tRNA-binding domains are also found in other aminoacyl-tRNA synthetases, such as yeast AspRS (*30*) or mammalian LysRS (*31*−*33*). In most cases, the catalytic parameters of the tRNA aminoacylation reaction are only slightly modified upon removal of these appended domains (see the cases of plant or human MetRS, *34* and *35*, or of mammalian LysRS, *31*−*33*). In contrast, the N-terminal appended domain of yeast AspRS and the interaction of yeast MetRS with Arc1p are associated to larger decreases of K_m values for tRNA (*9*, *30*). Therefore, the understanding of the biological roles of these supplementary domains deserves further in vivo functional studies, as well as structural characterization of the complexes with tRNA. One idea is that these domains can be also involved in autoregulation of the expression of the aaRS genes. Such a genetic control has recently been evidenced in the case of yeast aspartyl-tRNA synthetase (M. Frugier and R. Giegé, Asilomar Conference on Aminoacyl-tRNA Synthetases, Pacific grove, CA, January 13−18, 2002). In the case of MetRS, regulation of the expression of the corresponding gene might benefit from the anticooperative behavior of the dimeric form. Indeed, it was earlier proposed that MetRS might down regulate the transcription of its own gene by binding a tRNA-like structure at the 5′ end of its mRNA (*36*). A dimeric structure of MetRS might be important in such a mechanism. Indeed, upon conditions of limiting free tRNA where the weak tRNA site remains unoccupied, the MetRS would strongly bind its mRNA. In contrast, in the presence of an excess of free tRNA, dissociation of the repressor MetRS from mRNA would be facilitated. Eventually, such a behavior would couple repressor binding and MetRS expression to the degree of aminoacylation of $tRNA^{Met}$.

MATERIALS AND METHODS

Expression and Purification of MetRS Variants. In the case of *Pyrococcus abyssi*, the *metS* gene was PCR-amplified and cloned between the *Nco*I and *Sac*II sites of pET15lpa, a modified version of pET15b (Novagen). For the purification of MetRS−C12K, the resulting pET15bMet-pa was transformed into *E. coli* BL21-De3 cells. Cells were grown at 37 °C, in 1 L of 2xTY medium containing 50 μg/mL of ampicillin to an OD_{650} of 1. After induction with 1 mM IPTG, cells were further grown during 15 h at 37 °C. Cells were harvested and resuspended in 30 mL of buffer A (10 mM MOPS pH6.7, 10 mM 2-mercaptoethanol, 0.1 mM PMSF) and disrupted by ultrasonic disintegration. The crude extract was centrifuged and the supernatant heated for 20 min at 80 °C. After centrifugation, the supernatant was loaded onto a Q-HiLoad ion exchange column (1.6 × 10 cm; Amersham) equilibrated in buffer A containing 50 mM NaCl. The protein was eluted stepwise with 1 M NaCl. The collected fractions were dialyzed against buffer A containing 50 mM NaCl. The protein was then loaded on an S-Sepharose high-performance column (1.1 × 4 cm; Amersham) equilibrated in the same buffer. Protein was eluted stepwise with 500 mM NaCl. The collected fractions (6 mL) were then directly loaded onto a Superdex 75 molecular sieving column (1.6 × 60 cm, Amersham) equilibrated in buffer A containing 500 mM NaCl. The resulting protein was finally dialyzed against buffer A containing 25 mM NaCl before storage at 4 °C. About 12 mg homogeneous protein was obtained.

Variants of the *E. coli* MetRS were produced by site-directed mutagenesis of the pBSM676 plasmid (*37*) using PCR techniques. Mutated MetRS were produced and purified as previously described (*38*).

Mass Spectrometry and N-Terminal Sequencing. A sample of MetRS−C12K protein (60 μM) was dialyzed against water and diluted to 6 μM in 50% acetonitrile containing 0.1% formic acid. Mass spectrometry experiments were performed on an electrospray spectrometer (Quattro II, Micromass, Manchester, U.K.).

For N-terminal sequencing, 12 μg of MetRS−C12K protein was loaded on a 12% SDS−polyacrylamide gel. After electrophoresis, the protein was electrotransferred onto a ProBlot membrane (Perkin-Elmer) and stained with amido black. The band containing the MetRS−C12K protein was cut and submitted to six cycles of automatic Edman degradation (Applied Biosystems 473A sequencer).

Crystallization and Data Collection. Initial screening for crystallization conditions of MetRS−C12K protein was performed at 4 °C with the hanging drop technique using protein at 6 mg/mL in storage buffer and commercial Crystal Screen solutions (Hampton Research). Hexagonal crystals were obtained using condition 42 of SM1 (20% PEG8000, 50 mM K−phosphate). The size of the crystals could be increased by using sitting drops at 24 °C. For data collection, the crystals were soaked during 1 min in 24% PEG8000, 50 mM KH_2PO_4 containing 10% ethyleneglycol as cryo-protectant, and flash cooled in liquid ethane. Data were collected at 100K using a synchrotronic source at the ESRF (ID14-EH1 beamline). Diffraction images were analyzed using Mosflm (A. G. W. Leslie, Laboratory of Molecular Biology, Cambridge, U.K.). Data were processed further with programs of the CCP4 package (*39*).

Structure Determination and Refinement. The structure was solved by molecular replacement by using human EMAP II (residues 10−95) as a starting model and the AMORE program (*40*). The structure was refined by cycles of manual model building using O (*41*) and conjugate gradient minimization using CNS (*42*). The final model contains 106 residues out of the 107 composing the MetRS−C12K protein and 16 water molecules. Residue R74 (loop L2) is disordered and has not been included in the model. In addition, the side chains of K4, K22, N26, D28, K29, K50, K53, E55, K72, and R90 were disordered, and the corresponding residues were therefore modeled as alanines. The final working and free *R*-factors were 27.5% and 28.9%, respectively, for the data to 2 Å resolution.

Activity Measurements. Methionine-dependent [^{32}P]PPi-ATP exchange activity was measured at 25°C in standard buffer (20 mM Tris-HCl, 7 mM $MgCl_2$, 10 mM 2-mercapto-ethanol, 0.1 mM EDTA, pH 7.6) containing 2 mM [^{32}P]PPi, 2 mM ATP and 2 mM methionine (*43*). Aminoacylation assays (*21*) were performed in the same standard buffer plus 150mM KCl, in the presence of 2 mM ATP, 50 μM L-[^{14}C]methionine (Amersham; 1.6 TBq/mol), and 5 μM *E. coli* $tRNA_f^{Met}$. For K_m measurements, the concentration of tRNA was varied from 0.2 to 16 μM.

Fluorescence at Equilibrium. Variations of the intrinsic fluorescence of MetRS variants upon titration with $tRNA_f^{Met}$ were followed as described (*22*), at 25 °C, in 20 mM Tris-HCl (pH7.6), 10 mM 2-mercaptoethanol, 10 mM $MgCl_2$, and 0.1 mM EDTA. Saturation curves were obtained by varying the total concentration of $tRNA_f^{Met}$ (1250 pmol of methionine acceptance per A_{260} unit, prepared from overproducing cells, *44*) from 0.03 to 3.6 μM. Dissociation constants were determined from iterative nonlinear least-squares fits of the theoretical equations to the experimental values, after correction for the inner filter effect (*22*). Confidence limits and standard errors were determined by 100 Monte Carlo simulations from the experimental deviations on individual measurements (*45*).

Gel Shift Assay for tRNA Binding. The gel shift assays were performed as described in ref *7*. *E. coli* MetRS variants were incubated at 4 °C at increasing concentrations with 15 nM 5′-[^{32}P]-labeled $tRNA_f^{Met}$ (1200 pmole/A_{260} unit, purified as described in ref *44*), in a buffer containing 50 mM Tris−borate (pH 8.3), 5 mM $MgCl_2$, and 0.001% Triton X-100. The gel was run at 100 V for 3 h at 4 °C. After migration, the gel was dried onto a Nytran membrane and exposed on an X-ray film. Autoradiographies were then analyzed using the NIH-Image software. Silver staining was performed using the Bio-Rad kit, as recommended by the supplier.

ACKNOWLEDGMENT

The authors are grateful to Sophie Bourcier (Laboratoire de Chimie des Mécanismes Réactionnels, Ecole Polytechnique) for determination of the Mw of MetRS−C12K by mass spectrometry and to Jacques d'Alayer and Maryline Davi (Laboratoire de Microséquençage des Protéines, Institut Pasteur, Paris) for protein sequencing. Thibaut Crépin was a recipient of a doctoral fellowship from the Association pour la Recherche sur le Cancer.

REFERENCES

1. Cassio, D., and Waller, J. (1971) *Eur. J. Biochem. 20*, 283−300.
2. Mechulam, Y., Schmitt, E., Maveyraud, L., Zelwer, C., Nureki, O., Yokoyama, S., Konno, M., and Blanquet, S. (1999) *J. Mol. Biol. 294*, 1287−97.
3. Sugiura, I., Nureki, O., Ugaji-Yoshikawa, Y., Kuwabara, S., Shimada, A., Tateno, M., Lorber, B., Giege, R., Moras, D., Yokoyama, S., and Konno, M. (2000) *Struct. Fold. Des. 8*, 197−208.
4. Hountondji, C., Dessen, P., and Blanquet, S. (1986) *Biochimie 68*, 1071−1078.
5. Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990) *Nature 347*, 203−206.
6. Cusack, S., Härtlein, M., and Leberman, R. (1991) *Nucleic Acids Res. 19*, 3489−3498.
7. Morales, A. J., Swairjo, M. A., and Schimmel, P. (1999) *EMBO J. 18*, 3475−83.
8. Muller, J. P., Bron, S., Venema, G., and van Dijl, J. M. (2000) *Microbiology 146*, 77−88.
9. Simos, G., Segref, A., Fasiolo, F., Hellmuth, K., Shevchenko, A., Mann, M., and Hurt, E. C. (1996) *EMBO J. 15*, 5437−48.
10. Kao, J., Ryan, J., Brett, G., Chen, J., Shen, H., Fan, Y. G., Godman, G., Familletti, P. C., Wang, F., Pan, Y. C., and et al. (1992) *J. Biol. Chem. 267*, 20239−47.
11. Shalak, V., Kaminska, M., Mitnacht-Kraus, R., Vandenabeele, P., Clauss, M., and Mirande, M. (2001) *J. Biol. Chem. 276*, 23769−76.
12. Swairjo, M. A., Morales, A. J., Wang, C. C., Ortiz, A. R., and Schimmel, P. (2000) *EMBO J. 19*, 6287−98.
13. Kawaguchi, S.-I., Müller, J., Linde, D., Kuramitsu, S., Shibata, T., Inoue, Y., Vassylyev, D., and Yokoyama, S. (2001) *EMBO J. 20*, 562−569.
14. Kim, Y., Shin, J., Li, R., Cheong, C., Kim, K., and Kim, S. (2000) *J. Biol. Chem. 275*, 27062−8.
15. Renault, L., Kerjan, P., Pasqualato, S., Menetrey, J., Robinson, J. C., Kawaguchi, S., Vassylyev, D. G., Yokoyama, S., Mirande, M., and Cherfils, J. (2001) *EMBO J. 20*, 570−8.
16. Quevillon, S., Agou, F., Robinson, J. C., and Mirande, M. (1997) *J. Biol. Chem. 272*, 32573−9.
17. Shine, J., and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 1342−1346.
18. Lecompte, O., Ripp, R., Puzos-Barbe, V., Duprat, S., Heilig, R., Dietrich, J., Thierry, J. C., and Poch, O. (2001) *Genome Res. 11*, 981−93.
19. Tompa, M. (1999) *Proc. Int. Conf. Intell Syst. Mol. Biol*. 262−71.
20. Murzin, A. G. (1993) *EMBO J. 12*, 861−7.
21. Lawrence, F., Blanquet, S., Poiret, M., Robert-Gero, M., and Waller, J. P. (1973) *Eur. J. Biochem. 36*, 234−43.
22. Blanquet, S., Petrissant, G., and Waller, J. P. (1973) *Eur. J. Biochem. 36*, 227−33.
23. Blanquet, S., Iwatsubo, M., and Waller, J. P. (1973) *Eur. J. Biochem. 36*, 213−26.
24. Gale, A. J., Shi, J. P., and Schimmel, P. (1996) *Biochemistry 35*, 608−15.
25. Blanquet, S., and Dessen, P. (1976) *J. Mol. Biol. 103*, 765−84.
26. Dessen, P., Blanquet, S., Zaccai, G., and Jacrot, B. (1978) *J. Mol. Biol. 126*, 293−313.
27. Nomanbhoy, T., Morales, A. J., Abraham, A. T., Vortler, C. S., Giege, R., and Schimmel, P. (2001) *Nat. Struct. Biol. 8*, 344−8.
28. Kalogerakos, T., Dessen, P., Fayat, G., and Blanquet, S. (1980) *Biochemistry 19*, 3712−23.
29. Simos, G., Sauer, A., Fasiolo, F., and Hurt, E. C. (1998) *Mol. Cell 1*, 235−42.
30. Frugier, M., Moulinier, L., and Giege, R. (2000) *EMBO J. 19*, 2371−80.
31. Shiba, K., Stello, T., Motegi, H., Noda, T., Musier-Forsyth, K., and Schimmel, P. (1997) *J. Biol. Chem. 272*, 22809−16.
32. Stello, T., Hong, M., and Musier-Forsyth, K. (1999) *Nucleic Acids Res. 27*, 4823−9.
33. Francin, M., Kaminska, M., Kerjan, P., and Mirande, M. (2002) *J. Biol. Chem. 277*, 1762−9.
34. Kaminska, M., Deniziak, M., Kerjan, P., Barciszewski, J., and Mirande, M. (2000) *EMBO J. 19*, 6908−17.
35. Kaminska, M., Shalak, V., and Mirande, M. (2001) *Biochemistry 40*, 14309−14316.
36. Dardel, F., Panvert, M., and Fayat, G. (1990) *Mol. Gen. Genet. 223*, 121−33.
37. Mellot, P., Mechulam, Y., Le Corre, D., Blanquet, S., and Fayat, G. (1989) *J. Mol. Biol. 208*, 429−43.
38. Fourmy, D., Meinnel, T., Mechulam, Y., and Blanquet, S. (1993) *J. Mol. Biol. 231*, 1068−77.
39. Collaborative computational project No.4. (1994) *Acta Cristallogr. D50*, 760−763.
40. Navaza, J. (1994) *Acta Cryst A50*, 157−163.
41. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
42. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−21.
43. Blanquet, S., Fayat, G., Waller, J. P., and Iwatsubo, M. (1972) *Eur. J. Biochem. 24*, 461−9.
44. Meinnel, T., and Blanquet, S. (1995) *J. Biol. Chem. 270*, 15908−15914.
45. Dardel, F. (1994) *Comput. Applic. Biosci. 10*, 273−275.
46. Evans, S. V. (1993) *J. Mol. Graphics. 11*, 134−138.

BI026343M

doi:10.1016/S0022-2836(03)00917-3 J. Mol. Biol. (2003) 332, 59–72

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Use of Analogues of Methionine and Methionyl Adenylate to Sample Conformational Changes During Catalysis in *Escherichia coli* Methionyl-tRNA Synthetase

Thibaut Crepin[1], Emmanuelle Schmitt[1], Yves Mechulam[1]* Peter B. Sampson[2], Mark D. Vaughan[2], John F. Honek[2] and Sylvain Blanquet[1]

[1]*Laboratoire de Biochimie Unité Mixte de Recherche n°7654, CNRS-Ecole Polytechnique, F-91128 Palaiseau cedex, France*

[2]*Department of Chemistry University of Waterloo Waterloo, Ontario, Canada N2L3G1*

**Corresponding author*

Binding of methionine to methionyl-tRNA synthetase (MetRS) is known to promote conformational changes within the active site. However, the contribution of these rearrangements to enzyme catalysis is not fully understood. In this study, several methionine and methionyl adenylate analogues were diffused into crystals of the monomeric form of *Escherichia coli* methionyl-tRNA synthetase. The structures of the corresponding complexes were solved at resolutions below 1.9 Å and compared to those of the enzyme free or complexed with methionine. Residues Y15 and W253 play key roles in the strength of the binding of the amino acid and of its analogues. Indeed, full motions of these residues are required to recover the maximum in free energy of binding. Residue Y15 also controls the size of the hydrophobic pocket where the amino acid side-chain interacts. H301 appears to participate to the specific recognition of the sulphur atom of methionine. Complexes with methionyl adenylate analogues illustrate the shielding by MetRS of the region joining the methionine and adenosine moieties. Finally, the structure of MetRS complexed to a methionine analogue mimicking the tetrahedral carbon of the transition state in the aminoacylation reaction was solved. On the basis of this model, we propose that, in response to the binding of the 3′-end of tRNA, Y15 moves again in order to deshield the anhydride bond in the natural adenylate.

Keywords: methionine analogues; X-ray crystallography; catalysis

Introduction

Aminoacyl-tRNA synthetases (aaRS) catalyze the aminoacylation of their cognate tRNAs through a two-step mechanism involving activation of the amino acid by ATP to yield aminoacyl adenylate, followed by transfer onto the 3′-end of the tRNA molecule. As a result, the specificity and catalytic efficiency of aminoacyl-tRNA synthetases are key features of the translation process.

aaRS can be divided into two classes.[1,2] Class 1 aaRS have a catalytic domain built around a Rossmann fold and display two signature sequences: HIGH and KMSKS.[3,4] Class 2 aaRS possess a catalytic centre constructed around an antiparallel β-sheet containing three conserved sequence motifs.[5,6] Among class 1 aaRS, class 1a enzymes correspond to amino acids Leu, Val, Ile, Cys, Met, and Arg. LeuRS, ValRS and IleRS display relaxed specificities towards their cognate amino acid substrate and use proof-reading mechanisms to insure accurate tRNA aminoacylation.[7] Editing sites, distinct from the activating sites, have been

Abbreviations used: aaRS, aminoacyl-tRNA synthetases; DFM, L-difluoromethionine; TFM, L-trifluoromethionine.

E-mail address of the corresponding author: yves@botrytis.polytechnique.fr

acquired during evolution to enhance the specificity of these three aaRS.[8–12] In the case of ArgRS, precise recognition of the cognate amino acid is insured by electrostatic interactions with the guanidium group of arginine.[13,14] CysRS selects cysteine through coordination of an enzyme-bound zinc ion by the sulphur atom of the amino acid.[15] Methionine is recognized by MetRS through an induced-fit mechanism,[16] while homocysteine, a metabolic precursor of methionine, is rejected through an editing reaction.[17] The latter reaction involves cyclization of the misactivated homocysteine into homocysteine thiolactone.[18] However, MetRS remains able to transfer various amino acid analogues onto tRNA. The resulting misesterified tRNAs can direct incorporation of non-natural amino acids into proteins *in vivo*.[19–22]

Crystals of a monomeric fragment of *Escherichia coli* methionyl-tRNA synthetase (551 N-terminal residues) have been used to determine the 3D-structure of the free enzyme[23] and of the E:Met complex.[16] In the present study, a series of methionine and methionyl adenylate analogues were used in crystal soaking experiments. Analyses of the corresponding 3D structures have enabled us to describe the conformational changes of the synthetase and to make further progress in the description of methionine recognition and of methionyl adenylate shielding and deshielding. Information gained from this study may be useful to design efficient inhibitors against MetRS.[20,24–26] Recent studies have indeed identified ligands highly selective for bacterial synthetases.[27–29]

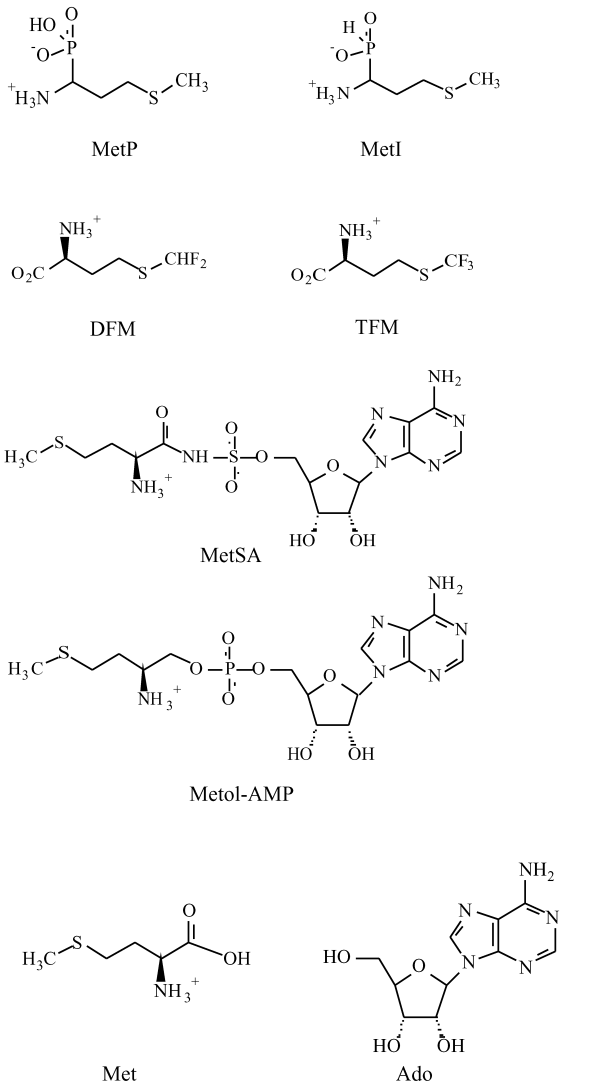


Figure 1. Analogues of methionine and of methionyl adenylate used in this study. MetP: methionine phosphonate (DL mixture); MetI: methionine phosphinate (DL mixture); DFM: L-difluoromethionine; TFM: L-trifluoromethionine; MetSA: 5′-*O*-[*N*-(L-methionyl)-sulphamoyl] adenosine; Metol-AMP: L-methioninyl adenylate; Met Ado: L-methionine plus D-adenosine.

Results

Binding of methionine and methionyl adenylate analogues to MetRS

Two groups of methionine analogues were used in this study. In the first group, MetP and MetI display a phosphonate or a phosphinate group, respectively, in place of the carboxyl group. In the case of the DFM and TFM analogues, the side-chain is modified, with the replacement of hydrogen atoms on the ε carbon by two or three fluorine atoms, respectively (Figure 1).

Interaction of methionine with MetRS is known to cause an increase in the intrinsic enzyme fluorescence.[30] By following the fluorescence change (37% at saturation of methionine), a K_d value of 43 μM could be determined for the E:Met complex. The binding of the above analogues to MetRS was therefore studied at equilibrium by using fluorescence spectroscopic techniques. With MetI, the fluorescence was increased by 36%. The K_d of this compound complexed to MetRS, is increased by a factor of 44 as compared to that of E:Met. In the three other cases (MetP, DFM and TFM), since addition of concentrations of these analogues up to 2 mM did not produce any measurable fluorescence change, dissociation constants could not be determined (Table 1). The binding experiments were repeated in the presence of either adenosine or adenosine plus pyrophosphate and magnesium. These compounds were expected to increase the affinity of the methionine analogues, because of known synergistic coupling between the ATP and methionine sites (see Blanquet *et al.*[31] and Table 1). Indeed the K_d of the various analogues complexed to the E:Ado:PP_i complex could be measured using this approach. The affinities were decreased by 130-fold (MetI), 60-fold (DFM) and 360-fold (TFM) as compared to that of the E:Met:Ado:PP_i complex (Table 1).

Table 1. Binding of methionine analogues to MetRS

	$\Delta\phi_{aa}$ (%)	K_{aa} (μM)	$\Delta\phi_{aa}^{Ado}$ (%)	K_{aa}^{Ado} (μM)	C^{Ado}	$\Delta\phi_{aa}^{Ado:PP_i}$ (%)	$K_{aa}^{Ado:PP_i}$ (μM)	C^{PP_i}
L-Met	36	43 ± 6	46	11 ± 2	4	65	0.6 ± 0.1	18
DL-MetP	NM	NM	NM	NM	–	NM	NM	–
DL-MetI	37	1900 ± 350	51	1600 ± 500	1.2	52	76 ± 8	20
L-DFM	NM	NM	19	800 ± 70	–	42	38 ± 5	21
L-TFM	NM	NM	NM	NM	–	37	220 ± 40	–

Dissociation constants were measured by following the variation of MetRS fluorescence ($\Delta\phi$) as a function of increasing amounts of the studied analogue. K_{aa} is the dissociation constant of the analogue from its complex with MetRS, K_{aa}^{Ado} is the dissociation constant of the analogue from its complex with MetRS:adenosine and $K_{aa}^{Ado:PP_i}$ is the dissociation constant of the analogue from its complex with MetRS:adenosine:pyrophosphate. The coupling constants are calculated as follows: $C^{Ado} = K_{aa}/K_{aa}^{Ado}$; $C^{PP_i} = K_{aa}^{Ado}/K_{aa}^{Ado:PP_i}$. NM: not measurable. $\Delta\phi_{aa}$ designates the relative variation of fluorescence observed at saturating concentrations of methionine, or of the methionine analogue under study. $\Delta\phi_{aa}^{Ado}$ and $\Delta\phi_{aa}^{Ado:PP_i}$ are the relative variations of fluorescence measured in the presence of adenosine, or adenosine plus PP_i-Mg^{2+}, respectively.

However, no binding of MetP could be observed by this method. Finally, two analogues of methionyl adenylate were also investigated: MetSA and Metol-AMP (Figure 1). Titration of MetRS with MetSA showed that this compound was tightly bound by MetRS, with a dissociation constant lower than 1 μM. This K_d value is of the same order as that previously measured with Metol-AMP.[30] While this manuscript was being completed, an independent study of MetSA has appeared, indicating that indeed this compound is quite potent ($IC_{50} \sim 7$ nM) against *E. coli* MetRS.[32]

The binding modes of the various methionine and methionyl adenylate analogues were then determined by X-ray crystallography after soaking native protein crystals in mother liquor containing millimolar concentrations of the studied ligands. Resolutions of the refined structures ranged from 1.7 Å to 1.93 Å. Each structure displayed very good geometry and the electronic density of the ligand could be unambiguously assigned (Table 2 and Figure 2).

Specificity of binding of the methionine side-chain

The methionine-binding pocket is made of the loop between βA and αA (A12, L13 and Y15), of αC (W253, A256, P257, Y260), of the $_{294}$GKD$_{296}$ loop and of αD (I297, H301 and W305). When using the free enzyme for comparison, binding of methionine induces the rearrangement of several aromatic residues. Among these residues, the side-chains of Y15, W253 and F300 rock towards the liganded amino acid. These rotations are propagated away from the active site, up to F304 and W229. All these movements create new stacking interactions between F300 and W253, on the one

Table 2. Data collection and refinement parameters

Crystal	DL-MetP	DL-MetI	L-Dfm	L-Tfm	MetSA	MetolAMP	Met-Ado-PP_i
Soaking time (minutes)	120	20	120	120	40	120	240
Concentration (mM)	5	1	10	10	10	10	3.75/2
Wavelength (Å)	0.964	0.9674	0.9340	0.964	0.9674	0.934	0.9793
X-ray source	LURE DW32	LURE DW32	ESRF-id14eh1	LURE DW32	LURE DW32	ESRF-id14eh1	ESRF-id29
Cell parameters							
a (Å)	78.4	79.0	78.6	78.5	78.8	78.0	77.8
b (Å)	45.0	45.5	45.2	45.3	45.2	45.6	45.7
c (Å)	85.9	86.3	86.0	86.3	86.4	87.3	87.5
β (deg.)	107.5	107.3	107.4	107.4	107.3	108.4	108.5
Unique reflections	52,574	45,381	63,973	53,727	43,035	45,389	53,467
Resolution (Å)	1.8	1.91	1.7	1.78	1.93	1.9	1.75
Completeness (%)	97.8 (93.4)	99.0 (98.4)	99.9 (99.8)	95.9 (91.1)	98.0 (97.7)	98.0 (93.0)	89.0[a] (89.0)
Redundancy	2.2	2.5	4.2	2.9	2.2	3.8	2.5
$R_{sym}(I)$ (%)[b]	5.9 (28.0)	7.7 (26.0)	5.1 (19.0)	5.4 (26.4)	7.2 (31.7)	9.3 (28.9)	6.4 (22.2)
R/Free-*R*[c] (%)	19.5/22.5	19.5/22.4	18.6/20.4	18.4/20.9	18.8/22.4	22.3/25.0	20.1/22.6
rmsd Bonds (Å)/angles (deg.)	0.0050/1.19	0.0063/1.22	0.0047/1.18	0.0049/1.22	0.0056/1.18	0.0056/1.18	0.0050/1.16
Water molecules	483	443	541	489	459	349	374
B Values (Å^2) protein/water/ligand	15.6/26.0/14.7	18.5/26.2/23.4	16.13/28.1/14.6	16.0/27.7/22.5	17.8/26.8/10.3	28.2/33.8/18.5	25.8/35.2/21.6/17.9

[a] The relatively low completeness of this dataset is due to a technical problem during data collection.

[b] $R_{sym}(I) = \frac{\sum_{hkl}\sum_i |\langle I_{hkl}\rangle - I_{hkl,i}|}{\sum_{hkl}\sum_i |I_{hkl}|}$, where *i* is the number of reflections *hkl*.

[c] Free-*R* was calculated using 5% of the reflections left out of refinement.

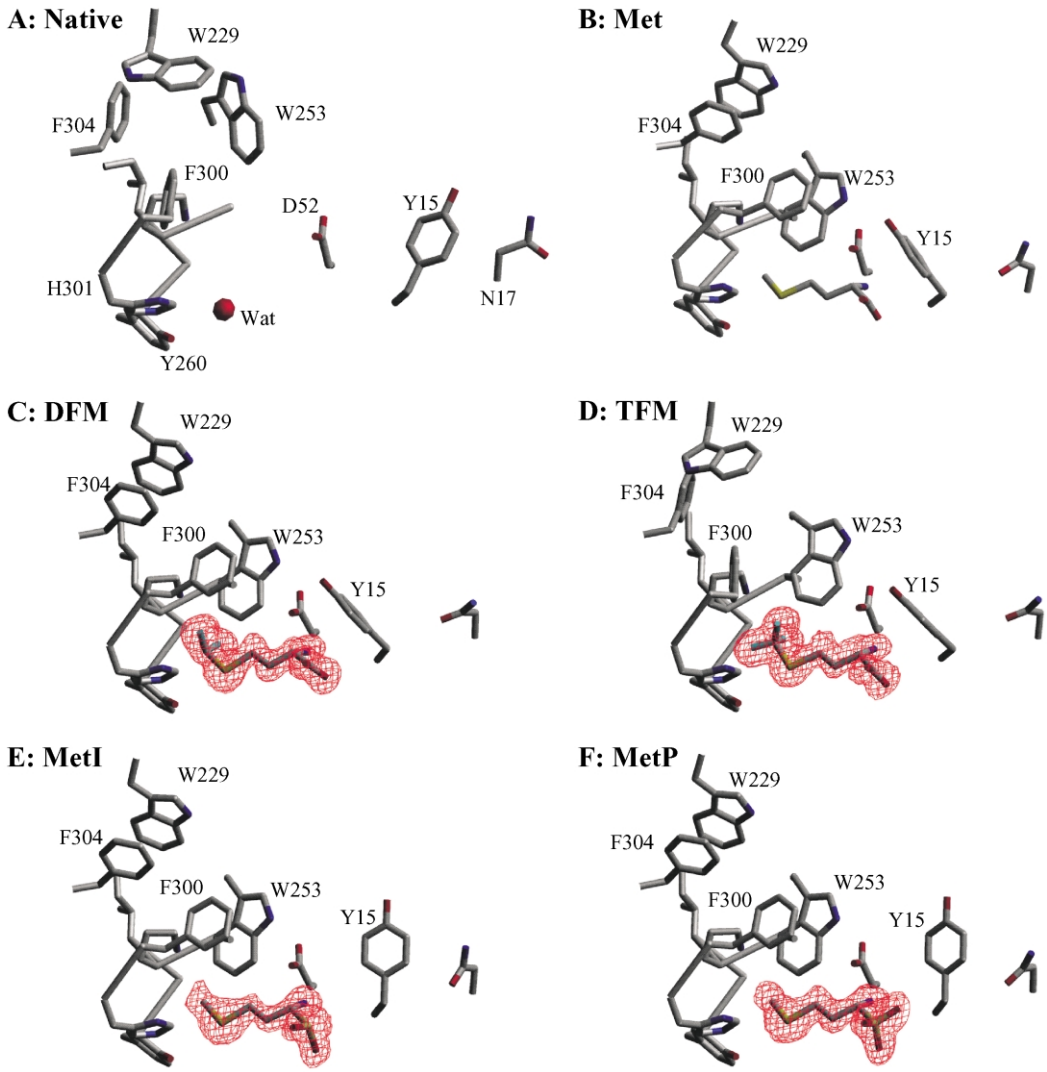


Figure 2. Rearrangement of the methionine-binding cavity upon binding of the various methionine analogues. Drawn are close views of the methionine-binding cavities in: A, free MetRS;[23] B, the MetRS:Met complex;[16] C, the MetRS:DFM complex; D, the MetRS:TFM complex; E, the MetRS:MetI complex; F, the MetRS:MetP complex. In the free enzyme, a water molecule occupies the position where the sulphur atom of methionine will bind. This water molecule is represented by a red ball. C–F, The omit $F_o - F_c$ electron density map, contoured at 3σ, is represented for each complex. During map calculations, the ligand atoms were omitted from the refined models. This Figure and those following were drawn with Setor.[57]

hand, and between F304 and W229, on the other hand, thereby contributing to a reduction of the volume of the binding pocket.[16]

Binding of DFM causes conformational changes identical with those induced by the binding of methionine. The structures of the two resulting complexes can be superimposed with an rms value of 0.208 Å for 545 C^{α} atoms compared. In both cases, the sulphur atom of the ligand appears to interact with the main-chain NH group of L13, the η oxygen of Y260 and the N^{ε} of His301. The role of His301 in amino acid recognition has already been highlighted by site-directed mutagenesis experiments.[33] We conclude that the presence in DFM of two fluorine atoms does not impair the movements leading to the creation of the amino acid-binding pocket. However, the fluorine atoms encounter unfavourable contacts with main-chain atoms of A256 and P257 that border the cavity. This observation may account for an affinity of DFM for MetRS lower than that of authentic methionine (Table 1).

In the case of the binding of TFM, the side-chains of W229, F300 and F304 remain in the open state characteristic of the free enzyme (Figure 2). W253 is rocked towards the analogue but does not reach the exact position encountered in the E:Met or the E:DFM complexes. Otherwise, steric hindrance by one fluorine atom of TFM would have occurred. In parallel, Y15 flips towards the NH group of TFM, but as above with W253, the movement of this residue is not complete. With TFM, the

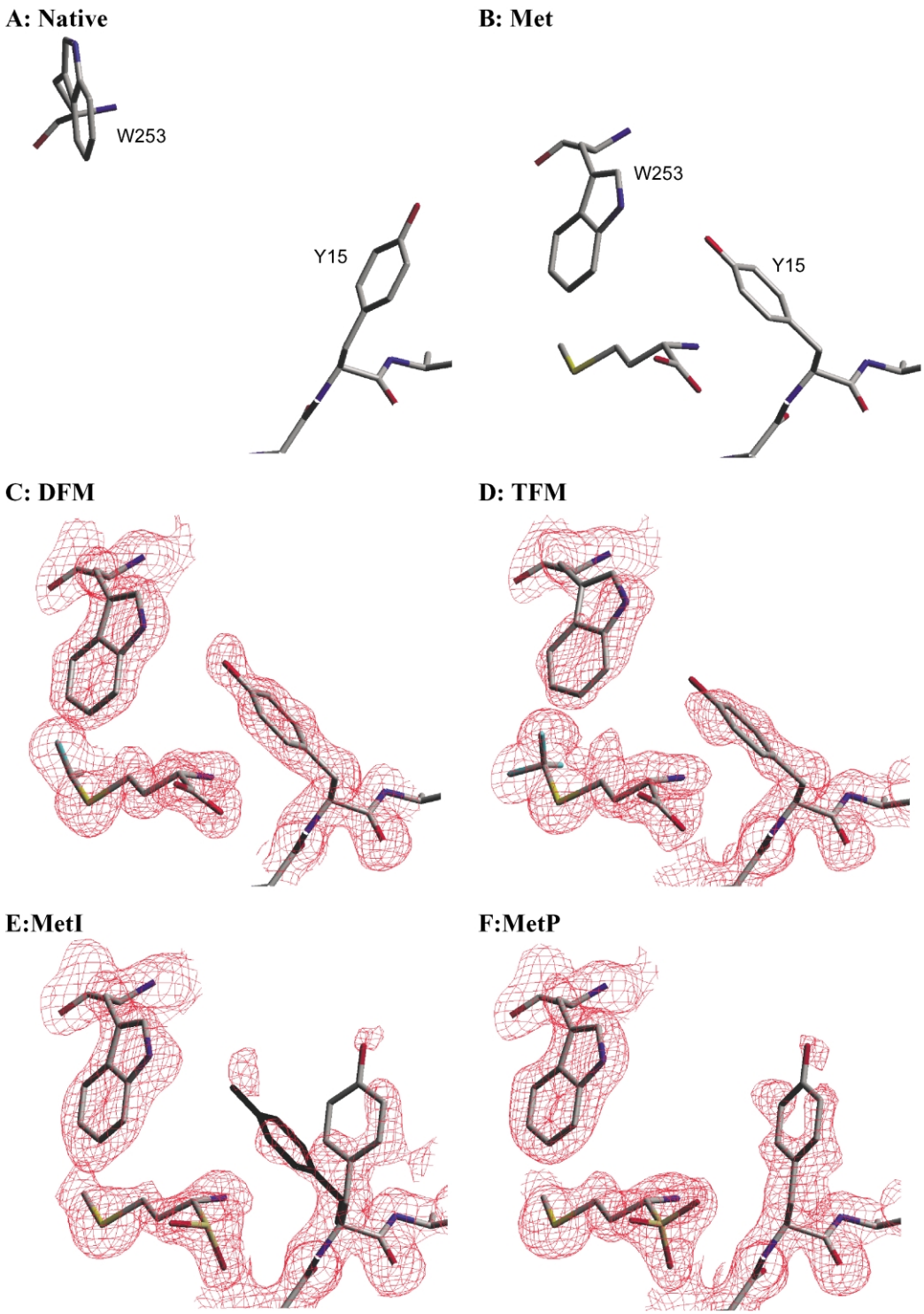


Figure 3. Electronic densities associated with Y15 and W253. A and B show the positions of W253 and Y15 in the free enzyme[23] (A) and in the MetRS:Met complex[16] (B). C–F, The final $2F_o - F_c$ electron density maps are represented. Each map is contoured at 1σ. C, MetRS:DFM complex; D, MetRS:TFM complex; E, MetRS:MetI complex; F, MetRS:MetP complex. E, The two alternative conformations of Y15 in the MetRS:MetI complex are shown.

distance between the OH group of the side-chain of Y15 and the NH group of the side-chain of W253 is 3.6 Å instead of 3.3 Å with DFM. As a consequence, the strength of the hydrogen bond between these two residues must be lowered. In good agreement with this view, the electronic densities of Y15 and of W253 in the MetRS:TFM complex are weakened as compared to those in the MetRS:DFM one (Figure 3). Therefore, W253 and Y15 control the size of the pocket for the

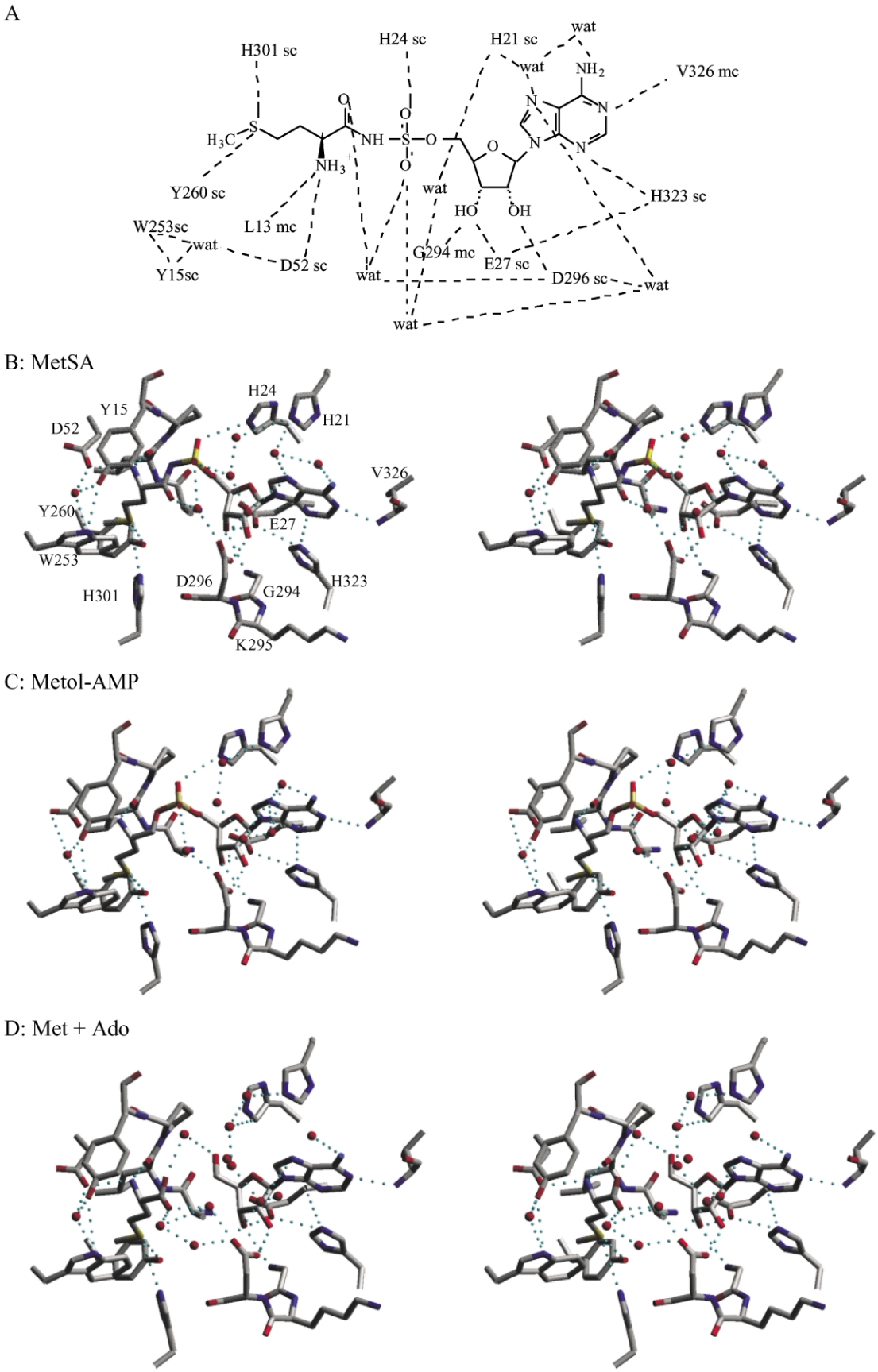


Figure 4. Binding of MetSA in the active site of MetRS. A, Schematic representation of the main electrostatic interactions between the enzyme and MetSA. Bottom: stereo views of active site-bound MetSA (B), Metol-AMP (C) and methionine plus adenosine (D). Only the enzyme residues relevant to the discussion in the text are drawn.

methyl group and hence participate in critically determining the specificity of MetRS for the methionine side-chain.

Role of the carboxylate moiety of methionine in the formation of the amino acid-binding pocket

In MetI and MetP, a phosphinate or a phosphonate group replaces the carboxylate group of methionine, respectively. The phosphonate group of MetP may resemble the transition state in the reaction of esterification of tRNA. Binding of MetI or of MetP on MetRS triggers conformational changes identical with those observed upon the complexation of methionine. The only deviation concerns the position of the side-chain of Y15. Examination of the electronic density shows that the position of this residue in the MetI and MetP complexes is not well defined (Figure 3). With the MetRS:MetP complex, the electronic density of Y15 mainly corresponds to the position encountered in the native enzyme. With the MetRS:MetI complex, the density is equally shared between two conformations. One conformation corresponds to that of Y15 in the free enzyme. The other one can be superimposed to that in the MetRS:Met complex. We conclude that the steric hindrance caused by the presence of the phosphonate group of MetP or the phosphinate group of MetI prevents Y15 from adopting the stable conformation reached in the MetRS:Met complex. With MetP, the destabilization of Y15 could reflect enzyme motions accompanying ester bond formation between the methionyl adenylate and the 3′-adenosine end of the tRNA.

Methionyl adenylate-binding site

Two analogues of methionyl adenylate were diffused into the MetRS crystals: MetSA and Metol-AMP (Figure 1). The binding pockets of MetRS for the two compounds are identical (Figure 4). The crevice for the adenine ring is on the N-side of βD. It is formed by βA and αA on the one side, and by β7 belonging to the KMSKS domain on the other side. The adenine base of ATP is held by hydrogen bonds involving the main-chain NH group of V326 and N^{δ} of H323. As shown in Figure 5, V326 is strictly conserved in class 1a aaRS. In IleRS and ValRS, a histidine residue plays the same role as that of H323 in MetRS. The ribose moiety adopts a C2′ *endo* conformation and is held by E27, D296 and the main-chain NH group of G294. The two latter residues belong to the GKD sequence downstream from the βD strand of the Rossmann fold. Both the glycine and the acidic residues in this GKD sequence are well conserved in class 1 aaRS (Figure 5). Several water molecules participate in the binding of the adenosine moiety of the methionyl adenylate analogues. They form a network involving the two histidine residues of the HLGH motif. The sulphamoyl group of MetSA (or the phosphate group of Metol-AMP) is held by H24 (second histidine residue of HLGH) and by D296 *via* a water molecule. Finally, the pocket for the methionyl moiety of MetSA or of Metol-AMP is identical with that found in the E:Met complex. All flipping motions of the aromatic residues characteristic of methionine binding are visible. The only significant difference is a rotation of the D296 side-chain. As a consequence, the interaction of D296 with the side-chain of Y325 is broken, and the latter residue is slightly displaced. Finally, the position of the side-chain of H24 is adjusted to obtain maximum interaction with the adenine ring.

Affinity of methionine is reinforced by fourfold in the presence of adenosine and by 72-fold in the presence of both adenosine and PP_i-Mg^{2+} (Table 2). To take advantage of this strong positive coupling, a crystal soaking experiment was attempted into a solution containing methionine, adenosine and PP_i-Mg^{2+}. The corresponding structure was refined to 1.75 Å resolution (Figure 4). Whereas methionine and adenosine were clearly visible in the electronic density, no trace of PP_i-Mg^{2+} association could be observed. Therefore, it seems that the crystals of MetRS do not accept PP_i-Mg^{2+}. Possibly, conformational changes required for the binding of the PP_i moiety are precluded by the packing of the molecules within the crystal. When the crystals were soaked in the presence of Met plus ATP-Mg^{2+}, only methionine was visible. Attempts to diffuse ATP-Mg^{2+} alone into the crystals were also unsuccessful. Diffraction was not altered by such soaking but ATP was not visible in electronic density. Again, crystal packing can be suspected to block motions in the enzyme required for ATP-Mg^{2+} binding and further adenylate formation. Residues of the KMSKS loop and of the zinc-binding domain (residues D138, R139) of MetRS are known to be involved in the catalysis.[34–36] However, in the complexes studied here, the side-chains of these critical residues (K335, K332, R139, D138) stay far away from the substrates. One may assume that they have to move to directly participate in catalysis (Figure 6). Flexibility of the zinc-binding domain is evidenced by the disappearance of its electronic density (residues 130–184) upon prolonged soaking in the presence of MetSA or of Metol-AMP. Clearly, these steps of the catalysis require further conformational changes which might be evidenced upon co-crystallization in another space group.

Discussion

Mechanism of formation of the methionine-binding pocket

The 3D structures of MetRS complexed with TFM or DFM show that the presence of fluorine atoms does not block the overall conformational rearrangement observed upon methionine binding.

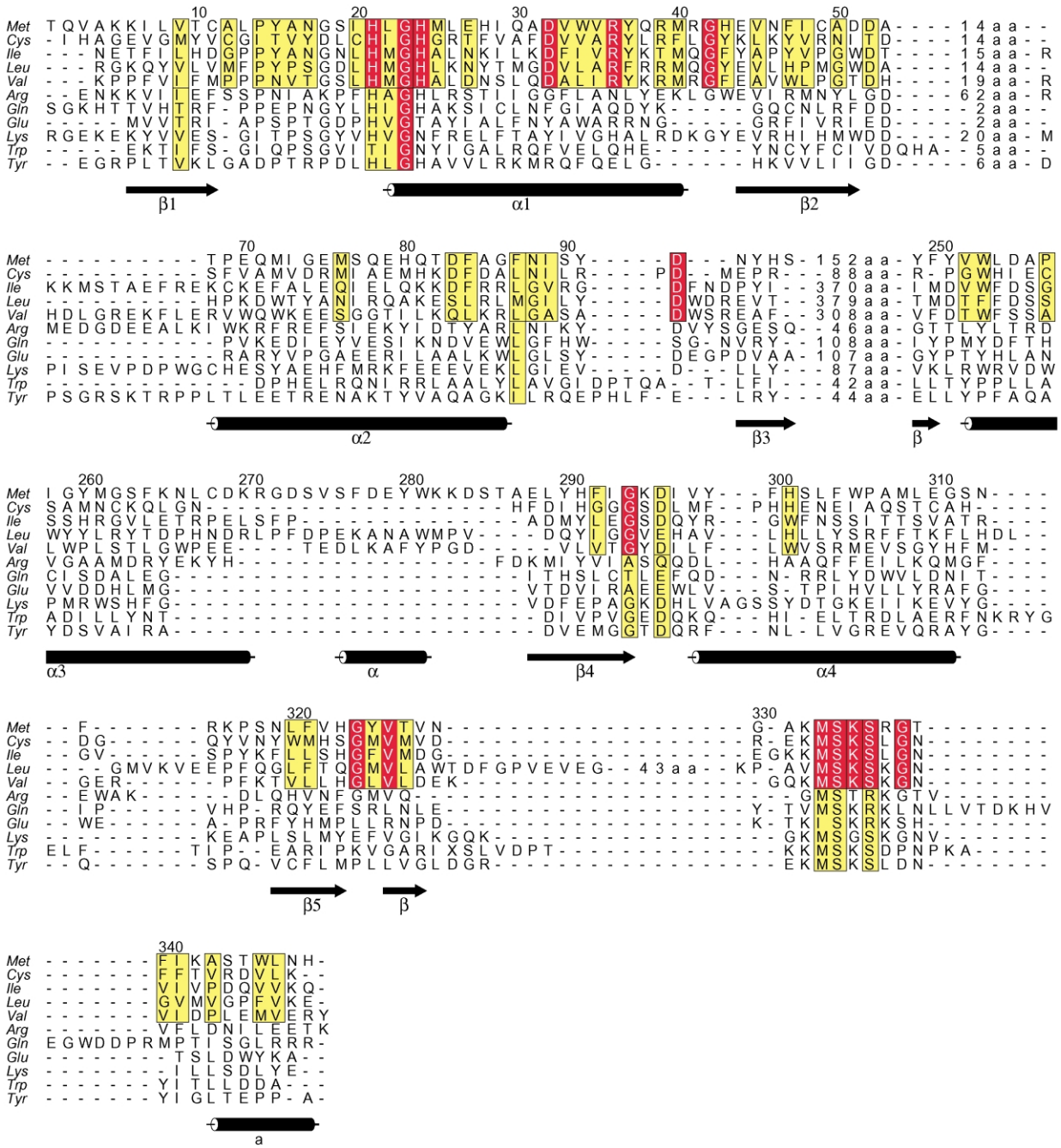


Figure 5. Alignment of the amino acid sequences of the Rossmann folds of class 1 aaRS based on their 3D structures. The secondary structures in MetRS are drawn below the alignment. Regions corresponding to variable insertions are omitted. Conservative replacements are boxed either for the whole set of class 1 aaRS or for the subset corresponding to class 1a aaRS excluding ArgRS. Conserved residues are shown with red boxes. Met: *E. coli* MetRS; Cys: *E. coli* CysRS; Ile: *T. thermophilus* IleRS; Leu: *T. thermophilus* LeuRS; Val: *T. thermophilus* ValRS; Arg: yeast ArgRS; Gln: *E. coli* GlnRS; Glu: *T. thermophilus* GluRS; Lys: *P. horikoshii* class 1 LysRS; Trp: *B. stearothermophilus* TrpRS; Tyr: *B. stearothermophilus* TyrRS. Numbering refers to the *E. coli* MetRS sequence. The Figure was drawn with Alscript.[58]

However, some steric hindrance occurs between A256, P257 and the fluorine groups. This effect may account for the reduced affinity of DFM for the enzyme, as compared to that of methionine. *In vivo*, DFM can be incorporated into proteins much more efficiently than TFM.[19,37] This behaviour suggests that TFM is less efficient than DFM in $tRNA^{Met}$ aminoacylation. The presence in TFM of one more fluorine atom indeed impairs the complete motion of W253. Moreover, rearrangements of several aromatic side-chains around the methionine-binding pocket are hindered. These observations suggest a picture of the methionine-binding process where the concerted movements of all the aromatic residues belonging to the amino acid pocket originate from a switch

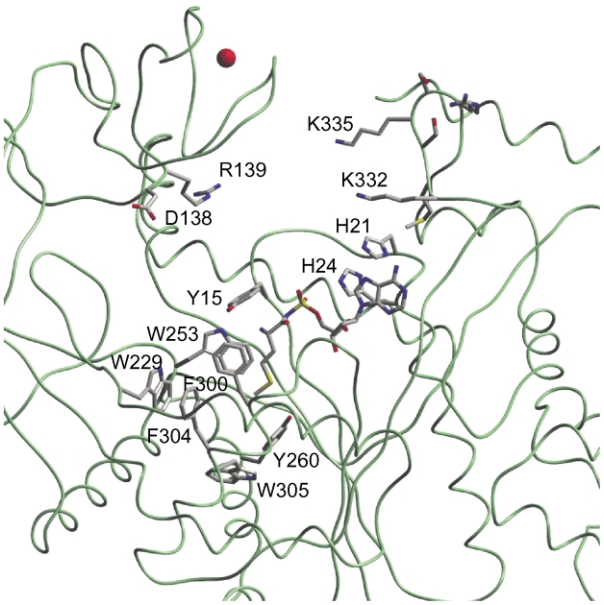


Figure 6. C^{α} Trace of MetRS in its complex with MetSA drawn with sticks. Side-chains of the aromatic residues contributing to the methionine-binding pocket are shown. Some residues known from site-directed mutagenesis experiments to participate in amino acid activation are also represented.[34–36,59] These include H21 and H24 (HIGH signature sequence), the $_{332}$KMSKS$_{336}$ loop, and D138 plus R139 in the zinc-binding region. The Figure shows how distant the $_{332}$KMSKS$_{336}$ and D138-R139 regions are from the centre of the MetSA ligand.

of W253. The movement of W253 allows F300 to stack on it, with concomitant release of F304. The cavity created by the displacement of W253 opens a new position for W229, stacked on F304. In parallel, the rotation of the Y15 side-chain stabilizes W253 in its final conformation.

Flexibility of the amino acid-binding pocket

Accommodation within the active site of MetRS of several methionine analogues indicates promiscuity of the amino acid-binding pocket of this synthetase.[17,21,22] Such a promiscuity might have played an important role in evolution to enable class 1a aaRS to recognize the modern hydrophobic amino acids involved in translation. In the cases of LeuRS, ValRS and IleRS, the amino acid-binding pockets also display flexibility and non-cognate natural amino acids can be activated efficiently.[7–12] To overcome their lack of selectivity, the three above aaRS have adopted editing strategies. In the cases of MetRS and CysRS, specificity of the activation reaction takes advantage of the occurrence of a sulphur atom in the amino acid substrate. With CysRS, complexation of cysteine does not promote strong rearrangement of the active site. Rather, tight specific recognition is achieved through the sulphur atom of the amino acid making the fifth coordination of a zinc ion in the active site.[15] As a consequence, the fidelity of Cys-AMP synthesis by CysRS does not require an editing mechanism.[38] Methionyl-tRNA synthetase appears as a hybrid case between CysRS and the three Leu, Ile and ValRS. The sulphur atom of methionine is recognized, though not as tightly as that of cysteine. Indeed, non-natural analogues devoid of sulphur atom, such as norleucine, can be activated.[17] It is quite interesting that a few residues involved in the specificity of binding of the methionine side-chain are conserved in other class 1a aaRS. For instance, W253 and H301 of MetRS correspond to W205 and H235 in *E. coli* CysRS (Figure 5). However, the roles of these two residues differ in the two aaRS. W205 of CysRS interacts with the sulphur atom of cysteine, while H235 appears not to be directly involved in the recognition of the amino acid.[15] In MetRS, as shown in this study, W253 is crucial to trigger structural rearrangements upon methionine binding and H301 interacts with the sulphur atom. Similar comments can be made with LeuRS. In this enzyme, H545, the histidine corresponding to H301 in MetRS, makes part of the hydrophobic pocket for leucine.[11] The variability of the functions of such conserved residues, which all belong to the amino acid recognition site, illustrates how a common ancestor with a promiscuous amino acid-binding pocket might have evolved to give the modern set of class 1a aaRS.

Roles of D296 and Y15 in the stability of the enzyme product complex and in catalysis

The structures of MetRS complexed to methionyl adenylate analogues highlight the movement of D296 and the precise adjustment of H24 for the docking of the α-phosphate of the ligand. Upon methionine binding, D296 is rejected. On the other hand, in the presence of a methionyl adenylate analogue, the side-chain of this residue flips back and interacts *via* a water molecule with the carbonyl oxygen of the methionine moiety and with a phosphoryl (or sulphamoyl) oxygen (Figure 4). In addition, the side-chain of D296 interacts with the 2′-OH group of the ribose. The latter interaction has been described with the corresponding residue, D194, in the case of TyrRS.[39] In the free enzyme, D296 bridges E27 *via* a water molecule positioned in place of the 2′-OH group of the ribose of the analogue. A plausible mechanism is the repulsion of the side-chain of D296 by the carboxylate of methionine. In agreement with this view, the structure of the MetRS:L-methioninol complex shows a position of D296 close to that which this residue holds in the free enzyme (data not shown). Such movements of D296 in the conserved GKD sequence suggest a particular role of this residue during the steps leading to catalysis. In agreement with this idea, the corresponding D194 in *Bacillus stearothermophilus* TyrRS makes an important contribution to the stabilization of the transition state of the tyrosine activation reaction.[40] These conformational changes may account for the

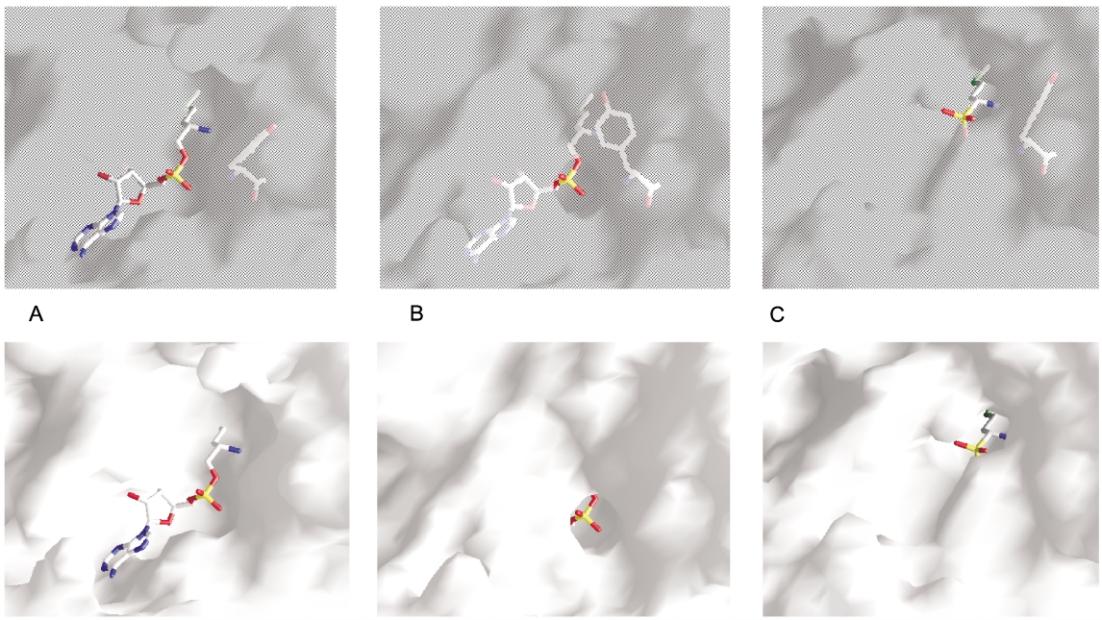


Figure 7. Molecular surface representation of the active site. In the top views, the surface are transparent, whereas they are opaque in the bottom views. In the top views, Y15 is drawn with sticks, thereby showing its contribution to shielding. A, Free MetRS containing Metol-AMP fitted within the catalytic pocket. This view is deduced from the superimposition of the 3D structure of the free enzyme with that of the MetRS:Metol-AMP complex. B, Same representation in the case of the MetRS:Metol-AMP complex. C, Same representation in the case of the MetRS:MetP complex. Comparison between A and B shows the burying of the adenylate in B. MetP is an analogue of the transition state of the aminoacyl transfer reaction. In this case, movement of Y15 renders the reactive group accessible. The Figure was drawn with Grasp.[60]

coupling free energy recovered upon simultaneous binding of methionine and adenosine. Indeed, D296 interacts with both the methionine and ATP moieties of the adenylate analogues, and might therefore be directly involved in synergistic bindings of the two substrates.

Another important residue for catalysis is Y15.[41] The movement of Y15 originates from the recognition of the methionine side-chain. It is impaired if the carboxylate of the amino acid is substituted by a bulky group such as a phosphonate or a phosphinate. The flexible switch of Y15 closes the cavity of methionine. The bound amino acid becomes completely buried within the enzyme. Such a motion of Y15 and the ensuing locking of the amino acid pocket account for the decreased affinities of the TFM and MetP analogues. Indeed, with these ligands, the position reached by Y15 must be less stable than that with L-methionine. In the case of the adenylate analogues, the adenine moiety of the ligands is shielded by a network of tightly bound water molecules (Figure 7). Notably, the MetP analogue appears to mimic the tetrahedral carbon of the transition state of the aminoacylation reaction. In the MetP:enzyme complex, Y15 is open, which suggests a similar motion upon attack of the adenylate (Figure 7). Y15 appears therefore to achieve multiple tasks. First, the tyrosine moves while the methionine pocket is formed. Motion of this residue is triggered by the amino acid itself and is mediated by the displacement of W253. This picture may explain the participation of Y15 in the editing reaction of MetRS as suggested from the study of a Y15A mutant. This mutant is able to cyclize methionyl adenylate into *S*-methyl homocysteine thiolactone as efficiently as it cyclizes homocysteinyl adenylate.[41] One can imagine that, in the absence of the Y15 side-chain, the active site cavity is large enough to allow the reactive anhydride bond in methionyl adenylate to be attacked by the S–CH_3 group of the activated methionine itself. With homocysteine as the substrate, W253 would no longer flip and the motion of Y15 would not occur. As a consequence, the recovery of free energy of binding would be limited. In agreement with this view, the affinity of homocysteine is weak[17,18] and our attempts to diffuse this amino acid inside MetRS crystals were unsuccessful. To deshield the reactive anhydride bond of aminoacyl adenylate and to obtain aminoacylation of the $tRNA^{Met}$, Y15 must move again. Such a motion may correspond to that observed in the presence of MetP (Figure 7). Possibly, the free energy recovered from the binding of the ribose group of the cognate tRNA helps Y15 to switch into such a productive open conformation.

With increasing detection of clinical isolates of bacteria that are multidrug resistant, there is a critical need for the investigation of bacterial enzymes that could be targeted for the development of new therapeutics. One such area has been the development of inhibitors of pathways involving methionine.[20] Methionyl-tRNA synthetase is an

important target in this regard. The well-known commercial success of the Ile-tRNA synthetase inhibitor mupirocin provides precedent for targeting the amino acid-tRNA synthetase class of enzymes in bacteria.[10] Recent work indicates that potent inhibitors of Met-tRNA synthetase can indeed be created; some of which have exhibited potent and selective *in vivo* antibacterial activity.[24–28] Therefore, a thorough knowledge of the detailed chemical mechanism and structural changes associated with catalysis for a bacterial Met-tRNA synthetase, as has been outlined herein, should serve as a foundation for future developments in this area of drug design.

Materials and Methods

Enzyme and ligands

M551 MetRS was purified as described.[23] Metol-AMP was from laboratory stocks and synthesized as described.[42] L-Trifluoromethionine (TFM) and L-difluoromethionine (DFM) were synthesized as described.[37,43,44] Methionine analogues, 1-amino-3-(methylmercapto)-propylphosphonic acid (MetP) and 1-amino-3-(methylmercapto)propylphosphinic acid (MetI) were prepared as racemic DL mixtures using described methods.[45–47] MetSA was prepared as follows.

2′,3′-*O*-Isopropylidene-5′-*O*-[(*N*-Boc-L-methionyl)-sulfamoyl]adenosine

Into anhydrous dimethylformamide (10 ml) was dissolved 2′,3′-*O*-isopropylidene-5′-*O*-sulfamoyladenosine[48] (377 mg, 0.98 mmol) followed by *N*-Boc-methionine succinimide ester[49] (368 mg, 1.18 mmol) and 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) (0.36 ml, 2.4 mmol). The mixture was stirred at room temperature for two hours under an argon atmosphere. Dimethylformamide was then removed *in vacuo* by repeated azeotropic removal with toluene. The viscous oil was then purified on a silica gel column eluting with EtOAc/MeOH 12:1. The residue was then chromatographed on a Sephadex LH-20 column with 70:30 MeOH/H_20 to give a white powder (432 mg, 77%).

R_f 0.25 (EtOAc/MeOH 10:1); mp 108–112 °C; ^{1}H NMR (250 MHz, CD_3OD) δ 8.38 (s, 1H, H-2), 8.22 (s, 1H, H-8), 6.23 (d, 1H, $J = 2.9$ Hz, H-1′), 5.37 (dd, 1H, $J = 6.1$ Hz, 2.8 Hz, H-2′), 5.11 (dd, 1H, $J = 6.1$ Hz, 2.4 Hz, H-3′), 4.53 (m, 1H, H-4′), 4.28 (m, 2H, H_2-5′), 4.10 (t, $J = 8.2$ Hz, C*H*α), 2.48 (t, 2H, $J = 7.1$ Hz, CH_2γ), 2.07 (m, 1H CH_2β), 2.01 (s, S-CH_3), 1.85 (m, 1H, CH_2β), 1.59 (s, 3H, C-CH_3), 1.40 [s, 9H, C-(CH_3)$_3$], 1.37 (s, 3H, C-CH_3); ^{13}C NMR (62.5 MHz,) δ 172.1 (*C*=O), 167.4 (*C*=O), 157.2 (C-6), 154.0 (C-2), 150.3 (C-4), 141.5 (C-8), 120.2 (C-5), 115.4 [*C*(CH$_3$)$_2$], 91.6 (C-1′), 85.5 (C-2′), 83.0 (C-3′), 80.2 (C-4′), 69.9 (C-5′), 57.5 (*C*Hα), 33.8 (*C*H$_2$β), 31.2 (*C*H$_2$γ), 28.7 [C(*C*H$_3$)$_3$], 27.4 (*C*H$_3$), 25.5 (*C*H$_3$), 15.3 (SC*H*$_3$). Fast atom bombardment high resolution mass spectrometry: m/z [618.20193 (M + H); {calcd for $C_{23}H_{36}N_7O_9S_2 + H^+$} 618.20160].

5′-*O*-[(L-Methionyl)-sulfamoyl]adenosine (MetSA)

The preceding compound (390 mg, 0.62 mmol) was dissolved in 8:1 trifluoroacetic acid/H_2O (2 ml) at room temperature and stirred for two hours. The solvents were removed and the residue co-evaporated four times with ethanol. The product was then dissolved in water and chromatographed on a Sephadex LH-20 column, eluting with MeOH/H_2O (7:3) to give a white powder (65 mg, 22%).

Mp 153–154 °C; ^{1}H NMR (250 MHz, ^2H_2O) δ 8.18 (s, 1H, H-2), 8.00 (s, 1H, H-8), 5.89 (d, 1H, $J = 5.3$ Hz, H-1′), 4.54 (t, 1H, $J = 5.3$ Hz, H-2′), 4.31 (t, 1H, $J = 4.0$ Hz, H-3′), 4.21 (m, 3H, H-4′ and H-5′), 3.70 (t, 1H, $J = 6.1$ Hz, C*H*α), 2.31 (t, 2H, $J = 7.5$ Hz, CH_2γ), 1.89 (m, 2H CH_2β), 1.80 (s, S–CH_3); ^{13}C NMR (62.5 MHz, DMSO) δ 172.6 (*C*=O), 158.1 (C-6), 156.5 (C-2), 149.2 (C-4), 139.3 (C-8), 119.4 (C-5), 93.2 (C-1′), 83.3 (C-2′), 83.2 (C-3′), 75.4 (C-4′), 69.9 (C-5′), 58.1 (*C*Hα), 30.7 (*C*H$_2$β), 25.1 (*C*H$_2$γ), 14.3 (SC*H*$_3$); electrospray mass spectrometry: (CH$_3$CN/H_2O 1:1) [478.06 (M + H); {calcd for $C_{15}H_{24}N_7O_7S_2 + H^+$} 478.117].

Fluorescence at equilibrium

Variations of the intrinsic fluorescence of MetRS upon titration with substrates were followed as described,[30,35,50–52] at 25 °C, in 20 mM Tris–HCl (pH7.6), 10 mM 2-mercaptoethanol, 2 mM $MgCl_2$ and 0.1 mM EDTA. Saturation curves were obtained by varying the total concentration of ligands up to 2 mM.

Dissociation constants were determined from iterative non-linear least-squares fits of the theoretical equations to the experimental values. Confidence limits and standard errors were determined by 100 Monte Carlo simulations from the experimental deviations on individual measurements.[53]

Crystallization and data collection

Crystals of MetRS M551 were obtained as described.[23] For data collection, crystals were first stabilized in 1.4 M ammonium citrate plus 30 mM potassium phosphate (pH 7.0) and then soaked in the same solution plus the studied ligand (Table 2). Diffraction images were analyzed with the MOSFLM program (A. G. W. Leslie, Laboratory of Molecular Biology, Cambridge, UK) and the data further processed using programs from the CCP4 package.[54]

Refinement

The coordinates of the free enzyme structure were used as the starting model. After rigid body refinement and conjugate gradient minimization using the CNS program, initial maps were calculated.[55] In all cases, the electronic density of the added ligand could be unambigously attributed. For the MetI and MetP analogues, although DL mixtures were used, density corresponding to the L stereoisomer could unambiguously be recognized. Manual adjustments of the models were then performed using O.[56] Models of the ligand and water molecules were then added. Several rounds of conjugate gradient minimization, alternated with addition of water molecules, allowed refinement of each structure. Final statistics regarding the liganded structures are given in Table 1. For the E:MetI and E:MetP structures, refinement evidenced alternative conformations for the side-chain of Y15. Only the most frequent conformation was used in the calculation of the final model. The electronic density maps obtained are shown in Figure 3. In

the case of MetI, refinement of the relative rates of occupancy of the two alternative Y15 conformations indicated 50% for each. With all studied enzyme complexes, residues 4–549 of MetRS could be modelled and refined. The only exceptions were residues 126–184 in the Met:Ado:enzyme complex and 130–184 in the Metol-AMP:enzyme complex. This region encompasses the zinc-binding peptide of MetRS.

Co-ordinates have been deposited to the PDB with IDs 1P7P, 1PFU, 1PFV, 1PFW, 1PFY, 1PG0 and 1PG2.

Acknowledgements

We thank Dr Steve Cusack for helpful discussions, and the staff of the LURE DW32, ESRF ID14 and ID29 beamlines for assistance during data collection. T.C. was a recipient of a doctoral fellowship from the Association pour la Recherche sur le Cancer. P.S. was a recipient of an OGS graduate scholarship. M.V. was a recipient of a NSERC graduate scholarship. The authors gratefully acknowledge research support from NSERC (Canada).

References

1. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, **347**, 203–206.
2. Cusack, S., Hartlein, M. & Leberman, R. (1991). Sequence, structural and evolutionary relationships between class 2 aminoacyl-tRNA synthetases. *Nucl. Acids Res.* **19**, 3489–3498.
3. Webster, T., Tsai, H., Kula, M., Mackie, G. A. & Schimmel, P. (1984). Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthetase. *Science*, **226**, 1315–1317.
4. Hountondji, C., Lederer, F., Dessen, P. & Blanquet, S. (1986). *Escherichia coli* tyrosyl- and methionyl-tRNA synthetases display sequence similarity at the binding site for the 3′-end of tRNA. *Biochemistry*, **25**, 16–21.
5. Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N. & Leberman, R. (1990). A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature*, **347**, 249–255.
6. Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A. *et al.* (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science*, **252**, 1682–1689.
7. Fersht, A. R. & Kaethner, M. M. (1976). Enzyme hyperspecificity. Rejection of threonine by the valyl-tRNA synthetase by misacylation and hydrolytic editing. *Biochemistry*, **15**, 3342–3346.
8. Lin, L., Hale, S. P. & Schimmel, P. (1996). Aminoacylation error correction. *Nature*, **384**, 33–34.
9. Nureki, O., Vassylyev, D. G., Tateno, M., Shimada, A., Nakama, T., Fukai, S. *et al.* (1998). Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science*, **280**, 578–582.
10. Silvian, L. F., Wang, J. & Steitz, T. A. (1999). Insights into editing from an Ile-tRNA synthetase structure with $tRNA^{Ile}$ and mupirocin. *Science*, **285**, 1074–1077.
11. Cusack, S., Yaremchuk, A. & Tukalo, M. (2000). The 2 Å crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. *EMBO J.* **19**, 2351–2361.
12. Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J., Vassylyev, D. G. & Yokoyama, S. (2000). Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA(Val) and valyl-tRNA synthetase. *Cell*, **103**, 793–803.
13. Cavarelli, J., Delagoutte, B., Eriani, G., Gangloff, J. & Moras, D. (1998). L-arginine recognition by yeast arginyl-tRNA synthetase. *EMBO J.* **17**, 5438–5448.
14. Delagoutte, B., Moras, D. & Cavarelli, J. (2000). tRNA Aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding. *EMBO J.* **19**, 5599–5610.
15. Newberry, K. J., Hou, Y. M. & Perona, J. J. (2002). Structural origins of amino acid selection without editing by cysteinyl-tRNA synthetase. *EMBO J.* **21**, 2778–2787.
16. Serre, L., Verdon, G., Choinowski, T., Hervouet, N., Risler, J. L. & Zelwer, C. (2001). How methionyl-tRNA synthetase creates its amino acid recognition pocket upon L-methionine binding. *J. Mol. Biol.* **306**, 863–876.
17. Fersht, A. R. & Dingwall, C. (1979). An editing mechanism for the methionyl-tRNA synthetase in the selection of amino acids in protein synthesis. *Biochemistry*, **18**, 1250–1256.
18. Jakubowski, H. & Fersht, A. R. (1981). Alternative pathways for editing non-cognate amino acids by aminoacyl-tRNA synthetases. *Nucl. Acids Res.* **9**, 3105–3117.
19. Duewel, H., Daub, E., Robinson, V. & Honek, J. F. (1997). Incorporation of trifluoromethionine into a phage lysozyme: implications and a new marker for use in protein ^{19}F NMR. *Biochemistry*, **36**, 3404–3416.
20. Vaughan, M. D., Sampson, P. B. & Honek, J. F. (2002). Methionine in and out of proteins: targets for drug design. *Curr. Med. Chem.* **9**, 385–409.
21. Kiick, K., van Hest, J. & Tirrell, D. (2000). Expanding the scope of protein biosynthesis by altering the methionyl-tRNA synthetase activity of bacterial expression host. *Angew. Chem. Int. Ed.* **39**, 2148–2152.
22. Kiick, K. L., Weberskirch, R. & Tirrell, D. A. (2001). Identification of an expanded set of translationally active methionine analogues in *Escherichia coli*. *FEBS Letters*, **502**, 25–30.
23. Mechulam, Y., Schmitt, E., Maveyraud, L., Zelwer, C., Nureki, O., Yokoyama, S. *et al.* (1999). Crystal structure of *Escherichia coli* methionyl-tRNA synthetase highlights species-specific features. *J. Mol. Biol.* **294**, 1287–1297.
24. Lee, J., Kang, S. U., Kim, S. Y., Kim, S. E., Job, Y. J. & Kim, S. (2001). Vanilloid and isovanilloid analogues as inhibitors of methionyl-tRNA and isoleucyl-tRNA synthetases. *Bioorg. Med. Chem. Letters*, **11**, 965–968.
25. Lee, J., Kang, S. U., Kang, M. K., Chun, M. W., Jo, Y. J., Kwak, J. H. & Kim, S. (1999). Methionyl adenylate analogues as inhibitors of methionyl-tRNA synthetase. *Bioorg. Med. Chem. Letters*, **9**, 1365–1370.
26. Lee, J., Kang, S. U., Kim, S. Y., Kim, S. E., Kang, M. K., Jo, Y. J. & Kim, S. (2001). Ester and hydroxamate analogues of methionyl and isoleucyl adenylates as

inhibitors of methionyl-tRNA and isoleucyl-tRNA synthetases. *Bioorg. Med. Chem. Letters*, **11**, 961–964.
27. Jarvest, R. L., Berge, J. M., Brown, M. J., Brown, P., Elder, J. S., Forrest, A. K. *et al.* (2003). Optimisation of aryl substitution leading to potent methionyl tRNA synthetase inhibitors with excellent gram-positive antibacterial activity. *Bioorg. Med. Chem. Letters*, **13**, 665–668.
28. Jarvest, R. L., Berge, J. M., Berry, V., Boyd, H. F., Brown, M. J., Elder, J. S. *et al.* (2002). Nanomolar inhibitors of *Staphylococcus aureus* methionyl tRNA synthetase with potent antibacterial activity against gram-positive pathogens. *J. Med. Chem.* **45**, 1959–1962.
29. Forrest, A. K., Jarvest, R. L., Mensah, L. M., O'Hanlon, P. J., Pope, A. J. & Sheppard, R. J. (2000). Aminoalkyl adenylate and aminoacyl sulfamate intermediate analogues differing greatly in affinity for their cognate *Staphylococcus aureus* aminoacyl tRNA synthetases. *Bioorg. Med. Chem. Letters*, **10**, 1871–1874.
30. Blanquet, S., Fayat, G., Waller, J. P. & Iwatsubo, M. (1972). The mechanism of reaction of methionyl-tRNA synthetase from *Escherichia coli*. Interaction of the enzyme with ligands of the amino acid activation reaction. *Eur. J. Biochem.* **24**, 461–469.
31. Blanquet, S., Fayat, G. & Waller, J. P. (1975). The amino acid activation reaction catalyzed by methionyl-transfer RNA synthetase: evidence for synergistic coupling between the sites for methionine, adenosine and pyrophosphate. *J. Mol. Biol.* **94**, 1–15.
32. Lee, J., Kim, S. E., Lee, J. Y., Kim, S. Y., Kang, S. U., Seo, S. H. *et al.* (2003). *N*-Alkoxysulfamide, *N*-hydroxysulfamide, and sulfamate analogues of methionyl and isoleucyl adenylates as inhibitors of methionyl-tRNA and isoleucyl-tRNA synthetases. *Bioorg. Med. Chem. Letters*, **13**, 1087–1092.
33. Fourmy, D., Mechulam, Y., Brunie, S., Blanquet, S. & Fayat, G. (1991). Identification of residues involved in the binding of methionine by *Escherichia coli* methionyl-tRNA synthetase. *FEBS Letters*, **292**, 259–263.
34. Fourmy, D., Mechulam, Y. & Blanquet, S. (1995). Crucial role of an idiosyncratic insertion in the Rossmann fold of class 1 aminoacyl-tRNA synthetases: the case of methionyl-tRNA synthetase. *Biochemistry*, **34**, 15681–15688.
35. Mechulam, Y., Dardel, F., Le Corre, D., Blanquet, S. & Fayat, G. (1991). Lysine 335, part of the KMSKS signature sequence, plays a crucial role in the amino acid activation catalysed by the methionyl-tRNA synthetase from *Escherichia coli*. *J. Mol. Biol.* **217**, 465–475.
36. Schmitt, E., Meinnel, T., Blanquet, S. & Mechulam, Y. (1994). Methionyl-tRNA synthetase needs an intact and mobile 332KMSKS336 motif in catalysis of methionyl adenylate formation. *J. Mol. Biol.* **242**, 566–576.
37. Vaughan, M., Cleve, P., Robinson, V., Duewel, H. & Honek, J. F. (1999). Difluoromethionine as a novel ^{19}F NMR structural probe for internal amino acid packing in proteins. *J. Am. Chem. Soc.* **121**, 8475–8478.
38. Fersht, A. R. & Dingwall, C. (1979). Cysteinyl-tRNA synthetase from *Escherichia coli* does not need an editing mechanism to reject serine and alanine. High binding energy of small groups in specific molecular interactions. *Biochemistry*, **18**, 1245–1249.
39. Brick, P., Bhat, T. N. & Blow, D. M. (1989). Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution. Interaction of the enzyme with the tyrosyl adenylate intermediate. *J. Mol. Biol.* **208**, 83–98.
40. Fersht, A. R. (1987). Dissection of the structure and activity of the tyrosyl-tRNA synthetase by site-directed mutagenesis. *Biochemistry*, **26**, 8031–8037.
41. Kim, H. Y., Ghosh, G., Schulman, L. H., Brunie, S. & Jakubowski, H. (1993). The relationship between synthetic and editing functions of the active site of an aminoacyl-tRNA synthetase. *Proc. Natl Acad. Sci. USA*, **90**, 11553–11557.
42. Cassio, D., Lemoine, F., Waller, J. P., Sandrin, E. & Boissonnas, R. A. (1967). Selective inhibition of aminoacyl ribonucleic acid synthetases by aminoalkyl adenylates. *Biochemistry*, **6**, 827–836.
43. Duewel, H., Daub, E., Robinson, V. & Honek, J. (1997). Incorporation of trifluoromethionine into a phage lysozyme: implications and a new marker for use in protein ^{19}F NMR. *Biochemistry*, **36**, 3404–3416.
44. Houston, M. E. & Honek, J. F. (1989). Facile synthesis of fluorinated methionines. *J. Chem. Soc., Chem. Commun.*, 761–762.
45. Lowther, W. T., Zhang, Y., Sampson, P. B., Honek, J. F. & Matthews, B. W. (1999). Insights into the mechanism of *Escherichia coli* methionine aminopeptidase from the structural analysis of reaction products and phosphorus-based transition-state analogues. *Biochemistry*, **38**, 14810–14819.
46. Kudzin, Z. H. & Stec, W. J. (1980). Phosphohomocysteine derivatives. *Synthesis*, **12**, 1032–1034.
47. Bayliss, E. K., Campbell, C. D. & Dingwall, J. G. (1984). 1-Aminoalkylphosphonous acids. Part 1. Isosteres of the protein amino acids. *J. Chem. Soc., Perkin Trans. 1*, **12**, 2845–2853.
48. Heathcock, D., Forsyth, C. J., Shiba, K. & Musier-Forsyth, K. (1996). Synthesis and aminoacyl-tRNA synthetase inhibitory activity of prolyl adenylate analogs. *Bioorg. Chem.* **24**, 273–289.
49. Anderson, G. W., Zimmerman, J. E. & Callahan, F. M. (1964). The use of esters of *N*-hydroxysuccinimide in peptide synthesis. *J. Am. Chem. Soc.* **86**, 1839–1842.
50. Blanquet, S., Fayat, G. & Waller, J. P. (1974). The mechanism of action of methionyl-tRNA synthetase from *Escherichia coli*. Mechanism of the amino acid activation reaction catalyzed by the native and the trypsin-modified enzymes. *Eur. J. Biochem.* **44**, 343–351.
51. Fayat, G., Fromant, M. & Blanquet, S. (1977). Couplings between the sites for methionine and adenosine 5′- triphosphate in the amino acid activation reaction catalyzed by trypsin-modified methionyl-transfer RNA synthetase from *Escherichia coli*. *Biochemistry*, **16**, 2570–2579.
52. Blanquet, S., Fayat, G., Poiret, M. & Waller, J. P. (1975). The mechanism of action of methionyl-tRNA synthetase from *Escherichia coli*. Inhibition by adenosine and 8-aminoadenosine of the amino-acid activation reaction. *Eur. J. Biochem.* **51**, 567–571.
53. Dardel, F. (1994). MC-Fit: using Monte-Carlo methods to get accurate confidence limits on enzyme parameters. *Comput. Appl. Biosci.* **10**, 273–275.
54. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs from protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
55. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.

56. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for the building of proteins model in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
57. Evans, S. V. (1993). Setor: hardware lighted three-dimensional solid model representation of macromolecules. *J. Mol. Graph.* **11**, 134–138.
58. Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37–40.
59. Schmitt, E., Panvert, M., Blanquet, S. & Mechulam, Y. (1995). Transition state stabilization by the 'high' motif of class I aminoacyl-tRNA synthetases: the case of *Escherichia coli* methionyl-tRNA synthetase. *Nucl. Acids Res.* **23**, 4793–4798.
60. Nicholls, A. & Honig, B. (1991). A rapid finite difference algorithm utilizing successive over-relaxations to solve the Poisson–Boltzmann equation. *J. Comp. Chem.* **12**, 435–445.

Edited by R. Huber

(Received 6 May 2003; received in revised form 9 July 2003; accepted 14 July 2003)

Structure *14*, 1511–1525, October 2006 DOI 10.1016/j.str.2006.08.007

Structures of Two Bacterial Prolyl-tRNA Synthetases with and without a *cis*-Editing Domain

Thibaut Crepin,[1,3] Anna Yaremchuk,[1,2,3] Mikhail Tukalo,[1,2] and Stephen Cusack[1,*]
[1] European Molecular Biology Laboratory
Grenoble Outstation
6 rue Jules Horowitz
BP 181
F-38042 Grenoble Cedex 9
France
[2] Institute of Molecular Biology and Genetics
NAS of Ukraine
252627 Kiev 03143
Ukraine

Summary

Prolyl-tRNA synthetases (ProRSs) are unique among synthetases in that they have diverse architectures, notably the variable presence of a *cis*-editing domain homologous to the freestanding deacylase proteins YbaK and ProX. Here, we describe crystal structures of two bacterial ProRSs from the pathogen *Enterococcus faecalis*, which possesses an editing domain, and from *Rhodopseudomonas palustris*, which does not. We compare the overall structure and binding mode of ATP and prolyl-adenylate with those of the archael/eukaryote-type ProRS from *Thermus thermophilus*. Although structurally more homologous to YbaK, which preferentially hydrolyzes Cys-tRNA^{Pro}, the editing domain of *E. faecalis* ProRS possesses key elements similar to ProX, with which it shares the activity of hydrolyzing Ala-tRNA^{Pro}. The structures give insight into the complex evolution of ProRSs, the mechanism of editing, and structural differences between prokaryotic- and eukaryotic-type ProRSs that can be exploited for antibiotic design.

Introduction

Aminoacyl-tRNA synthetases (aaRSs) catalyze the aminoacylation of their cognate tRNAs through a two-step mechanism involving activation of the amino acid by ATP to yield the aminoacyl-adenylate, followed by transfer to the 3′ terminal adenosine of the tRNA molecule. Analysis of primary sequences as well as the determination of three-dimensional structures has allowed the partition of aaRSs into two evolutionarily distinct classes, each characterized by a class-defining catalytic domain containing short conserved sequence motifs (Cusack et al., 1990; Eriani et al., 1990). The active site of class I enzymes is built around a Rossmann fold harboring the two signature sequences, HIGH and KMSKS, while the catalytic center of class II enzymes is built around an antiparallel β sheet and is characterized by three conserved motifs (motifs 1, 2, and 3). Within each class, aaRSs can be subdivided into subclasses that usually share a common anticodon-binding module (Cusack, 1995).

*Correspondence: cusack@embl-grenoble.fr
[3] These authors contributed equally to this work.

The accuracy of the aminoacylation reaction is essential to insure the fidelity of protein synthesis. However, where amino acids differ by, for instance, only a methyl group, aaRSs are unable to achieve sufficient discrimination for accurate protein synthesis due to misactivation and/or mischarging of the similar noncognate amino acid. Several aaRSs have therefore evolved editing or proofreading activities to correct such errors (Jakubowski, 2004). In pretransfer editing, the noncognate aminoacyl-adenylate intermediate is hydrolyzed prior to transfer, whereas in posttransfer editing, the noncognate aminoacyl-tRNA is hydrolyzed. Pretransfer editing may be tRNA dependent or independent and may occur via the synthetic active site (e.g., as shown in recent studies of GlnRS [Gruic-Sovulj et al., 2005] and ProRS [Hati et al., 2006]) or a distinct editing site, as proposed for the class Ia enzymes such as IleRS (Nordin and Schimmel, 2005) (although in this case the mechanism of translocation of the noncognate adenylate between sites remains obscure). In contrast, posttransfer editing exclusively depends on an additional "editing" domain that can specifically deacylate mischarged tRNAs while leaving cognate aminoacyl-tRNA untouched. Extensive biochemical and structural studies are now available on seven aaRSs that possess an additional editing domain. These include three class Ia aaRSs, IleRS (Fukunaga and Yokoyama, 2006; Nordin and Schimmel, 2005; Silvian et al., 1999), LeuRS (Lincecum et al., 2003; Tukalo et al., 2005), and ValRS (Fukai et al., 2000), and four class IIa systems, ProRS (Beuning and Musier-Forsyth, 2000), ThrRS (Dock-Bregeon et al., 2000, 2004; Dwivedi et al., 2005), AlaRS (Beebe et al., 2003; Sokabe et al., 2005; Swairjo and Schimmel, 2005), and PheRS (Kotik-Kogan et al., 2005; Roy et al., 2004). These studies show that in both synthetase classes posttransfer editing occurs as a result of a conformational change in the 3′ end of the tRNA that allows translocation of the aminoacyl moiety of the tRNA from the synthetic to the editing active sites, which are typically separated by about 35 Å. Furthermore, in most cases, the structural basis of amino acid discrimination in the editing site and a mechanism of hydrolysis have been proposed. Here, we focus on the prolyl system, where less structural information is available on the editing mechanism.

Primary sequence alignments and phylogenetic analysis initially revealed the existence of two diverged forms of prolyl-tRNA synthetase (ProRS): an eukaryote/archae-type and a prokaryote-type (Yaremchuk et al., 2000; Woese et al., 2000; Beuning and Musier-Forsyth, 2001; Musier-Forsyth et al., 2004). The main difference between these forms is the presence, only in eukaryote/archae-like ProRSs, of a C-terminal zinc domain appended to the class IIa anticodon binding domain, and, only in prokaryote-like ProRSs, of a large insertion domain (INS) of about 180 residues inserted between motifs 2 and 3 of the catalytic domain. In addition, divergent sequence motifs of the catalytic domain are sufficient to distinguish between the two main types (Yaremchuk et al., 2000). Subsequently, systematic genome sequencing has necessitated a modification of this

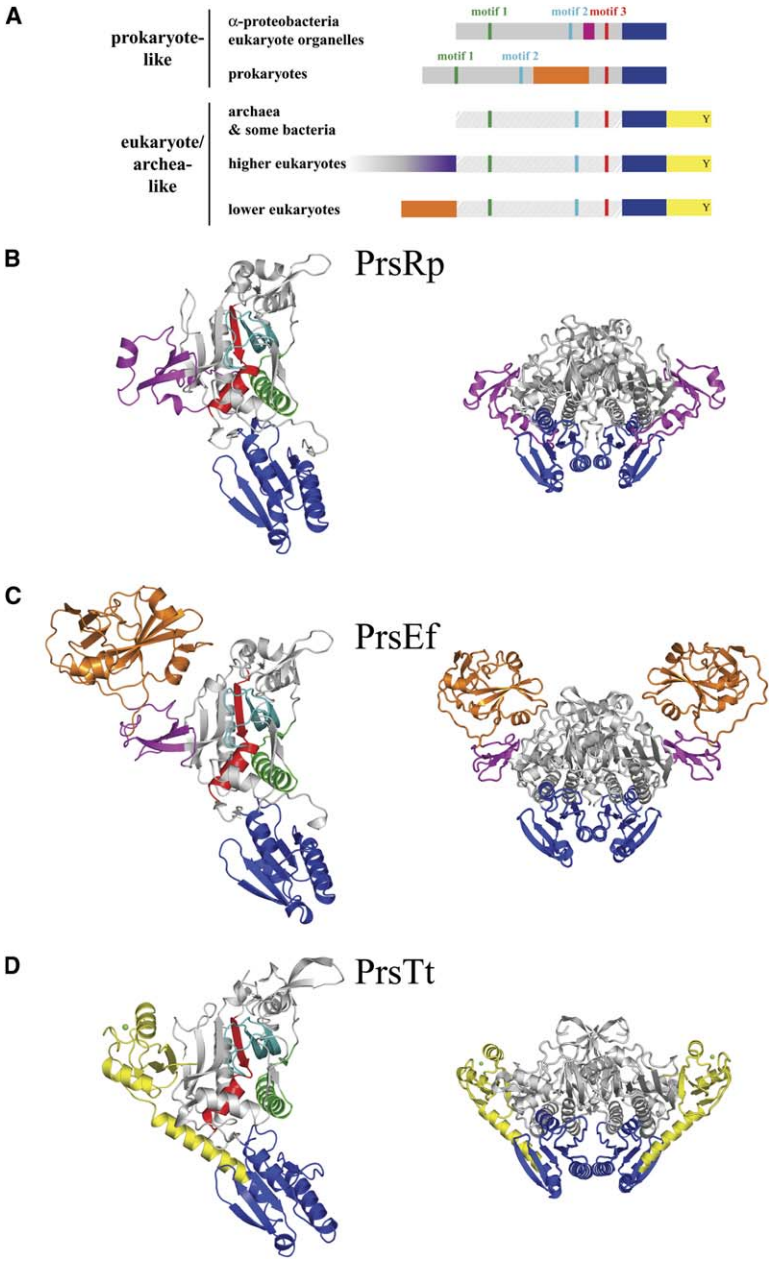


Figure 1. Varied Architecture of Prolyl-tRNA Synthetases

(A) Schematic diagram of ProRS architectures. To reflect systematic differences in sequence, the catalytic domains of prokaryote- and eukaryote/archae-like ProRSs are shown in dark and light gray, respectively. In this and other figures, the class II motifs 1, 2, and 3 are green, cyan, and red, respectively. The class IIa anticodon-binding domain is blue. The eukaryote/archae-type-specific C-terminal domain is yellow. The editing domain (orange) is inserted between motifs 2 and 3 in many bacterial ProRSs, but it is an N-terminal extension in yeast (e.g., *S. cerevisaie*, *S. pombe*) and parasites (e.g., *P. falciparum*, *C. parvum*, *T. brucei*). The α-proteobacteria/organelles-specific truncated insertion domain is magenta. The GluRS part of the higher-eukaryote bifunctional Glu-ProRS is shaded violet.
(B–D) Equivalent colors are used for the ribbon representation of ProRS monomers (left) and dimers (right) of (B) *R. palustris*, (C) *E. faecalis*, and (D) *T. thermophilus* (the zinc atom of PrsTt is green).
The figure was drawn with PyMOL (DeLano, 2002).

classification, as at least five distinct architectures of ProRSs can now be distinguished (Figure 1A). First, the INS domain is not exclusively eubacterial; it is also found at the N terminus of the eukaryotic/archae-like ProRSs of yeasts and single-celled parasites (also see Figure S1 in the Supplemental Data available with this article online). Second, the INS domain is absent from certain prokaryote-like ProRSs, such as those of α-proteobacteria and eukaryotic organelles. In addition, isolated paralogs of the ProRS INS domain, designated YbaK and ProX, exist in the genomes of almost all organisms sequenced (Ahel et al., 2003; Zhang et al., 2000).

ProRS has been shown to significantly misactivate and mischarge alanine (Beuning and Musier-Forsyth, 2000) and cysteine (Ahel et al., 2002; Ambrogelly et al., 2002; Beuning and Musier-Forsyth, 2001; Kamtekar et al., 2003). Furthermore, the large INS domain of *E. coli* ProRS plays a critical role in the hydrolysis of mischarged Ala-tRNA^{Pro}, but not Cys-tRNA^{Pro} (Wong et al., 2002). The fact that many eukaryote/archae-like and prokaryote-like ProRSs do not possess an INS domain suggests that either editing activity is not required in these organisms or other means of deacylating mischarged tRNA^{Pro} exist. Indeed, it has been shown that the paralogs ProX and YbaK are able to preferentially deacylate in *trans* Ala-tRNA^{Pro} (Ahel et al., 2003; Ruan and Soll, 2005; Wong et al., 2003) and Cys-tRNA^{Pro} (An and Musier-Forsyth, 2004; Ruan and Soll, 2005), respectively. Furthermore, the mode of action of YbaK may depend on the formation of a transient synthetase:YbaK:aminoacyl-tRNA ternary complex (An and Musier-Forsyth, 2005).

To understand the complex evolution of the ProRS system and, in particular, the mechanism of posttransfer

Table 1. Data Collection Statistics

	Native PrsRp	SeMet PrsRp Absorption Peak	SeMet PrsRp Inflection Point	PrsRp: ProAMS	PrsRp: CysAMS	PrsRp: ATP:Mg	PrsEf:ATP: Mn:Prolinol	PrsEf: ProAMS
X-ray source	ID14eh1	ID29	ID29	ID14eh3	ID14eh1	ID14eh1	ID14eh4	ID14eh4
Wavelength (Å)	0.9340	0.97917	0.97927	0.9310	0.97935	0.934	1.072	0.939
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$	C2	$C222_1$	$P2_1$	$P4_12_12$
Cell dimensions (Å)								
a (Å)	110.8	107.4	107.4	110.4	217.9	114.2	68.6	121.44
b (Å)	212.6	211.6	211.6	211.9	107.5	211.2	92.7	121.44
c (Å)	150.6	149.9	149.9	150.2	110.7	148.2	101.2	178.84
β (°)	90	90	90	90	120.43	90	106.10	90
Resolution range[a] (Å)	20–2.0 (2.1–2.0)	50–3.5 (3.7–3.5)	50–3.5 (3.7–3.5)	30–2.8 (3.0–2.8)	50–2.85 (3–2.85)	20–2.4 (2.5–2.4)	30–2.3 (2.4–2.3)	30–2.3 (2.4–2.3)
Completeness[a] (%)	94.5 (82.7)	98.0 (99.9)	98.0 (99.8)	96.2 (91.6)	94.3 (92.7)	96.1 (98.8)	99.0 (98.6)	97.7 (82.5)
R_{sym} I[a,b] (%)	7.5 (41.3)	8.8 (31.8)	7.2 (45.8)	7.7 (36.3)	10.0 (49.0)	8.5 (49.6)	10.2 (45.0)	8.4 (43.0)
Total reflections[a]	595,892 (32,275)	113,600 (12,780)	114,005 (12,656)	176,476 (37,057)	122,917 (12,325)	268,259 (29,662)	199,679 (23,173)	224,204 (14,090)
Unique reflections[a]	120,017 (10,237)	21,525 (2,373)	4,083 (4,541)	43,155 (9,308)	48,780 (6,805)	67,428 (7,912)	54,112 (6,344)	58,612 (5,803)
I/σI[a]	15.4 (1.9)	10.7 (2.8)	8.6 (2.2)	13.1 (3.6)	6.8 (1.5)	11.7 (2.4)	6.44 (1.95)	11.2 (2.4)
Multiplicity[a]	5.0 (3.2)	5.3 (5.4)	2.8 (2.8)	4.1 (4.0)	2.5 (1.8)	4.0 (3.7)	3.7 (3.7)	3.8 (2.4)

[a] Values in parentheses are for the highest-resolution shell.
[b] R_{sym} (I) = $(\sum_{hkl}\sum_i|<I_{hkl}> - I_{hkl,i}|)/(\sum_{hkl}\sum_i|I_{hkl}|)$, where i is the number of reflection hkl.

editing by the INS domain, we have determined crystal structures of ProRSs from the pathogenic gram positive bacterium *Enterococcus faecalis* (PrsEf) and the α-proteobacterium *Rhodopseudomonas palustris* (PrsRp), two prokaryote-like ProRSs with and without, respectively, a *cis*-editing domain. We also report structures of various complexes with ATP and cognate prolyl- and noncognate cysteinyl-adenylate analogs. To our knowledge, comparison with our previous work on the eukaryote/archae-like ProRS from *Thermus thermophilus* gives new insights into the distinct structural and mechanistic features of these two types of ProRSs, which could be exploited for the design of bacterial-specific antibiotics. We also compare the structure of the INS domain of PrsEf with the known structures of ProX and YbaK and draw conclusions about the evolutionary relatedness and specificity of these homologous domains.

Results

Structure Determination

The structure of *R. palustris* ProRS (PrsRp, 438 residues) was initially solved at 3.5 Å resolution by using the MAD method and a single crystal containing selenomethionine-substituted enzyme. This was necessary, as the homology with the known structure of the eukaryotic-like *T. thermophilus* ProRS was too low for molecular replacement to work. With the resulting phases, the model was built and refined by using native crystal data at 2 Å resolution to an R factor of 19.1% (R_{free} = 22.1%). The PrsRp model permitted the structure of *E. faecalis* ProRS (PrsEf, 572 residues) to be readily solved by molecular replacement by using native data at 2.36 Å resolution in a crystal form with a dimer in the asymmetric unit. The model of PrsEf has been refined to an R factor of 21.9% (R_{free} = 28.2%). These two initial models were subsequently used to solve the structures of various small substrate complexes such as ATP and cognate and noncognate aminoacyl-adenylate analogs by molecular replacement (see Experimental Procedures and Tables 1 and 2 for the crystallographic details).

Overview of the Structures

The subunit and dimeric structures of the two prokaryote-like ProRSs together with the known eukaryote/archae-like ProRS structure from *T. thermophilus* are

Table 2. Refinement Statistics

	Native PrsRp	PrsRp:ProAMS	PrsRp:CysAMS	PrsRp:ATP:Mg^{2+}	PrsEf:ATP:Mn^{2+}:prolinol	PrsEf:ProAMS
R factor (%)	19.1	20.6	21.1	19.7	19.6	21.9
R_{free} (%)	22.1	26.9	27.7	26.1	26.0	28.2
Bond length (Å)	0.015	0.012	0.013	0.013	0.016	0.011
Bond angle (°)	1.416	1.419	1.531	1.566	1.694	1.352
Mean B factor	29.87	44.78	59.18	36.85	21.84	32.15
Total nonhydrogen atoms	11,093	10,683	10,672	10,865	9,248	9,328
Residues	1,323	1,321	1,325	1,323	1,112	1,135
Ligands	—	2xProAMS	3xCysAMS	3xATP	2xATP, 2xprolinol	2xProAMS
Water molecules	623	76	80	279	364	203
Ions	—	—	—	9	6 manganese	21 sulphate
Ramachandran						
Favored	100%	100%	99.9%	100%	99.8%	99.6%
Additional disallowed	—	—	Arg436 (A)	—	2xAsp373	2xAsp373

shown in Figures 1B–1D. All structures present a common core comprising the N-terminal class II catalytic domain harboring the three conserved motifs linked by a structured random coil to a class IIa anticodon-binding domain. Multiple structural alignments with SSM (Krissinel and Henrick, 2004) show that a core of 347 residues, featuring 23 secondary-structure elements, can be aligned. The corresponding Cα position root-mean-square deviations (rmsds) (adenylate-bound structures) and sequence identities (shown in parentheses) are 2.13 Å (19.9%) for PrsTt and PrsEf, 1.85 Å (23.6%) for PrsTt and PrsRp, and 1.14 Å (46.1%) for PrsEf and PrsRp. The corresponding sequence alignments and secondary-structure assignments are given in Figure 2A. Different domains are added to this core according to the divergences observed in the primary sequence and thus giving rise to distinct overall architectures. PrsRp contains a small inserted domain of 63 residues between motifs 2 and 3 of the conserved active site (Figure 2A). This module is composed of a small three-stranded antiparallel β sheet, contiguous to the seven-stranded β sheet of the catalytic domain, and four short helices (Figure 2B). PrsEf contains a truncated version of this domain, comprising the antiparallel β sheet onto which the large INS domain is grafted (Figure 2C). In PrsTt, there are also three β strands contiguous to the active site β sheet that occupy the same three-dimensional position (Yaremchuk et al., 2000), but they belong to the unique C-terminally appended domain characteristic of eukaryote/archae-like ProRSs, and the strands are in the antiparallel orientation. Despite this overall structural divergence between the eukaryote/archae- and prokaryote-like ProRSs, one notable feature is functionally conserved. The carboxylate group of the absolutely conserved final tyrosine of the C-terminal domain of eukaryote/archae-like ProRSs (Tyr477 of PrsTt) points directly into the active site and stabilizes the position of basic residues interacting with the ATP pyrophosphate (Yaremchuk et al., 2000). In the case of prokaryote-like ProRSs, the carboxylate group of the strictly conserved glutamate residue (Glu218 in both PrsRp and PrsEf) perfectly superimposes on that of PrsTt Tyr477 and plays the same role (see ATP binding section below, Figure 3).

cis-Editing Domain of PrsEf

The most striking feature of PrsEf is the large INS domain (residues 237–390), which is linked by ordered random coil to the minimal INS common with PrsRp (Figures 1C and 2C). This, in turn, is anchored to the catalytic domain as described above. The INS domain is composed of eight β strands in two mixed sheets surrounded by five short peripheral α helices, forming a compact globular structure with virtually no direct contact with the rest of the enzyme (in some structures, the side chain of Asn307 on helix α8 forms a single hydrogen bond with Asp198). The fold of this domain is, as expected, similar to that of the *H. influenzae* YbaK protein, whose crystal structure determination gave the first structural insight into editing in the prolyl system (Zhang et al., 2000). Lys279 in PrsEf (equivalent to Lys46 in *H. influenzae* Ybak, the absolutely conserved lysine residue in this family of deacylases; see Discussion), required for the hydrolysis of misacylated $tRNA^{Pro}$ by prokaryote-like ProRSs (Wong et al., 2003), is in a shallow cavity, and its Cα atom is located 32 Å from adenylate α-phosphate in the catalytic domain active site. This configuration of an independently folded editing domain poised over the synthetic catalytic domain, with 30–40 Å between the two active sites, is highly reminiscent of the other editing synthetases of known structure, either of class 1a IleRS, ValRS, and LeuRS or of class IIa ThrRS (Dock-Bregeon et al., 2000). However, the deacylase fold common to the three class 1a enzymes is unrelated to that of ThrRS, which is, again, quite distinct from that of ProRSs. Indeed, eukaryotic and prokaryotic ThrRSs have a distinct deacylase domain from that of many archael ThrRSs (which resembles a D-amino acid deacylase [Dwivedi et al., 2005]), showing that, during evolution, at least four structurally unrelated deacylase domains have been coupled in *cis* to aaRSs. In the case of the class 1a enzymes, significant rotation of the editing domain is required during the editing mechanism, as shown, for instance, in the cases of LeuRS (Fukunaga and Yokoyama, 2005; Tukalo et al., 2005), IleRS (Silvian et al., 1999), and ValRS (Fukai et al., 2000). The absence of strong interaction between the INS domain of PrsEf and the catalytic core suggests that the domain could also be flexibly linked and could, hence, possibly rotate during the posttransfer editing process. Indeed, small variations in the orientation are observed in the various structures.

Small-Substrate Recognition by Prokaryote-Type ProRSs

Complete data sets were collected for cocrystals of both PrsRp and PrsEf with different substrates, including ATP and cognate and noncognate aminoacyl-adenylate analogs (see Tables 1 and 2 for crystallographic data). In each case, difference maps unambiguously revealed the different ligands, allowing visualization in detail of the molecular basis of substrate recognition by the two prokaryote-like ProRSs and enabling comparison to previously published work on eukaryote-like ProRSs (Kamtekar et al., 2003; Yaremchuk et al., 2001).

ATP Complexes

The complex of PrsRp with Mg^{2+}-ATP has been determined at 2.4 Å resolution, and that of PrsEf with Mn^{2+}-ATP and prolinol (a nonreactive analog of proline) has been determined at 2.3 Å resolution (Table 1). These structures give a detailed view of the conformation of the preactivation enzyme-substrate complex in the two types of prokaryotic ProRSs (Figures 3A and 3B) and can be compared to the equivalent binary complex of eukaryote-like PrsTt (Figure 3C). The structures of the PrsRp and PrsEf ATP complexes are very similar. Both show the ATP bound in the canonical class II bent form and coordinated with three divalent cations (either magnesium or manganese), as previously observed in several other, but not all, class II systems. The adenine ring is stacked between motif 2 Phe155 and the invariant motif 3 arginine residue (Arg322 in PrsRp and Arg447 in PrsEf). The ribose adopts a C3′ endo conformation and interacts with the main chains of the motif 2 β strand and the fifth β strand and the side chain of the prokaryote-like conserved glutaminyl residue (Glu282 in PrsRp and Glu407 in PrsEf). The principal divalent cation (M1) is coordinated by the α- and β-phosphate and residues

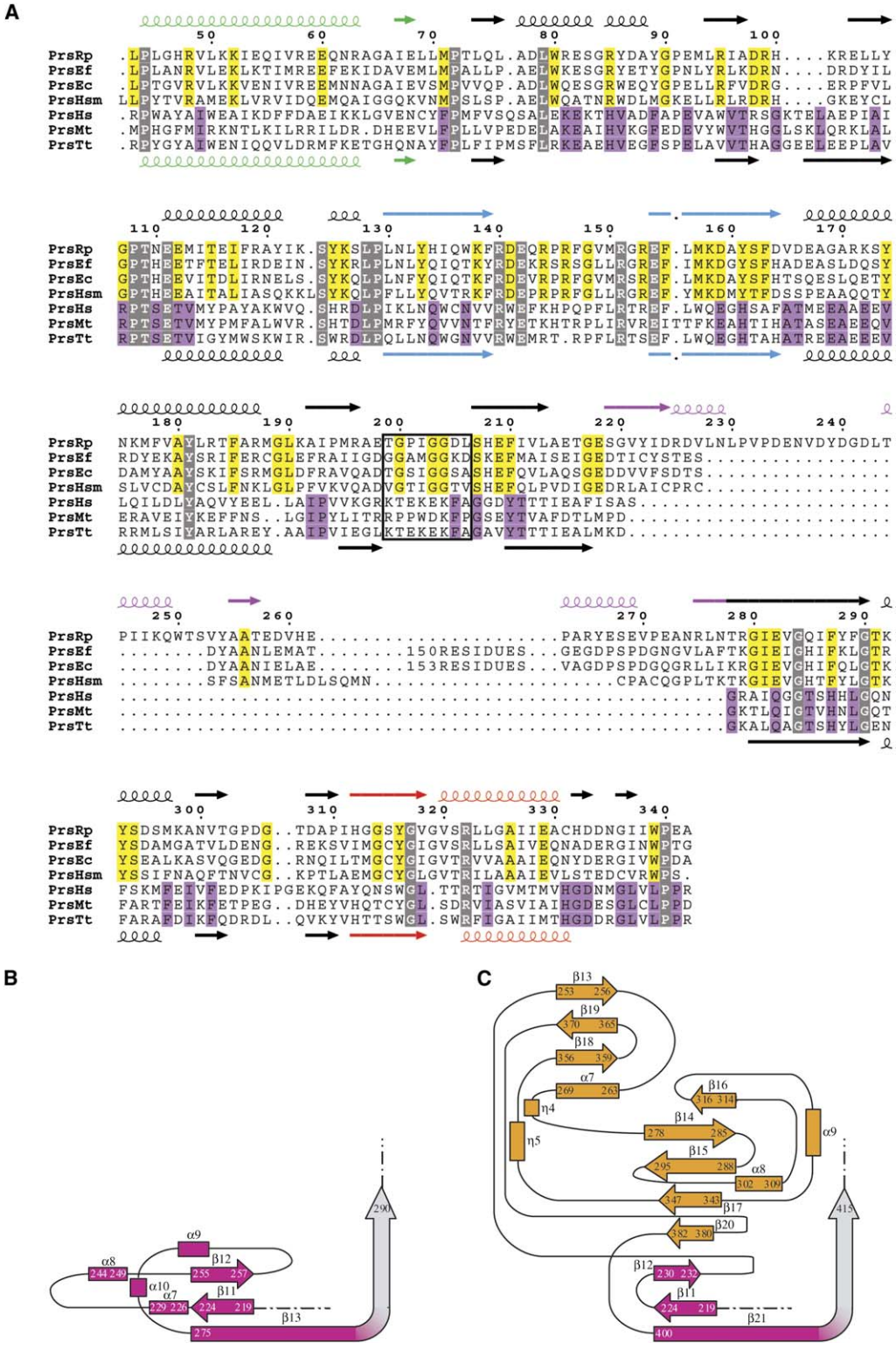


Figure 2. Sequence Alignments and Insertion Domain Topology of Prolyl-tRNA Synthetases

(A) Sequence alignment of the catalytic domain of representative ProRSs. The upper four sequences in each block are prokaryote-like ProRSs, and the lower three sequences are eukaryote/archae-like enzymes. PrsRp, *R. palustris* (Q6N5P6); PrsEf, *E. faecalis* (Q831W7); PrsEC, *E. coli* (Q8X8W2); PrsHsm, *H. sapiens* mitochondrial (Q7L3T8); PrsHs, *H. sapiens* (P07814); PrsMt, *M. thermoautotrophicum* (O26708); PrsTt, *T. thermophilus* (Q72GF9). The secondary-structure elements correspond to the PrsRp (above) and PrsTt (below) structures. Conserved residues are color coded as follows: universal, gray; prokaryote-specific, yellow; eukaryote/archae-specific, violet. Proline-binding loops are outlined with a dark box. The editing domain of PrsEf and PrsEC has been omitted (see Figure S1).

(B) Topology of the truncated insertion domain of PrsRp.

(C) Topology of the inserted editing domain of PrsEf.

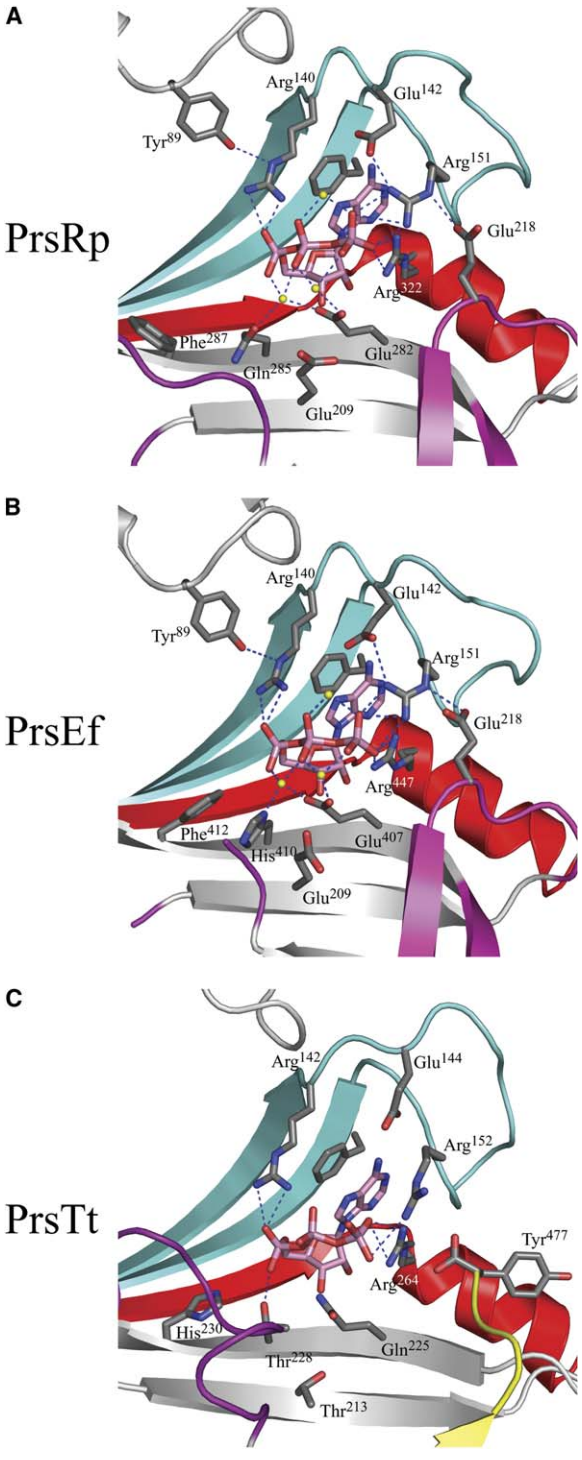


Figure 3. Interactions of ProRSs with ATP

(A–C) Active sites of (A) *R. palustris*, (B) *E. faecalis*, and (C) *T. thermophilus* ProRS with bound ATP (predominantly pink molecule), showing hydrogen bonds with key interacting residues. The insertion domains are in magenta (in [A] and [B]), and the eukaryote/archae-type-specific C-terminal domain is in yellow (in [C]). Note the functionally equivalent roles of Glu218 in PrsRp and PrsEf with the conserved carboxy terminus (Tyr477) in PrsTt. In each case, the proline-binding loop is in the open conformation.

Gln285 (His410 in PrsEf) and Glu282 (Glu407) from the same β strand. It is relatively unusual for Mg to be coordinated by a histidine, but less so for Mn. The two other divalent cations bridge either side of the β- and γ-phosphates; one of them (M2) is also coordinated by Glu282 (Glu407 in PrsEf), and the other (M3) is coordinated by water molecules positioned by Glu142 (Glu142 in PrsEf). Glu209 and Asp219 position water molecules that complete the M2 octahedral coordination shell. Arg322 (Arg447) and Arg151 (Arg151), the motif 3 arginine and the arginine from the motif 2 loop, respectively, interact with the γ-phosphate, and Arg151 is positioned by Glu142 and Glu218 (which is functionally equivalent to the C-terminal carboxyl group of Tyr477 in PrsTt, Figure 3C). Residues 199–206, which form the proline-binding loop, are out of the active site, either disordered or in an ordered open conformation in a similar configuration in both ATP structures. The same open conformation of the loop is observed in the native unliganded structure of PrsRp (there is no unliganded structure of PrsEf). In the PrsEf-ATP complex, which also contains prolinol (although only present in low occupancy), Phe412 is flipped into the active site in one subunit of the dimer, but not the other in the crystallographic asymmetric unit. The corresponding residue, Phe287, is flipped out of the active site in the native PrsRp and PrsRp:Mg^{2+}-ATP structures (see below).

Cognate and Noncognate Adenylate Analog Complexes

Structures of complexes with 5′-O-(N-[prolyl]-sulphamoyl) adenosine (ProAMS), a nonhydrolyzable analog of the prolyl-adenylate, have been refined at 2.4 and 2.3 Å resolution for two crystal forms of PrsEf and at 2.8 Å resolution for PrsRp. For both enzymes, the analog lies in a crevice delimited by motifs 2 and 3 and by the contiguous fifth β strand of the conserved β sheet. Interactions between both prokaryote-like ProRSs and ProAMS are shown in Figures 4A and 4B and are compared to the equivalent binary complex of eukaryote-like PrsTt in Figure 4C. The negatively charged sulphamoyl group is fixed between the class II conserved Arg140 on one side and either His410 in PrsEf or Gln285 in PrsRp on the other side (Figures 4A and 4B). Interestingly there are no divalent cations in this complex, and correspondingly, unlike in the ATP case, there are direct hydrogen bond interactions between the sulfate and the functionally equivalent His410 imidazol Nδ1 of PrsEf or the Gln285 amine function of PrsRp. Concerning the prolyl moiety of the adenylate, the hydrophobic part of the amino acid fills a cavity bordered by the methionyl, the aspartic acyl, and the tyrosyl residues of the strictly conserved prokaryote-like 157-MKDxYSF-163 motif and the main chain of the 314-GSYG-317 motif (439-GCYG-442 in PrsEf). The imino group is held by hydrogen bonds to Thr109 and Glu111 of the class IIa conserved TXE loop, and the proline carbonyl-oxygen interacts with class II conserved Arg140. As discussed below, binding of the adenylate analog to PrsEf and PrsRp is accompanied by a conformational change and ordering of the proline-binding loop, similar, but of greater magnitude, to that previously observed in PrsTt.

Most ProRSs, whether prokaryote- or eukaryote/archael-like, are unable to discriminate efficiently against

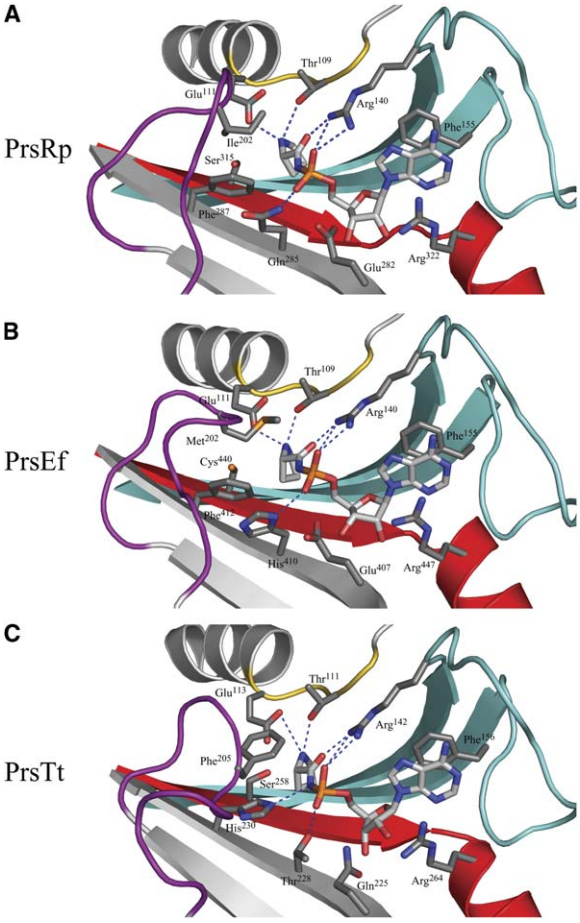


Figure 4. Interactions of ProRSs with the Prolyl-Adenylate Analog, ProAMS

(A–C) ProAMS (predominantly gray molecule) bound in the active sites of (A) *R. palustris*, (B) *E. faecalis*, and (C) *T. thermophilus* ProRS, showing key hydrogen bonds to the proline and sulfate moieties. Class II synthetase conserved motifs 1, 2, and 3 are shown in green, cyan, and red, respectively, and the TXE loop is shown in gold. On the proline-binding loop (violet), which is in the closed conformation, hydrophobic residues Ile202, Met202, and Phe205 play equivalent roles in PrsRp, PrsEf, and PrsTt, respectively.

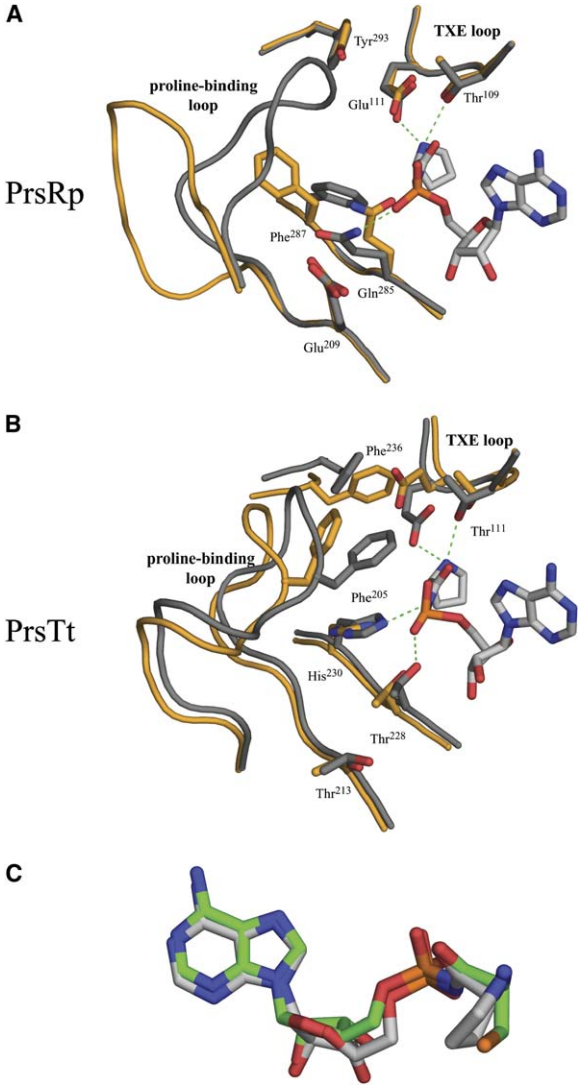


Figure 5. Conformational Changes Induced by Prolyl-Adenylate Binding

(A and B) Conformational changes induced in (A) *R. palustris* and (B) *T. thermophilus* ProRS active sites by the interaction with the prolyl-adenylate analog. The ligand-free and -bound states are shown in yellow and gray, respectively. In PrsRp (and PrsEf, not shown), there is a large change in the proline-binding loop and a rotamer flip of Phe287.

(C) Comparison of the conformations of ProAMS (predominantly gray molecule) and CysAMS (predominantly green molecule) bound in the *R. palustris* active site, after superposition of the Cα atoms of the corresponding structures.

cysteine (Ahel et al., 2002) and alanine (Beuning and Musier-Forsyth, 2001). The structure of the PrsRp:CysAMS binary complex has been solved and refined at 2.85 Å resolution. CysAMS takes up very nearly the same conformation as ProAMS (Figure 5C), with the same induced fit movement of the proline-binding loop. The cysteine side chain partially occupies the volume available to the proline ring without introducing any steric clash. Similar observations have been made from structural studies of the CysAMS and AlaAMS bound in the active site of the eukaryote/archael-like *Methanothermobacter thermautotrophicus* ProRS (Kamtekar et al., 2003).

Discussion

Induced-Fit Substrate Recognition of Proline in Prokaryote-like and Eukaryote/Archael-like ProRSs

Several class II aaRSs, notably prolyl- and histidyl-tRNA synthetases, undergo induced-fit conformational changes upon amino acid binding (Yaremchuk et al., 2001). In the case of the eukaryote-like PrsTt, it has been previously described that tight proline binding is achieved by the closing of the proline-binding loop over the bound amino acid (Yaremchuk et al., 2001). This is also true for both prokaryote-like ProRSs, although in these cases the displacement of the loop is of considerably larger magnitude (Figures 5A and 5B). In the absence of bound ligand or in the presence of ATP, the proline-binding loop of PrsRp and PrsEf, residues 199-xGxI/MGGxx-206, is either disordered or observed in an open conformation position well removed from the active site. In the presence of ProAMS, the

loop is well ordered and closed over the proline moiety with an average movement of 7.7 Å for the eight Cα atoms comprising the loop. A conserved, medium-sized hydrophobic residue, Ile202 in PrsRp and Met202 in PrsEf, is closest to the proline and is analogous to the absolutely conserved phenylalanine in the proline-binding loop of eukaryote-like ProRSs (Phe205 in PrsTt). In prokaryote-like ProRSs, two highly conserved flanking motifs, 174-YxxxxxAYxxxFxR-187 and 208-H/KEF-210, firmly anchored the proline-binding loop in the catalytic domain, thus allowing its large movement. In prokaryote-like PrsRS, loop closure necessitates a rotamer flip of the conserved Phe287 (PrsRp)/Phe412 (PrsEf) (Figure 5A). In PrsTt, the equivalent of this residue is His230, which interacts with the proline carboxyl group and does not change conformation. Instead, in PrsTt, Phe236 flips the rotamer upon proline binding, whereas the equivalent residue in prokaryote-like ProRSs, Tyr293 (PrsRp)/Tyr418 (PrsEf), is stably anchored by hydrogen bonding to conserved Arg85 (Figures 5A and 5B). There is a more significant difference in the induced-fit mechanism than these small differences: in eukaryote-like PrsTt, segment 76–86 is disordered, except in the presence of the adenylate, where it takes a helical conformation (Yaremchuk et al., 2001), but the equivalent region always forms an ordered helix in prokaryote-like PrsRS.

Despite the induced-fit mechanism of substrate amino acid recognition, the proline-binding pocket of all types of ProRSs is also able to accommodate cysteine, sometimes with very similar affinity (Ahel et al., 2002). In the case of PrsRp, using the same concentrated protein solution, it was easier to cocrystallize the PrsRp:CysAMS complex (100 nM final of compound) than the PrsRp:ProAMS one (500 nM of compound with the active sites half occupied). Thus, the affinity seems to be better for CysAMS than for ProAMS, paralleling the K_M values for the activation of cysteine and proline of 0.17 mM and 0.28 mM, respectively, measured by pyrophosphate exchange for PrsRp (Ahel et al., 2002). Despite this, the specificity of the activation reaction has been shown to be in favor of the cognate amino acid, with proline being activated 333 times better than cysteine by PrsRp (Ahel et al., 2002). In the same study, it was shown that the ratio of the initial overall rate of proline to cysteine aminoacylation by PrsRp was about 460. The equivalent specificity ratios for activation and aminoacylation, respectively, are 369 and 660 for PrsEc. These results suggest that subtle differences in the way the 3′ end of the tRNA is bound in the presence of cognate or noncognate amino acid may also enhance overall specificity, at least for some bacterial-type enzymes.

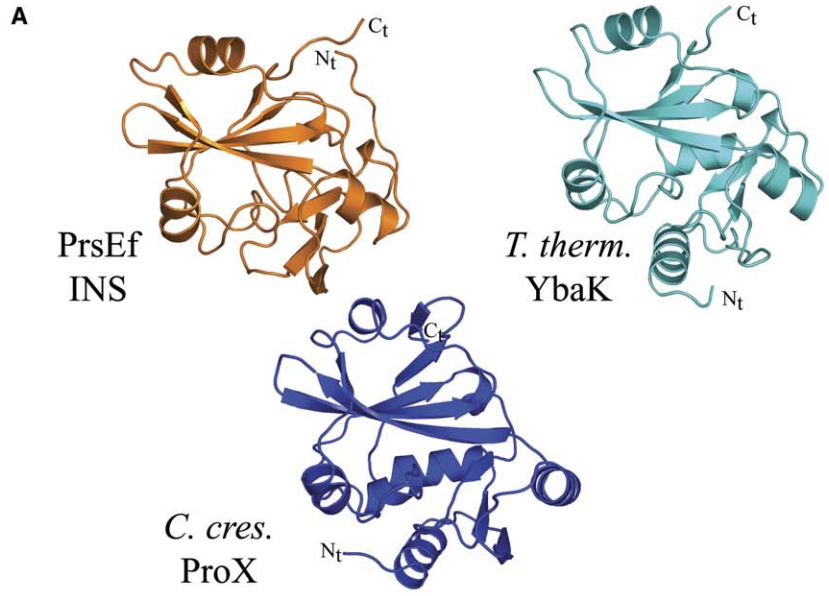
The ProRS-ProX-YbaK Family of Deacylase Domains

As discussed in the Introduction, the *cis*-editing domain of bacterial ProRSs is homologous to the N-terminal extension of lower eukaryotic ProRSs as well as to the two paralogs, YbaK and ProX, which can be distinguished on the basis of sequence alignments (Figure S1). YbaK-like proteins are thus far exclusively found in bacteria (including one example to date in an archaebacterium, *Aeropyrum pernix*), yet there appears to be no correlation with whether or not the organism has a eukaryote-like ProRS (e.g., *T. thermophilus*, *A. pernix*), a bacterial-like ProRS with an inserted editing domain (e.g., *E. coli*, *E. faecalis*), or a bacterial-like ProRS without an editing domain (e.g., *R. palustris*). ProX-like homologs on the other hand are found more broadly in prokaryotes and eukaryotes, including plants and mammals, but not so far in archae. Some organisms such as *E. coli* and *P. aeruginosa* possess three homologous deacylases, YbaK, ProX, and the ProRS INS. Recent results showed that YbaK and ProX proteins are moderately general aminoacyl-tRNA deacylases with preferences for certain amino acids, but not for tRNA. Thus, *Haemophilus influenzae* and *E. coli* YbaK proteins preferentially hydrolyze Cys-tRNA^{Pro} and Cys-tRNA^{Cys} (An and Musier-Forsyth, 2004, 2005; Ruan and Soll, 2005), but they also have weak deacylase activity (i.e., requiring high protein concentration) against Gly-tRNA^{Gly}, Ala-tRNA^{Ala}, Ser-tRNA^{Ser}, Pro-tRNA^{Pro}, and Met-tRNA^{Met} (in that order), whereas the ProX homologs only deacylate Ala-tRNA^{Pro}, Ala-tRNA^{Ala}, and Gly-tRNA^{Pro} (in that order of preference) (Ahel et al., 2003; Ruan and Soll, 2005). The activity of ProX is most similar to that of the *cis*-editing domain of ProRSs, which efficiently deacylates Ala-tRNA^{Pro}, but neither Pro-tRNA^{Pro} nor Cys-tRNA^{Pro} (Ahel et al., 2003; Beuning and Musier-Forsyth, 2000).

Atomic structures of three YbaK and two ProX proteins are known: *Aeropyrum pernix* Ape2540 YbaK (PDB ID 1WDV; Dali-lite z-score, 18.0; 135 aligned residues; rmsd compared to PrsEf, 2.1 Å), *Thermus thermophilus* YbaK (PDB ID 2CX5; 16.3; 136; 2.3 Å), *Haemophilus influenzae* YbaK (PDB ID 1DBX; 15.1; 133; 2.7 Å), *Caulobacter crescentus* ProX (PDB ID 1VJF; 12.0; 132; 3.0 Å), and *Agrobacterium tumefaciens* Atu3699 ProX (PDB ID 1VKI; 11.8; 128; 2.8 Å). Representative structures and sequence alignments of these six domains are shown in Figures 6A and 6C. Note that the N-terminal extension of the domain in PrsEf wraps around the protein to emerge close to the C terminus (Figure 6A), allowing the insertion of the domain into the body of the enzyme. Using SSM, a core of 116 residues can be aligned in three dimensions (featuring eight secondary-structure elements); pairwise rmsds and sequence identities are as shown in Figure 6B. More extensive structure-based sequence alignments (Figure S1) show that, although clearly evolutionarily related, the four classes of domain—bacterial ProRS insertion, ProX, YbaK, and N-terminal extension of lower eukaryotic ProRSs—have distinctive conserved motifs, and that the only strictly conserved residue is the functionally important lysine, Lys279, in PrsEf (Figures 6C and 7). The N-terminal extensions in the yeast ProRSs present a number of significant deviations from consensus sequences that may indicate that they are nonfunctional (Ahel et al., 2003).

Structural similarity (e.g., Dali z-score and rmsd) and phylogenetic analysis (Figure S2) suggest that the bacterial ProRS insertion is most closely related to YbaK proteins, whereas the N-terminally appended domain found in lower eukaryote ProRSs is closer to ProX. However, this does not correlate with the apparent closer functional similarity of ProX and the bacterial ProRS INSs, which have a preference for deacylation of Ala-tRNA^{Pro} and neither of which deacylate Cys-tRNA^{Pro} as YbaK does. A closer examination of the three-dimensional alignment suggests a possible rationale for this

A



B

% identity/RMSD	ProXCc	ProXAt	YbaKAp	YbaKTt	YbaKHi	PrsEf
ProXCc 1vjf:A		2.14Å	2.33Å	2.46Å	2.68Å	2.76Å
ProXAt 1vki:A	43.1%		2.30Å	2.38Å	2.43Å	2.72Å
YbaKAp 1wdv:A	14.7%	19.0%		1.52Å	1.70Å	1.89Å
YbaKTt 2cx5:C	18.1%	14.7%	37.1%		1.87Å	2.17Å
YbaKHi 1dbx:A	11.2%	17.2%	26.7%	21.6%		2.51Å
PrsEf INS	12.1%	11.2%	18.1%	23.2%	16.4%	

C

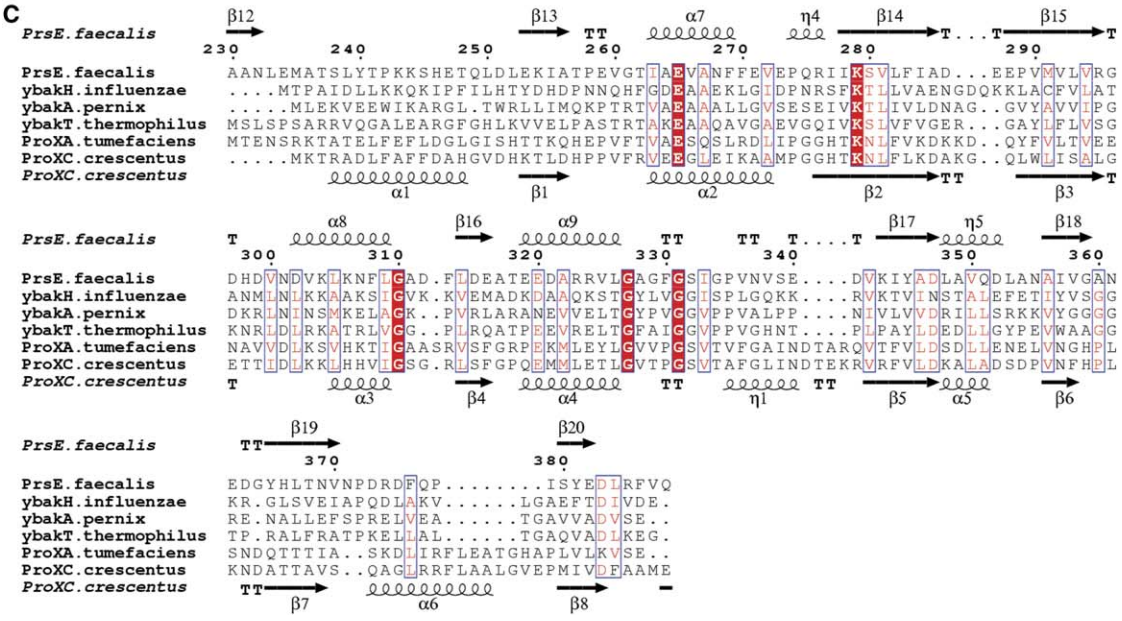


Figure 6. Sequence and Structure Alignment of the ProRS-ProX-YbaK Family of Deacylase Domains

(A) Ribbon representation of the PrsEf insertion domain (red), *T. thermophilus* YbaK (cyan), and *C. crescentus* ProX (blue) superimposed to be in the same orientation.

(B) Pairwise rmsd of Cα positions and sequence identity between the six known structures of members of the ProRS-ProX-YbaK family of deacylase domains.

(C) Structure-based sequence alignment of members of the ProRS-ProX-YbaK family of deacylase domains of known structure. PrsE.faecalis, insertion domain of PrsEf (accession number Q831W7); YbaKH.influenzae (accession number P45202, PDB entry 1DBX); YbaKA.pernix (Q9Y8U3, 1WDV); YbaKT.thermophilus (Q5SHN1, 2CX5); ProXA.tumefaciens (Q8U9M7, 1VKI); ProXC.crescentus (Q9ABV9, 1VJF). The secondary-structure elements correspond to the PrsEf insertion domain (top) and *C. crescentus* ProX (bottom).

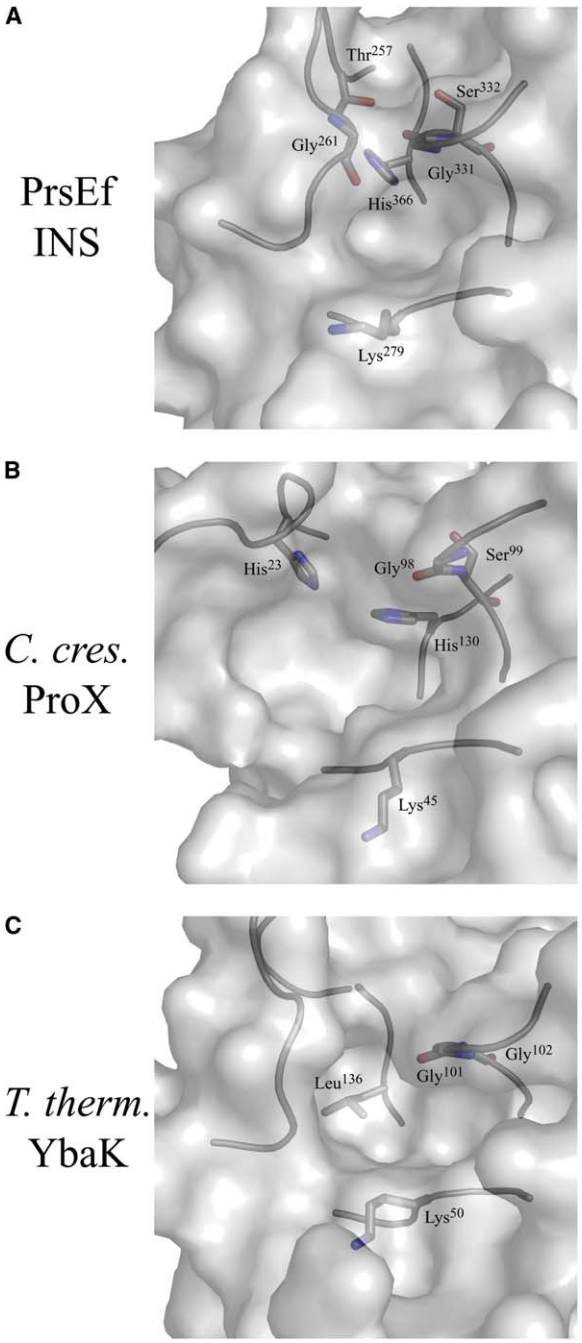


Figure 7. Putative Deacylase Active Site

(A–C) Surface representation of the putative deacylase active site of (A) the PrsEf insertion domain, (B) *C. crescentus* ProX protein, and (C) *T. thermophilus* YbaK protein, showing key residues mentioned in the text. Note the absolutely conserved lysine (Lys279, Lys45, and Lys50, respectively) pointing away from the active site and the conserved glycine (Gly331, Gly98 and Gly101, respectively) with its carbonyl oxygen pointing into the active site. In PrsEf and ProX, His366 and His130, respectively, are equivalently positioned.

discrepancy. Both the bacterial ProRS INS and ProX contain an absolutely conserved histidine, His366, in PrsEf and His130 in *C. crescentus* ProX (Figure S1). Although these residues do not correspond in the sequence alignment, their side chains almost exactly superpose in three dimensions, both emerging from the floor of the putative editing active site, close to the conserved lysine (Figure 7). The equivalent residue in PrsEc (His369) has been mutated to an alanine or cysteine with the result that deacylation activity, although reduced, notably loses specificity such that Pro-tRNA^{Pro} is also deacylated (Wong et al., 2002). This suggests that this histidine forms part of the amino acid specificity pocket of the editing site and, in the case of ProX and bacterial ProRS INS, restricts access to small amino acids such as alanine. Equivalent mutations that modulate the amino acid specificity of the editing active site are also known in the class I leucyl (Mursinna et al., 2001) and isoleucyl systems (Hendrickson et al., 2002).

Unfortunately, we have not been able to observe any electron density corresponding to either alanine or the nonhydrolyzable analog of mischarged Ala-tRNA^{Pro}, 3′-(L-alanyl)amino-3′-deoxyadenosine (Ala3AA or Ala-A76) when these were cocrystallized with PrsEf. Thus, the mode of binding to and mechanism of action of the editing active site remain unknown. However, as previously pointed out (Zhang et al., 2000), it seems likely that a functionally important element is Gly331 in PrsEf, conserved in all homologous editing domains except those of yeasts (Gly98 in *C. crescentus* ProX and Gly101 in *H. influenzae* YbaK) and usually forming a GXXXP motif (see alignments in Figure 6C and Figure S1). This glycine is at the apex of a tight turn, and its carbonyl oxygen prominently points into the editing active site (Figure 7). The PrsEf editing domain has a particularly closed active site, compared to the other known structures of YbaK and ProX, due to the occurrence of a second glycine loop containing the nonconserved Gly261, whose Cα is separated by just 3.3 Å from that of Gly361; these two glycine loops form a lid over the active site (Figure 7). The conformation of the conserved lysine (Lys279 in PrsEf) is of note. In the six known structures of the homologous editing domains, the lysine side chain is in an extended conformation and points away from other elements previously highlighted in the editing active site (Figure 7). In PrsEf, a sulfate ion from the crystallization medium is close to Lys279. This may indicate that the role of this lysine is to position the 3′ end of the tRNA by binding to the phosphate of C75 or A76. However, a more direct role in catalysis is suggested by the fact that a lysine-to-arginine mutation in *H. influenzae* YbaK abolishes activity (Musier-Forsyth, personal communication). For this reason, it is possible that the N-terminal editing-like domain of the ProRS from the bovine pathogen *Theileria parva* (accession number EAN33410), in which an arginine replaces the conserved lysine, is inactive. Indeed, it has been shown that the yeast N-terminal editing-like domain is nonfunctional (Ahel et al., 2003), probably due to the lack of several otherwise highly conserved features, including the GXXXP and HP motifs (see above), whereas the N-terminal extension of *Plasmodium falciparum* ProRS does have editing activity (Ahel et al., 2003).

Model for Posttransfer Editing Conformation of PrsEf

To gain some insight into the mechanism of editing, we have modeled the PrsEf-tRNA^{Pro} complex in the posttransfer editing conformation (Figure 8A). This is based

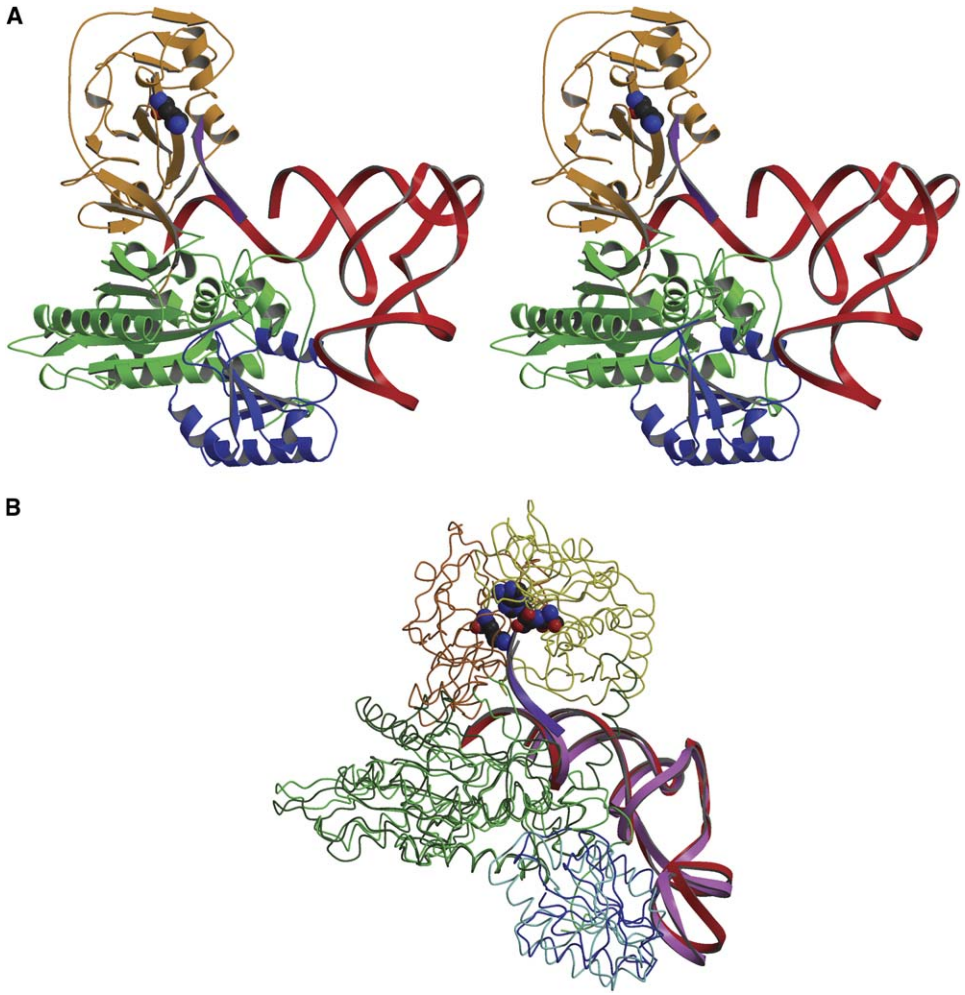


Figure 8. Model of the Posttransfer Editing Conformation of PrsEf

(A) Stereofigure of a hypothetical model of one subunit of PrsEf with bound tRNA in the aminoacylation configuration (red) and the posttransfer-editing conformation (3′ end in purple). The catalytic domain is green, the anticodon-binding domain is blue, and the editing domain is orange. The conserved lysine in the editing domain is in CPK representation. The aminoacylation complex is based on unpublished data of the PrsRp-tRNA^{pro} complex.

(B) Superposition of aminoacylation and posttransfer tRNA complexes of subunits of PrsEf (the catalytic domain is green, the anticodon-binding domain is blue, the editing domain is orange, and tRNA is red) and ThrEc (the catalytic domain is dark green, the anticodon-binding domain is light blue, the editing domain is yellow, and tRNA is pink). The modeled 3′ end of the tRNA entering the editing site is purple in each case. The ThrEc complex in the aminoacylation conformation is a crystal structure (PDB entry 1QF6), while all other complexes are modeled. Lysine 279 is shown in the PrsEf editing domain, and SerA76 (posttransfer-editing substrate analog) is shown in the ThrEc editing domain (larger molecule to right, position taken from PDB entry 1TKY) in CPK representation.

on the crystal structure of the PrsRp-tRNA^{Pro} complex that we have determined in the aminoacylation conformation (data not shown). In Figure 8B, we compare the results with the equivalent exercise of modeling the posttransfer editing conformation of ThrEc-tRNA^{Thr} (Dock-Bregeon et al., 2004). In both cases, an eventual rotation of the editing domain is not accounted for. As in the case of ThrEc-tRNA^{Thr}, a smooth change in direction of the four single-stranded nucleotides 73-CCCA is sufficient to bring the 3′ end of the tRNA^{Pro} into the PrsEf editing site. Although the orientation of the AlaA76 is as yet unknown in the editing site, it is possible to model a plausible conformation in which the conserved lysine is adjacent to the phosphate of A76 and the noncognate charged alanine enters the active site cleft with the side chain approaching His366 (see above), in such a way that a larger side chain, such as cysteine, would not bind. Interestingly, in comparing the overall arrangement between PrsEf-tRNA^{Pro} and ThrEc-tRNA^{Thr}, it is apparent that the PrsEf- and ThrEc editing domains approach the 3′ end of the tRNA from opposite sides (Figure 8B), with little steric overlap, reminiscent of the different orientations of the editing domains of LeuRSTt and ValRSTt (compare Figure 4A in Tukalo et al. [2005]).

Conclusions

The ProRSs are intriguing and unique among synthetases due to their particularly diverse architectures, notably the variable presence of a *cis*-editing domain (Figure 1A). Here, we compared the structure and small-substrate-binding site of two bacterial-type ProRSs, with and without an editing domain, with the previously

determined structure of the archael/eukaryote-type ProRSs. These results show that whereas there are significant systematic differences between the two types of ProRSs, even including the amino acids interacting with the substrates (which could be exploited for design of specific antibiotics), other features such as the induced-fit mechanism of proline recognition and the presence of an additional carboxyl group in the pyrophosphate binding site are conserved. Previous results as well as our own show that there is no structural hindrance to the binding of noncognate alanine and cysteine, which have smaller or equivalent volumes to proline, in the proline-binding site. This intrinsic lack of discrimination can give rise to significant and unacceptable levels of mischarged Ala-tRNA^{Pro} and Cys-tRNA^{Pro}, to an extent dependent on the particular source of the ProRS, but not strongly correlated with ProRS type (Ahel et al., 2002). Some bacterial and unicellular eukaryotic ProRSs have YbaK-like *cis*-editing domains that can specifically hydrolyze Ala-tRNA^{Pro} but are inactive against Cys-tRNA^{Pro}. A possible rationale for this, as pointed out previously (Ahel et al., 2002), is that a classical second sieve editing mechanism to discriminate against the smaller alanine is easier to implement than one to discriminate similarly sized cysteine from proline. It thus seems that in organisms in which the relative levels of free amino acids lead to unacceptable mischarging by ProRS, elimination of Cys-tRNA^{Pro} and Ala-tRNA^{Pro} (in cases in which there is no *cis*-editing domain) has been left to "third" parties such as the YbaK and ProX, present in most organisms (Ahel et al., 2003; An and Musier-Forsyth, 2004). This mechanism has the potential disadvantage that it is not necessarily tRNA^{Pro} specific, since, for instance, YbaK alone lacks specific tRNA recognition capabilities and in vitro hydrolyzes both Cys-tRNA^{Pro} and Cys-tRNA^{Cys} (An and Musier-Forsyth, 2005; Ruan and Soll, 2005). Upon overexpression, YbaK can be observed to regulate the level of tRNAs charged with cysteine in vivo (Ruan and Soll, 2005). However, in vitro results show that, whereas in the presence of ProRS, Cys-tRNA^{Pro} is indeed efficiently hydrolyzed by Ybak, in the presence of CysRS, synthesis of cognate Cys-tRNA^{Cys} is, if anything, enhanced by the presence of Ybak (An and Musier-Forsyth, 2005). Efficient hydrolysis of Cys-tRNA^{Pro} probably occurs through the formation of a ternary Ybak:ProRS:cys-tRNA^{Pro} complex (An and Musier-Forsyth, 2005), which not only restores the tRNA specificity to the editing mechanism (as it is guaranteed by the synthetase), but strongly increases the likelihood of the mischarged species being hydrolyzed rather than being taken up by Ef-Tu and used in protein synthesis.

Finally, it is interesting to consider possible evolutionary scenarios that led to the current diversity of ProRS architectures and distribution of YbaK-like proteins (also see Ahel et al., 2002, 2003). The core elements (catalytic- and anticodon-binding domains) of the archael- and bacterial-type ProRSs presumably arose from a common proline-specific ancestor, which earlier had diverged from the other class IIa enzymes, notably SerRS, ThrRS, and GlyRS. Both types of ProRSs then probably acquired, for stability reasons, new structural elements contiguous to the catalytic domain β sheet, either as an additional C-terminal domain (archael-type ProRS) or as a minimal insertion into the catalytic domain (bacterial-type ProRS; i.e., the purple elements in Figures 2A, 2B, and 3). Subsequently, to increase specificity, notably against alanine, ProRSs of both types may have acquired a *cis*-editing domain, either as an N-terminal extension (present day lower eukaryotes) or as an additional insertion into the catalytic domain (as in certain bacteria today) of the ancestral freestanding YbaK-like protein. Then, in certain organisms, perhaps those with favorable proline amino acid levels relative to alanine and cysteine or those in which the intrinsic ProRS specificity became sufficient, the *cis*-editing domain became first defunct (as in current-day yeasts) or was totally lost (as in certain bacteria and eukaryotic organelles). In these organisms, either deacylation of mischarged tRNA^{Pro} is not necessary (as levels are not life threatening), or it is achieved by the scavanger deacylases ProX and YbaK. However the possibility that acquisition of the *cis*-editing domain was a later event, restricted to lower eukaryotes and to certain bacterial lineages and neither archae nor metozoan ProRS, which had become fused with GluRS, cannot be excluded. Finally, it is interesting to recall that YbaK proteins, which preferentially deacylate Cys-tRNA^{Pro}, are widespread in bacteria, exceptional in archae (only *Aeropyrum pernix* to date), and nonexistent in eukaryotes. Similarly, ProX is absent from archae. Thus, there may still be as yet unidentified proteins that can deacylate Cys-tRNA^{Pro} in archae and eukaryotes and Ala-tRNA^{Pro} in archae.

Experimental Procedures

Expression and Purification

R. palustris ProRS (PrsRp) was expressed in *E. coli* strain BL21-CodonPlus-RIL (Stratagene) from the pET15b expression vector (Invitrogen) as an uncleavable His-tag fusion protein. After a 7 hr induction at 23°C with 1 mM isopropyl thiogalactopyronose (IPTG), the cells were harvested, resuspended in 20 mM Tris·HCl (pH 8.0), 100 mM NaCl, 5 mM $MgCl_2$, 2 mM 2-mercaptoethanol, and sonicated. After centrifugation, the supernatant was complemented with imidazol to reach 15 mM and loaded on a nickel affinity column. The Ni-NTA resin was washed with a 20 mM Tris·HCl (pH 8.0), 1 M NaCl, 15 mM imidazol, 5 mM $MgCl_2$, 2 mM 2-mercaptoethanol buffer. The purified protein was then eluted with a 20 mM Tris·HCl (pH 8.0), 100 mM NaCl, 250 mM imidazol, 5 mM $MgCl_2$, 5 mM 2-mercaptoethanol buffer and dialyzed against the same buffer without imidazol. The native protein was finally concentrated to obtain a 10 mg·ml^{-1} protein solution.

SeMet-substituted PrsRp was produced in a defined medium containing seleno-methionine as described (Van Duyne et al., 1993). Protein purification was initiated with the protocol described above. After dialysis, the SeMet PrsRp was loaded on a 1 ml MonoQ column (Amersham Biosciences). The protein was eluted with a linear gradient of 0.1–0.6 M NaCl (1 ml·min^{-1}; 0.6 M·hr^{-1}) in a 20 mM Tris·HCl (pH 8.0), 5 mM $MgCl_2$, 2 mM 2-mercaptoethanol buffer. The SeMet ProRS was finally dialyzed against a 20 mM Tris·HCl (pH 8.0), 100 mM NaCl, 5 mM $MgCl_2$, 5 mM 2-mercaptoethanol buffer and concentrated.

The *E. faecalis* ProRS (PrsEf) gene was cloned into pQE-60 plasmid (QIAGEN). *E. coli* M15 strain transformed with the corresponding plasmid was grown at 37°C in LB in the presence of 140 μg·ml^{-1} ampicillin and 35 μg·ml^{-1} kanamycin. At an OD_{600} of 1.5, the expression was induced with 1 mM IPTG, and the culture was incubated for an additional 4 hr. Cells were then harvested, resuspended in 20 mM Tris·HCl (pH 8.0), 1 mM EDTA, 5 mM DTT (buffer A), and disrupted by sonication. After centrifugation, the PrsEf-containing protein fraction was selectively precipitated with saturated $(NH_4)_2SO_4$ (from 40% to 80%), and after intensive dialysis against buffer A, it was loaded on a Q-Sepharose-FF column equilibrated with the same

buffer. Elution occurred with a linear gradient of 20–800 mM NaCl. The active fractions were dialyzed and loaded on a MonoQ20 column equilibrated with 20 mM HEPES (pH 7.5), 5 mM $MgCl_2$, 0.5 mM EDTA, 5 mM DTT. The protein was eluted with a gradient of 20–600 mM NaCl. Fractions containing PrsEf were concentrated to 10 mg·ml^{-1}. The buffer was changed to 10 mM HEPES (pH 7.0), 5 mM DTT before the protein was stored at −80°C.

Crystallization

Hanging-drop vapor diffusion experiments were carried out at 20°C. Native and SeMet-labeled PrsRp crystals were obtained by mixing protein solution at 5 mg·ml^{-1} with an equal volume of the optimized solution containing 0.1 M citric acid (pH 5.5), 15%–17% PEG 3000. Crystals were frozen in liquid nitrogen after transfer for a few seconds in 0.1 M citric acid (pH 5.5), 15%–17% PEG 3000, 20% ethylene glycol. For PrsRp:adenylate analog complex cocrystals, the enzyme was mixed first with 500 nM ProAMS or 100 nM CysAMS, and then with an equal volume of a solution containing 0.1 M citric acid (pH 5.5), 10%–11% PEG 3000, 15%–20% ethylene glycol. The drops were equilibrated for about 21 hr and streak seeded from crushed native crystals with a cat's whisker. Crystals were frozen in liquid nitrogen after transfer for a few seconds in mother liquor containing 20% ethylene glycol. ProAMS and CysAMS were purchased from RNA-TEC NV (Leuven, Belgium).

Monoclinic crystals of PrsEf were obtained by mixing 5 μl of the protein at 3 mg/ml in 10 mM MES (pH 6.5), 2 mM DTT, 6 mM $MgCl_2$, and 500 μM prolyl-adenylate analog (or 5 mM prolinol and 2 mM ATP) with 5 μl of the reservoir solution containing 20% PEG 6K, 1 M LiCl, 100 mM MES (pH 6.5). They were of space group $P2_1$, had cell dimensions of a = 70.5 Å, b = 101.4 Å, c = 104.2 Å, β = 106.1 (cocrystallized with pro-AMS, data not shown) or a = 68.6 Å, b = 92.7 Å, c = 101.2 Å, β = 106.1 (cocrystallized with ATP, Mn^{2+}, and prolinol), and diffracted to 2.3 Å resolution. Tetragonal crystals (space group $P4_12_12$) of PrsEf grew in 52% ammonium sulfate, only in the presence of the prolyl-adenylate analog. Frozen tetragonal crystals were found to have variable cell dimensions, e.g., a = b = 121.4 Å, c = 178.8 Å (best diffraction at 2.3 Å resolution, as described here), a = b = 122.1 Å, c = 183.2 Å (diffraction to 2.9 Å resolution), and a = b = 128.1 Å, c = 180.7 Å (mercury- or platinum-soaked crystals). This hindered structure determination by MIR methods. Crystals were flash cooled in the nitrogen stream after transfer for a few seconds in mother liquor containing 22% glycerol (monoclinic crystals) or 80% lithium sulfate (tetragonal crystals).

Structure Determination and Refinement

The structure of PrsRp was solved by a two-wavelength (inflection point and peak of the Se K-edge) SeMet MAD experiment with crystals diffracting to 3.5 Å resolution. Data were processed by using XDS (Kabsch, 1993). Selenium sites (30 of 36, for 3 monomers in the asymmetric unit) were identified with SHELXD (Uson and Sheldrick, 1999), refined with SHARP (de la Fortelle and Bricogne, 1997), and used for phasing in SOLVE (Terwilliger et al., 1987). A truncated *T. thermophilus* ProRS model was used to guide building of an initial structure. Phases were then extended to an isomorphous native data set at 2 Å resolution, allowing most of the model to be automatically built by using ARP-wARP (Perrakis et al., 1999). The model was completed in O (Jones et al., 1991) and refined (R_{free} = 20.1; R factor = 16.9) with CNS (Brunger et al., 1998) and then REFMAC5 with TLS refinement (Murshudov, 1997). The crystallographic asymmetric unit contains 3 molecules (1 dimer plus 1 monomer); each subunit comprised 441 residues (438 for the enzyme and 3 additional residues from the uncleaved N-terminal His-tag). PrsRp:ProAMS diffraction data were isomorphous to the native enzyme data (space group $C222_1$), whereas PrsRp:CysAMS crystals belonged to space group C2. The structure was solved by molecular replacement with MOLREP (Vagin and Teplyakov, 2000), followed by model building and refinement.

The tetragonal crystal form of PrsEf was solved by molecular replacement with the PrsRp structure. Building of the extra editing domain was assisted by running RESOLVE by using the 2-fold noncrystallographic symmetry and automatic sequence fitting. $P2_1$ structures were determined by molecular replacement, followed by refinement. All PrsEf structures contain a dimer in the asymmetric unit. In most structures, residues 12–18 are poorly ordered, and electron density for the extreme C-terminal residues 566–572 (571–572 in one subunit) is absent. Residues 199–205 of the proline-binding loop are either in an ordered closed conformation (structure with ProAMS), ordered open conformation (subunit B of ATP, Mn^{2+}, and prolinol complex), or disordered (subunit A of ATP, Mn^{2+}, and prolinol complex).

Data collection and refinement statistics are summarized in Table 1.

Supplemental Data

Supplemental Data include two figures and are available at http://www.structure.org/cgi/content/full/14/10/1511/DC1/.

Acknowledgments

The authors thank the members of the European Synchrotron Radiation Facility (ESRF)-European Molecular Biology Laboratory Joint Structural Biology Group for access to ESRF synchrotron facilities and, in particular, Pierre Legrand for help during the multiwavelength anomalous dispersion data collection for PrsRp and the subsequent structure solution. The work on PrsEf was initially a collaboration with Arnulf Dorn, Franck Darnel, and Malcolm Page (Basilea, Basel, Switzerland), from whom the expression clone was obtained. We are very grateful to Dieter Söll (Yale) for a gift of the expression clone of *Rhodopseudomonas palustris* ProRS and to Susan Fridd for initial crystallization work on PrsRp. T.C., whose PI is Michael Kron (Medical College of Wisconsin), was supported by National Institutes of Health grant U01 AI53877-01 ("Microfilaricidal aminoacyl-tRNA synthetase inhibitors").

Received: July 12, 2006
Revised: August 18, 2006
Accepted: August 19, 2006
Published: October 10, 2006

References

Ahel, I., Stathopoulos, C., Ambrogelly, A., Sauerwald, A., Toogood, H., Hartsch, T., and Soll, D. (2002). Cysteine activation is an inherent in vitro property of prolyl-tRNA synthetases. J. Biol. Chem. *277*, 34743–34748.

Ahel, I., Korencic, D., Ibba, M., and Soll, D. (2003). Trans-editing of mischarged tRNAs. Proc. Natl. Acad. Sci. USA *100*, 15422–15427.

Ambrogelly, A., Ahel, I., Polycarpo, C., Bunjun-Srihari, S., Krett, B., Jacquin-Becker, C., Ruan, B., Kohrer, C., Stathopoulos, C., RajBhandary, U.L., and Soll, D. (2002). Methanocaldococcus jannaschii prolyl-tRNA synthetase charges tRNA(Pro) with cysteine. J. Biol. Chem. *277*, 34749–34754.

An, S., and Musier-Forsyth, K. (2004). Trans-editing of Cys-tRNAPro by *Haemophilus influenzae* YbaK protein. J. Biol. Chem. *279*, 42359–42362.

An, S., and Musier-Forsyth, K. (2005). Cys-tRNAPro editing by *Haemophilus influenzae* YbaK via a novel synthetase YbaK tRNA ternary complex. J. Biol. Chem. *280*, 34465–34472.

Beebe, K., Ribas De Pouplana, L., and Schimmel, P. (2003). Elucidation of tRNA-dependent editing by a class II tRNA synthetase and significance for cell viability. EMBO J. *22*, 668–675.

Beuning, P.J., and Musier-Forsyth, K. (2000). Hydrolytic editing by a class II aminoacyl-tRNA synthetase. Proc. Natl. Acad. Sci. USA *97*, 8916–8920.

Beuning, P.J., and Musier-Forsyth, K. (2001). Species-specific differences in amino acid editing by class II prolyl-tRNA synthetase. J. Biol. Chem. *276*, 30779–30785.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Cusack, S. (1995). Eleven down and nine to go. Nat. Struct. Biol. *2*, 824–831.

Cusack, S., Berthet-Colominas, C., Hartlein, M., Nassar, N., and Leberman, R. (1990). A second class of synthetase structure revealed by

X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. Nature *347*, 249–255.

de la Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Macromol. Crystallogr. A *276*, 472–494.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (San Carlos, CA: DeLano Scientific).

Dock-Bregeon, A., Sankaranarayanan, R., Romby, P., Caillet, J., Springer, M., Rees, B., Francklyn, C.S., Ehresmann, C., and Moras, D. (2000). Transfer RNA-mediated editing in threonyl-tRNA synthetase. The class II solution to the double discrimination problem. Cell *103*, 877–884.

Dock-Bregeon, A.C., Rees, B., Torres-Larios, A., Bey, G., Caillet, J., and Moras, D. (2004). Achieving error-free translation; the mechanism of proofreading of threonyl-tRNA synthetase at atomic resolution. Mol. Cell *16*, 375–386.

Dwivedi, S., Kruparani, S.P., and Sankaranarayanan, R. (2005). A D-amino acid editing module coupled to the translational apparatus in archaea. Nat. Struct. Mol. Biol. *12*, 556–557.

Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature *347*, 203–206.

Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J., Vassylyev, D.G., and Yokoyama, S. (2000). Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA(Val) and valyl-tRNA synthetase. Cell *103*, 793–803.

Fukunaga, R., and Yokoyama, S. (2005). Aminoacylation complex structures of leucyl-tRNA synthetase and tRNA(Leu) reveal two modes of discriminator-base recognition. Nat. Struct. Mol. Biol. *12*, 915–922.

Fukunaga, R., and Yokoyama, S. (2006). Structural basis for substrate recognition by the editing domain of isoleucyl-tRNA synthetase. J. Mol. Biol. *359*, 901–912.

Gruic-Sovulj, I., Uter, N., Bullock, T., and Perona, J.J. (2005). tRNA-dependent aminoacyl-adenylate hydrolysis by a nonediting class I aminoacyl-tRNA synthetase. J. Biol. Chem. *280*, 23978–23986.

Hati, S., Ziervogel, B., Sternjohn, J., Wong, F.C., Nagan, M.C., Rosen, A.E., Siliciano, P.G., Chihade, J.W., and Musier-Forsyth, K. (2006). Pre-transfer editing by class II prolyl-tRNA synthetase: Role of aminoacylation active site in "selective release" of noncognate amino acids. J. Biol. Chem. *281*, 27862–27872.

Hendrickson, T.L., Nomanbhoy, T.K., de Crecy-Lagard, V., Fukai, S., Nureki, O., Yokoyama, S., and Schimmel, P. (2002). Mutational separation of two pathways for editing by a class I tRNA synthetase. Mol. Cell *9*, 353–362.

Jakubowski, S. (2004). Accuracy of aminoacyl-tRNA synthetases: proofreading of amino acids. In Aminoacyl-tRNA Synthetases, M. Ibba, C. Francklyn, and S. Cusack, eds. (Austin, TX: Landes Biosciences), pp. 384–396.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. *26*, 795–800.

Kamtekar, S., Kennedy, W.D., Wang, J., Stathopoulos, C., Soll, D., and Steitz, T.A. (2003). The structural basis of cysteine aminoacylation of tRNAPro by prolyl-tRNA synthetases. Proc. Natl. Acad. Sci. USA *100*, 1673–1678.

Kotik-Kogan, O., Moor, N., Tworowski, D., and Safro, M. (2005). Structural basis for discrimination of L-phenylalanine from L-tyrosine by phenylalanyl-tRNA synthetase. Structure *13*, 1799–1807.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. *60*, 2256–2268.

Lincecum, T.L., Jr., Tukalo, M., Yaremchuk, A., Mursinna, R.S., Williams, A.M., Sproat, B.S., Van Den Eynde, W., Link, A., Van Calenbergh, S., Grotli, M., et al. (2003). Structural and mechanistic basis of pre- and posttransfer editing by leucyl-tRNA synthetase. Mol. Cell *11*, 951–963.

Murshudov, G.N. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Mursinna, R.S., Lincecum, T.L., Jr., and Martinis, S.A. (2001). A conserved threonine within *Escherichia coli* leucyl-tRNA synthetase prevents hydrolytic editing of leucyl-tRNALeu. Biochemistry *40*, 5376–5381.

Musier-Forsyth, K., Burke, B., and Cusack, S. (2004). Prolyl-tRNA Synthetases. In The Aminoacyl-tRNA Synthetases, M. Ibba, C. Francklyn, and S. Cusack, eds. (Austin, TX: Landes Bioscience), pp. 149–161.

Nordin, B.E., and Schimmel, P. (2005). Isoleucyl-tRNA synthetase. In The Aminoacyl-tRNA Synthetases, M. Ibba, C. Francklyn, and S. Cusack, eds. (Austin, TX: Landes Biosciences), pp. 24–35.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. Nat. Struct. Biol. *6*, 458–463.

Roy, H., Ling, J., Irnov, M., and Ibba, M. (2004). Post-transfer editing in vitro and in vivo by the beta subunit of phenylalanyl-tRNA synthetase. EMBO J. *23*, 4639–4648.

Ruan, B., and Soll, D. (2005). The bacterial YbaK protein is a Cys-tRNAPro and Cys-tRNACys deacylase. J. Biol. Chem. *280*, 25887–25891.

Silvian, L.F., Wang, J., and Steitz, T.A. (1999). Insights into editing from an ile-tRNA synthetase structure with tRNAile and mupirocin. Science *285*, 1074–1077.

Sokabe, M., Okada, A., Yao, M., Nakashima, T., and Tanaka, I. (2005). Molecular basis of alanine discrimination in editing site. Proc. Natl. Acad. Sci. USA *102*, 11669–11674.

Swairjo, M.A., and Schimmel, P.R. (2005). Breaking sieve for steric exclusion of a noncognate amino acid from active site of a tRNA synthetase. Proc. Natl. Acad. Sci. USA *102*, 988–993.

Terwilliger, T.C., Kim, S.H., and Eisenberg, D. (1987). Generalized-method of determining heavy-atom positions using the difference patterson function. Acta Crystallogr. A *43*, 1–5.

Tukalo, M., Yaremchuk, A., Fukunaga, R., Yokoyama, S., and Cusack, S. (2005). The crystal structure of leucyl-tRNA synthetase complexed with tRNA(Leu) in the post-transfer-editing conformation. Nat. Struct. Mol. Biol. *12*, 923–930.

Uson, I., and Sheldrick, G.M. (1999). Advances in direct methods for protein crystallography. Curr. Opin. Struct. Biol. *9*, 643–648.

Vagin, A., and Teplyakov, A. (2000). An approach to multi-copy search in molecular replacement. Acta Crystallogr. D Biol. Crystallogr. *56*, 1622–1624.

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J. Mol. Biol. *229*, 105–124.

Woese, C.R., Olsen, G.J., Ibba, M., and Soll, D. (2000). Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol. Mol. Biol. Rev. *64*, 202–236.

Wong, F.C., Beuning, P.J., Nagan, M., Shiba, K., and Musier-Forsyth, K. (2002). Functional role of the prokaryotic proline-tRNA synthetase insertion domain in amino acid editing. Biochemistry *41*, 7108–7115.

Wong, F.C., Beuning, P.J., Silvers, C., and Musier-Forsyth, K. (2003). An isolated class II aminoacyl-tRNA synthetase insertion domain is functional in amino acid editing. J. Biol. Chem. *278*, 52857–52864.

Yaremchuk, A., Cusack, S., and Tukalo, M. (2000). Crystal structure of a eukaryote/archaeon-like protyl-tRNA synthetase and its complex with tRNAPro(CGG). EMBO J. *19*, 4745–4758.

Yaremchuk, A., Tukalo, M., Grotli, M., and Cusack, S. (2001). A succession of substrate induced conformational changes ensures the amino acid specificity of *Thermus thermophilus* prolyl-tRNA synthetase: comparison with histidyl-tRNA synthetase. J. Mol. Biol. *309*, 989–1002.

Zhang, H., Huang, K., Li, Z., Banerjei, L., Fisher, K.E., Grishin, N.V., Eisenstein, E., and Herzberg, O. (2000). Crystal structure of YbaK

protein from *Haemophilus influenzae* (HI1434) at 1.8 Å resolution: functional implications. Proteins *40*, 86–97.

Accession Numbers

Structure factors and coordinates for the six structures described here have been deposited at the PDB for immediate release with codes 2I4L (native PrsRp), 2I4M (PrsRp in complex with ProAMS), 2I4N (PrsRp in complex with CysAMS), 2I4O (PrsRp in complex with ATP), 2J3L (PrsEf in complex with ATP and prolinol), and 2J3M (PrsEf in complex with ProAMS).

nature structural & molecular biology

Structural dynamics of the aminoacylation and proofreading functional cycle of bacterial leucyl-tRNA synthetase

Andrés Palencia[1,3], Thibaut Crépin[1,3], Michael T Vu[2], Tommie L Lincecum Jr[2], Susan A Martinis[2] & Stephen Cusack[1]

Leucyl-tRNA synthetase (LeuRS) produces error-free leucyl-tRNA^{Leu} by coordinating translocation of the 3′ end of (mis-)charged tRNAs from its synthetic site to a separate proofreading site for editing. Here we report cocrystal structures of the *Escherichia coli* LeuRS–tRNA^{Leu} complex in the aminoacylation or editing conformations, showing that translocation involves correlated rotations of four flexibly linked LeuRS domains. This pivots the tRNA to guide its charged 3′ end from the closed aminoacylation state to the editing site. The editing domain unexpectedly stabilizes the tRNA during aminoacylation, and a large rotation of the leucine-specific domain positions the conserved KMSKS loop to bind the 3′ end of the tRNA, promoting catalysis. Our results give new insight into the structural dynamics of a molecular machine that is essential for accurate protein synthesis.

LeuRS is a large, multidomain, class Ia aminoacyl-tRNA synthetase (aaRS) whose essential function in all organisms is to synthesize Leu-tRNA^{Leu} for use in protein synthesis. Similar to several other synthetases, notably the other two class 1a synthetases, valyl-tRNA synthetase (ValRS) and isoleucyl-tRNA synthetase (IleRS), LeuRS possesses an error-correction mechanism to enhance the specificity of aminoacylation and, thus, the accuracy of protein synthesis. This post-transfer editing mechanism hydrolytically deacylates tRNA^{Leu} that has been mischarged with noncognate amino acids that are similar to leucine, such as isoleucine, methionine or non-canonical norvaline[1].

LeuRS comprises a main enzyme body (Rossmann-fold catalytic domain and class 1a anticodon–binding domain) and four flexibly linked additional domains, denoted the zinc (ZN1), editing, leucine-specific and C-terminal domains (**Fig. 1a**). Proofreading requires that the 3′ end of the tRNA, which is initially charged (or mischarged) in the so-called 'synthetic' site of the enzyme, translocate to the editing site, located ~35 Å away in an independently folded editing, or CP1, domain[2]. Previously, we and others have characterized various structural and biochemical features of the editing state of LeuRS. These include (i) determination of the crystal structure of the *Thermus thermophilus* LeuRS (LeuRSTT) tRNA^{Leu} in complex with the 3′ end of the tRNA bound in the editing site[3], (ii) elucidation of the structural basis for the amino acid specificity of the LeuRS editing site, which binds noncognate amino acids but rejects cognate leucine[2] and (iii) a theoretical analysis of the hydrolytic mechanism[4]. Recently, a series of benzoxaborole compounds were shown to bind specifically in the LeuRS editing site in a tRNA-dependent fashion[5]. These compounds form a long-lived covalent adduct with the tRNA 3′-terminal ribose hydroxyl groups in the editing site, thus trapping the tRNA on the enzyme and inhibiting aminoacylation[5]. Benzoxaboroles have the potential to be potent antibiotics and are currently undergoing development for use against fungal, bacterial and parasitic pathogens[5–8].

Although the editing states of bacterial class 1a LeuRS[3], ValRS[9] and IleRS[10] are well characterized, the only published structure of any class 1a synthetase with tRNA bound in the aminoacylation active site is that of an archaeal LeuRS from *Pyrococcus horikoshii* (LeuRSPH)[11]. However, archaeal and eukaryotic cytoplasmic LeuRSs are architecturally distinct from bacterial LeuRS[12,13]. Therefore, the available structures of the editing and aminoacylation states, being, respectively, from bacterial and archaeal systems, are not directly comparable. Furthermore, the LeuRSPH–tRNA^{Leu} aminoacylation complex lacks any bound small substrates and does not represent the enzymatically active aminoacylation state. Indeed, few class 1 synthetase–tRNA complexes have been determined in this functional state, the only examples being the class 1b GlnRS[14–16] and GluRS[17,18] aminoacylation complexes. In other class I cocrystal structures, either the tRNA is directed to the editing site[9,10], the 3′ end is disordered or incorrectly bound[19–24], or the adenylate is not present[25].

Here we present the crystal structure of the functional aminoacylation complex of *E. coli* LeuRS (LeuRSEC). In this ternary complex structure, the 3′ end of *E. coli* tRNA$_5^{Leu}$(UAA) is bound in the synthetic site and poised to interact with leucyl-adenylate (present as a nonhydrolyzable analog), the enzyme-bound, activated intermediate of the two-step aminoacylation reaction (**Fig. 1b**). In addition,

[1]European Molecular Biology Laboratory (EMBL), Grenoble Outstation and Unit of Virus Host-Cell Interactions, University of Grenoble–EMBL–Centre National de la Recherche Scientifique, Grenoble, France. [2]Department of Biochemistry, University of Illinois, Roger Adams Laboratory, Urbana, Illinois, USA. [3]These authors contributed equally to this work. Correspondence should be addressed to S.C. (cusack@embl.fr).

Received 18 January; accepted 2 May; published online 10 June 2012; doi:10.1038/nsmb.2317

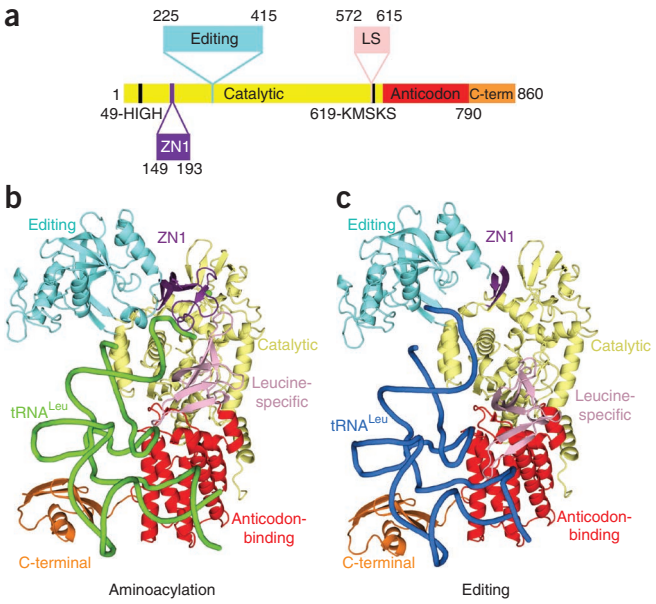


Figure 1 Structures of the *E. coli* LeuRS–tRNA^{Leu} complex in the aminoacylation and editing states. (**a**) The domain structure of LeuRSEC. Residue numbers indicate domain boundaries. The color code used throughout this paper for the various domains is as follows: yellow, catalytic domain; purple, ZN1 domain (zinc ion in green); cyan, editing domain; pink, leucine-specific domain; red, anticodon-binding domain; orange, C-terminal domain. (**b**) Aminoacylation conformation with the tRNA in green. (**c**) Editing conformation with the tRNA in blue. In this state, the ZN1 domain is partially disordered.

we present the high-resolution structure of LeuRSEC with the tRNA 3′ end bound in the editing site (**Fig. 1c**) and leucine or leucyl-adenylate analog bound in the synthetic site. We also report high-resolution structures of two different crystal forms of the complex with the tRNA trapped in the editing site by the simplest benzoxaborole compound 1-hydroxy-3*H*-2,1-benzoxaborole. Comparison of these structures enables us to describe the substantial domain and active site rearrangements that accompany tRNA translocation between the two functional states, providing new insight into the mechanism of aminoacylation and proofreading in class 1a synthetases.

RESULTS

Summary of structures determined

We subjected complexes of LeuRSEC with a tRNA$_5^{Leu}$(UAA) transcript (**Supplementary Fig. 1a**,**b**) and various combinations of small-molecule substrates and inhibitors, such as leucine, leucinol, ATP, AMP, LeuAMS (the sulfamoyl analog of leucyl-adenylate, LeuAMP) and 1-hydroxy-3*H*-2,1-benzoxaborole, to extensive crystallization screens with the aim of obtaining crystals of different functional states of the enzyme. We obtained two different crystal forms, both diffracting to up to 2-Å resolution, with the tRNA directed toward or bound in the editing site. These also grew with or without 1-hydroxy-3*H*-2,1-benzoxaborole bound in the editing site (where it makes a covalent adduct with the 3′ end of the tRNA[5]) and with or without leucine or LeuAMS bound

Table 1 Data collection and refinement statistics

E. coli LeuRS + *E. coli* tRNA (UAA)	Editing + leucine	Editing + benzoxaborole	Editing + benzoxaborole	Aminoacylation + LeuAMS
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_1$	$C2$
Cell dimensions				
a, *b*, *c* (Å)	77.08, 119.37, 141.10	76.18, 118.94, 141.03	89.68, 77.11, 91.14	158.66, 69.20, 228.84
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 102.24, 90.00	90.00, 104.35, 90.00
Resolution (Å)[a]	50–2.00 (2.00–2.10)	50–2.08 (2.08–2.15)	50–2.02 (2.02–2.09)	50–2.5 (2.50–2.60)
R_{sym}	7.1 (71.6)	5.6 (36.7)	6.6 (68.8)	8.4 (51.5)
$I / \sigma I$	10.5 (2.0)	20.1 (3.1)	14.2 (1.6)	11.1 (2.1)
Completeness (%)	99.3 (99.1)	96.7 (74.2)	98.2 (84.5)	90.3 (76.6)
Redundancy	3.90 (3.76)	3.91 (2.80)	3.65 (2.81)	2.89 (2.16)
Refinement				
Resolution (Å)	2.0	2.08	2.02	2.5
No. reflections (work/free)	83,506 / 4,411	71,134 / 3,759	74,330 / 3,977	71,885 / 3,826
R_{work} / R_{free}	0.210 / 0.257	0.199 / 0.244	0.208 / 0.249	0.188 / 0.250
No. atoms				
Protein	6,465	6,517	6,476	2 × 6,836
tRNA	1516	1711	1777	2 × 1,653
Ligand	9 (Leucine)	9 (AN2679)	9 (AN2679)	64 (2 × LeuAMS)
Mg^{2+}	1	2	1	2
Water/other	415	325/4 glycerol	402	169
B-factors				
Protein	42.1	30.4	42.3	46.2[A][b]/62.5[D]
tRNA	68.6	55.6	80.9	51.3[B]/74.1[E]
Ligand	35.5	29.9	31.9	23.9[F]/36.4[G]
Mg^{2+}	62.1	47.9	50.1	61.7[M]
Water	45.1	31.4	39.4	41.1[Z]
R.m.s. deviations				
Bond lengths (Å)	0.014	0.011	0.013	0.014
Bond angles (°)	1.68	1.50	1.64	1.77

[a]Values in parentheses are for highest-resolution shell. [b]Values are for each molecule in the asymmetric unit, with chain indicator given in square brackets.

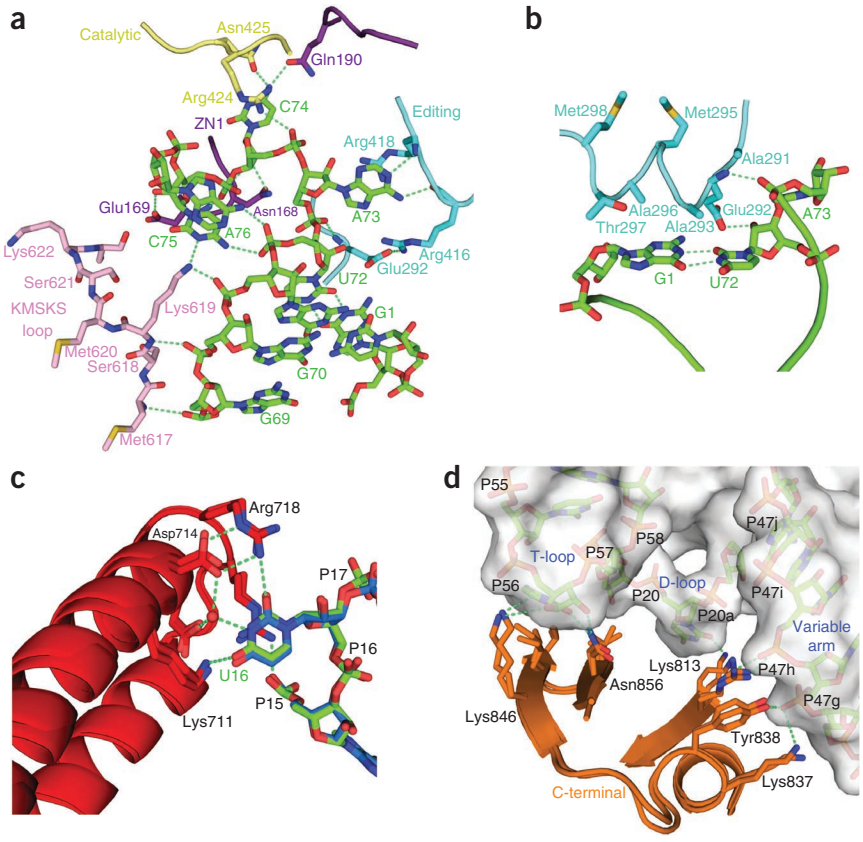


Figure 2 LeuRS–tRNA^{Leu} interactions in the aminoacylation complex. (**a**) Several domains (color coded as in **Fig. 1**) of LeuRS are involved in binding and stabilizing the conformation of nucleotides 69–76 of the 3′ end of tRNA^{Leu} (green). The base of G71 is omitted for clarity. (**b**) The α-helix at residues 291–298 of the editing domain stacks on the G1-U72 base pair and contacts the backbone of A73. (**c**) A network of interactions from the anticodon-binding domain, conserved between the two states (overlaid), specifically recognizes the base of U16 (see also **Supplementary Fig. 4a**,**b**). (**d**) Interactions of the C-terminal domain with the T-loop, D-loop and long variable arm of the tRNA (surface representation) are conserved in the editing and aminoacylation states (overlaid).

in the synthetic site. A different condition gave crystals diffracting to 2.5-Å resolution of the LeuRS–tRNA^{Leu}(UAA)–LeuAMS ternary complex with the tRNA in the aminoacylation conformation (**Supplementary Fig. 1c**). Details of the structure determination are given in the Online Methods and **Table 1**.

The LeuRS–tRNA^{Leu}–LeuAMS aminoacylation complex structure

The *E. coli* LeuRS–tRNA^{Leu}–LeuAMS ternary complex (**Fig. 1b**) shows the 3′ end of the tRNA bending back into the synthetic active site, bringing the 2′-OH of the A76 ribose into the required position to attack the nonreactive leucyl-adenylate analog; thus, the structure closely mimics the functional aminoacylation state. Notably, in only this state are all four flexibly linked domains of the enzyme fully ordered because of a complex network of mutually stabilizing interdomain and protein-tRNA interactions. Owing to extra contacts between bases 69–76 of the tRNA 3′ strand to multiple domains of LeuRSEC, the tRNA makes considerably more contacts with the synthetase in the aminoacylation state than in the editing state (**Fig. 2a**, **Supplementary Fig. 1a**,**b** and **Supplementary Table 1**). These extra contacts not only ensure correct selection of cognate tRNAs but also control the domain rearrangements that create the catalytically active aminoacylation state. In the editing state, discrimination of cognate tRNA is no longer essential, and the key A73 identity element[26,27] is not recognized by the protein.

The hairpin structure of the tRNA 3′ end is stabilized by both protein-tRNA and intra-tRNA interactions, the latter including base-base stacking and phosphate-base hydrogen bonds. A rotamer flip of the class 1a conserved tyrosine[28], Tyr43, into an open position avoids a steric clash with the bound ribose of A76. This tyrosine acts as a lid that can either pack down on the hydrophobic amino acid substrate, preventing premature hydrolysis of the leucyl-adenylate, or be open, to allow A76 binding and the aminoacylation reaction to proceed.

The discriminator base A73 is stacked between Trp223 and Arg416, making base-specific hydrogen bonds to the main chain of Arg416 and Arg418, which are part of the highly conserved (discussed later) 416-RLRDWGVSRQRYWG-429 motif that links the editing domain to the catalytic domain[3,11]. Other residues from this motif also interact with the tRNA, whereas C74 is flipped out into a pocket formed between the ZN1 domain and the catalytic domain (**Fig. 2a**). The ZN1 domain, ordered in the aminoacylation but not the editing state (see below), has a mutually stabilizing interface with the first four residues of the editing domain loop 286–298, also disordered in the editing state. Unexpectedly, a short helix in this loop stacks upon the G1-U72 base pair of the tRNA, which remains unbroken (**Fig. 2b**). Residue Glu292 in this helix forms a hydrogen bond with the phosphate of A73 and the 2′-OH of U72, and also forms a salt bridge with Arg416, which is part of the motif mentioned earlier. These interactions, respectively, block the acceptor stem in place and form a physical barrier separating the single-stranded region 73–76 from the acceptor-stem helix (**Fig. 2a**).

Indeed, for the 3′ end of the tRNA to translocate from the synthetic site into the editing site, the Glu292-Arg416 salt bridge would need to be broken. The unexpected observation that the CP1 editing domain loop 286–298 is directly involved in positioning the tRNA for aminoacylation provides an explanation for the required integrity of this loop for efficient aminoacylation by LeuRSEC[29,30]. This may also explain the preserved editing domain in metazoan mitochondrial LeuRS, even though the editing active site is defunct[31].

Comparison of the LeuRSEC aminoacylation complex with that of archaeal LeuRSPH[11] shows that the mode of binding of the tRNA 3′ end is similar, although there are marked differences in the orientations of the 74-C-C-A bases (**Supplementary Fig. 1d**). The ZN1 domain is in the identical position in the two complexes, but there is no equivalent to LeuRSEC's leucine-specific domain in the archaeal enzyme. Additionally, there is no leucyl-adenylate in the archaeal complex, and the KMSKS loop, discussed in more detail later, is in an open, inactive conformation.

Phylogenetic analysis of 27 LeuRS sequences from representative species covering the whole breadth of bacterial clades supports the importance of key residues and interactions cited here and elsewhere in the text (**Supplementary Fig. 2a**–**e**). For instance,

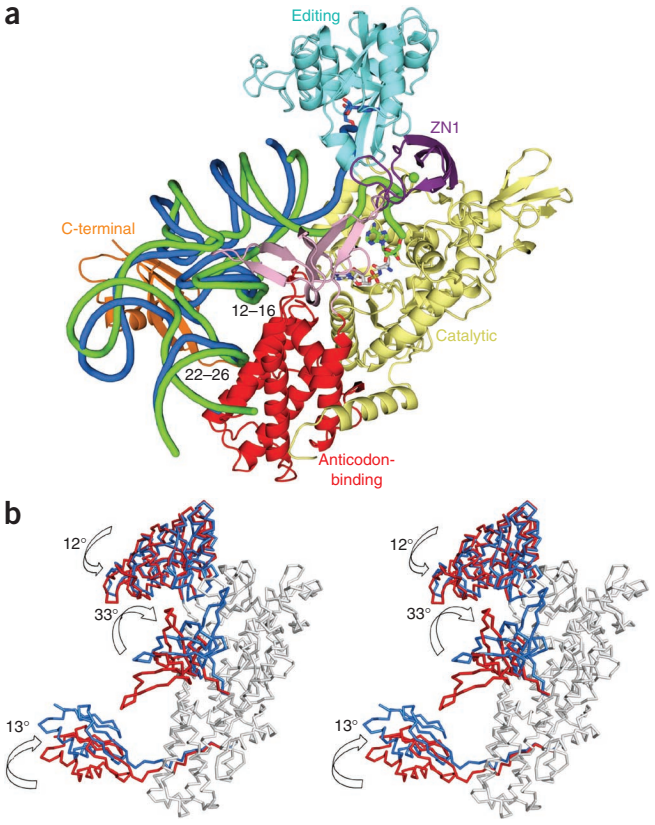


Figure 3 Comparison between the aminoacylation and the editing configurations. (**a**) Global comparison of the two states. The LeuRSEC–tRNA^{Leu} aminoacylation complex (tRNA green tube) is shown together with the tRNA in the editing conformation (tRNA blue tube) after superimposition of the catalytic and the anticodon-binding domains of the two complexes. (**b**) Stereo diagram showing rotations of the flexibly linked domains between the editing (red) and aminoacylation states (blue), after superposition of the body of the enzyme (gray).

the 416-R/KLRDWGVSRQRYWG-429 motif is highly conserved (**Supplementary Fig. 2c**), as are ZN1 domain residues Asn168, Glu169 and Gln190, which interact with the tRNA (not shown). The alignment also suggests that the editing-domain loop containing the helix that stacks on the first base pair of the tRNA is structurally, and probably functionally, conserved, although the sequence can diverge (**Supplementary Fig. 2b**). A glutamate or aspartate residue is found at position 292 in about half the sequences, suggesting that the 292–416 salt bridge is not essential.

Structure of the *E. coli* LeuRS–tRNA^{Leu} editing complex

The overall structure of the *E. coli* LeuRS–tRNA^{Leu} editing complex, determined at nearly 2-Å resolution in the orthorhombic form, is shown in **Figure 1c**. The orthorhombic and monoclinic forms of the complex have only minor differences in domain orientations, and the occurrence of similar structures in two different crystal forms suggests that they faithfully represent the post-transfer editing state of the enzyme. This is reinforced by the global similarity (apart from slight changes in the orientation of the flexibly linked editing and C-terminal domains) of the LeuRSEC editing complex to the previously published structure of the LeuRSTT editing complex (PDB 2BYT)[3] (**Supplementary Fig. 3**). However, there are some notable differences between the two editing complexes. First, the ZN1 domain is well ordered and closes over the leucine binding site in the LeuRSTT complex, whereas there is only weak, uninterpretable electron density for the corresponding region in the LeuRSEC editing complex. This flexibly linked domain has a more active role in the LeuRSEC aminoacylation complex, as described earlier. Second, the LeuRSEC leucine-specific domain is larger and has a different topology compared to that of LeuRSTT. Only in the LeuRSEC editing complex does the leucine-specific domain directly interact with the tRNA. This occurs via an extended β-hairpin, absent from LeuRSTT, that contacts bases 10 and 27 via Arg595 and Arg600 (**Supplementary Fig. 3**). Owing to the high variability in the sequence of the leucine-specific domain (and its absence in some cases), these contacts are likely to be idiosyncratic for *E. coli* and other closely related bacterial LeuRS enzymes.

A second notable difference in protein-tRNA interactions concerns tRNA base 16, which is not contacted in the *T. thermophilus* system. However, in both the aminoacylation and editing complexes of LeuRSEC, U16 is involved in a network of direct and water-mediated interactions with Lys711 (K), Asp714-Asp715 (DD) and Arg718-Arg719 (RR), referred to as the K/DD/RR motif (**Fig. 2c** and **Supplementary Fig. 4a**). Mutations of these residues reduce tRNA binding and catalytic efficiency for aminoacylation, and substitution of U16 by guanine or cytosine eliminates aminoacylation (**Supplementary Fig. 4b,c**). This suggests that U16 is a previously overlooked identity element, at least for *E. coli* LeuRS and related bacteria in which the K/DD/RR motif is conserved (**Supplementary Fig. 2d**). Phylogenetic analysis shows that U16 is conserved not only in all *E. coli* tRNA^{Leu} isoacceptors but also in those of most bacteria. Uridine is more favored than cytosine, which occurs occasionally, by the interaction of its O4 with well-conserved Lys711 (**Supplementary Fig. 4a**).

In the LeuRSTT-tRNA editing complex, we used a tRNA^{Leu} with a truncated long variable arm comprising two base pairs and a tetraloop[3]. In both LeuRSEC conformations, the long variable arm is as in the wild type (four base pairs and a tetraloop) and contacts the C-terminal domain via the variable-stem bases P47f-P47i (**Fig. 2d** and **Supplementary Fig. 1**). In all LeuRSEC complex structures, the extremity of the anticodon stem-loop does not contact the synthetase and is poorly disordered, as previously observed for the *T. thermophilus* system[3].

We obtained structures of *E. coli* LeuRS with the tRNA in the editing conformation either in the absence of other substrates, or with leucine or LeuAMS in the synthetic site, or with the benzoxaborole covalent adduct in the editing site. In the absence of benzoxaborole, the 3′ CCA end of the tRNA is less tightly bound in the editing active site, and the last four base pairs of the acceptor stem and discriminator base, which are not in contact with protein, are poorly ordered owing to flexibility. Formation of the benzoxaborole-tRNA adduct strongly stabilizes tRNA binding in the editing site and improves the ordering of the tRNA acceptor stem, although the mobility (as judged by the crystallographic *B*-factors) of the last two base pairs and discriminator base is still high. The interactions of the 1-hydroxy-3*H*-2,1-benzoxaborole in the editing active site are identical to those previously described for similar compounds[5,6]. Binding of leucine or LeuAMS results in little change of the structure apart from the closure of Tyr43, which is otherwise in an open configuration, over the substrate leucine (see earlier).

Global comparison of the editing and aminoacylation states

Both the tRNA and the flexibly linked domains markedly change their orientation during the transition between the aminoacylation state to the editing state (**Fig. 3a,b** and **Supplementary Video 1**). A major feature of the aminoacylation complex compared to the editing

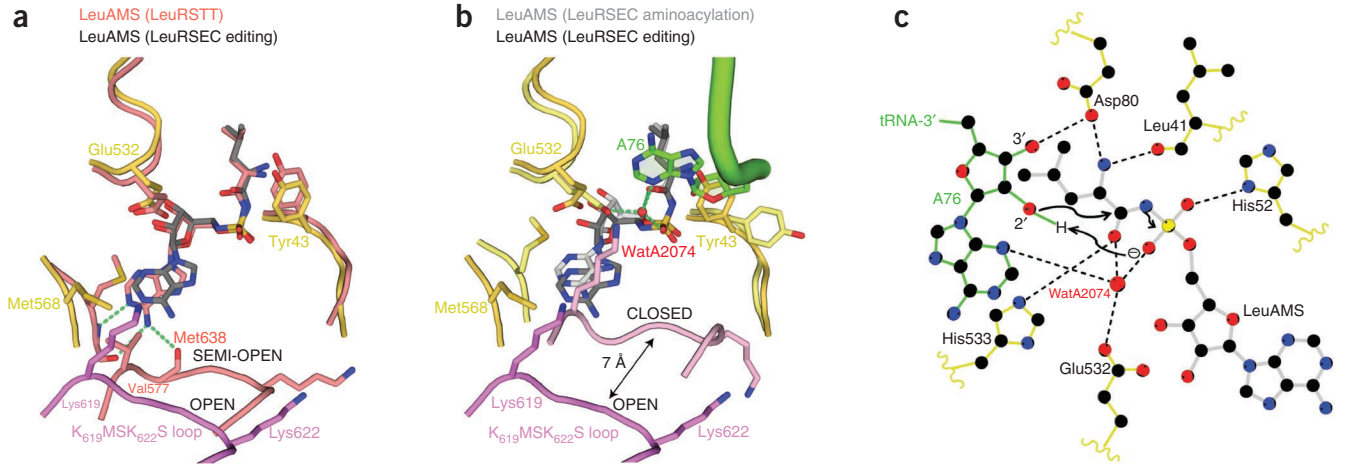


Figure 4 Binding of the tRNA 3′ end induces full closure of the KMSKS loop. (**a**) Comparison between the synthetic sites of the LeuRSEC–LeuAMS–$tRNA^{Leu}$ complex with the tRNA in the editing configuration (yellow, catalytic domain; purple, fully open KMSKS loop; dark gray, LeuAMS) and the LeuRSTT–LeuAMS complex (PDB 1H3N) with the KMSKS loop in the semi-open conformation (all protein chains and LeuAMS shown in salmon). Hydrogen bonds between the semi-open KMSKS loop conformation and the adenine base of LeuAMS are indicated by green dashed lines; these are absent in the open conformation. (**b**) Comparison between the synthetic sites of the LeuRS–LeuAMS–$tRNA^{Leu}$ complex in the aminoacylation configuration (pale yellow, catalytic domain; light purple, closed KMSKS loop; light gray, LeuAMS; green, tRNA) and editing configuration (colored as in **a**). In the aminoacylation state, the presence of A76 necessitates a side chain flip of Tyr43, and a water molecule (WatA2074, red sphere) is coordinated by A76 N3, the reoriented Glu532 carboxylate, a LeuAMS sulfate oxygen and the carbonyl-oxygen of the substrate leucine, as indicated by green dashed lines (see main text and **c**). (**c**) Schematic diagram of the LeuRSEC active site indicating the presumed substrate-assisted mechanism of aminoacylation (see main text).

conformation is that the entire leucine-specific domain rotates 33° toward the synthetic active site. This brings the β-hairpin 577–583 into contact with the ZN1 domain, at the same time breaking the contact that occurs in the editing state between the 595–601 hairpin and the tRNA. The entire tRNA pivots by 15° around conserved contacts with the anticodon-binding domain (**Fig. 3a**), tipping it toward the synthetic active site. The pivoting tRNA is accompanied by the C-terminal domain, maintaining its conserved interactions with the T-loop and long variable arm (**Fig. 2d** and **Supplementary Fig. 2e**). The editing domain rotates by 12° to open up a passage for translocation of the 3′ end of the tRNA from the otherwise closed aminoacylation state. Although in the *E. coli* system the ZN1 domain is ordered in only the aminoacylation state, in structures of LeuRSTT with bound adenylate (PDB 1H3N)[12] or with tRNA in the editing state (PDB 2BYT), the ZN1 domain is packed over the leucyl-adenylate in a position completely incompatible with tRNA 3′-end binding[3] (**Supplementary Fig. 3c**). This position may help to prevent premature hydrolysis of the adenylate, but the domain then needs to rotate by about 44° into a position that is compatible with 3′-end binding (**Supplementary Fig. 3d**,**e**). More detailed implications of these domain motions are discussed later.

Active site interactions of the KMSKS loop

Comparison of the *E. coli* LeuRS aminoacylation and editing complexes shows that the class 1 conserved and catalytically important 619-KMSKS loop moves as an integral part of the leucine-specific domain. The movement brings it some 6–7 Å closer into the active site in the aminoacylation state, allowing it to make crucial interactions with the 3′ strand that correctly position the tRNA extremity for the transfer reaction (**Figs. 2a** and **4a**,**b**). In contrast, in the editing conformation, there are no contacts to the 3′ strand of the acceptor stem (**Supplementary Fig. 1b**). In the closed aminoacylation state, Met568 flips to stack against the center of the adenine base of the LeuAMS (**Fig. 4a**). This correlates with a rotamer switch of Glu532, allowing it, the N3 of A76 and the LeuAMS sulfate to coordinate a water molecule, or possibly a magnesium ion, that could be important in the catalytic mechanism (**Fig. 4b**,**c**, discussed later). One result of this tightening of the active site is that the adenylate itself is slightly compressed into a strained conformation in comparison to the more extended conformation observed in structures without the tRNA in the aminoacylation state (**Fig. 4a**,**b**). Whereas the leucyl moiety is unchanged, the adenylate ribose is closer to the carbonyl group, partly owing to the maintenance of the interaction of its 2′-OH with Glu532, which, as mentioned, has changed side chain conformation (**Fig. 4b**). In contrast, in the editing conformation with leucine or LeuAMS in the synthetic site, the KMSKS loop is in an open, relaxed conformation (**Fig. 4a**). In this case, the adenine moiety of the LeuAMS is much less tightly bound, with no base-specific contacts to the backbone of Val569 and Met620 and no stacking with Met568 (**Fig. 4a**,**b**).

These observations are consistent with a previous suggestion based on TyrRS structures that the KMSKS loop could be in three states: open, semi-open and closed[32]. With the tRNA in the editing conformation, and either without synthetic site substrates or with bound leucine or LeuAMS, the KMSKS loop is in the fully open conformation and unable to form base-specific interactions to the adenine (**Fig. 4a**). With adenylate bound in the absence of tRNA (LeuRSTT–LeuAMS complex[12]; PDB 1H3N), the KMSKS loop is in a semi-open conformation with base-specific interactions to the adenine base (**Fig. 4a**). The fully closed conformation of the KMSKS loop has been observed only with the leucyl-adenylate bound together with the 3′ of the tRNA (**Fig. 4b**). However, a similar state might occur for the pre-activation state when ATP and leucine are both bound, as observed, for instance, for TyrRS[19] (PDB 1H3E) or GlnRS[14]. Nonetheless, the only known structure of a class 1a enzyme with an ATP analog bound, but no amino acid substrate—that of *P. horikoshii* ArgRS[24] (PDB 2ZUE)—has the KMSKS loop in the semi-open state. As suggested for the tyrosyl system, the semi-open post-activation state with bound LeuAMP (as observed with LeuRSTT; **Fig. 4a**) may be required to permit initial binding of the tRNA 3′ end before reclosure, induced by 3′-end interactions, to allow the aminoacylation reaction.

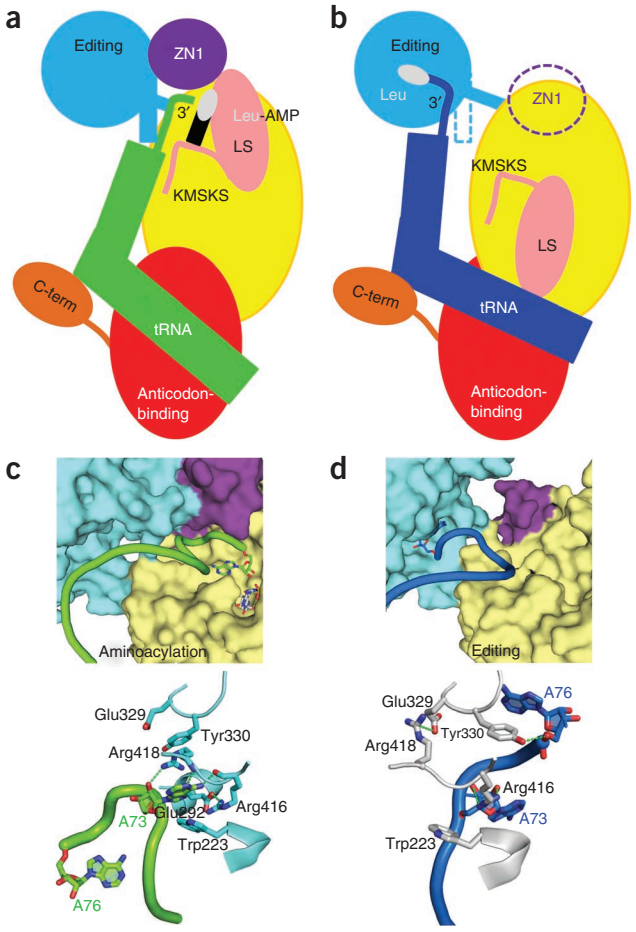


Figure 5 Dynamics of the functional cycle of bacterial LeuRS. (**a**,**b**) Schematic diagram of the structural changes between the aminoacylation and proofreading conformations of the LeuRSEC–tRNA^{Leu} complex (see main text). Color coding for each domain is as in **Figure 1**, with leucine represented as a white oval and AMP as a black rectangle. (**c**,**d**). Dynamic rearrangements in the interface between the editing, ZN1 and catalytic domains that allow translocation of the tRNA 3′ end from the synthetic site (**c**) to the editing site (**d**) for proofreading. Some elements of the catalytic domain have been removed for clarity. Upper panels show the cavities of the catalytic and editing sites in surface representation; lower panels show the residues important for the translocation of the tRNA (rotated 180° with respect to the upper panels, for clarity).

The crucial role of the KMSKS loop in activation and aminoacylation, and its coupling with the leucine-specific domain explain why a gross deletion of this domain (569–618 replaced by three or five alanines) abolishes aminoacylation activity of *E. coli* LeuRS[33]. However, the leucine-specific domain is highly variable in size and sequence, and some bacterial LeuRS lack it completely: for example, those of *Campylobacter jejuni* and *Helicobacter pylori*, and the second LeuRS (denoted B2) from *Agrobacterium radiobacter*, which is resistant to the natural antibiotic TM84 (an analog of LeuAMP)[34]. Notably, *Mycoplasma mobile* has evolved a minimal LeuRS that lacks both the leucine-specific domain and also, uniquely, the entire editing domain[35]. It remains to be seen how the absence of the leucine-specific domain affects the function of these bacterial enzymes, but one might hypothesize that the KMSKS loop is more dynamic, as it does not have the inertial mass of the rest of the leucine-specific domain to carry with it.

Aminoacylation catalytic mechanism

For class I synthetases, the aminoacylation reaction requires the nucleophilic attack of A76 ribose O2′ on the carbonyl carbon of the aminoacyl-adenylate. The reaction is most likely substrate assisted, with the phosphate of the adenylate acting as the general base to abstract the proton from the 2′-OH in a concerted mechanism[14,36] (**Fig. 4c**). This mechanism is potentially universal, as it does not depend on nearby, nonconserved residues, as proposed by some authors[15]. The LeuRSEC aminoacylation complex structure is generally consistent with this mechanism as the O2′ is 3.6 Å from the sulfate oxygen of the leucyl-adenylate analog LeuAMS, which is comparable to 3.8 Å for the equivalent distance in the GlnRS ternary complex[15] (PDB 1QTQ). However, these distances are substantially longer than a hydrogen bond, suggesting that the energy barrier for deprotonation might be too high for the direct reaction. This could reflect the fact that the structures do not correspond precisely to the real situation (for example, the sulfamoyl analog lacks the negative charge of leucyl-adenylate) or that additional elements might be required to explain the observed low energy barrier for the aminoacylation reaction.

One fully conserved LeuRS residue that could have a role in the aminoacylation reaction is Glu532, which is strategically placed in the aminoacylation complex (**Fig. 4b**,**c**). To test the role of this residue, we expressed an E532Q mutant of LeuRSEC and measured the temperature dependence of the pyrophosphate (PPi) exchange and overall leucylation reactions (**Supplementary Fig. 4c**). The E532Q mutant had similar activation energies for the first step and the overall reaction (93 kJ mol^{-1} and 94.9 kJ mol^{-1}, respectively) in contrast to the wild type, which had corresponding values of 25.2 kJ mol^{-1} and 55.0 kJ mol^{-1}. This suggests that the rate-limiting step for the mutant E532Q has shifted to the ATP-dependent activation of leucine (**Supplementary Fig. 4d**), consistent with an important role for Glu532 in activation, perhaps by coordinating a magnesium ion. Indeed a water molecule (Wat A2074) that is coordinated by Glu532 could occupy the magnesium position (**Fig. 4b**,**c**). In the presence of the authentic, negatively charged LeuAMP, it is plausible that this site could still bind a magnesium ion, and this might facilitate the aminoacylation reaction, even though the E532Q mutant does not show this. Combined quantum-molecular dynamics simulations are likely to be required to determine the most plausible mechanism for the aminoacylation reaction in LeuRS.

DISCUSSION

In the aminoacylation state, the tRNA 3′ end is deeply buried in the catalytic domain. How does aminoacyl transfer with concomitant formation of AMP lead to product release and translocation? With reference to the schematic diagram of the two functional states (**Fig. 5a**,**b**), we propose the following mechanism for the functional cycle of *E. coli* LeuRS. We hypothesize that the aminoacylation state is under strain with regard to the unusual conformation of the 3′ end of the tRNA, the distorted state of the adenylate and the fully closed conformation of the KMSKS loop. Formation of this state is possible because of the presumably high binding energy of the tRNA 3′ end. Once the covalent integrity of the adenylate is broken by completion of the aminoacylation reaction, the strain is released by relaxation of the KMSKS loop into its open conformation. This is accompanied by the dissociation of AMP and the flipping of the entire leucine-specific domain, whose inertia will prevent easy reversion to the closed state. Concomitant with this, the KMSKS loop interactions with the tRNA are broken, destabilizing the binding of the entire acceptor end of

the now charged tRNA. This leads to broader cooperative destabilization of the interactions holding the ZN1 domain and the editing domain in contact with each other and with the tRNA. The release of constraints allows the tRNA 3′ end to relax to its preferred conformation, with the discriminator base stacked on the 1:72 base pair and the CCA end now able to enter the editing active site. Locally, this is accompanied by a rearrangement of the interactions of Glu292, Arg416, Arg418 and Tyr330 with the tRNA, each of which has distinct roles in the two states (**Fig. 5c**,**d**). More globally, the editing domain rotates slightly as the tRNA pivots around its fixed contact points on the anticodon-binding domain, and the C-terminal domain moves synchronously, maintaining its contacts to the tRNA. In this conformation, the tRNA is already partially disassociated from the catalytic domain (**Supplementary Fig. 1a**,**b** and **Supplementary Table 1**), favoring total tRNA release after proofreading in the case of correctly charged Leu-tRNA^{Leu}, or possibly direct re-aminoacylation, without full release, of hydrolyzed mischarged tRNA.

Our structures provide detailed information on the end points of translocation; it might be difficult to obtain crystal structures of eventual intermediates, although an indication of such states was observed for the archaeal LeuRSPH aminoacylation complex[11]. Molecular dynamics simulations can potentially fill in the gap, as used in the analysis of the mechanism for Glu-tRNA^{Glu} release from the synthetic site of class 1b GluRS[18]. In another study, normal-mode analysis of LeuRSTT revealed both correlated and anticorrelated motions between various structural elements[37]. Correlated motions included the coupling of the KMSKS loop with the leucine-specific domain, and of the editing domain loop 286–298 with the ZN1 domain; anticorrelated motions were detected between the leucine-specific and editing domains, consistent with the structure-based results described here. In the future, a completed ensemble of structures (including the currently lacking preactivation complex with bound ATP) together with molecular dynamics simulations and fast kinetic studies coupled with structure-based mutagenesis will hopefully provide further insight into the dynamics of the functional cycle of bacterial leucyl-tRNA synthetase. Notably, recent results have revealed that eukaryotic, cytoplasmic LeuRS acts as an intracellular leucine sensor in the amino acid–induced mTORC1 signaling pathway, which regulates cell growth[38,39]. It is proposed that Rag GTPases, which activate mTORC1, bind to the LeuRS editing domain and sense the presence of bound leucine via the conformational state of the enzyme. This particularly exemplifies the need for further analysis and understanding of the conformational dynamics of LeuRS.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Protein Data Bank: Atomic coordinates and structure factors for the editing complex with L-leucine have been deposited with accession code 4ARC; for the orthorhombic and monoclinic forms of the complexes with benzoxaborole under accession codes 4ARI and 4AS1, respectively; and for the aminoacylation complex with accession code 4AQ7.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank the ESRF-EMBL Joint Structural Biology Group for access to ESRF beamlines and the EMBL-ESRF-ILL-IBS Partnership for Structural Biology for access to structural biology instrumentation, notably the high-throughput crystallization platform.

AUTHOR CONTRIBUTIONS

T.C. crystallized the editing-state complexes and A.P. crystallized the aminoacylation complex. T.C., A.P. and S.C. collected X-ray data and performed the structural analysis. M.T.V. and T.L.L. performed the mutagenesis and associated biochemical studies under the supervision of S.A.M. S.C. wrote the manuscript with input from all other authors. S.A.M. acknowledges funding from US National Institutes of Health grant GM63789.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nsmb.2317.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Karkhanis, V.A., Mascarenhas, A.P. & Martinis, S.A. Amino acid toxicities of *Escherichia coli* that are prevented by leucyl-tRNA synthetase amino acid editing. *J. Bacteriol.* **189**, 8765–8768 (2007).
2. Lincecum, T.L. Jr. *et al.* Structural and mechanistic basis of pre- and posttransfer editing by leucyl-tRNA synthetase. *Mol. Cell* **11**, 951–963 (2003).
3. Tukalo, M., Yaremchuk, A., Fukunaga, R., Yokoyama, S. & Cusack, S. The crystal structure of leucyl-tRNA synthetase complexed with tRNALeu in the post-transfer-editing conformation. *Nat. Struct. Mol. Biol.* **12**, 923–930 (2005).
4. Hagiwara, Y., Field, M.J., Nureki, O. & Tateno, M. Editing mechanism of aminoacyl-tRNA synthetases operates by a hybrid ribozyme/protein catalyst. *J. Am. Chem. Soc.* **132**, 2751–2758 (2010).
5. Rock, F.L. *et al.* An antifungal agent inhibits an aminoacyl-tRNA synthetase by trapping tRNA in the editing site. *Science* **316**, 1759–1761 (2007).
6. Seiradake, E. *et al.* Crystal structures of the human and fungal cytosolic leucyl-tRNA synthetase editing domains: A structural basis for the rational design of antifungal benzoxaboroles. *J. Mol. Biol.* **390**, 196–207 (2009).
7. Zhang, Y.K. *et al.* Synthesis and structure-activity relationships of novel benzoxaboroles as a new class of antimalarial agents. *Bioorg. Med. Chem. Lett.* **21**, 644–651 (2011).
8. Ding, D. *et al.* Design, synthesis, and structure-activity relationship of *Trypanosoma brucei* leucyl-tRNA synthetase inhibitors as antitrypanosomal agents. *J. Med. Chem.* **54**, 1276–1287 (2011).
9. Fukai, S. *et al.* Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNAVal and valyl-tRNA synthetase. *Cell* **103**, 793–803 (2000).
10. Silvian, L.F., Wang, J. & Steitz, T.A. Insights into editing from an ile-tRNA synthetase structure with tRNAIle and mupirocin. *Science* **285**, 1074–1077 (1999).
11. Fukunaga, R. & Yokoyama, S. Aminoacylation complex structures of leucyl-tRNA synthetase and tRNALeu reveal two modes of discriminator-base recognition. *Nat. Struct. Mol. Biol.* **12**, 915–922 (2005).
12. Cusack, S., Yaremchuk, A. & Tukalo, M. The 2 crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. *EMBO J.* **19**, 2351–2361 (2000).
13. Fukunaga, R. & Yokoyama, S. Crystal structure of leucyl-tRNA synthetase from the archaeon *Pyrococcus horikoshii* reveals a novel editing domain orientation. *J. Mol. Biol.* **346**, 57–71 (2005).
14. Perona, J.J., Rould, M.A. & Steitz, T.A. Structural basis for transfer RNA aminoacylation by *Escherichia coli* glutaminyl-tRNA synthetase. *Biochemistry* **32**, 8758–8771 (1993).
15. Rath, V.L., Silvian, L.F., Beijer, B., Sproat, B.S. & Steitz, T.A. How glutaminyl-tRNA synthetase selects glutamine. *Structure* **6**, 439–449 (1998).
16. Bullock, T.L., Uter, N., Nissan, T.A. & Perona, J.J. Amino acid discrimination by a class I aminoacyl-tRNA synthetase specified by negative determinants. *J. Mol. Biol.* **328**, 395–408 (2003).
17. Sekine, S. *et al.* ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding. *EMBO J.* **22**, 676–688 (2003).
18. Sekine, S. *et al.* Structural bases of transfer RNA-dependent amino acid recognition and activation by glutamyl-tRNA synthetase. *Structure* **14**, 1791–1799 (2006).
19. Yaremchuk, A., Kriklivyi, I., Tukalo, M. & Cusack, S. Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *EMBO J.* **21**, 3829–3840 (2002).
20. Kobayashi, T. *et al.* Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol.* **10**, 425–432 (2003).
21. Yang, X.L. *et al.* Two conformations of a crystalline human tRNA synthetase-tRNA complex: implications for protein synthesis. *EMBO J.* **25**, 2919–2929 (2006).
22. Shen, N. *et al.* Catalytic mechanism of the tryptophan activation reaction revealed by crystal structures of human tryptophanyl-tRNA synthetase in different enzymatic states. *Nucleic Acids Res.* **36**, 1288–1299 (2008).
23. Hauenstein, S., Zhang, C.M., Hou, Y.M. & Perona, J.J. Shape-selective RNA recognition by cysteinyl-tRNA synthetase. *Nat. Struct. Mol. Biol.* **11**, 1134–1141 (2004).
24. Konno, M. *et al.* Modeling of tRNA-assisted mechanism of Arg activation based on a structure of Arg-tRNA synthetase, tRNA, and an ATP analog (ANP). *FEBS J.* **276**, 4763–4779 (2009).

25. Delagoutte, B., Moras, D. & Cavarelli, J. tRNA aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding. *EMBO J.* **19**, 5599–5610 (2000).
26. Tocchini-Valentini, G., Saks, M.E. & Abelson, J. tRNA leucine identity and recognition sets. *J. Mol. Biol.* **298**, 779–793 (2000).
27. Larkin, D.C., Williams, A.M., Martinis, S.A. & Fox, G.E. Identification of essential domains for *Escherichia coli* tRNALeu aminoacylation and amino acid editing using minimalist RNA molecules. *Nucleic Acids Res.* **30**, 2103–2113 (2002).
28. Crepin, T. *et al.* Use of analogues of methionine and methionyl adenylate to sample conformational changes during catalysis in *Escherichia coli* methionyl-tRNA synthetase. *J. Mol. Biol.* **332**, 59–72 (2003).
29. Li, T., Guo, N., Xia, X., Wang, E.D. & Wang, Y.L. The peptide bond between E292–A293 of *Escherichia coli* leucyl-tRNA synthetase is essential for its activity. *Biochemistry* **38**, 13063–13069 (1999).
30. Du, X. & Wang, E.D. E292 is important for the aminoacylation activity of *Escherichia coli* leucyl-tRNA synthetase. *J. Protein Chem.* **22**, 71–76 (2003).
31. Lue, S.W. & Kelley, S.O. An aminoacyl-tRNA synthetase with a defunct editing site. *Biochemistry* **44**, 3010–3016 (2005).
32. Kobayashi, T. *et al.* Structural snapshots of the KMSKS loop rearrangement for amino acid activation by bacterial tyrosyl-tRNA synthetase. *J. Mol. Biol.* **346**, 105–117 (2005).
33. Vu, M.T. & Martinis, S.A. A unique insert of leucyl-tRNA synthetase is required for aminoacylation and not amino acid editing. *Biochemistry* **46**, 5170–5176 (2007).
34. Reader, J.S. *et al.* Major biocontrol of plant tumors targets tRNA synthetase. *Science* **309**, 1533 (2005).
35. Li, L. *et al.* Naturally occurring aminoacyl-tRNA synthetases editing-domain mutations that cause mistranslation in Mycoplasma parasites. *Proc. Natl. Acad. Sci. USA* **108**, 9378–9383 (2011).
36. Uter, N.T. & Perona, J.J. Active-site assembly in glutaminyl-tRNA synthetase by tRNA-mediated induced fit. *Biochemistry* **45**, 6858–6865 (2006).
37. Weimer, K.M., Shane, B.L., Brunetto, M., Bhattacharyya, S. & Hati, S. Evolutionary basis for the coupled-domain motions in *Thermus thermophilus* leucyl-tRNA synthetase. *J. Biol. Chem.* **284**, 10088–10099 (2009).
38. Han, J.M. *et al.* Leucyl-tRNA synthetase is an intracellular leucine sensor for the mTORC1-signaling pathway. *Cell* **149**, 410–424 (2012).
39. Bonfils, G. *et al.* Leucyl-tRNA synthetase controls TORC1 via the EGO complex. *Mol. Cell* **46**, 105–110 (2012).

ONLINE METHODS

Crystallization. *E. coli* LeuRS with an N-terminal six-histidine tag and *E. coli* $tRNA^{Leu}_{UAA}$ transcript were prepared and purified as described[40,41]. Protein was stored in buffer comprising 20 mM Tris-HCl (pH 7.4), 100 mM NaCl, 5 mM $MgCl_2$ and 5 mM 2-β-mercaptoethanol. Crystallization was performed at 20 °C by the hanging drop vapor diffusion method. For the editing complex, solutions were prepared with 33 μM LeuRS, 40 μM $tRNA^{Leu}$ and 5 mM L-leucine or 1 mM 1-hydroxy-3*H*-2,1-benzoxaborole (a gift from Anacor Pharmaceuticals). Crystals were obtained by mixing 2 μl of this solution with 1 μl of reservoir solution containing 0.1 M sodium acetate (pH 5.6), 14–18% (w/v) PEG 6000 and 200 mM NaCl. The crystals were frozen in liquid nitrogen after transfer for a few seconds in the mother liquor, which contained 22% (v/v) ethylene glycol as cryoprotectant and the corresponding small substrates. In the case of the aminoacylation complex, solutions were prepared with 33 μM LeuRS, 50 μM $tRNA^{Leu}$ and 1 mM leucyl-adenylate analog (LeuAMS, purchased from RNA-TEC). Crystals were obtained by mixing 2 μl of this solution with 2 μl of reservoir solution containing 0.1 M bis-TRIS (pH 5.5), 23–25% (w/v) PEG 3350 and 200 mM ammonium acetate. The crystals were frozen in liquid nitrogen before X-ray exposure without added cryoprotectant.

Structure determination and refinement. All diffraction data sets were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) using a wavelength of 1.072 Å for the complex with L-leucine, 0.931 Å and 0.934 Å for the orthorhombic- and monoclinic-crystal forms, respectively, of the complex with benzoxaborole and 0.938 Å for the complex with LeuAMS. Data were integrated and scaled with the XDS suite[42]. Further data analysis was performed with the CCP4 suite[43]. The structure of the LeuRS–$tRNA^{Leu}$–L-leucine editing complex was initially solved by molecular replacement with PHASER[44] using as models the *T. thermophilus* LeuRS structure (PDB 1H3N)[45] without its editing domain and the *E. coli* LeuRS editing domain structure (PDB 2AJG)[46]. The model was improved by automatic building using ARP-wARP[47], and manual adjustments were made with COOT[48]. The structure of the aminoacylation complex was solved by molecular replacement with PHASER using the core of the protein from the *E. coli* editing complex (residues 1–223, 416–568 and 795–860). The resulting map was used to manually position the core of the tRNA and the additional flexibly linked domains using COOT. Residues 159–189 of the ZN1 domain, not visible in the LeuRSEC editing complex, were built using the LeuRSTT ZN1 domain as a guide. All models were refined using REFMAC5 with TLS. The orthorhombic ($P2_12_12_1$) and monoclinic ($P2_1$) forms of the LeuRSEC editing complex have only minor differences involving a small coordinated rotation of the C-terminal domain tRNA and editing domain, without changing specific protein-tRNA interactions. In the monoclinic form, the C-terminal domain is partially disordered, and in general the *B*-factors are slightly higher. Interfaces were analyzed with the PISA server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) and domain motions with the program DynDom[49]. Structure quality was analyzed with MolProbity[50] (http://molprobity.biochem.duke.edu/) and showed all residues in allowed regions (with 95.1–98.0% of residues in favored regions) for the different models. Figures were drawn with PyMOL (http://www.pymol.org/), and a video showing the conformational changes between the aminoacylation and editing states was generated by CHIMERA (http://www.cgl.ucsf.edu/chimera/).

Enzyme assays. A reaction mixture consisting of 60 mM Tris (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT, 4 mM ATP, 21 μM [^{3}H]leucine (10 μCi nmol^{-1}) and 2.5 μM $tRNA^{Leu}$ was initiated by the addition of 25 nM enzyme and quenched[51]. Higher concentrations of 4 μM $tRNA^{Leu}$ and 1 μM enzyme concentrations were used for weak mutant activities (**Supplementary Fig. 4b**). Leucine-dependent PPi-ATP exchange reactions contained 50 mM HEPES (pH 8.0), 10 mM $MgCl_2$, 1 mM DTT, 1 mM ATP, 1 mM [^{32}P]PP_i (10 μCi μmol^{-1}) and 1 mM leucine and were initiated by 10 nM enzyme. Each reaction was quenched on Baker-Flex polyethyleneimine cellulose TLC plates and chromatographed in 750 mM KH_2PO_4 (pH 3.5) and 4 M urea at 25 °C.

40. Larkin, D.C., Williams, A.M., Martinis, S.A. & Fox, G.E. Identification of essential domains for *Escherichia coli* $tRNA^{Leu}$ aminoacylation and amino acid editing using minimalist RNA molecules. *Nucleic Acids Res.* **30**, 2103–2113 (2002).
41. Vu, M.T. & Martinis, S.A. A unique insert of leucyl-tRNA synthetase is required for aminoacylation and not amino acid editing. *Biochemistry* **46**, 5170–5176 (2007).
42. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
43. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
44. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. Likelihood-enhanced fast translation functions. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 458–464 (2005).
45. Cusack, S., Yaremchuk, A. & Tukalo, M. The 2 crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. *EMBO J.* **19**, 2351–2361 (2000).
46. Liu, Y., Liao, J., Zhu, B., Wang, E.D. & Ding, J. Crystal structures of the editing domain of *Escherichia coli* leucyl-tRNA synthetase and its complexes with Met and Ile reveal a lock-and-key mechanism for amino acid discrimination. *Biochem. J.* **394**, 399–407 (2006).
47. Perrakis, A., Morris, R. & Lamzin, V.S. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463 (1999).
48. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
49. Hayward, S. & Lee, R.A. Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J. Mol. Graph. Model.* **21**, 181–183 (2002).
50. Chen, V.B.. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
51. Lincecum, T.L. Jr. *et al.* Structural and mechanistic basis of pre- and posttransfer editing by leucyl-tRNA synthetase. *Mol. Cell* **11**, 951–963 (2003).

Vol 458|16 April 2009|doi:10.1038/nature07745

LETTERS

The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit

Alexandre Dias[1]*, Denis Bouvier[1]*, Thibaut Crépin[1]*, Andrew A. McCarthy[1,2], Darren J. Hart[1,2], Florence Baudin[1], Stephen Cusack[1,2] & Rob W. H. Ruigrok[1]

The influenza virus polymerase, a heterotrimer composed of three subunits, PA, PB1 and PB2, is responsible for replication and transcription of the eight separate segments of the viral RNA genome in the nuclei of infected cells. The polymerase synthesizes viral messenger RNAs using short capped primers derived from cellular transcripts by a unique 'cap-snatching' mechanism[1]. The PB2 subunit binds the 5′ cap of host pre-mRNAs[2–4], which are subsequently cleaved after 10–13 nucleotides by the viral endonuclease, hitherto thought to reside in the PB2 (ref. 5) or PB1 (ref. 2) subunits. Here we describe biochemical and structural studies showing that the amino-terminal 209 residues of the PA subunit contain the endonuclease active site. We show that this domain has intrinsic RNA and DNA endonuclease activity that is strongly activated by manganese ions, matching observations reported for the endonuclease activity of the intact trimeric polymerase[6,7]. Furthermore, this activity is inhibited by 2,4-dioxo-4-phenylbutanoic acid, a known inhibitor of the influenza endonuclease[8]. The crystal structure of the domain reveals a structural core closely resembling resolvases and type II restriction endonucleases. The active site comprises a histidine and a cluster of three acidic residues, conserved in all influenza viruses, which bind two manganese ions in a configuration similar to other two-metal-dependent endonucleases. Two active site residues have previously been shown to specifically eliminate the polymerase endonuclease activity when mutated[9]. These results will facilitate the optimisation of endonuclease inhibitors[10–12] as potential new anti-influenza drugs.

The exact role of the influenza polymerase PA subunit remains unclear, although it has been implicated in cap-binding, endonuclease activity, viral RNA binding and replication[9,13–15]. PA is separable by trypsinization[9] into a large carboxy-terminal domain—the crystal structure of which has recently been reported[16,17]—and a small N-terminal domain, which contains residues important for protein stability, promoter binding, cap-binding and endonuclease activity of the polymerase complex[9]. In particular, the importance of Asp 108 and Lys 134 for endonuclease function called into question a previous conclusion that three acidic residues of PB1 were exclusively responsible for the endonuclease activity[2]. The enzymology of the endonuclease within the context of intact viral ribonucleoprotein particles (RNPs) has been extensively studied. The endonuclease is activated by the binding of divalent cations, most effectively manganese followed by cobalt, and less efficiently magnesium, zinc and nickel, but not cadmium, calcium or trivalent ions[6]. RNPs bind four manganese ions in a highly cooperative fashion, presumably two each in the nucleotidyl-transferase and endonuclease sites[6]. It was concluded that the enzyme probably has a two-metal dependent mechanism in common with several other nucleases. Furthermore,

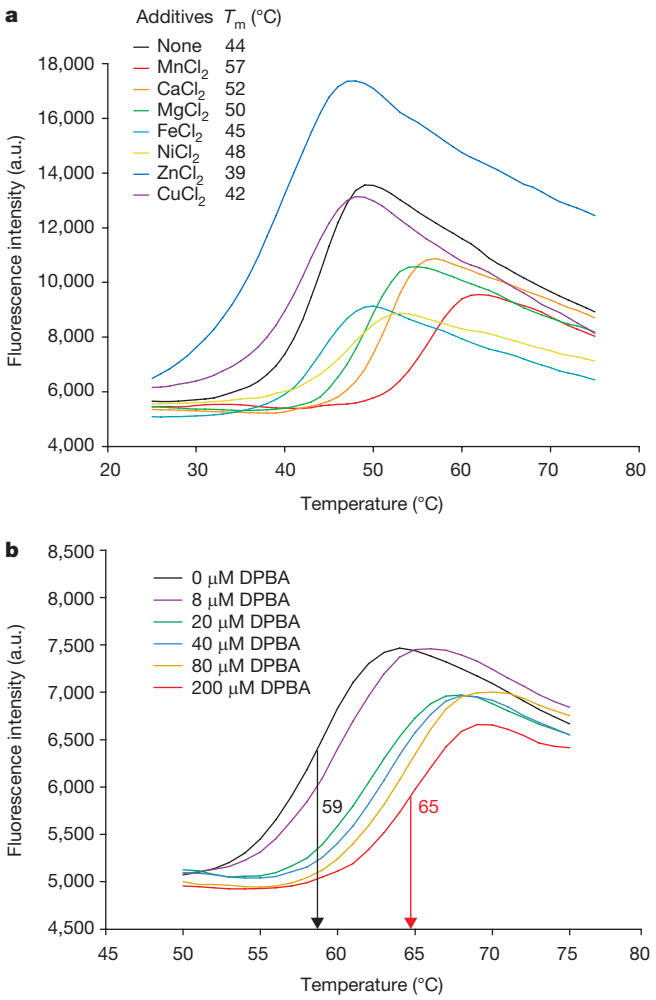


Figure 1 | Manganese ions stabilize PA-Nter. a, Assay of thermal stability using Thermofluor. Thermal shift assays were performed with different metal salts. The results obtained in the absence of metal salts are indicated with a black line and those in the presence of metal salts with coloured lines. The different melting points (T_m) extracted from the curves at pH 8.0 are given in the inset. The effect of $CoCl_2$ on protein stability at pH 7.0 was investigated but was not interpretable owing to quenching by the metal. **b,** Assay of thermal stability with DPBA. Thermofluor assays of PA-Nter were performed in the presence of $MnCl_2$ with increasing concentrations of DPBA. a.u., arbitrary units.

[1]Unit of Virus Host-Cell Interactions, UJF-EMBL-CNRS, UMR 5233, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France. [2]Grenoble Outstation, European Molecular Biology Laboratory, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France.
*These authors contributed equally to this work.

the endonuclease cleaves single-stranded (ss)DNA with only slightly reduced activity compared to RNA; indeed capped ssDNA oligomers are substrates for cap-snatching[7].

To clarify the functional role of the N-terminal domain of PA we studied a proteolytically stable fragment from residues 1 to 209 (denoted PA-Nter). To optimise the stability and mono-dispersity of PA-Nter we used dynamic light scattering and thermal shift assays with a variety of buffers and additives. These experiments showed a marked increase in thermal stability (apparent melting temperature, T_m, shifts from 44 to 57 °C) and improved mono-dispersity after the addition of manganese ions and to a lesser extent calcium and magnesium ions, but not other divalent metal ions (Fig. 1a). The structural effect of manganese binding, investigated by circular dichroism, demonstrated a small, but reproducible increase in helical content after the addition of 1 mM $MnCl_2$ (Supplementary Fig. 1). Because PA-Nter has been implicated in the viral endonuclease activity[9], which is known to depend on divalent cations, notably manganese[6], we assayed the domain for endonuclease activity using both structured and unstructured ssRNA (Fig. 2) and ssDNA substrates (Supplementary Fig. 2). Using a partially structured 81-nucleotide panhandle RNA (ph-RNA) we found that PA-Nter has intrinsic RNase activity that is strictly divalent-cation-dependent (Fig. 2a, b). Consistent with previous results on RNPs[6], strong activity was observed at pH 8 with manganese, weaker activity with magnesium and none with zinc, calcium or cobalt, whereas at pH 7 activity was also observed with cobalt (Fig. 2b). After 40 min incubation highly structured RNAs such as transfer RNA and signal recognition particle (SRP) *Alu* RNA are relatively resistant to degradation, partially structured ph-RNAs and short ph-RNAs are partially degraded and unstructured U-rich RNA is completely degraded, suggesting that the enzyme is single-strand specific (Fig. 2c). The enzyme also completely degrades circular ssDNA showing that it is a nonspecific endonuclease (Supplementary Fig. 2a). The endonuclease activity on both RNA and DNA is inhibited in a dose-dependent manner by the compound 2,4-dioxo-4-phenylbutanoic acid (DPBA), a known inhibitor of influenza endonuclease[8] (Fig. 2d and Supplementary Fig. 2b). The inhibition

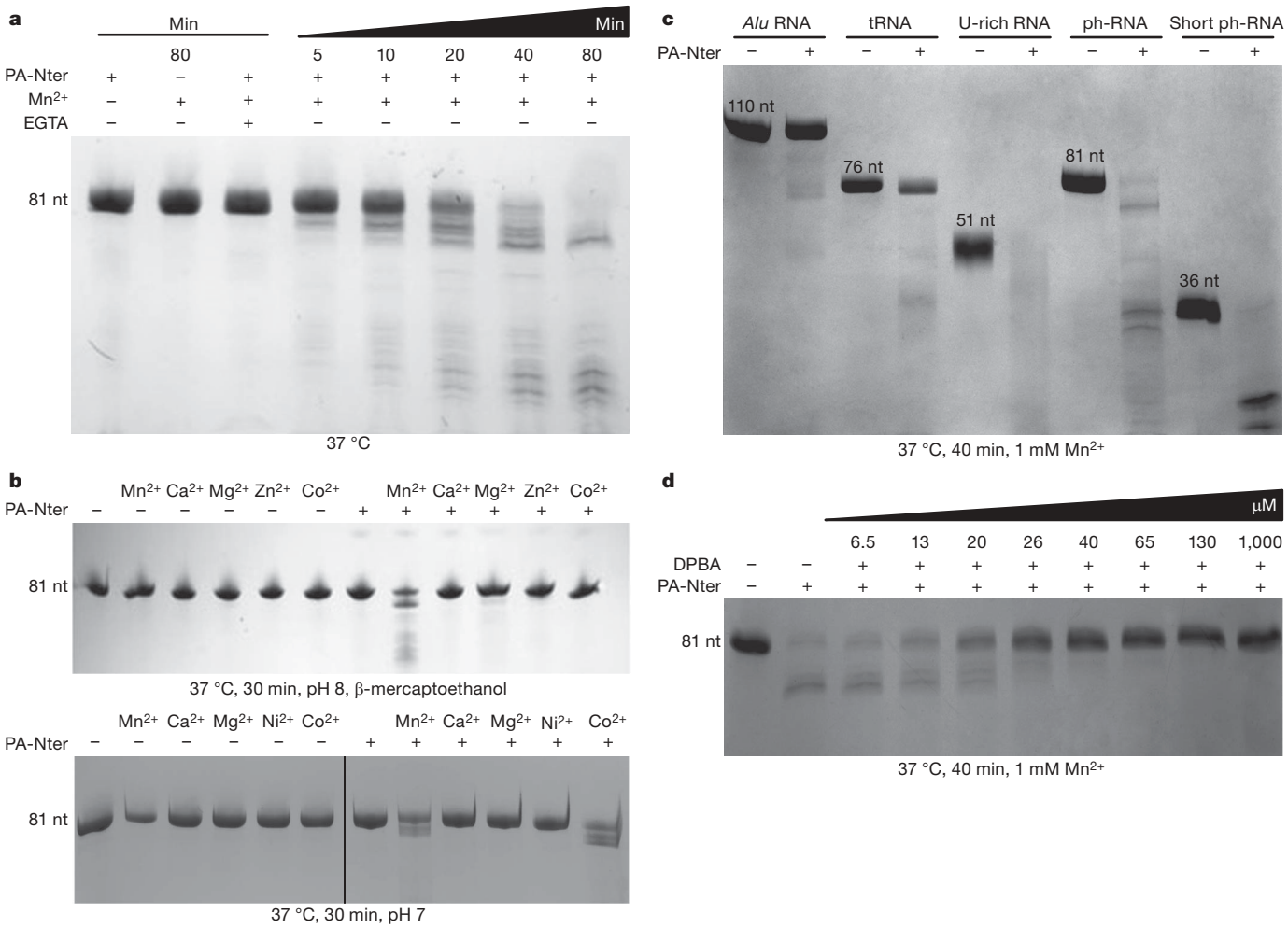


Figure 2 | Endonuclease activity of PA-Nter. a, A time series is shown. Purified ph-RNA was incubated with PA-Nter plus 1 mM $MnCl_2$. The incubation at 37 °C was stopped by adding 20 mM EGTA after 5, 10, 20, 40 and 80 min (lanes 4 to 8). As controls, ph-RNA was incubated for 80 min at 37 °C with PA-Nter only (lane 1), 1 mM $MnCl_2$ only (lane 2), or PA-Nter plus 1 mM $MnCl_2$ and 20 mM EGTA (lane 3). The reaction products were loaded on an 8% acrylamide and 8 M urea gel (15% for **b**, bottom panel and **c**) and stained with methylene blue. nt, nucleotides. **b**, The effect of divalent cations on PA-Nter RNase activity. In the top panel, purified ph-RNA plus PA-Nter were incubated at pH 8 in the presence of β-mercaptoethanol and 1.5 mM $MnCl_2$, $CaCl_2$, $MgCl_2$, $ZnCl_2$ or $CoCl_2$. In the bottom panel, ph-RNA and PA-Nter were incubated at pH 7 with 1.5 mM $MnCl_2$, $CaCl_2$, $MgCl_2$, $NiCl_2$ or $CoCl_2$. After 30 min the reactions were stopped by the addition of 20 mM EGTA. Controls were performed using either salts or PA-Nter alone as indicated. At pH 7, $CoCl_2$ stimulated the endonuclease stronger than $MnCl_2$, whereas at pH 8, $CoCl_2$ precipitates and thus does not activate the endonuclease[6]. **c**, PA-Nter RNase activity on different RNA substrates. SRP *Alu* RNA, tRNA, U-rich RNA, ph-RNA or short ph-RNA were incubated with 1 mM $MnCl_2$ plus PA-Nter (lanes 2, 4, 6, 8 and 10) or in the absence of PA-Nter (lanes 1, 3, 5, 7 and 9). The digestion was performed at 37 °C. After 40 min the reaction was stopped by the addition of 20 mM EGTA. **d**, The inhibition of PA-Nter endonuclease activity by DPBA. Cleavage of ph-RNA by PA-Nter was tested at 37 °C during 40 min in the presence of 1 mM $MnCl_2$ and increasing concentrations of DPBA (0, 6.5, 13, 20, 26, 40, 65, 130 and 1,000 μM). As a control, ph-RNA was incubated with 1 mM $MnCl_2$ alone (lane 1).

constant (K_i) for this compound is estimated at 26 μM, in excellent agreement with the half-maximal inhibitory concentration (IC_{50}) reported for the same compound inhibiting cleavage of capped RNA by the intact influenza virus polymerase[8]. Titrating the inhibitor to manganese-bound PA-Nter increases the thermal stability even further (apparent T_m shifts from 59 to 65 °C) (Fig. 1b), whereas the inhibitor has no effect on the stability of metal-free enzyme (data not shown).

We grew small square-plate crystals of PA-Nter only in the presence of both manganese and magnesium that diffracted to about 2 Å resolution, with three independent molecules in the asymmetric unit. The crystal structure reveals a single, folded domain with residues 1–196 visible, comprising seven α-helices and a mixed, five-stranded β-sheet (Fig. 3a). The structure-based sequence alignment among influenza A, B and C viruses (Fig. 3b) projected onto a surface representation shows a very highly conserved depression that is strongly negatively charged owing to a concentration of acidic residues (Supplementary Fig. 3a, b), suggestive of an active site. A structure similarity search gave no high scoring hits, indicating that the global fold is new. The most similar protein found is the archaeal Holliday junction resolvase Hjc from *Pyrococcus furiosus*[18]. The structural alignment of PA-Nter with Hjc superposes helix α3 and strands β1–β5 (Fig. 4a, b) encompassing a structural motif characteristic of many nucleases including resolvases and type II restriction enzymes. The motif includes catalytically important divalent metal ion binding acidic residues Asp 33 and Glu 46 of Hjc[18], on which Asp 108 and Glu 119 of PA-Nter exactly superpose. Structural alignment of PA-Nter with type II restriction endonucleases such as BamHI or EcoRV shows a similar superposition of active site elements (Fig. 4c). The catalytically important residues Glu 45, Asp 74, Asp 90 and Lys 92 of EcoRV align with His 41, Asp 108, Glu 119 and Lys 134 of PA-Nter, respectively, although the lysines are positioned differently in the primary sequence (Fig. 4d)[19]. The conserved lysine is implicated in stabilizing the attacking hydroxide nucleophile during catalysis[19]. Thus PA-Nter is a new member of the PD-(D/E)XK nuclease superfamily which encompasses a diversity of enzymes involved in various aspects of DNA metabolism[20]. In PA-Nter, the characteristic motif occurs at 107-PDLYDYK, although the separation between the two acidic residues is unusually short and the putative catalytically important lysine (Lys 134) has 'migrated' to an alternative position, as in some other members of the superfamily[20]. In this family, PA-Nter is unusual in that it is biologically functional as an RNase and has a histidine in the active site.

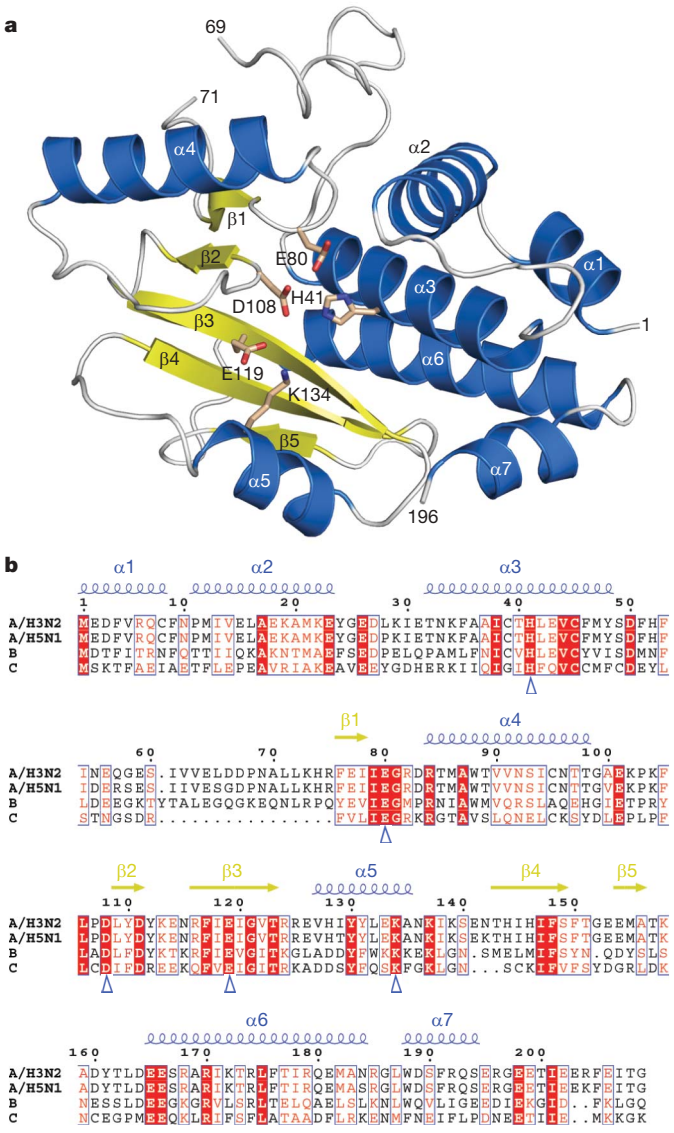


Figure 3 | Three-dimensional structure of PA-Nter. **a**, Ribbon diagram of the structure of influenza A/Victoria/3/1975 PA-Nter with α-helices in blue and β-strands in yellow, labelled as in **b**. The key active site residues are indicated in stick representation. **b**, Sequence alignment of PA-Nter from representative influenza strains: A/Victoria/3/1975 (human H3N2), A/Duck/Vietnam/1/2007 (avian H5N1), B/Ann Arbor/1/1966 (B) and C/Johannesburg/1/1966 (C). The secondary structure of PA-Nter of A/Victoria/3/1975 is shown over the sequence alignment in accordance with colours used for **a**. Residues in a solid red background are identical between the four sequences. The blue triangles indicate the key active site residues.

Our structural and biochemical results combined with previous observations on the trimeric polymerase provide compelling evidence that PA-Nter is the endonuclease that cleaves host mRNAs during cap-snatching. First, the domain has intrinsic RNA and DNA endonuclease activity that is preferentially activated by manganese, in accordance with observations reported for the viral RNPs (Fig. 2b)[6,7]. Second, this activity is inhibited by a compound known to inhibit influenza endonuclease activity with a nearly identical K_i (Fig. 2d)[8]. Third, the domain contains a structural motif characteristic of the catalytic core of a broad family of nucleases, including type II endonucleases. The active site features a cluster of three acidic residues (Glu 80, Asp 108 and Glu 119) and a putative catalytic lysine (Lys 134) (Fig. 4). Fourth, these acidic residues, together with His 41,

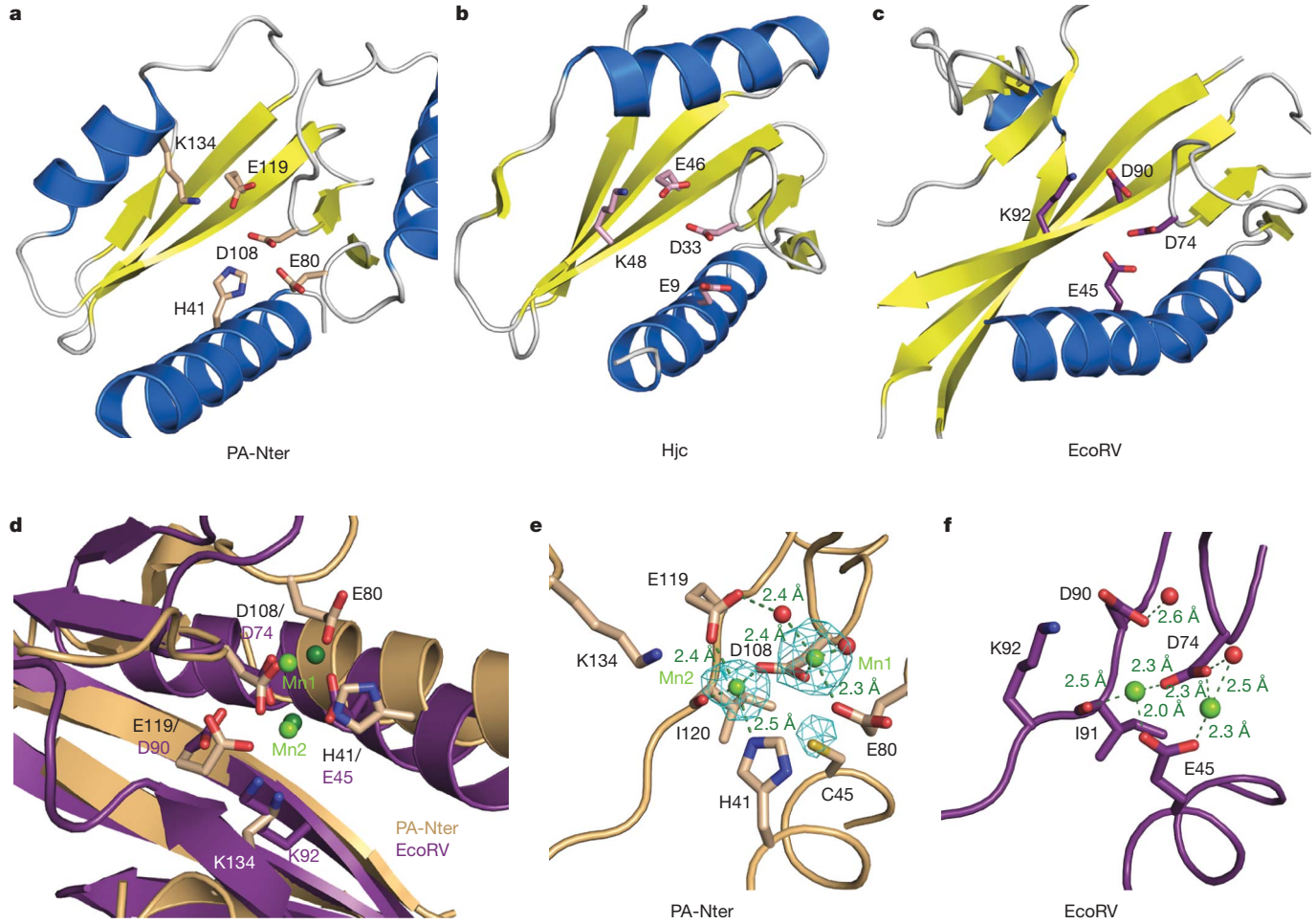


Figure 4 | Comparison of PA-Nter with other nucleases of the PD-(D/E)XK superfamily. **a**–**c**, Comparison of PA-Nter (**a**), *P. furiosus* Holliday junction resolvase Hjc (PDB accession 1GEF)[18] (**b**), and *E. coli* EcoRV restriction enzyme (PDB accession 1STX, product complex with DNA and manganese)[19] (**c**) after superposition of the conserved core active site structural motif. The root mean squared deviations are 2.9 Å for 77 aligned Cα atoms of Hjc (*Z*-score 3.3), and 2.46 (3.1) Å for 55 (72) aligned Cα atoms of EcoRV. Secondary-structure elements are as in Fig. 3a with key active sites residues in stick representation. **d**–**f**, Details of the metal-binding active sites of influenza PA-Nter (molecule A) (beige) and *E. coli* EcoRV restriction enzyme product complex (purple). **d**, Superposition of the active sites showing overlap of key active site metal binding and catalytic functional groups of the two proteins. Manganese ions are light green for PA-Nter and dark green for EcoRV. **e**, **f**, Manganese ion interactions within the active sites of PA-Nter (**e**) and the EcoRV product complex (**f**). Manganese ions and water molecules are shown as green and red spheres, respectively. In **e** the anomalous difference map contoured at 3σ, calculated using manganese K edge diffraction data and model phases, is in cyan. Peak heights are 14.1, 10.1 and 5.0σ for Mn1, Mn2 and the sulphur of Cys 45, respectively.

all absolutely conserved in influenza viruses, coordinate two manganese ions in a configuration consistent with a two-metal-dependent reaction mechanism as proposed for many nucleases (Fig. 4e)[19,21,22]. Fifth, mutations of important active site residues D108A and K134A have been shown to specifically abolish endonuclease activity of the trimeric polymerase[9]. In addition, the mutation H41A abolished all polymerase activity, consistent with a crucial role in both the stability of PA and the activity of the endonuclease[9].

Our results establish for the first time, to our knowledge, a unique role for the PA subunit of influenza virus polymerase and contradict the widely held view that the endonuclease active site is located in the PB1 subunit. This was on the basis of data showing that PB1 residues 508–522 could be cross-linked to capped RNAs by means of a thiouridine adjacent to the endonucleolytic cleavage site[2]. The cross-linked region includes acidic residues Glu 508, Glu 519 and Asp 522, which when individually mutated to alanine specifically abolished endonuclease activity, although their exact function was not determined. These data can be reconciled with our observations if the cross-linked region of PB1 is in spatial proximity to the actual endonuclease active site in PA-Nter. We note that similar cross-linking experiments[2] led to the incorrect identification of the cap-binding site on PB2 (refs 4, 23). Indeed it is very likely that other regions of the polymerase regulate the intrinsic endonuclease activity of PA-Nter, because it is known that there is a cooperative functional interaction between capped RNA binding, 3′ and 5′ end viral RNA (vRNA) binding, and endonuclease activity[24,25]. It is therefore surprising that, as with the cap-binding domain of PB2, a small independently folded domain should have so many of the functional properties of the endonuclease reported for the trimeric complex, as both cap-snatching activities were thought to be detectable only in the trimeric complex. Nonetheless, full understanding of the concerted mechanism of cap-dependent transcription will undoubtedly require high resolution structures of the intact polymerase with bound vRNA and capped RNA.

Influenza can be a serious illness and there is a continuing risk of a devastating pandemic resulting from the eventual acquisition by highly pathogenic avian strains (for example, H5N1) of the ability to be transmissible from human to human. Whereas mutations in the viral polymerase are known to have an important role in virulence and interspecies transmission[26,27], none of the systematic differences in PA-Nter between human and avian strains (Fig. 3b) has been implicated in these processes nor seems to be critically placed as to affect endonuclease function. Several specific influenza virus endonuclease inhibitors

have been described[8,10,11] including some designed to chelate metal ions bound in the active site[12]. Our results show that at least one of these inhibitors is active against PA-Nter and that it binds and stabilizes the metal-bound form of the enzyme (Fig. 1b). These observations will be helpful in developing potential new antivirals using a structure-based approach.

METHODS SUMMARY

PA-Nter, residues 1–209 from A/Victoria/3/1975 (H3N2), was expressed in *Escherichia coli* and purified by affinity and gel filtration chromatography. The influence of metal ions on thermal stability was tested by Thermofluor assays[28]. The endonuclease activity was tested by incubation at 37 °C of 13 μM PA-Nter with 10 μM of various RNA substrates: *Alu* RNA, 110 nucleotides of the *Alu* domain of *Pyrococcus horikoshii* SRP RNA, *Candida albicans* $tRNA^{Asn}$, U-rich RNA (5′-GGCCAUCCUGU$_7$ CCCU$_{11}$CU$_{19}$-3′)[29], ph-RNA of 81 nucleotides[30], and short ph-RNA of 36 nucleotides comprising just the conserved 3′ and 5′ ends with a short linker and circular single stranded DNA (M13mp18) (Fermentas). Crystals diffracting to a maximum of 2 Å resolution were obtained at 20 °C by the hanging-drop method using a protein solution of 5–10 mg ml^{-1} in 20 mM Tris-HCl, pH 8.0, 100 mM NaCl and 2.5 mM $MnCl_2$ and a reservoir composition of 1.2 M Li_2SO_4, 100 mM MES, pH 6.0, 10 mM Mg acetate and 3% ethylene glycol. Diffraction data were collected on beamlines ID14-4 and ID23-1 at the European Synchrotron Radiation Facility (ESRF). The structure was solved by the single-wavelength anomalous diffraction (SAD) method using a gadolinium-chloride-soaked crystal, and initial phases were improved by three-fold non-crystallographic symmetry averaging. Data were also measured on a native crystal at the manganese K edge to reveal the location and identity of bound manganese ions through anomalous difference Fourier synthesis. The structure described is that of a native crystal soaked with additional $MnCl_2$ at 10 mM. The final *R* factor (R_{free}) is 0.217 (0.268). Of the three molecules in the asymmetric unit, A has the best defined metal structure and D is the most disordered. Further details are available in the Supplementary Methods including a table of crystallographic statistics (Supplementary Table 1).

Received 10 November; accepted 22 December 2008.
Published online 1 February 2009.

1. Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. A unique cap(m^{7}GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23,** 847–858 (1981).
2. Li, M. L., Rao, P. & Krug, R. M. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. *EMBO J.* **20,** 2078–2086 (2001).
3. Fechter, P. *et al.* Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding. *J. Biol. Chem.* **278,** 20381–20388 (2003).
4. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Struct. Mol. Biol.* **15,** 500–506 (2008).
5. Shi, L., Summers, D. F., Peng, Q. & Galarz, J. M. Influenza A virus RNA polymerase subunit PB2 is the endonuclease which cleaves host cell mRNA and functions only as the trimeric enzyme. *Virology* **208,** 38–47 (1995).
6. Doan, L., Handa, B., Roberts, N. A. & Klumpp, K. Metal ion catalysis of RNA cleavage by the influenza virus endonuclease. *Biochemistry* **38,** 5612–5619 (1999).
7. Klumpp, K., Doan, L., Roberts, N. A. & Handa, B. RNA and DNA hydrolysis are catalyzed by the influenza virus endonuclease. *J. Biol. Chem.* **275,** 6181–6188 (2000).
8. Tomassini, J. *et al.* Inhibition of cap (m^{7}GpppXm)-dependent endonuclease of influenza virus by 4-substituted 2,4-dioxobutanoic acid compounds. *Antimicrob. Agents Chemother.* **38,** 2827–2837 (1994).
9. Hara, K., Schmidt, F. I., Crow, M. & Brownlee, G. G. Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. *J. Virol.* **80,** 7789–7798 (2006).
10. Hastings, J. C., Selnick, H., Wolanski, B. & Tomassini, J. E. Anti-influenza virus activities of 4-substituted 2,4-dioxobutanoic acid inhibitors. *Antimicrob. Agents Chemother.* **40,** 1304–1307 (1996).
11. Tomassini, J. E. *et al.* A novel antiviral agent which inhibits the endonuclease of influenza viruses. *Antimicrob. Agents Chemother.* **40,** 1189–1193 (1996).
12. Parkes, K. E. *et al.* Use of a pharmacophore model to discover a new class of influenza endonuclease inhibitors. *J. Med. Chem.* **46,** 1153–1164 (2003).
13. Fodor, E. *et al.* A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *J. Virol.* **76,** 8989–9001 (2002).
14. Lee, M. T. *et al.* Definition of the minimal viral components required for the initiation of unprimed RNA synthesis by influenza virus RNA polymerase. *Nucleic Acids Res.* **30,** 429–438 (2002).
15. Honda, A., Mizumoto, K. & Ishihama, A. Minimum molecular architectures for transcription and replication of the influenza virus. *Proc. Natl Acad. Sci. USA* **99,** 13166–13171 (2002).
16. He, X. *et al.* Crystal structure of the polymerase PA_C–$PB1_N$ complex from an avian influenza H5N1 virus. *Nature* **454,** 1123–1126 (2008).
17. Obayashi, E. *et al.* The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* **454,** 1127–1131 (2008).
18. Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y. & Morikawa, K. Crystal structure of the archaeal Holliday junction resolvase Hjc and implications for DNA recognition. *Structure* **9,** 197–204 (2001).
19. Horton, N. C. & Perona, J. J. DNA cleavage by EcoRV endonuclease: two metal ions in three metal ion binding sites. *Biochemistry* **43,** 6841–6857 (2004).
20. Knizewski, L., Kinch, L. N., Grishin, N. V., Rychlewski, L. & Ginalski, K. Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. *BMC Struct. Biol.* **7,** 40 (2007).
21. Beese, L. S. & Steitz, T. A. Structural basis for the 3′-5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10,** 25–33 (1991).
22. Viadiu, H. & Aggarwal, A. K. The role of metals in catalysis by the restriction endonuclease BamHI. *Nature Struct. Biol.* **5,** 910–916 (1998).
23. Tarendeau, F. *et al.* Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog.* **4,** e1000136 (2008).
24. Hagen, M., Chung, T. D., Butcher, J. A. & Krystal, M. Recombinant influenza virus polymerase: requirement of both 5′ and 3′ viral ends for endonuclease activity. *J. Virol.* **68,** 1509–1515 (1994).
25. Rao, P., Yuan, W. & Krug, R. M. Crucial role of CA cleavage sites in the cap-snatching mechanism for initiating viral mRNA synthesis. *EMBO J.* **22,** 1188–1198 (2003).
26. Gabriel, G. *et al.* The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc. Natl Acad. Sci. USA* **102,** 18590–18595 (2005).
27. Salomon, R. *et al.* The polymerase complex genes contribute to the high virulence of the human H5N1 influenza virus isolate A/Vietnam/1203/04. *J. Exp. Med.* **203,** 689–697 (2006).
28. Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357,** 289–298 (2006).
29. Saito, T., Owen, D. M., Jiang, F., Marcotrigiano, J. & Gale, M. Jr. Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. *Nature* **454,** 523–527 (2008).
30. Baudin, F., Bach, C., Cusack, S. & Ruigrok, R. W. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO J.* **13,** 3158–3165 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the ESRF, EMBL and MRC-France for access to synchrotron facilities and the Partnership for Structural Biology for an integrated structural biology environment. The work was partially funded by the EU FLUPOL contract (SP5B-CT-2007-044263) and the ANR FLU INTERPOL contract (ANR-06-MIME-014-02). A.D. has a PhD fellowship from the MENRT. We thank J.-L. Decout for advice about the inhibitor, E. Kowalinski and C. Swale for RNA transcripts, and C. Petosa and J. Perona for a critical reading of the manuscript.

Author Contributions A.D., D.B., T.C. and A.A.M. performed the experiments, D.J.H. and F.B. supervised the search for a soluble fragment of PA, R.W.H.R. and S.C. supervised the biochemical characterization of PA-Nter, and S.C. supervised the crystallography. S.C. wrote the paper with the help of A.D., D.B., T.C. and R.W.H.R.

Author Information Atomic coordinates and structure factors have been deposited with the Protein Data Bank (PDB) under accession codes 2W69 and R2W69SF. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.C. (cusack@embl.fr).

JOURNAL OF VIROLOGY, Sept. 2010, p. 9096–9104 Vol. 84, No. 18
0022-538X/10/$12.00 doi:10.1128/JVI.00995-10

Mutational and Metal Binding Analysis of the Endonuclease Domain of the Influenza Virus Polymerase PA Subunit[▽]

Thibaut Crépin,[1] Alexandre Dias,[1] Andrés Palencia,[1,2] Christopher Swale,[1] Stephen Cusack,[1,2] and Rob W. H. Ruigrok[1]*

UJF-EMBL-CNRS, UMI 3265, Unit of Virus Host-Cell Interactions, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France,[1] *and European Molecular Biology Laboratory, Grenoble Outstation, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France*[2]

Received 7 May 2010/Accepted 23 June 2010

Influenza virus polymerase initiates the biosynthesis of its own mRNAs with capped 10- to 13-nucleotide fragments cleaved from cellular (pre-)mRNAs. Two activities are required for this cap-snatching activity: specific binding of the cap structure and an endonuclease activity. Recent work has shown that the cap-binding site is situated in the central part of the PB2 subunit and that the endonuclease activity is situated in the N-terminal domain of the PA subunit (PA-Nter). The influenza endonuclease is a member of the PD-(D/E)XK family of nucleases that use divalent metal ions for nucleic acid cleavage. Here we analyze the metal binding and endonuclease activities of eight PA-Nter single-point mutants. We show by calorimetry that the wild-type active site binds two Mn^{2+} ions and has a 500-fold higher affinity for manganese than for magnesium ions. The endonuclease activity of the isolated mutant domains are compared with the cap-dependent transcription activities of identical mutations in trimeric recombinant polymerases previously described by other groups. Mutations that inactivate the endonuclease activity in the isolated PA-Nter knock out the transcription but not replication activity in the recombinant polymerase. We confirm the importance of a number of active-site residues and identify some residues that may be involved in the positioning of the RNA substrate in the active site. Our results validate the use of the isolated endonuclease domain in a drug-design process for new anti-influenza virus compounds.

Influenza virus is a segmented negative-strand RNA virus that replicates in the nucleus of infected cells. Its eight viral RNA (vRNA) segments are covered by the viral nucleoprotein (NP) with a stoichiometry of 24 nucleotides per nucleoprotein protomer (25). The 3′ and 5′ ends of the vRNA are bound to the RNA-dependent RNA polymerase, a heterotrimeric complex composed of PB1, PB2, and PA. The complex between vRNA, NP, and the polymerase is called a ribonucleoprotein particle (RNP). Both ends of the viral RNA are necessary for polymerase activity, and together they form the promoter for transcription (12, 18, 33). Influenza virus RNA polymerase differs from the polymerases of the nonsegmented negative-strand RNA viruses, in that it does not carry the enzymatic functions required for the 5′ capping of its own mRNAs (guanylyl- and methyltransferase activities). Instead, influenza virus mRNAs are capped by a unique cap-snatching mechanism (27). The polymerase binds to cellular pre-mRNAs via their cap structure and then cleaves them to generate capped 10- to 13-residue oligonucleotides, which are used as primers to initiate the transcription of viral mRNAs. The viral mRNAs are terminated by a 3′-end poly(A) sequence generated by the stuttering of the RNA polymerase at an oligo(U) motif located at the 5′ end of the template (28, 29). The endonuclease activity for cleavage of the host mRNAs is not active in recombinant trimeric polymerase in the absence of vRNA. For viral transcription, the binding of the 5′ end of the genomic vRNA is necessary to stimulate the cap binding (4). The binding of both ends of the vRNA is also required to stimulate the endonuclease activity (12), and direct binding of annealed 3′ and 5′ vRNA ends stimulates cap binding and endonuclease activity even more strongly (22). These results suggest that significant conformational changes may take place in the RNA polymerase complex upon binding of the vRNA. When the three-dimensional (3D) electron microscopy model of the polymerase on a recombinant RNP was compared with that of free recombinant polymerase, structural differences were indeed observed (1, 5, 35).

The cap-binding site was known to be located within the PB2 subunit (23) but was only recently shown by structural analysis to reside in an independently folding domain of the PB2 subunit, between amino acids 320 and 483 (10). This isolated domain binds to cap analogues in the absence of other parts of the RNA polymerase and of vRNA. The position of the endonuclease site was also controversial but has recently been proven to reside in an also independently folding amino-terminal domain of the PA subunit (6, 37). The crystal structure of the first 196 residues of PA (PA-Nter) (Fig. 1A) shows structural homology with nucleases of the PD-(D/E)XK superfamily, which contains bacterial restriction enzymes such as *Escherichia coli* EcoRV and *Pyrococcus furiosus* Holiday junction resolvase. This family of enzymes binds to one or two divalent metal ions, in particular, Mg^{2+} or Mn^{2+}. One of the ions is clearly involved in catalysis, whereas the role of the second one is not clearly established and may have a modulatory role (inhibition or stimulation, depending on the concentration and the nature of the ion) (17, 26). PA-Nter shows endonuclease activity in the absence of the rest of the poly-

* Corresponding author. Mailing address: UJF-EMBL-CNRS, UMI 3265, Unit of Virus Host-Cell Interactions, 6 rue Jules Horowitz, BP 181, 38042 Grenoble Cedex 9, France. Phone: 33 4 76 20 72 73. Fax: 33 4 76 29 71 99. E-mail: ruigrok@embl.fr.

▽ Published ahead of print on 30 June 2010.

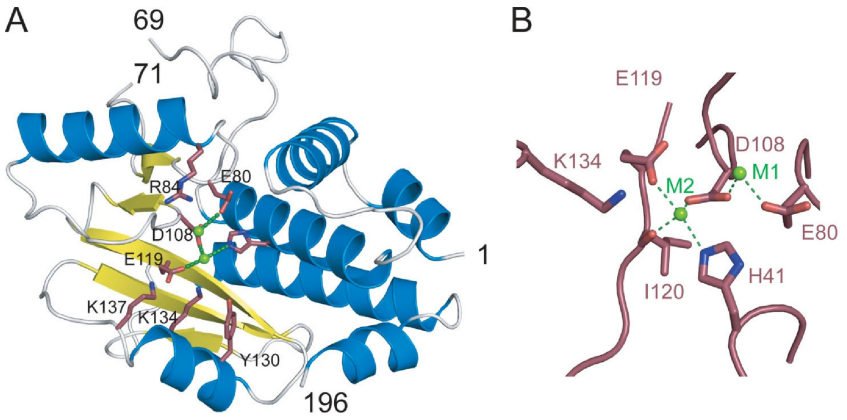


FIG. 1. Active site of PA-Nter. (A) Ribbon diagram of the structure of influenza A/Victoria/3/1975 PA-Nter (Protein Data Bank accession number 2W69) with α helices in blue and β strands in yellow. The key active-site residues mutated in this study are indicated in pink, and the manganese ions are in green. (B) Blowup of the active site of PA-Nter indicating the two metal binding sites and the metal binding ligands.

merase and has the same metal ion dependence as the polymerase in intact RNPs. The domain cleaves most efficiently with manganese ions, followed by cobalt ions (6, 7). In the crystal structure of PA-Nter determined by Dias et al., two divalent cation sites were identified, and both are occupied by manganese (6); metal site 1 (M1) is liganded by Glu80, Asp108, and Glu119 (through a water molecule); and metal site 2 (M2) is liganded by His41, Asp108, Glu119, and the carbonyl oxygen of Ile120 (Fig. 1A and B). On the other hand, Yuan and coworkers crystallized PA-Nter in the presence of only $MgCl_2$ and observed a single Mg^{2+} ion in site M1 (37). The active site also contains Lys134, which could correspond to the catalytic lysine of the PD-(D/E)XK motif (6, 37, 38).

Both the isolated cap-binding domain of PB2 and the endonuclease domain of PA have significant advantages compared to intact viral RNPs for inhibitor screening and structure-based antiviral drug optimization. However, it needs to be ascertained that the isolated domains have the same function and activity as the domains inside the intact trimeric RNA polymerase. A mutational analysis of the cap-binding domain confirmed that mutations in the isolated domain have the same effects as those in the context of the intact polymerase (10). Here we present the results of a mutational analysis of the active site of the isolated endonuclease domain. The mutations include the active-site residues involved in metal binding (His41, Glu80, Asp108, Glu119), the putative catalytic Lys134 residue, and three strictly conserved residues located on the rim surrounding the active site: Arg84, Tyr130, and Lys137 (Fig. 1A). The effects of these mutations introduced in the isolated domain on endonuclease activity are compared to those already tested in the intact recombinant polymerase (13, 37). We also tested the affinities of Mn^{2+} and Mg^{2+} ions for the wild-type (wt) and mutant endonuclease domains and correlate metal binding with enzymatic activity.

MATERIALS AND METHODS

Mutants. The clones containing the mutated DNA coding for PA-Nter (residues 1 to 209), cloned in pETM11, were obtained from Geneart (Germany). The clones code for a poly(His) sequence separated from the protein-coding sequence by a TEV cleavage site. The vectors were used to transform the *E. coli* BL21(DE3) RIL CodonPlus strain (Stratagene). The mutant proteins were expressed in LB medium overnight at 15°C after induction with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). All PA-Nter mutants were expressed in the soluble fraction of the bacterial cells and were purified by an immobilized metal affinity column (IMAC). A second IMAC step was performed after cleavage by the His-tagged TEV protease, followed by gel filtration on a Superdex 200 column (GE Healthcare). Finally, the protein was concentrated to 5 to 10 $mg \cdot ml^{-1}$. For H41A and H41E, the yields were only about 5% of those obtained for the other mutants. The elution profile during the final gel filtration step showed the presence of aggregated material for these mutants.

Biophysical characterization. (i) Far-UV CD spectra. Far-UV circular dichroism (CD) spectra were recorded with a 1-mm path length at 20°C on a Jasco model J-810 CD spectropolarimeter equipped with a Peltier thermostat, as described previously (6). The PA-Nter concentration was 10 μM in 10 mM Tris-HCl, pH 7.0–10 mM NaCl.

(ii) Thermal shift assays. Thermal shift assays were performed with 10 μM PA-Nter in 20 mM Tris-HCl, pH 7.0–100 mM NaCl and a 5× dilution of SYPRO orange dye (Invitrogen), as described. The dye was excited at 490 nm, and the emission light was recorded at 575 nm while the temperature was increased by increments of 1°C per minute from 25 to 75°C. Control assays were carried out in the absence of protein or dye to check that no fluorescence signal was recorded. The experiments gave virtually identical results when they were performed under the same conditions. The variation between experiments came only from the estimation of the flexion point of the curve and was less than 0.5°C in triplicate experiments. For this reason, the results shown in Fig. 2C do not have error bars.

(iii) Isothermal titration calorimetry. Isothermal titration calorimetry (ITC) experiments were performed using a high-precision VP-ITC system (Microcal Inc., Northampton, MA). Proteins were first extensively dialyzed against the titration buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl). All solutions were filtered, degassed to avoid bubble formation, and equilibrated to the corresponding temperature before each experiment. Protein solutions at about 60 μM in the calorimetric cell were titrated with the appropriate metal (0.8 to 9 mM) dissolved in dialysis buffer. Depending on the binding affinities, titrations were carried out either by constant-volume injections (30 injections of 7.5 μl) or by increasing-volume injections (27 injections from 4 to 20 μl) in order to better define the titration curves. The heat evolved after each metal injection was obtained from the integral of the calorimetric signal. The resulting binding isotherms were analyzed by nonlinear least-squares fitting of the experimental data to models corresponding to a single set of identical sites or corresponding to two sets of independent sites. Analysis of the data was performed using the Microcal Origin program (OriginLab Corporation, Northampton, MA).

Endonuclease assays. Endonuclease assays were carried out using an unstructured U-rich RNA probe of 51 nucleotides (6, 31) or a short panhandle RNA of 36 nucleotides comprising just the conserved 3′ and 5′ ends of the viral RNA with a short linker (6). RNA cleavage was performed by incubating 13 μM PA-Nter with various RNA substrates (all at 10 μM) at 37°C in a final volume of 50 μl. The reaction buffer was 20 mM Tris-HCl, pH 8, 100 mM NaCl, and 10 mM β-mercaptoethanol plus 1 mM $MnCl_2$ or 1 mM $MgCl_2$. Incubations were stopped by addition of EGTA at a final concentration of 20 mM. The reaction products

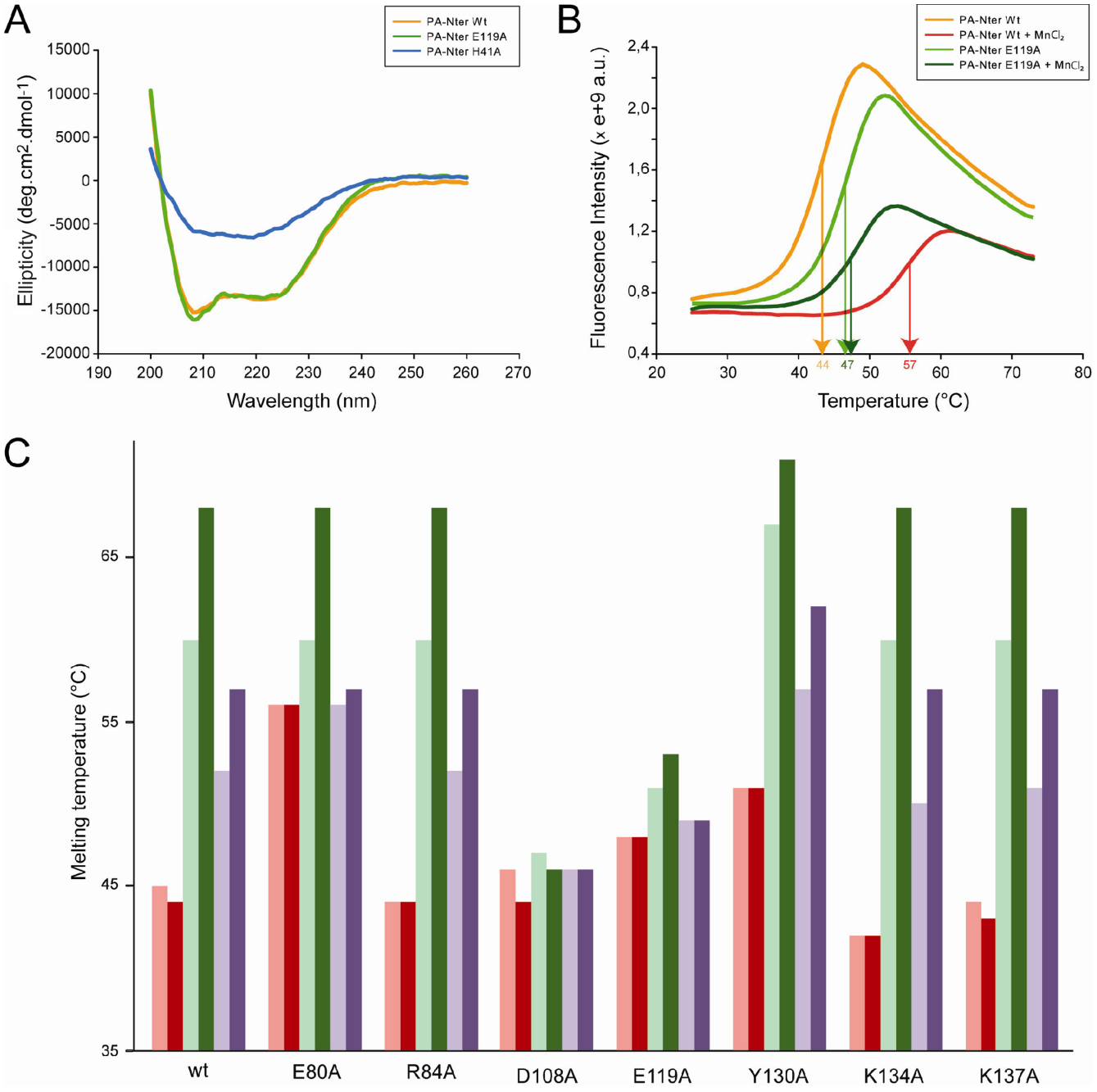


FIG. 2. Biophysical characterization of PA-Nter mutants. (A) Circular dichroism on wt PA-Nter and two of the eight mutants. CD spectra of the wt, H41A, and E119A PA-Nters are in orange, blue, and green, respectively. (B) Thermal stabilization by $MnCl_2$. Thermal shift assays on wt and E119A PA-Nters were done in the presence (red and dark green, respectively) or absence (orange and green, respectively) of 0.5 mM $MnCl_2$. (C) Effects of metals and DPBA binding on the thermal stability. Thermal shift assays to test the metal ion stabilization were performed on wt PA-Nter and the corresponding mutants in the presence (dark colors) and absence (light colors) of DPBA. The proteins were incubated with 0.1 mM $MnCl_2$ (green) or 5 mM $MgCl_2$ (violet) or without any metal (red). All experiments were repeated at least three times on different occasions and with different protein preparations. The results were identical when the same experimental conditions were used.

were loaded on 8 M urea–15% polyacrylamide gels and stained with methylene blue. The results of these assays were not quantified. The activities were estimated from the disappearance of the band of the intact panhandle RNA after incubation for 80 min and 6 h as ++ for wt enzyme, + for mutants like R84A and Y130A that digested less than 50% of the substrate RNA after 80 min but all of it after 6 h, ± for mutants like K137A that still had a significant amount of intact RNA after 6 h, and − for mutants like D108A, E119A, and K134A that were inactive.

RESULTS

All PA-Nter mutants were expressed and purified, and all mutants except the H41A mutant had yields comparable to the yield of the wt. This mutant yielded only about 5% of that for the wild type, and the elution profile of the final gel filtration step showed the presence of aggregates. The folding of the

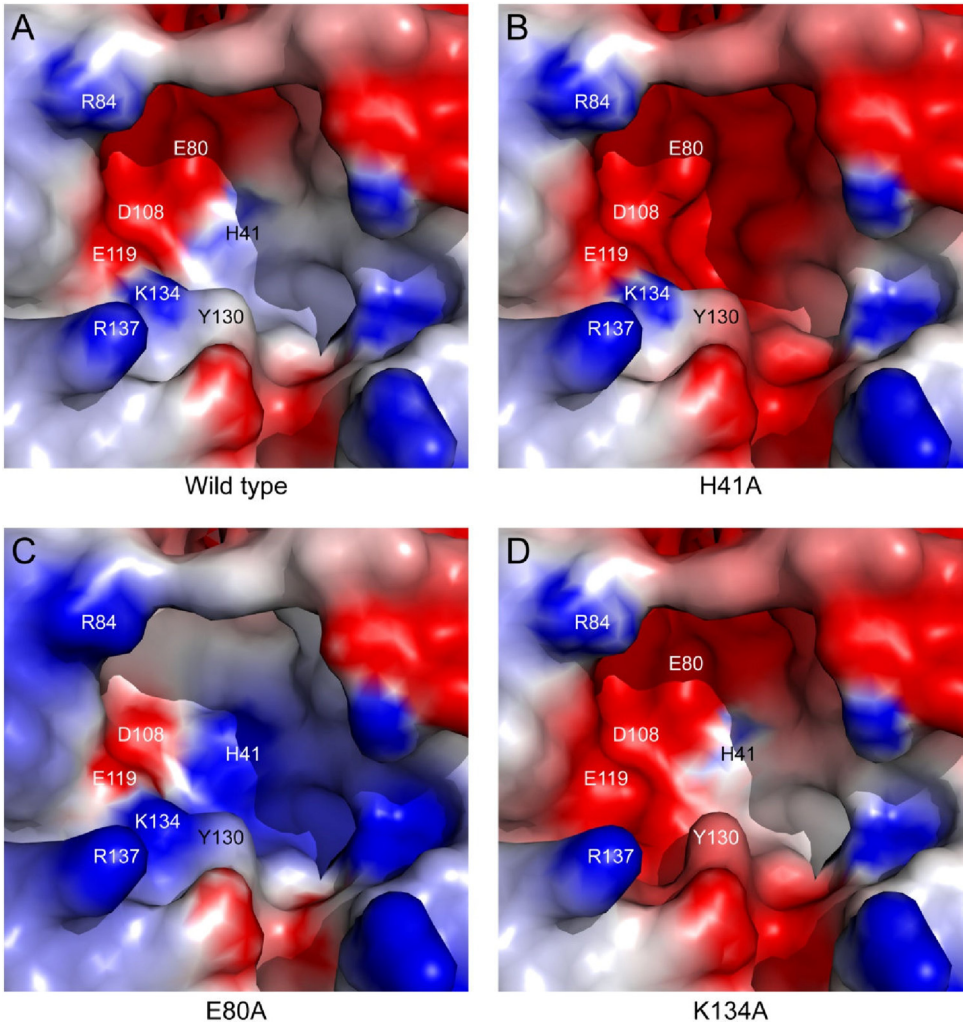


FIG. 3. Electrostatic surface potentials of the active sites of wt and three mutant PA-Nters. The surface of wt PA-Nter (A) was calculated from the crystal structure (Protein Data Bank accession number 2W69), whereas those of mutant PA-ters H41A (B), E80A (C), and K134A (D) correspond to models generated *in silico*. The electrostatic surfaces were calculated using the DelPhi program (30) with H41 fully protonated. The potential scales range from −5.0 kT/e (red) to 5.0 kT/e (blue).

mutants was checked by circular dichroism. All mutants showed a spectrum identical to that of the wt PA-Nter, as shown in Fig. 2A only for the E119A mutant, except that the H41A mutant showed little secondary structure (Fig. 2A).

Thermal stability. The thermal stability of the mutants was tested by Thermofluor assays in which a hydrophobic fluorophore has little affinity for native proteins but binds to denatured proteins, leading to an increase of the fluorescence (9). The apparent melting temperature (T_m) of denaturation can be obtained from the temperature dependence of the fluorescence (Fig. 2B). The thermal stability of the H41A mutant could not be derived since the fluorescent probe bound to the protein at room temperature, another indication that it is not properly folded. The R84A, D108A, and K137A PA-Nters have the same T_m as wt PA-Nter (Fig. 2C, pink bars) and the K134A PA-Nter is slightly less stable, whereas the E80A, E119A, and Y130A PA-Nters are more stable (Fig. 2C). The active site of wt PA-Nter is strongly negatively charged (Fig. 3A). In general, mutations that increase the negative charge in the active-site pocket destabilize the domain (e.g., see Fig. 3B and D for the H41A and K134A PA-Nters, respectively), whereas mutations that reduce the negative charge increase the stability of the domain (e.g., Fig. 3C for the E80A PA-Nter). The fact that the H41A PA-Nter does not fold properly suggests that a basic residue at position 41 is important to provide the necessary electrostatic compensation for correct folding.

Nuclease activity. The RNase activity of the mutants was tested in the presence of 1 mM $MnCl_2$ or 1 mM $MgCl_2$ using short panhandle RNA (36 nucleotides) and unstructured U-rich RNA (51 nucleotides) as substrates (6). The substrate RNA was analyzed on gels after digestion for 80 min or 6 h, as shown in Fig. 4 and summarized in Table 1. Two of three mutants with mutations of acidic residues directly involved in metal binding (E80A, D108A, and E119A) are inactive both with $MnCl_2$ and with $MgCl_2$. The exception was the E80A mutant, which retained activity only in the presence of manganese ions. The K134A mutant is also completely inactive,

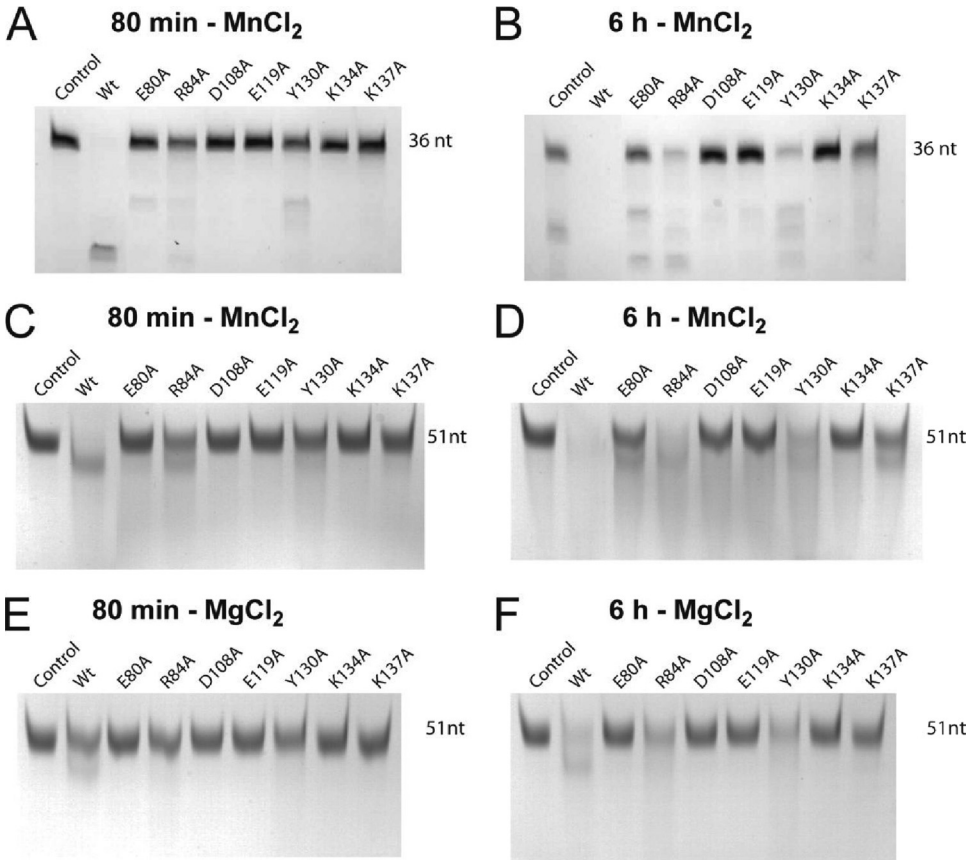


FIG. 4. Endonuclease activities of PA-Nter mutants.(A and B) RNase activities of wild-type and mutant PA-Nters in the presence of 1 mM $MnCl_2$ using a short panhandle RNA of 36 nucleotides (6) for 80 min (A) and 6 h (B). (C to F) The same experiment described for panels A and B but using U-rich RNA (51 nucleotides) plus 1 mM $MnCl_2$ (C and D) and using U-rich RNA plus 1 mM $MgCl_2$ (E and F). RNA cleavage was performed by incubating 13 μM PA-Nter for 80 min (A, C, and E) or 6 h (B, D, and F) with 10 μM RNA at 37°C in a final volume of 50 μl.

which is consistent with it being the catalytic lysine. The R84A, Y130A, and K137A mutants situated on the active-site rim are all less active than the wild type to various degrees but are not inactive. Also note that the activities of the wild type and all active mutants were higher in the presence of 1 mM $MnCl_2$ than in the presence of 1 mM $MgCl_2$ (Fig. 4C to F).

Metal binding monitored by thermal stabilization. We previously showed that the addition of 0.5 mM Mn^{2+} ions significantly enhances the thermal stability of wt PA-Nter and that 0.5 mM dioxo-4-phenylbutanoic acid (DPBA), a known inhibitor of the influenza virus endonuclease (34), supershifts the T_m but only in the presence of metal ions (6). It is likely that the three oxygens on this inhibitor ligate the two resident metal ions in a similar manner, as has recently been observed in the integrase-inhibitor complex of retroviruses (14). To monitor metal binding to the mutant proteins, we therefore used the Thermofluor assay to measure the T_m in the presence of 0.1 mM $MnCl_2$ (Fig. 2C, light green), 5 mM $MgCl_2$ (Fig. 2C, light violet), or no metal (Fig. 2C, pink). We also measured the T_m in the presence of metal ions plus DPBA (Fig. 2C, red, dark green, and dark violet). Two of the three mutations that knock out the metal-ligating acidic residues (E80A, D108A, and E119A) also abolish metal ion binding. The exception was the E80A mutant, which was stabilized by Mn^{2+} but not by Mg^{2+} ions, consistent with the results of the endonuclease assay. All other mutations were stabilized by both types of cations and were further stabilized when DPBA was added in the presence of metal ions, similar to the result for the wt. In the absence of ions, the inhibitor did not stabilize the proteins.

TABLE 1. Effects of mutations on RNase activity of PA-Nter and comparison with transcription and replication activities of the recombinant trimeric polymerase

Mutation	Role	RNase activity		Transcription[a]	Replication[a]
		$MnCl_2$	$MgCl_2$		
wt		++	++	+	+
H41A	Ligand Mn2	ND[b]	ND	−	−
E80A	Ligand Mn1	±	−	−	+
R84A	RNA positioning	+	+	+	+
D108A	Ligands Mn1 and Mn2	−	−	−	+
E119A	Ligand Mn2	−	−	−	+
Y130A	RNA positioning	+	+	+	+
K134A	Catalytic	−	−	−	+
K137A	RNA positioning	±	±	+	+

[a] The results obtained with influenza virus recombinant RNP (13, 37).
[b] ND, not determined.

Metal binding measured by ITC. In the PD-(D/E)XK nuclease superfamily, the positions equivalent to Asp108,

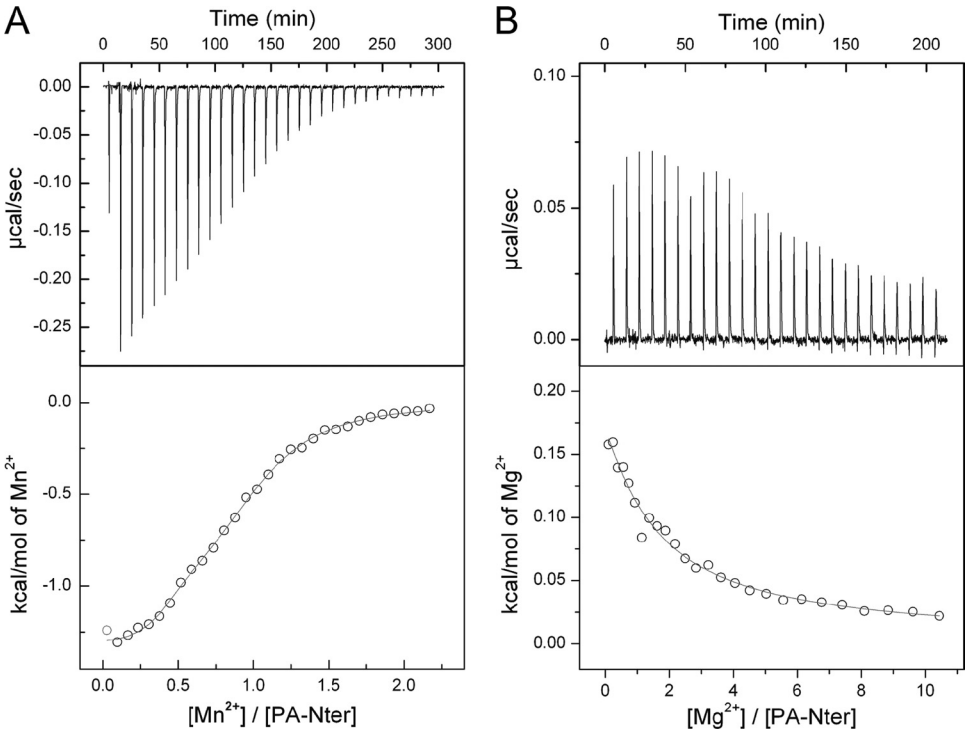


FIG. 5. Isothermal titration calorimetry of wild-type PA-Nter with metal ions. (A) $MnCl_2$ at 0.3 mM was added to 60 μM PA-Nter at 25°C in 20 mM Tris-HCl (pH 8.0) plus 100 mM NaCl; (B) $MgCl_2$ at 9 mM was added to 60 μM PA-Nter at 25°C in 20 mM Tris-HCl (pH 8.0) plus 100 mM NaCl. In the lower panels, the circles represent experimental data and the continuous lines correspond to the best fit to a model with two binding sites.

Glu119, and Lys134 are always present, giving the family its name. Glu80 and His41 are not always conserved, and indeed, this combination seems to be specific for the influenza virus endonuclease (6, 37). Glu80 is particularly interesting, as E80A shows binding and residual nuclease activity with Mn^{2+} ions but not with Mg^{2+} ions. In order to understand these observations, we directly measured Mn^{2+} and Mg^{2+} ion binding to wt PA-Nter and E80A by ITC (Fig. 5 and Table 2), which is the most appropriate technique for directly measuring the thermodynamics of protein-ligand binding (21). We find that binding of Mn^{2+} ions to wt PA-Nter is exothermic (Fig. 5A), whereas binding of Mg^{2+} ions is endothermic (Fig. 5B). The dissociation constants for the Mn^{2+} and Mg^{2+} ions are obtained by model fitting (Table 2), in which the number of binding sites and the values for the disassociation complex (K_d^{-1}) and enthalpy change (ΔH) are variables.

TABLE 2. Metal specificity of PA-Nter[a]

Ion	Site	K_d (μM)	
		wt	E80A
Mn^{2+}	High affinity	0.3	5.2
	Low affinity	6.5	46
Mg^{2+}	High affinity	148	
	Low affinity	4,000	

[a] Isothermal titration calorimetry experiments were performed on a VP-ITC calorimeter using protein solutions at a concentration of 60 μM and metal solution concentrations between 0.8 and 9 mM.

For the wt protein, the titration data for both Mn^{2+} and Mg^{2+} are best fitted with a two-site model since the quality of the fit (by the chi-square test) is better by a factor of 2 than that for a single-site model. Note that at this stage that we make no assumption about whether the two Mn^{2+} sites revealed by ITC are the same as the two Mg^{2+} sites or whether they correspond to sites M1 and M2 defined by the crystal structure (6). The affinity of wt PA-Nter for Mn^{2+} ions (K_ds for the two sites, 0.3 and 6.5 μM) is 500 to 600 times higher than that for Mg^{2+} ions (K_ds for the two sites, 148 and 4,000 μM). This is consistent with the higher endonuclease activity shown by wt PA-Nter in the presence of 1 mM Mn^{2+} compared to that in the presence of 1 mM Mg^{2+} ions (compare Fig. 4C and E). The measured affinities lie in the range of those obtained for other divalent cation binding proteins, such as RNAses and proteases (3, 20, 36). The fact that the binding enthalpy is exothermic for Mn^{2+} ions and endothermic for Mg^{2+} ions suggests a more optimal coordination for bound Mn^{2+} ions, which is reflected in a more favorable enthalpy.

Addition of $MgCl_2$ to the E80A mutant produced no heat effect, in agreement with the absence of thermal stabilization and nuclease activity upon addition of Mg^{2+} ions. The binding data for Mn^{2+} could be modeled using a single site with a K_d of 77 μM and an occupancy of 2.1 or with two sites with K_d values of 5.2 and 46 μM. Because the chi-square test of the single-site model provides a value twice that of the two-site-model, we favor the two-site model. Although the Mn^{2+} binding affinities for the two-site model are only 10 times lower than those for wt PA-Nter, the endonuclease activity of this

mutant is impaired in the presence of 1 mM $MnCl_2$, which may be due to a difference in the geometry of metal binding due to the absence of Glu80.

DISCUSSION

This work was undertaken with dual aims: first, to determine whether mutations in the isolated endonuclease domain had the same effects on activity as the equivalent mutations made in the trimeric polymerase and, second, to clarify whether the nuclease activity of the domain depends on binding to one or two Mg^{2+} or Mn^{2+} ions.

Concerning the first aim, all the alanine mutations that we made in PA-Nter have already been studied in the context of the trimeric recombinant polymerase (13, 37). As Table 1 shows, mutations that inactivate the endonuclease activity in the isolated PA-Nter domain knock out transcription but not replication in the intact polymerase. The experiments with the intact recombinant polymerase were performed in the presence of $MgCl_2$ without added $MnCl_2$. Bearing in mind our results for this mutant, this could explain why the intact polymerase containing the E80A mutation was inactive in transcription in the assay of Hara and coworkers (13). The results for this mutation presented by Yuan and coworkers (37) are more ambiguous; although the mutant was inactive in the endonuclease assay, some globin mRNA-primed transcription activity was retained. Concerning the conserved Arg84, Tyr130, and Lys137 residues on the active-site rim, we hypothesized that these could be necessary for the binding and correct positioning of the substrate RNA in the active site. In particular, Arg84 is bound to a well-ordered sulfate ion in the crystal structure of PA-Nter (6). This sulfate is in the same position as one of the phosphates of the EcoRV restriction enzyme product complex (16). Tyr130 and Lys137 coordinate water molecules that bind to the monophosphate group in a complex of PA-Nter with nucleoside monophosphates (38). In the isolated endonuclease domain, the enzymatic activities of R84A, Y130A, and K137A were reduced, whereas in the context of the intact trimeric polymerase, the mutations retain full transcription activity (13). We may explain this difference by the fact that in the intact polymerase the prebinding of the cellular mRNA to the PB2 cap-binding domain probably considerably enhances substrate affinity by increasing the local concentration. Furthermore, the presence of other positively charged polymerase domains such as the highly basic surfaces of the PB2 627 domain (19, 32) and the C-terminal two-thirds of PA (15, 24) may also assist with the correct positioning of the substrate RNA over the endonuclease active site.

An additional observation is that the active site of the endonuclease is very acidic, like that of other endonucleases (Fig. 3A); and mutations that increase the negative charge destabilize the protein, whereas those that decrease the acidity stabilize PA-Nter. The single exception is the D108A mutant PA-Nter, which has the same stability as wt PA-Nter, for which we do not have an explanation. The H41A mutation in the context of the intact polymerase results in disruption not only of transcription but also of replication (13, 37). Therefore, it seems that the correct folding or the stability of PA-Nter is a prerequisite to the correct formation of the active site of the polymerase on the PB1 subunit. These observations also suggest a rationale for why PA-Nter has a histidine at position 41 rather than a glutamate, as in EcoRV (5).

In conclusion, the accordance of the activities of PA-Nter with those of the intact recombinant trimer suggests that isolated PA-Nter has the same structure in the context of the intact trimer and validates the use of the isolated domain for drug screening and structure-based design. A similar conclusion was drawn for the PB2 cap-binding domain, i.e., that the isolated domain has the same structure as that in recombinant RNP (10). This would imply that activation of the cap-binding and endonuclease functions in the intact polymerase, by the binding of the 5′ end of the vRNA or a 3′-5′ duplex (12, 22), are likely not due to the induced formation of the two active sites but rather to the removal of inhibition (perhaps steric) of these functions. The structural differences that are observed between free polymerase and RNP-bound polymerase (1, 5, 35) suggest important domain rearrangements that may change the disposition of the cap-binding and endonuclease sites and the overall binding of host mRNA by the intact polymerase.

Concerning the second aim of our study, in a careful enzymatic analysis of the endonuclease activity of purified RNPs, Doan et al. showed that there are two interacting metal binding sites that need to be occupied with divalent metal ions for full nuclease activity (7). They found that manganese ions are two times stronger than magnesium ions at activating the endonuclease and showed that the affinity for Mn^{2+} ions is stronger than for Mg^{2+} ions. The metal dependence of the endonuclease activity shown here is in agreement with these results. The nuclease activity of PA-Nter and the mutants is higher in the presence of 1 mM $MnCl_2$ than in the presence of 1 mM $MgCl_2$, and the shift in T_m is higher in the presence of 0.1 mM Mn^{2+} ions than in the presence of 5 mM Mg^{2+} ions.

Yuan and coworkers grew crystals of PA-Nter in the presence of 100 mM $MgCl_2$ and observed a single metal ion only in position M1, even in the presence of mononucleotide phosphates (38). Therefore, in the absence of substrate, Mg^{2+} ions seem to bind only to the M1 site even at concentrations that lay several orders of magnitude above the low affinity K_d (4 mM). Because a second magnesium ion was never seen in the active site, it is possible that the low-affinity binding site for Mg^{2+} lies outside the active site. Histidine is one of the ligands of the M2 site in the crystal structure (6). Whereas manganese ions can be favorably coordinated by both acidic residues and histidine (2, 11), ligation of Mg^{2+} ions by histidine is uncommon (8). This agrees with the biochemical and enzymatic data presented here for the E80A mutant (Glu80 is an M1 ligand), which showed neither RNase activity nor binding of Mg^{2+} ions, although the mutant could still bind two manganese ions. All crystallographic data (37, 38) and our enzymatic and ITC data presented here suggest that only a single magnesium ion can bind to the enzymatic cavity of PA-Nter in the absence of substrate. However, Doan et al. (7) found a Hill coefficient of 2 for the endonuclease activity of intact viral RNPs in the presence of Mg^{2+}, suggesting that two ions can bind in the presence of substrate. The finding that the stability of wt PA-Nter supershifts in the presence of Mg^{2+} plus the inhibitor also suggests that two ions can bind when PA-Nter is stabilized by the inhibitor.

Dias et al. grew crystals in a mixture of 2.5 mM $MnCl_2$ and

5 mM $MgCl_2$ and located two Mn^{2+} ions in positions M1 and M2 (6). Two manganese ions were also found by Zhao and coworkers upon adding $MnCl_2$ (38). Our ITC results explain these structural observations, since they show that Mn^{2+} can bind with a high affinity to two sites (K_ds, 0.3 and 6.5 μM) and that Mn^{2+} binding to its second site is 20-fold higher than that of Mg^{2+} ions to their high-affinity site (K_d, 148 μM). Although with these data we cannot prove that the two Mn^{2+} binding sites obtained through ITC correlate with the two binding sites observed in the crystal structure, it is likely that it is the case. This is supported by the results on activity and Mn^{2+} binding of the E80A mutant, which has an active site that resembles that of EcoRV, which also binds to two manganese ions.

In conclusion, our results are consistent with previous results on the metal dependence of the endonuclease in intact RNPs. We show quantitatively that the endonuclease active site binds to two Mn^{2+} ions and has a significantly higher affinity for Mn^{2+} ions than for Mg^{2+} ions. As was mentioned by Zhao et al., the cellular concentration of free magnesium ions is in the millimolar range, whereas that of manganese ions is in the micromolar range (38), making roles for both ions in the endonuclease activity during infection by influenza virus entirely possible. This suggestion is strengthened by the findings of Doan et al. that indicated a synergistic activation of cleavage activity with combinations of different metal ions (7).

ACKNOWLEDGMENTS

We acknowledge the Partnership for Structural Biology for an integrated structural biology environment.

The work was partially funded by the EU FLUPOL contract (SP5B-CT-2007-044263), the ANR FLU INTERPOL contract (ANR-06-MIME-014-02), and Lyon Biopôle. Alexandre Dias was the recipient of a Ph.D. fellowship from the French MENRT.

REFERENCES

1. **Area, E., J. Martin-Benito, P. Gastaminza, E. Torreira, J. M. Valpuesta, J. L. Carrascosa, and J. Ortin.** 2004. 3D structure of the influenza virus polymerase complex: localization of subunit domains. Proc. Natl. Acad. Sci. U. S. A. **101:**308–313.
2. **Bewley, M. C., and J. M. Flanagan.** 2001. Arginase, p. 952–962. *In* A. Messerschmidt, R. Huber, T. Poulos, and K. Wieghart (ed.), Handbook of metalproteins. John Wiley & Sons, Ltd., Chichester, United Kingdom.
3. **Brazier, M. W., P. Davies, E. Player, F. Marken, J. H. Viles, and D. R. Brown.** 2008. Manganese binding to the prion protein. J. Biol. Chem. **283:** 12831–12839.
4. **Cianci, C., L. Tiley, and M. Krystal.** 1995. Differential activation of the influenza virus polymerase via template RNA binding. J. Virol. **69:**3995–3999.
5. **Coloma, R., J. M. Valpuesta, R. Arranz, J. L. Carrascosa, J. Ortin, and J. Martin-Benito.** 2009. The structure of a biologically active influenza virus ribonucleoprotein complex. PLoS Pathog. **5:**e1000491.
6. **Dias, A., D. Bouvier, T. Crepin, A. A. McCarthy, D. J. Hart, F. Baudin, S. Cusack, and R. W. Ruigrok.** 2009. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature **458:**914–918.
7. **Doan, L., B. Handa, N. A. Roberts, and K. Klumpp.** 1999. Metal ion catalysis of RNA cleavage by the influenza virus endonuclease. Biochemistry **38:**5612–5619.
8. **Dupureur, C. M.** 2008. Roles of metal ions in nucleases. Curr. Opin. Chem. Biol. **12:**250–255.
9. **Ericsson, U. B., B. M. Hallberg, G. T. Detitta, N. Dekker, and P. Nordlund.** 2006. Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal. Biochem. **357:**289–298.
10. **Guilligay, D., F. Tarendeau, P. Resa-Infante, R. Coloma, T. Crepin, P. Sehr, J. Lewis, R. W. Ruigrok, J. Ortin, D. J. Hart, and S. Cusack.** 2008. The structural basis for cap binding by influenza virus polymerase subunit PB2. Nat. Struct. Mol. Biol. **15:**500–506.
11. **Guss, J. M., and H. C. Freeman.** 2001. Aminopeptidase, p. 973–980. *In* A. Messerschmidt, R. Huber, T. Poulos, and K. Wieghart (ed.), Handbook of metalproteins. John Wiley & Sons, Ltd., Chichester, United Kingdom.
12. **Hagen, M., T. D. Chung, J. A. Butcher, and M. Krystal.** 1994. Recombinant influenza virus polymerase: requirement of both 5′ and 3′ viral ends for endonuclease activity. J. Virol. **68:**1509–1515.
13. **Hara, K., F. I. Schmidt, M. Crow, and G. G. Brownlee.** 2006. Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. J. Virol. **80:**7789–7798.
14. **Hare, S., S. S. Gupta, E. Valkov, A. Engelman, and P. Cherepanov.** 2010. Retroviral intasome assembly and inhibition of DNA strand transfer. Nature **464:**232–236.
15. **He, X., J. Zhou, M. Bartlam, R. Zhang, J. Ma, Z. Lou, X. Li, J. Li, A. Joachimiak, Z. Zeng, R. Ge, Z. Rao, and Y. Liu.** 2008. Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. Nature **454:**1123–1126.
16. **Horton, N. C., and J. J. Perona.** 2004. DNA cleavage by EcoRV endonuclease: two metal ions in three metal ion binding sites. Biochemistry **43:**6841–6857.
17. **Imhof, P., S. Fischer, and J. C. Smith.** 2009. Catalytic mechanism of DNA backbone cleavage by the restriction enzyme EcoRV: a quantum mechanical/molecular mechanical analysis. Biochemistry **48:**9061–9075.
18. **Klumpp, K., R. W. Ruigrok, and F. Baudin.** 1997. Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. EMBO J. **16:**1248–1257.
19. **Kuzuhara, T., D. Kise, H. Yoshida, T. Horita, Y. Murazaki, A. Nishimura, N. Echigo, H. Utsunomiya, and H. Tsuge.** 2009. Structural basis of the influenza A virus RNA polymerase PB2 RNA-binding domain containing the pathogenicity-determinant lysine 627 residue. J. Biol. Chem. **284:**6855–6860.
20. **Lai, B., Y. Li, A. Cao, and L. Lai.** 2003. Metal ion binding and enzymatic mechanism of Methanococcus jannaschii RNase HII. Biochemistry **42:**785–791.
21. **Leavitt, S., and E. Freire.** 2001. Direct measurement of protein binding energetics by isothermal titration calorimetry. Curr. Opin. Struct. Biol. **11:** 560–566.
22. **Lee, M. T., K. Klumpp, P. Digard, and L. Tiley.** 2003. Activation of influenza virus RNA polymerase by the 5′ and 3′ terminal duplex of genomic RNA. Nucleic Acids Res. **31:**1624–1632.
23. **Li, M. L., P. Rao, and R. M. Krug.** 2001. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. EMBO J. **20:**2078–2086.
24. **Obayashi, E., H. Yoshida, F. Kawai, N. Shibayama, A. Kawaguchi, K. Nagata, J. R. Tame, and S. Y. Park.** 2008. The structural basis for an essential subunit interaction in influenza virus RNA polymerase. Nature **454:**1127–1131.
25. **Ortega, J., J. Martin-Benito, T. Zurcher, J. M. Valpuesta, J. L. Carrascosa, and J. Ortin.** 2000. Ultrastructural and functional analyses of recombinant influenza virus ribonucleoproteins suggest dimerization of nucleoprotein during virus amplification. J. Virol. **74:**156–163.
26. **Pingoud, V., W. Wende, P. Friedhoff, M. Reuter, J. Alves, A. Jeltsch, L. Mones, M. Fuxreiter, and A. Pingoud.** 2009. On the divalent metal ion dependence of DNA cleavage by restriction endonucleases of the EcoRI family. J. Mol. Biol. **393:**140–160.
27. **Plotch, S. J., M. Bouloy, I. Ulmanen, and R. M. Krug.** 1981. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell **23:**847–858.
28. **Poon, L. L., D. C. Pritlove, E. Fodor, and G. G. Brownlee.** 1999. Direct evidence that the poly(A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. J. Virol. **73:**3473–3476.
29. **Robertson, J. S., M. Schubert, and R. A. Lazzarini.** 1981. Polyadenylation sites for influenza virus mRNA. J. Virol. **38:**157–163.
30. **Rocchia, W., S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, and B. Honig.** 2002. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. J. Comput. Chem. **23:**128–137.
31. **Saito, T., D. M. Owen, F. Jiang, J. Marcotrigiano, and M. Gale, Jr.** 2008. Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. Nature **454:**523–527.
32. **Tarendeau, F., T. Crepin, D. Guilligay, R. W. Ruigrok, S. Cusack, and D. J. Hart.** 2008. Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. PLoS Pathog. **4:**e1000136.
33. **Tiley, L. S., M. Hagen, J. T. Matthews, and M. Krystal.** 1994. Sequence-specific binding of the influenza virus RNA polymerase to sequences located at the 5′ ends of the viral RNAs. J. Virol. **68:**5108–5116.
34. **Tomassini, J., H. Selnick, M. E. Davies, M. E. Armstrong, J. Baldwin, M. Bourgeois, J. Hastings, D. Hazuda, J. Lewis, W. McClements, et al.** 1994. Inhibition of cap (m7GpppXm)-dependent endonuclease of influenza virus by 4-substituted 2,4-dioxobutanoic acid compounds. Antimicrob. Agents Chemother. **38:**2827–2837.
35. **Torreira, E., G. Schoehn, Y. Fernandez, N. Jorba, R. W. Ruigrok, S. Cusack, J. Ortin, and O. Llorca.** 2007. Three-dimensional model for the isolated

recombinant influenza virus polymerase heterotrimer. Nucleic Acids Res. **35:**3774–3783.

36. **Wehenkel, A., M. Bellinzoni, F. Schaeffer, A. Villarino, and P. M. Alzari.** 2007. Structural and binding studies of the three-metal center in two mycobacterial PPM Ser/Thr protein phosphatases. J. Mol. Biol. **374:**890–898.
37. **Yuan, P., M. Bartlam, Z. Lou, S. Chen, J. Zhou, X. He, Z. Lv, R. Ge, X. Li, T. Deng, E. Fodor, Z. Rao, and Y. Liu.** 2009. Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. Nature **458:**909–913.
38. **Zhao, C., Z. Lou, Y. Guo, M. Ma, Y. Chen, S. Liang, L. Zhang, S. Chen, X. Li, Y. Liu, M. Bartlam, and Z. Rao.** 2009. Nucleoside monophosphate complex structures of the endonuclease domain from the influenza virus polymerase PA subunit reveal the substrate binding site inside the catalytic center. J. Virol. **83:**9024–9030.

OPEN ACCESS Freely available online

PLOS | PATHOGENS

Monomeric Nucleoprotein of Influenza A Virus

Sylvie Chenavas[1], Leandro F. Estrozi[1,2], Anny Slama-Schwok[3], Bernard Delmas[3], Carmelo Di Primo[4,5], Florence Baudin[1,6], Xinping Li[6,7], Thibaut Crépin[1], Rob W. H. Ruigrok[1]*

1 UJF-EMBL-CNRS UMI 3265, Unit of Virus Host Cell Interactions, Grenoble, France, **2** CEA-CNRS-UJF UMR 5075, Institut de Biologie Structurale, Grenoble, France, **3** INRA UR 892, Virologie et Immunologie Moléculaires, Jouy en Josas, France, **4** Université de Bordeaux, Institut Européen de Chimie et de Biologie, ARNA laboratory, Pessac, France, **5** INSERM, U869, ARNA, Bordeaux, France, **6** European Molecular Biology Laboratory, Heidelberg, Germany, **7** MPI for Biology of Ageing, Bio-MS Facility, Cologne, Germany

Abstract

Isolated influenza A virus nucleoprotein exists in an equilibrium between monomers and trimers. Samples containing only monomers or only trimers can be stabilized by respectively low and high salt. The trimers bind RNA with high affinity but remain trimmers, whereas the monomers polymerise onto RNA forming nucleoprotein-RNA complexes. When wild type (wt) nucleoprotein is crystallized, it forms trimers, whether one starts with monomers or trimers. We therefore crystallized the obligate monomeric R416A mutant nucleoprotein and observed how the domain exchange loop that leads over to a neighbouring protomer in the trimer structure interacts with equivalent sites on the mutant monomer surface, avoiding polymerisation. The C-terminus of the monomer is bound to the side of the RNA binding surface, lowering its positive charge. Biophysical characterization of the mutant and wild type monomeric proteins gives the same results, suggesting that the exchange domain is folded in the same way for the wild type protein. In a search for how monomeric wt nucleoprotein may be stabilized in the infected cell we determined the phosphorylation sites on nucleoprotein isolated from virus particles. We found that serine 165 was phosphorylated and conserved in all influenza A and B viruses. The S165D mutant that mimics phosphorylation is monomeric and displays a lowered affinity for RNA compared with wt monomeric NP. This suggests that phosphorylation may regulate the polymerisation state and RNA binding of nucleoprotein in the infected cell. The monomer structure could be used for finding new anti influenza drugs because compounds that stabilize the monomer may slow down viral infection.

Citation: Chenavas S, Estrozi LF, Slama-Schwok A, Delmas B, Di Primo C, et al. (2013) Monomeric Nucleoprotein of Influenza A Virus. PLoS Pathog 9(3): e1003275. doi:10.1371/journal.ppat.1003275

Editor: Félix A. Rey, Institut Pasteur, France

Received November 29, 2012; **Accepted** February 11, 2013; **Published** March 28, 2013

Funding: This work was funded in part by the French Agency for Research (ANR) (FLUNUCLEOVIR N° ANR 2010 Blan 1307 01), the European Commission under grant agreement no259751 (Flupharm project: New drugs targeting influenza virus polymerase, 11/2010-05/2014), a grant from the French A(H1N1) research program, coordinated by the Institut de Microbiologie et Maladies Infectieuses (INSERM, France) and by Lyon Biopôle. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ruigrok@embl.fr

Introduction

Negative strand RNA viruses have an RNA genome in the opposite sense of that of messenger RNA. Therefore, the first viral activity after entering the host cell is transcription by the viral RNA-dependent RNA polymerase. The template for transcription is a complex between the viral RNA and the nucleoprotein (NP) that binds to the RNA sugar-phosphate backbone [1,2]. NP is necessary for RNA elongation by the polymerase [3,4]. However, its main function may be to separate the newly made mRNA from the template RNA because the infecting viral replication complexes do not contain helicases and purified influenza virus NP melts dsRNA [1].

Negative strand RNA viruses include non-segmented viruses like the *Rhabdoviridae* (ex. vesicular stomatitis virus (VSV) and rabies virus) and the *Paramyxoviridae* (ex. Sendai and measles virus) and segmented viruses like the *Arenaviridae* (Lassa fever virus), the *Bunyaviridae* (Rift Valley fever virus (RVFV)) and the *Orthomyxoviridae* (influenza viruses). When expressed in a transfected cell in the absence of other viral components, the nucleoproteins of most of these viruses bind to cellular RNA and form nucleoprotein-RNA complexes that are indistinguishable from the viral complexes [5]. The formation of such complexes results from two coupled activities of the nucleoproteins: RNA binding and self polymerisation. In infected cells, these nucleoproteins bind almost exclusively to their viral RNAs and, therefore, all these viruses have developed a mechanism to stop their NPs from binding to cellular RNA and from polymerizing. The non-segmented viruses code for another viral protein, the phosphoprotein (P), that binds with its N-terminal end to RNA-free nucleoprotein, indicated by N^0 [6–8]. The structure of the N^0P complex of VSV shows how the P binding site overlaps with the RNA binding groove on the nucleoprotein and with one of the sites involved in nucleoprotein polymerisation, thus blocking both activities [9].

The segmented viruses do not code for an equivalent of a phosphoprotein and solve the problem in different ways. The nucleoprotein of RVFV has been crystallised in two forms; as a monomer and as a hexameric ring [10,11]. In the ring, two N-terminal α helices of NP swing out to the back of a neighbouring protomer for self polymerisation. Inside the ring there is a continuous positively charged surface that binds the RNA [12]. In the monomeric form the two N-terminal helices fold onto the positively charged surface of their own protomer. Thus, the monomeric, closed form avoids at the same time RNA binding and polymerisation. It is likely that, in the infected cell, a signal on the newly produced RNA or the polymerase itself changes the

Author Summary

The RNAs of negative strand RNA viruses are encapsidated by their specific viral nucleoproteins, forming helical nucleoprotein-RNA structures that are the template for transcription and replication. All these nucleoproteins have two activities in common: RNA binding and self-polymerisation, and it is likely that these activities are coupled. All these viruses have to keep their nucleoprotein from binding to cellular RNA and from polymerisation before viral RNA binding. The non-segmented viruses solve this by coding for a phosphoprotein that binds to the nucleoprotein, blocking both activities. The segmented viruses, such as influenza and Bunyaviruses, do not code for a phosphoprotein and need to solve this problem differently. Here we present the atomic structure of monomeric influenza virus nucleoprotein. Although the structures of the influenza virus and the Rift Valley Fever Virus (Bunya virus) nucleoproteins are different, there are functional similarities when the monomer and polymer structures are compared. Both nucleoproteins have a core structure that is identical in the monomer and the polymer. They contain a flexible arm that moves over to a neighbouring protomer in the polymer structure but that folds onto the core in the monomer structure, hiding the RNA binding groove in the Rift valley Fever Virus nucleoprotein and modifying the electrostatic potential of the RNA binding platform of the influenza virus protein.

conformation of the nucleoprotein so that it binds to the viral RNA and polymerizes. The Lassa fever virus nucleoprotein was crystallised in its intact form that shows a C-terminal exonuclease domain [13,14] and an N-terminal domain. The intact protein crystallised as a circular trimer with the C-terminal domain of one protomer binding to the N-terminal domain of the next one [13]. Neither of the domains had RNA bound. When the N-terminal domain was expressed alone it did bind RNA [15] and, compared with the closed structure of the intact protein, secondary structure elements had moved away to open the RNA binding cleft. The C-terminal domain may control the opening and closing of the RNA binding groove and a linear association of nucleoprotein protomers could permit the binding of RNA. Thus, although the structures of the RVFV and Lassa virus nucleoproteins are totally different, both proteins show inactive conformations with a closed RNA binding site and conformations in which secondary structure elements have moved away to open the RNA binding site [5]. For both proteins it was also suggested that polymerisation and RNA binding are coupled, like for the nucleoproteins of the non-segmented viruses.

The nucleoproteins of influenza A H1N1 and H5N1 crystallised as trimers [16,17] and NP of influenza B virus as a tetramer [18]. In these structures, each protomer binds with a domain exchanged C-terminal tail loop (residues 402–428 for A/H1N1) in a groove on the core of a neighbouring protomer. On the opposite side of the loop, on the core of the protein, there is a large and shallow positively charged surface. Mutating residues on this surface lowers RNA binding [17]. NP purified from virus by CsCl gradient centrifugation exists in an equilibrium between monomers and oligomers going from trimers and tetramers to large structures resembling viral ribonucleoprotein complexes [19]. Although recombinant NP is generally considered to exist only as trimers [16,17], it was recently shown that there exists an equilibrium between monomers and trimers/tetramers [20,21]. This equilibrium is shifted to the oligomeric state at 300 mM salt whereas a stable population of monomers was found at 50 mM salt [20]. For some mutants like R416A and E339A the equilibrium is shifted to the monomeric form [16,20,22–24] whereas for the Y148A mutant the equilibrium is shifted to the trimeric form [20].

Here we show the structure of the monomeric R416A mutant and describe the RNA binding characteristics of monomeric wild type NP. In the infected cell NP may be kept monomeric by post translational modification. Mass spectroscopy analysis on NP isolated from viral RNP showed phosphorylation of serine 165. By mutagenesis we generated the S165D mutant to mimic this phosphorylation. S165D was monomeric with the same biophysical characteristics as the R416A mutant but showed high cooperativity for RNA binding at concentrations above the Kd.

Results

We previously showed that influenza virus NP can be stabilised as monomers in 50 mM salt and as trimers and tetramers in 300 mM salt [20] and wanted to further study their RNA binding behaviour. Wild type NP monomers and trimers/tetramers were incubated with an RNA oligonucleotide of 51 nucleotides [20]. Monomeric NP bound rapidly to the RNA and formed circular structures within 1 hour (Figure 1 left). These structures resemble circular recombinant mini RNPs [25]. The trimer/tetramer binds rapidly to RNA with a three-fold higher affinity than monomeric NP [20] but upon binding remained as trimers and tetramers up to 18 hours incubation (Figure 1 middle). However, when oligomeric NP was added to the RNA and then diluted to 50 mM NaCl, higher order polymers and rings formed within 1 hour like for the monomeric NP (Figure 1 right), most likely through dissociation of the oligomers upon which the monomers polymerised onto the RNA. In agreement with the results published in Tarus *et al.* [20] our results suggest that influenza NP is in equilibrium between monomers and oligomers but only monomers can form circular NP-RNA complexes.

In order to study monomeric NP we tried to crystallise monomeric wt NP but we obtained the same crystals and structure as was obtained for the wt trimer [16]. Thus, the crystallisation conditions seem to push monomeric NP to form trimers in the crystal. We then crystallised the obligate monomeric R416A mutant of NP. Influenza A/WSN/33 R416A mutant NP was concentrated to 10 mg/ml and crystallised by vapour diffusion using the sitting drop method. Crystals with space group $C222_1$ diffracted to 2.7 Å (Table 1). The structure was solved by molecular replacement starting from the structure of wt A/WSN/33 NP (PDB 2IQH) in which residues 400 to 498 had been removed. Even though the crystals of wt and mutant NP have the same space group, wt NP crystallises as a trimer whereas mutant NP crystallised as a monomer but with 3 molecules in the asymmetric unit; the cell dimensions are different and the positions of the monomers are different from those of the protomers in the trimer. The structure (Figures 2B, C and Figure S1 in Text S1) shows residues 21–391 and 408–498. Most of the mutant protein structure is identical to that of the wt, from residue 22 to 385, the rmsd is 1.09 Å for 364 aligned Cα's. From residue 386 onward the structure is different because the exchange domain folds into the groove of its own protomer rather than in the groove of a neighbouring protomer. Figure 2 shows the comparison of the structures of trimeric wt and monomeric R416A NP. The core of the protein is in grey because this part of the structure is the same for the wt trimer and the mutant monomer, residues 386 to 401 are in green, the trimer exchange domain (residues 402–428) is in yellow, and residues 429–498 are in red (Figure S1 in Text S1). Figure 2A shows the trimer structure and only the residues 402–489 of the dark grey protomer are shown. In the orientation shown

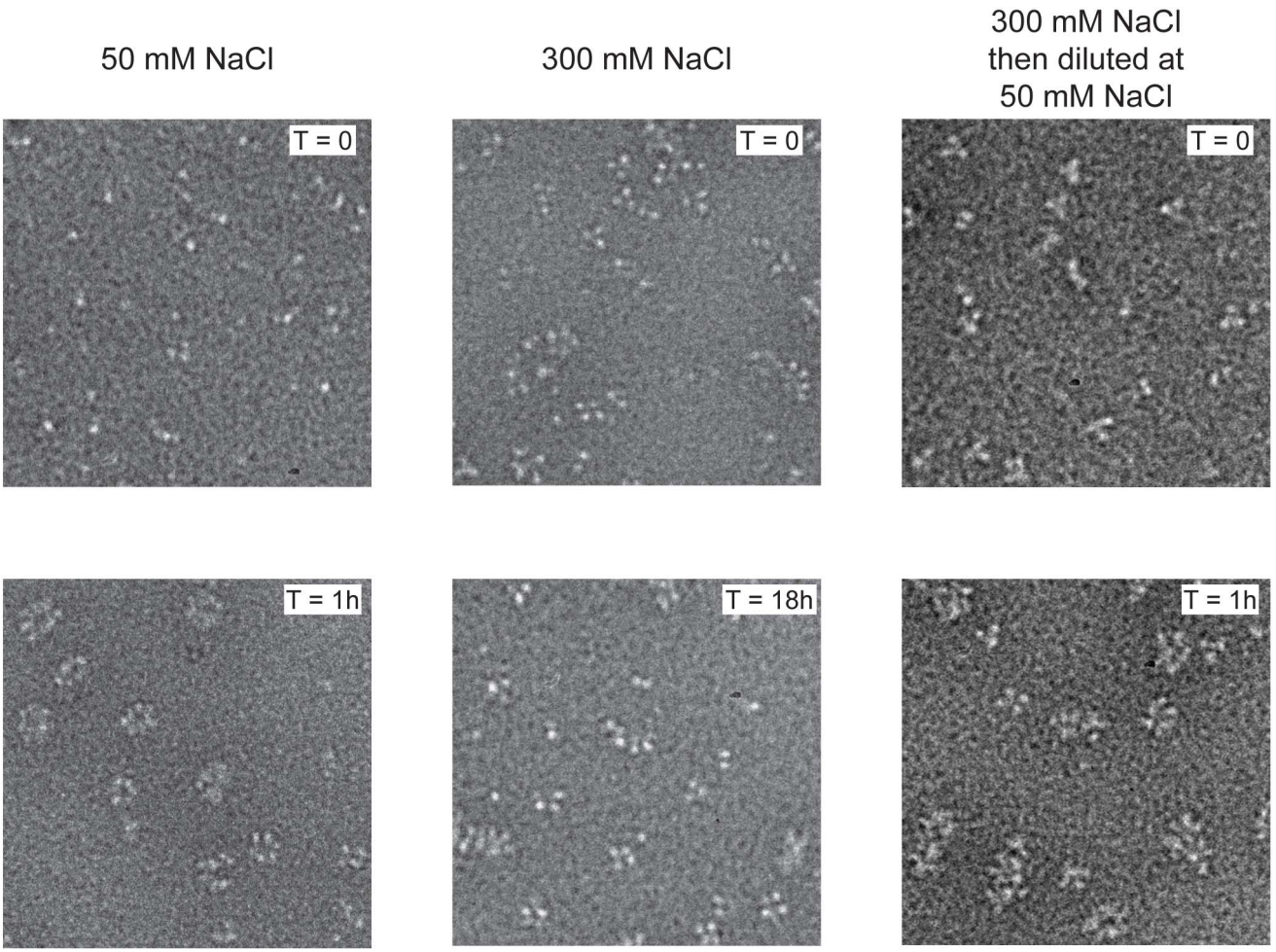


Figure 1. Kinetics of nucleoprotein-RNA ring formation. Negative stain electron micrographs of a 51 nucleotide RNA incubated with monomeric wt NP at 50 mM NaCl (left), oligomeric wt NP at 300 mM NaCl (middle) or oligomeric NP first at 300 mM NaCl which was subsequently diluted to 50 mM (right). The incubation times are indicated on the micrographs. T = 0 indicated the moment of mixing the RNA and NP and taking the first sample after 10–15 seconds. All magnifications are the same, the bar corresponds to 20 nm.
doi:10.1371/journal.ppat.1003275.g001

residues 386–401, the green part, are behind its own protomer and, thus, not visible. In the wt trimer (Figures 2A, 3D, E and Figure S1 in Text S1), the "green" part of the chain forms a random coil to residue 402 of the exchange domain. Then, residues 402–421 (yellow, figure 2A) of the dark grey protomer form a hairpin that binds in a groove on the surface of the light grey protomer [16,17]. This domain exchange is terminated by α-helix 19 that is also bound to the light grey protomer. Then the chain (in red) loops back to its own protomer starting with α-helix 20. The remainder of the C-terminal domain forms a random coil bound to the surface of the core of NP.

In the monomeric mutant protein (Figures 2B, C, 3A, B and Figure S1 in Text S1), residues 386–390 form a β-strand (β7) and residues 392–407 form a flexible loop that points towards the RNA binding surface. Then, from the mutated R416 up to α-helix 19 the chain binds in the groove on its own surface making very similar interactions as the domain exchange hairpin of the wt trimer. α-Helix 19 binds at exactly the same place as in the trimer structure but on its own protomer rather than on its neighbour. α-Helix 20 binds in exactly the same position to its own protomer in the monomer structure and in the trimer structure (rmsd of 0.5 Å between the two structures). The remaining "red" strand goes towards β-strand 7 to fold into β-strand 9 forming a two stranded β-sheet.

Figure 3 compares the H1N1 R416A monomer structure with the trimer structures of H1N1 and H5N1. The major difference between the H1N1 and H5N1 trimer structures is that the exchange domains don't bind to the same neighbours and, thus, point into different directions (Figure 3D and G). For trimeric H1N1 the last visible C-terminal residue is 489 and for H5N1 this is residue 496 and in both structures the C-termini point away from the RNA binding surface. However, in the structure of R416A, the chain is visible until the penultimate residue and points into the RNA binding surface, reducing the space for RNA binding and changing the electrostatic characteristics of this surface (compare figures 3C, F and I, blue is positively charged). At the other side of the monomer the 392–407 loop also points into the RNA binding surface. Although we could not model this loop because its density is missing, the RNA binding surface may be reduced from both sides.

Mutations in the domain exchange loop and in the surface groove in which the loop binds do not only influence the stability of the trimer but also the stability of the monomer. In the trimer structures, R416 makes an ionic bond with E339 and it was

Table 1. Data collection and refinement statistics.

	NP R416A
Data Collection	
X-ray source	ESRF ID14-EH4
Wavelength (Å)	0.97625
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	165.4, 285.4, 118.3
Resolution range[(a)] (Å)	40–2.7 (2.82–2.7)
Completeness[(a)] (%)	99.4 (99.7)
Rsym I[(a)(b)] (%)	9.9 (69.2)
I/σI	11.7 (1.9)
Total reflections[(a)]	254,298 (32,620)
Unique reflections[(a)]	77,415 (10,163)
Multiplicity[(a)]	3.3 (3.2)
Refinement statistics	
R-factor (%)	20.0
R_{free} (%)	24.9
Mean *B*-factor	
Protein	50.8
Ligand	50.4
No. Residue atoms	11 051
No. Ligand atoms	3
Ramachandran plot	
Allowed regions (%)	98.9
Disallowed regions (%)	1.1
R.m.s. deviations	
Bond lengths (Å)	0.015
Bond angles (°)	1.715

[(a)]Values in parentheses are for the highest-resolution shell.
[(b)]*R*sym $(I) = [\sum_{hkl}\sum_i|<I_{hkl}> - I_{hkl,i}|]/[\sum_{hkl}\sum_i|I_{hkl}|]$, where *i* is the number of reflection *hkl*.
doi:10.1371/journal.ppat.1003275.t001

assumed that R416A and E339A formed monomers because this bond in the trimer was disrupted [16,23,26]. In the R416A structure, E339 makes hydrogen bonds with R461 and the mutated R416A points towards R461; if the wt monomer structure were the same as the R416A structure there could be a clash between arginines 416 and 461. Therefore, the R416A mutation may both stabilise the monomer and destabilise the trimer.

Because we did not succeed in crystallising monomeric wt NP, we used a variety of biophysical methods to compare the wt and mutant monomers. Monomeric R416A and monomeric wt NP stabilised at low salt have an identical sedimentation behaviour with an $S_{20,w}$ of 4.3 S [20]. This S-value corresponds to a hydrodynamic radius of 3.3 nm. This means that the exchange domain of the wild type monomer does not extend out in solution in the same conformation as in the oligomeric form but must be close to the core of the protein like in the R416A structure. The circular dichroism spectrum of the trimer form of the wt protein is identical to that of the obligate Y148A trimer, both at 50 and 300 mM NaCl (Figure 4A). The spectra could not be measured below 200 nm because of the high NaCl concentration and the secondary structure content could not reliably be determined. The spectra had a clear α-helical signature with a strong minimum at 222 nm. All spectra of monomeric NP (wt NP at 50 mM salt and R416A mutant NP at both 50 and 300 mM salt) were identical. They showed the same value at 222 nm as oligomeric NP but had a more pronounced minimum at 207 nm (Figure 4B and C). Each of these experiments was repeated with at least three independent protein preparations and always gave the same results. This indicates that all monomeric proteins have the same secondary structure content, slightly different from the oligomeric proteins. We also used CD to determine the melting temperature by heating the proteins while measuring the helical content at 208 and 222 nm. Again, the wt and mutant monomers showed an identical behaviour with an apparent denaturation midpoint value of 43.5 ± 0.5°C (average of 6 independent measurements) whereas the wt trimer denatured at 47 ± 0.5°C. All results for the monomeric wt and mutant proteins are identical and different from those for the oligomeric form thus, it is likely that monomeric wt NP has a similar structure as the R416A mutant NP.

Because only monomeric NP forms RNP-like NP-RNA structures, NP newly produced in the infected cell should remain monomeric and free of nucleic acid until binding to viral RNA. However, accumulation of NP in the nucleus at physiological temperatures would likely result in the formation of trimers/tetramers and the affinity of monomeric NP for RNA is very high; Kd = 41 nM [20]. NP was reported to be phosphorylated on several serine residues [27,28]. Phosphorylation was found to be a highly dynamic process and phosphorylated NP was also detected in RNPs [29]. We analyzed the NP isolated from influenza virus A/PR/8/34 RNPs by Mass Spectrometry. Several phosphorylated serines were detected, but only one residue is conserved in all A and B viruses: Serine 165. In the monomeric structure, this serine is situated at the interface between the two lobes of the core of NP, between α-helix 19 and residue Phe488, close to the C-terminus (Figure 5A). There is space for a phosphate group that could be stabilised by residue R267 nearby. In the trimer structure there does not seem to be enough space to locate a phosphate group because it would clash with main chain Ser407 and the negatively charged side chain of Glu405. We produced the S165D mutant mimicking the phosphorylation of Ser165. The mutant protein was monomeric at low and high salt when investigated by EM (Figure 5C), had the CD signature of a monomeric protein (Figure 4D), had a hydrodynamic radius of 3.3 nm and a thermal denaturation midpoint as determined by CD of 41 ± 0.5°C and, thus, resembled the monomeric form of NP. The Kd for RNA of S165D was determined by surface plasmon resonance (SPR) using a 24 nt oligoribonucleotide [20] and by a filter binding assay using a radioactive panhandle RNA [1]. The SPR analysis gave a Kd of 730 nM (Figure S2 in Text S1); much higher than the Kd for wt monomeric NP of 41 nM [20]. The Kd derived from the filter binding experiment was in the same order of magnitude; 250 nM compared with 30 nM for wt recombinant NP and for NP isolated from virus (Figure S3 in Text S1). Thus, the affinity of the mutant was 10–20 times lower than that of wt protein. However, when studied by DLS at higher concentrations, the S165D mutant showed a much higher cooperativity upon RNA binding than wt NP. S165D NP polymerisation onto a 24-mer RNA oligonucleotide reached a plateau in 30 minutes compared to 2 hours for wt NP (Figure 5B). The same observation was made for the formation of NP-RNA rings as followed by negative staining EM. NP-RNA complexes formed immediately upon mixing of S165D NP with a 51 nt RNA oligo [20] (Figure 5C, T = 0) whereas complex formation with wt NP was slower (Figure 1). Thus, although the Kd for RNA is lower for the S165D mutant than for wt NP, the kinetics of assembly on RNA is more rapid.

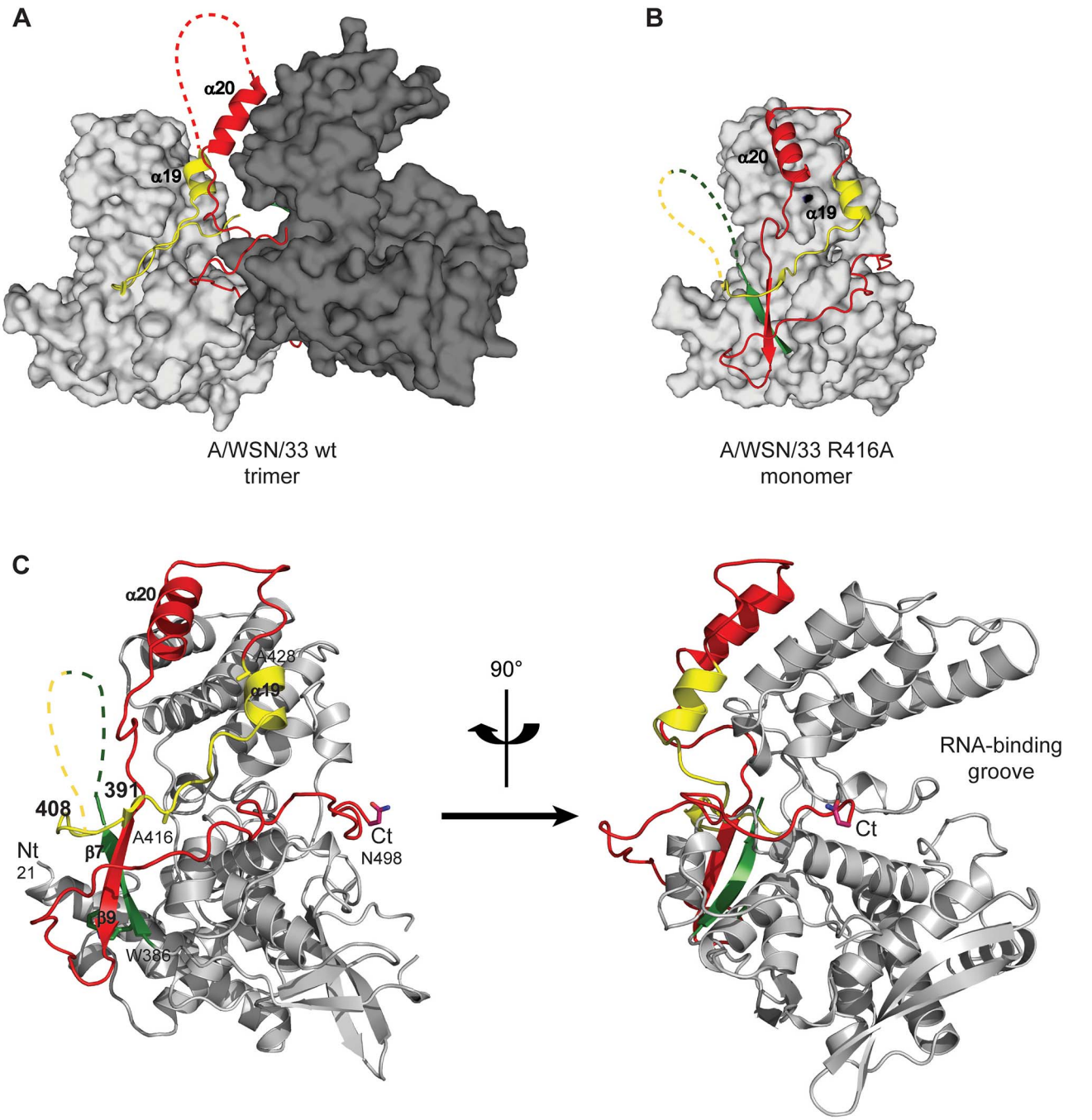


Figure 2. Structure of monomeric NP and comparison with the trimer structure. For the 4 panels residues 21–385 that have an identical structure between the monomer and the trimer are shown in grey, 386–401 in green, 402–428 in yellow and 429–498 in red. **A**. Two NP protomers in the trimer structure (PDB ID code 2IQH). The two bodies of the proteins (in atomic surface representation) are shown light and dark grey and for clarity only the exchange domain (in yellow) and C-terminal sequences (in red) of the dark grey protomer are shown. Residues 386–401(green) are behind the core of the dark grey protomer and not visible in this orientation. **B**. Monomer structure, the green, yellow and red parts are bound to their own protomer. **C**. Ribbon diagrams of two views of a single protomer in the monomer structure using the same colours as in A and B.
doi:10.1371/journal.ppat.1003275.g002

Discussion

Like for other negative strand RNA viruses, here we argue that influenza virus NP also exists in a monomeric form that is free of RNA and that only monomeric NP can form virus-like NP-RNA complexes. The structure of the monomer can perhaps best be described as self-inhibited in which the exchange domain that is involved in trimer formation takes up equivalent positions on its own core rather than on the core of a neighbouring protomer. A flexible loop formed by residues 392–407 and the C-terminus point towards the RNA binding surface which may be the reason for the lowered affinity for RNA, in particular for the obligate monomeric mutant R416A [20,22,24]. Wild type monomeric NP can bind oligonucleotides of 8 residues while remaining monomeric. However, when the RNA binding site is saturated with 24+ nucleotides [30,31], the monomer oligomerises [20], possibly

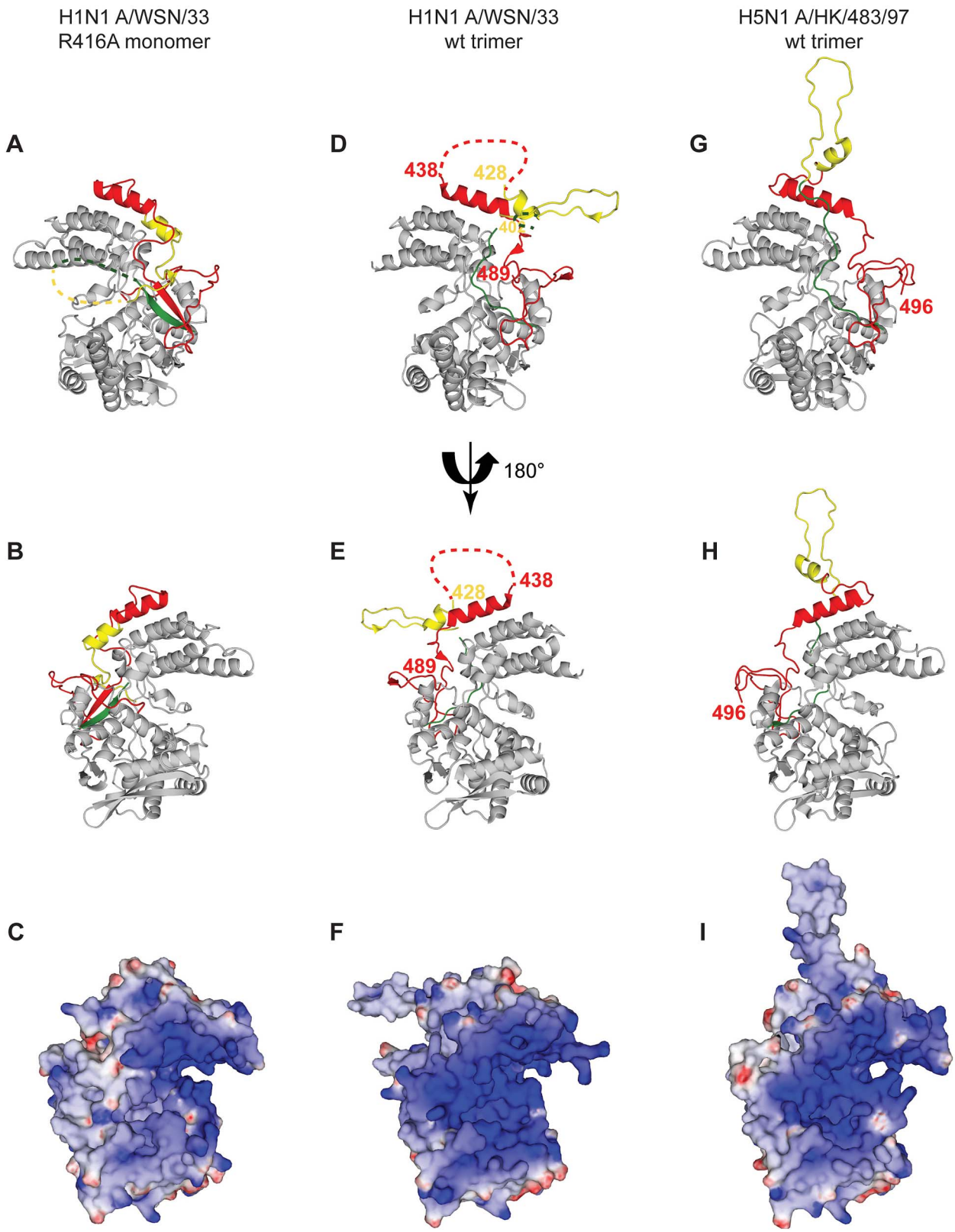


Figure 3. Comparison of the R416A monomer structure with the H1N1 and H5N1 wt trimer structures. A and B. Ribbon diagrams of two views (180° turn) of the monomer and **C** representation of the electrostatic surface of the monomer (from −10 kT/e (red) to +10 kT/e (blue)) calculated using DelPhi [52]. The views in **B** and **C** are the same. **D, E, F.** Similar representations for the wt H1N1 trimer structure (PDB ID code 2IQH). The beginning and end of the flexible loop between helices 19 and 20 are indicated as is the last visible residue at the C-terminus, residue 489. **G, H, I.** Similar representation for the H5N1 trimer structure (PDB ID code 2Q06) with last residue 496. Note that the exchange domains of the two trimeric structures point into different directions because these are bound to different neighbours in the trimer. Also note that the RNA binding surface of the

monomer is less positively charged (less "blue") than the trimeric structures due to the fact that the C-terminus is bound onto this surface whereas the C-terminal ends of the trimeric structures point away from this surface.
doi:10.1371/journal.ppat.1003275.g003

because the RNA pushes out secondary structure elements that stabilise the monomer. The stability of the wt monomer is enhanced by post-translational modification of serine 165. The biochemical behaviour of the monomeric S165D mutant is different with a lowered affinity for RNA but with an enhanced polymerisation at RNA concentrations above the Kd compared to wt monomeric NP. Recently, the lab of Ervin Fodor published the phosphoproteome of influenza A and B viruses [32]. Among various sites they also observed phosphorylation of S165. Recombinant virus with a S165A mutation could not be recovered indicating the importance of this phosphorylation site for the activity of the protein. Using RNP reconstitution experiments in 293 T cells they could measure transcription and replication for the S165A but not for a S165E mutant. As mentioned above the Ser to Asp mutation can be accommodated into the monomer structure but a mutation to Glu would lead to steric hindrance. While this paper was under review, two papers were published describing the structure of the intact viral RNP [33,34]. Both papers describe that the distance between the NP protomers in the RNP is larger than the distance between the protomers in the trimers that are free from RNA [16] (Figure S4 in Text S1). This could imply that the insertion of the exchange domain into a neighbouring NP protomer in the RNP is not equivalent to that in the RNA-free trimer and this may be related to our findings that the S165D mutant does not form trimers but can polymerize onto RNA and that trimers can bind RNA but do not form NP-RNA rings.

Influenza NP binds to RNA without sequence specificity [1]. The only RNA sequence specificity for influenza virus proteins seems to reside in the polymerase that binds 5′ viral RNA and 5′-3′ panhandle structures [35–39]. In a recent model for influenza virus replication it was suggested that soluble, newly produced polymerase binds to the newly replicated 5′ end after which NP polymerises onto the elongated replicate [40,41]. NP binds to the polymerase with a loop containing residues R204, W207 and R208 that is disordered in both the trimer and the monomer structures [42]. The mobility of this loop was recently suggested to influence RNA binding affinity [43]. It is possible that binding of NP to the polymerase leads to opening up of the RNA binding site after which the NP binds cooperatively to the primed NP-RNA site. Phosphorylation and de-phosphorylation of NP probably plays a regulatory role in vRNA encapsidation.

Because of its importance during the viral life cycle, NP is widely recognised as an antiviral target. Through high-throughput testing Kao *et al.* and Su *et al.* identified "nucleozin" that has antiviral activity and appears to aggregate NP in the cell leading to interference with nuclear import [44,45]. A later crystal structure and mutational analysis showed that two NP trimers stick together into a hexamer by six nucleozin-derived molecules, with two complementary sites per NP protomer [46]. The crystal structure

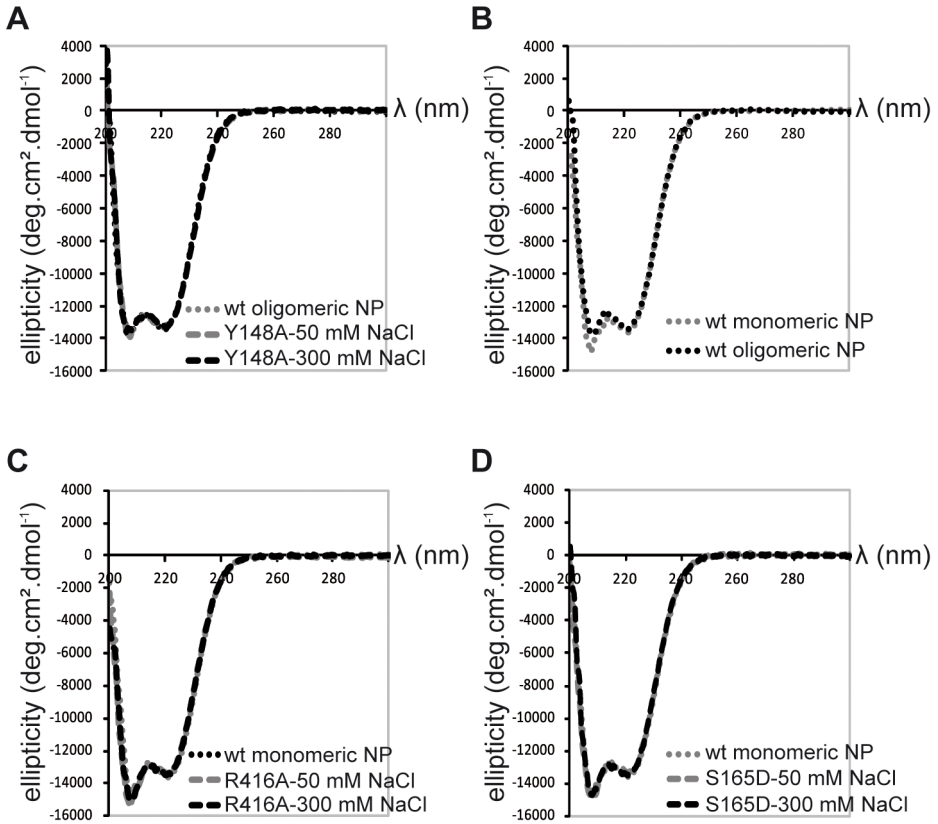


Figure 4. Circular dichroism on NP trimers and monomers. A. Trimeric wt and obligate trimeric Y148A mutant NP. **B.** Comparison of wt trimeric and monomeric NP. **C.** Monomeric wt and R416A NP. **D.** Monomeric wt and S165D NP.
doi:10.1371/journal.ppat.1003275.g004

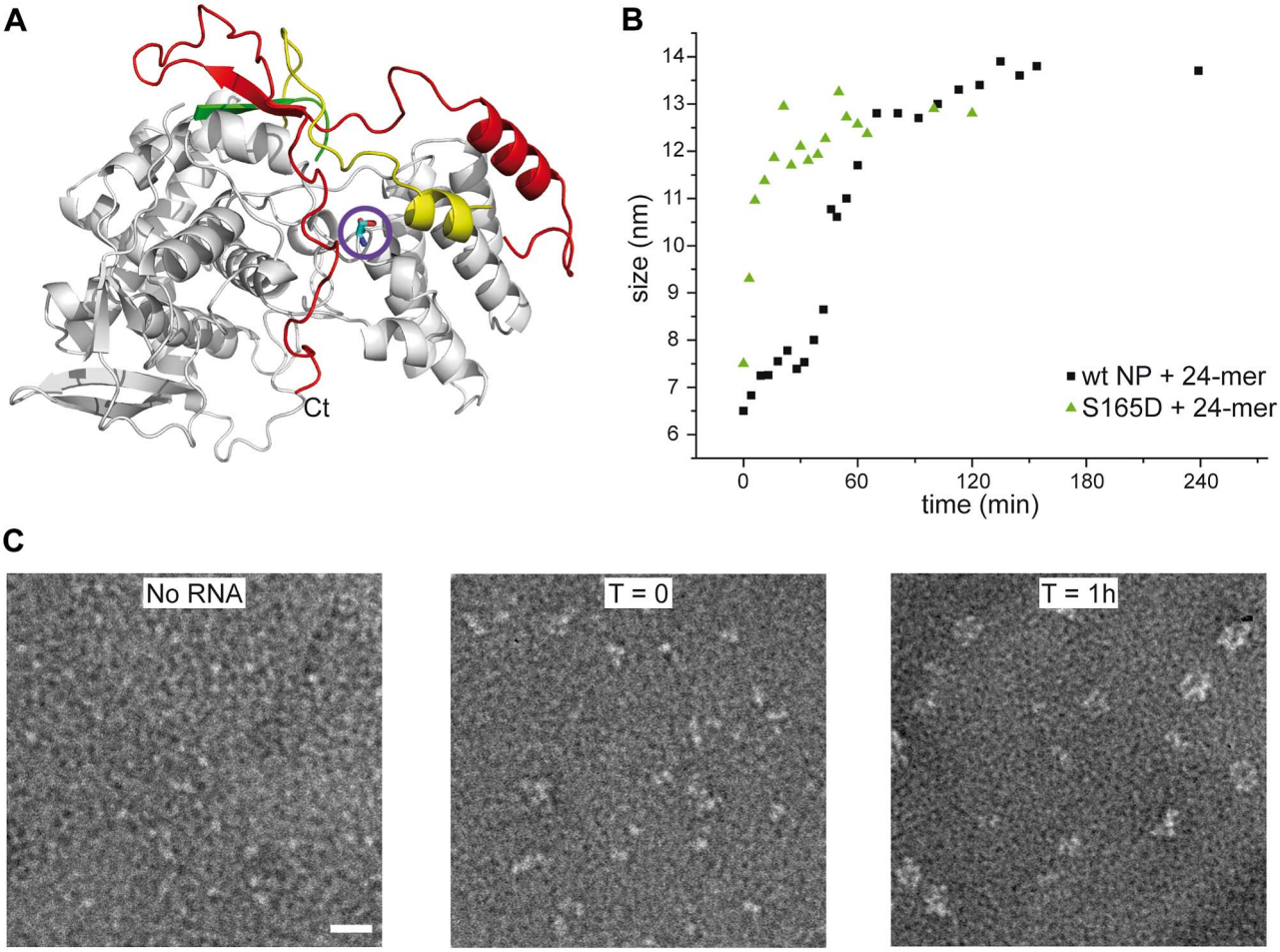


Figure 5. S165D mutant NP. A. Ribbon diagram in the same colors as in Figure 2 with serine 165 as a stick model encircled. **B.** Comparison of the kinetics of the formation of NP-RNA complexes using a 24 nt ribonucleotide by wt monomeric NP and the S165D mutant monitored by dynamic light scattering. **C.** Kinetics of formation of NP-RNA rings by the S165D mutant using a 51 nt ribonucleotide as followed by negative staining EM. In the T = 0 panel, with a sample preparation immediately after mixing the protein and the RNA, NP polymerisation onto the RNA can already be distinguished.
doi:10.1371/journal.ppat.1003275.g005

of the trimer has also been used for structure-aided drug design [26]. Peptides derived from the exchange domain (residues 402–428) interfere with polymerisation of NP and viral replication and molecules designed to interfere with the E339-R416 salt bridge have antiviral effect. The structure of the NP monomer presented here may serve for the design of new antiviral molecules. In particular, drugs that stabilise the monomer should have an antiviral effect.

Materials and Methods

Protein expression and purification of wt and mutant NP

The full-length NP gene of the H1N1 strain A/WSN/33 with a 6-His-tag at its C-terminal end was cloned in the pET22 vector (Novagen) under the control of a T7 promotor. The R416A, Y148A and S165D mutations were introduced by using *PfuUltra* DNA polymerase with the QuikChange II site-directed mutagenesis kit (Stratagene). *Escherichia coli* BL21 (DE3) cells carrying the plasmids were induced 4 hours by adding 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37°C or 12 h at 28°C (mutant proteins) and collected by centrifugation. The pellet was resuspended and sonicated in a lysis buffer composed of 50 mM Tris at pH 7.4 with 300 mM NaCl, 15 to 30 mM imidazole, 1 M NDSB201 (Sigma), 5 mM β-ME and 5 mM $MgCl_2$. The protein was purified by Ni^{2+} affinity chromatography (Ni-NTA, Qiagen) followed by a Heparin column (GE Healthcare). The protein was then dialyzed against 20 mM Tris pH 7.4 and 50 mM or 300 mM NaCl. The last purification step was size-exclusion chromatography using a Superdex S200 column. The protein was eluted in high or low salt according to the required polymerisation state. The protein concentration was determined by using the extinction coefficient $\varepsilon = 55500\ M^{-1}.cm^{-1}$ at 280 nm.

Crystallization, structure determination and refinement

The R416A mutant protein was crystallized by vapor diffusion using the sitting drop method. The crystals were obtained in 0.1 M Hepes pH 7.5, 1.2 M potassium sodium tartrate with a protein concentration of about 10 mg/ml. Data were collected at the ESRF (beamline ID14-4) and processed with the XDS package [47,48]. The structure was solved by molecular replacement using the wild-type H1N1 nucleoprotein structure (PDB ID code 2IQH) as a model. Model building and refinement were performed using CCP4i suite program for crystallography (MOLREP, REFMAC, COOT) [49]. The coordinates have been deposited in the Protein

Data Bank under PDB ID code 3zdp. The protein structure figures were made using PyMOL [50].

Electron microscopy

Samples were applied between a carbon and a mica layer and negatively stained with 2% (w/v) sodium silicotungstate (pH 7.0). The carbon film was covered by a copper grid and air dried. Micrographs were recorded with a JEOL 1200 EX II microscope at 100 kV with a nominal magnification of 40,000×. Micrographs were taken on a 2K×2K CCD camera (Gatan Inc.). The RNA binding kinetics was performed with a protein solution at 100 μM and an RNA of 51 nucleotides [20] with a final ratio NP/RNA of 3/1. For the time points for EM analysis a fraction of the mix was diluted to have a final protein concentration between 10 and 20 μg/ml. The dilutions were made with a buffer with the same salt concentration except for the test of salt dilution on the monomer-oligomer equilibrium where the dilution was done with 20 mM Tris-HCl pH 7.4 in order to reduce the salt concentration from 300 to 50 mM NaCl.

Dynamic light scattering

The measurements were performed on a Malvern nanosizer instrument thermostated at 20°C, in 20 mM Tris pH7.5, 50 or 300 mM NaCl according to the sample. The protein concentrations were in a range of 5 to 50 μM and for each sample, 12 to 18 scans were averaged. All experiments were repeated at least 3 times. The scattering intensity data were processed using the instrumental software to obtain the hydrodynamic radius and the size distribution of scatterers in each sample.

Circular dichroism

Protein samples were centrifuged at 16000× g and the protein concentration in the supernatant was adjusted to 0.5 mg/ml. CD experiments were performed on a JASCO-810 spectrometer at 20°C with a 1 mm path-length quartz cell using a bandwidth of 1 nm, an integration time of 1 second and a scan rate of 50 nm/min. Each spectrum is the average of 10 scans. All spectra were corrected by subtracting the buffer spectrum acquired under the same conditions. All data were normalized to mean residue ellipticity. The melting temperature was obtained for each protein sample by measuring the CD signal at 208 nm and 222 nm from 20°C to 80°C every 2°C. At 80°C all proteins precipitated. The value of the denaturation midpoint (°C) is the average of at least 3 measurements on independent protein preparations.

Determination of the dissociation equilibrium constant for binding of NP to RNA

Surface plasmon resonance experiments. The RNA binding kinetics were performed on a Biacore 3000 apparatus using streptavidin coated sensor chips (SA, Biacore) prepared as indicated by the manufacturer. Immobilization of the biotinylated 24 nt oligoribonucleotide [20] (100 to 200 RU) on the sensor chip was carried out in PBS buffer. The oligoribonucleotide was denatured at 80°C and renatured slowly at room temperature for one hour before each experiment. The S165D mutant was injected at concentrations between 0 and 8 μM. Measurements were conducted at 25°C and samples were injected at 25 μl/min flow rate. One flow cell left blank was used as a reference. To reduce the non-specific response to minimal values, 300 mM NaCl and 0.025% P20 surfactant were added to the running buffer. Because of non-specific polymerization of the S165D mutant NP, the sensorgrams could not be fitted by global analysis to determine the rate constants. The signals at the end of the injection were plotted as the function of the concentration. In order to determine an apparent dissociation equilibrium constant, a linear function that describes a non-specific binding mode was added to the standard Langmuir equation [51].

Filter binding experiments. The filter binding experiments were performed with a radiolabelled 81 nt panhandle RNA as described [1].

Mass spectrometry identification of phosphorylation

NP from egg grown influenza A/PR/8/34 virus was cut out from a Coomassie-stained 1D SDS-PAGE gel and digested with Trypsin (Roche Applied Science). The digest was analyzed by LC-ESI-MS/MS with an Ion trap MS HCT ultra PTM Discovery System (Bruker Daltonics, Bremen, Germany) coupled with a Nano-LC 2D HPLC system (Eksigent, Dublin, CA, USA). A CapRod monolithic C18 column (100 μm×15 cm, Merck, Darmstadt, Germany) was used to separate the peptides. The gradient was 10–70% ACN within 20 min at 300 nl/min flow rate. The top two precursor ions were selected over m/z range from 400 to 1200 for fragmentation. Fragmentation was performed subsequently for 60 ms over 300 to 2000 m/z. The raw data were processed with DataAnalysis (Version 3.1, Bruker, Bremen Germany). The extracted MS/MS data were submitted to MASCOT (version 2.103, Matrix Science, London, UK) in-house server via Biotools (version3.0, Brucker, Bremen, Germany). Proteins were identified by searching the peptide lists against SwissProt. The following parameters were used: Taxonomy: Viruses; Enzyme: trypsin; Max Missed Cleavages: 2; Variable modifications: oxidation (M); carbamidomethyl (C); phospho (ST), phospho(Y); Peptide Mass Tolerance: ±0.5 Da: Fragment Mass Tolerance: ±0.5 Da.

Supporting Information

Text S1 Supporting Figures S1, S2, S3, S4.
(PPTX)

Acknowledgments

We thank Thomas Franz (MPI for Biology of Ageing, Bio-MS Facility, Cologne, Germany) for his expert help in interpreting the mass spectroscopy results and Stephen Cusack (EMBL Grenoble Outstation) for discussions and support. We thank the Partnership for Structural Biology for the excellent structural biology environment.

Author Contributions

Conceived and designed the experiments: SC ASS BD TC RWHR. Performed the experiments: SC LFE ASS CDP FB XL TC. Analyzed the data: SC ASS XL TC RWHR. Wrote the paper: TC RWHR.

References

1. Baudin F, Bach C, Cusack S, Ruigrok RW (1994) Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. Embo J 13: 3158–3165.
2. Iseni F, Baudin F, Blondel D, Ruigrok RW (2000) Structure of the RNA inside the vesicular stomatitis virus nucleocapsid. Rna 6: 270–281.
3. Arnheiter H, Davis NL, Wertz G, Schubert M, Lazzarini RA (1985) Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. Cell 41: 259–267.
4. Honda A, Ueda K, Nagata K, Ishihama A (1988) RNA polymerase of influenza virus: role of NP in RNA chain elongation. J Biochem 104: 1021–1026.

5. Ruigrok RW, Crepin T, Kolakofsky D (2011) Nucleoproteins and nucleocapsids of negative-strand RNA viruses. Curr Opin Microbiol 14: 504–510.
6. Masters PS, Banerjee AK (1988) Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. J Virol 62: 2658–2664.
7. Curran J, Marq JB, Kolakofsky D (1995) An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. J Virol 69: 849–855.
8. Mavrakis M, Mehouas S, Real E, Iseni F, Blondel D, et al. (2006) Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. Virology 349: 422–429.
9. Leyrat C, Yabukarski F, Tarbouriech N, Ribeiro EA, Jr., Jensen MR, et al. (2011) Structure of the vesicular stomatitis virus N(0)-P complex. PLoS Pathog 7: e1002248.
10. Raymond DD, Piper ME, Gerrard SR, Smith JL (2010) Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. Proc Natl Acad Sci U S A 107: 11769–11774.
11. Ferron F, Li Z, Danek EI, Luo D, Wong Y, et al. (2011) The hexamer structure of Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes. PLoS Pathog 7: e1002030.
12. Raymond DD, Piper ME, Gerrard SR, Skiniotis G, Smith JL (2012) Phleboviruses encapsidate their genomes by sequestering RNA bases. Proc Natl Acad Sci U S A 109: 19208–19213.
13. Qi X, Lan S, Wang W, Schelde LM, Dong H, et al. (2010) Cap binding and immune evasion revealed by Lassa nucleoprotein structure. Nature 468: 779–783.
14. Hastie KM, Kimberlin CR, Zandonatti MA, MacRae IJ, Saphire EO (2011) Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3′ to 5′ exonuclease activity essential for immune suppression. Proc Natl Acad Sci U S A 108: 2396–2401.
15. Hastie KM, Liu T, Li S, King LB, Ngo N, et al. (2011) Crystal structure of the Lassa virus nucleoprotein-RNA complex reveals a gating mechanism for RNA binding. Proc Natl Acad Sci U S A 108: 19365–19370.
16. Ye Q, Krug RM, Tao YJ (2006) The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. Nature 444: 1078–1082.
17. Ng AK, Zhang H, Tan K, Li Z, Liu JH, et al. (2008) Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. Faseb J 22: 3638–3647.
18. Ng AK, Lam MK, Zhang H, Liu J, Au SW, et al. (2012) Structural basis for RNA binding and homo-oligomer formation by influenza B virus nucleoprotein. J Virol 86: 6758–6767.
19. Ruigrok RW, Baudin F (1995) Structure of influenza virus ribonucleoprotein particles. II. Purified RNA-free influenza virus ribonucleoprotein forms structures that are indistinguishable from the intact influenza virus ribonucleoprotein particles. J Gen Virol 76 (Pt 4): 1009–1014.
20. Tarus B, Bakowiez O, Chenavas S, Duchemin L, Estrozi LF, et al. (2012) Oligomerization paths of the nucleoprotein of influenza A virus. Biochimie 94: 776–785.
21. Ye Q, Guu TS, Mata DA, Kuo RL, Smith B, et al. (2013) Biochemical and structural evidence in support of a coherent model for the formation of the double-helical influenza a virus ribonucleoprotein. MBio 4: e00467–00412.
22. Elton D, Medcalf E, Bishop K, Digard P (1999) Oligomerization of the influenza virus nucleoprotein: identification of positive and negative sequence elements. Virology 260: 190–200.
23. Chan WH, Ng AK, Robb NC, Lam MK, Chan PK, et al. (2010) Functional analysis of the influenza virus H5N1 nucleoprotein tail loop reveals amino acids that are crucial for oligomerization and ribonucleoprotein activities. J Virol 84: 7337–7345.
24. Boulo S, Akarsu H, Lotteau V, Muller CW, Ruigrok RW, et al. (2011) Human importin alpha and RNA do not compete for binding to influenza A virus nucleoprotein. Virology 409: 84–90.
25. Coloma R, Valpuesta JM, Arranz R, Carrascosa JL, Ortin J, et al. (2009) The structure of a biologically active influenza virus ribonucleoprotein complex. PLoS Pathog 5: e1000491.
26. Shen YF, Chen YH, Chu SY, Lin MI, Hsu HT, et al. (2011) E339…R416 salt bridge of nucleoprotein as a feasible target for influenza virus inhibitors. Proc Natl Acad Sci U S A 108: 16515–16520.
27. Petri T, Dimmock NJ (1981) Phosphorylation of influenza virus nucleoprotein in vivo. J Gen Virol 57: 185–190.
28. Kistner O, Muller K, Scholtissek C (1989) Differential phosphorylation of the nucleoprotein of influenza A viruses. J Gen Virol 70 (Pt 9): 2421–2431.
29. Almond JW, Felsenreich V (1982) Phosphorylation of the nucleoprotein of an avian influenza virus. J Gen Virol 60: 295–305.
30. Compans RW, Content J, Duesberg PH (1972) Structure of the ribonucleoprotein of influenza virus. J Virol 10: 795–800.
31. Ortega J, Martin-Benito J, Zurcher T, Valpuesta JM, Carrascosa JL, et al. (2000) Ultrastructural and functional analyses of recombinant influenza virus ribonucleoproteins suggest dimerization of nucleoprotein during virus amplification. J Virol 74: 156–163.
32. Hutchinson EC, Denham EM, Thomas B, Trudgian DC, Hester SS, et al. (2012) Mapping the phosphoproteome of influenza a and B viruses by mass spectrometry. PLoS Pathog 8: e1002993.
33. Arranz R, Coloma R, Chichon FJ, Conesa JJ, Carrascosa JL, et al. (2012) The Structure of Native Influenza Virion Ribonucleoproteins. Science 338:1634–1637.
34. Moeller A, Kirchdoerfer RN, Potter CS, Carragher B, Wilson IA (2012) Organization of the Influenza Virus Replication Machinery. Science 338:1631–1634.
35. Fodor E, Pritlove DC, Brownlee GG (1994) The influenza virus panhandle is involved in the initiation of transcription. J Virol 68: 4092–4096.
36. Tiley LS, Hagen M, Matthews JT, Krystal M (1994) Sequence-specific binding of the influenza virus RNA polymerase to sequences located at the 5′ ends of the viral RNAs. J Virol 68: 5108–5116.
37. Hagen M, Chung TD, Butcher JA, Krystal M (1994) Recombinant influenza virus polymerase: requirement of both 5′ and 3′ viral ends for endonuclease activity. J Virol 68: 1509–1515.
38. Cianci C, Tiley L, Krystal M (1995) Differential activation of the influenza virus polymerase via template RNA binding. J Virol 69: 3995–3999.
39. Klumpp K, Ruigrok RW, Baudin F (1997) Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. Embo J 16: 1248–1257.
40. Jorba N, Coloma R, Ortin J (2009) Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. PLoS Pathog 5: e1000462.
41. Resa-Infante P, Jorba N, Coloma R, Ortin J (2011) The influenza virus RNA synthesis machine: advances in its structure and function. RNA Biol 8: 207–215.
42. Marklund JK, Ye Q, Dong J, Tao YJ, Krug RM (2012) Sequence in the influenza A virus nucleoprotein required for viral polymerase binding and RNA synthesis. J Virol 86: 7292–7297.
43. Tarus B, Chevalier C, Richard CA, Delmas B, Di Primo C, et al. (2012) Molecular dynamics studies of the nucleoprotein of influenza A virus: role of the protein flexibility in RNA binding. PLoS One 7: e30038.
44. Kao RY, Yang D, Lau LS, Tsui WH, Hu L, et al. (2010) Identification of influenza A nucleoprotein as an antiviral target. Nat Biotechnol 28: 600–605.
45. Su CY, Cheng TJ, Lin MI, Wang SY, Huang WI, et al. (2010) High-throughput identification of compounds targeting influenza RNA-dependent RNA polymerase activity. Proc Natl Acad Sci U S A 107: 19151–19156.
46. Gerritz SW, Cianci C, Kim S, Pearce BC, Deminie C, et al. (2011) Inhibition of influenza virus replication via small molecules that induce the formation of higher-order nucleoprotein oligomers. Proc Natl Acad Sci U S A 108: 15366–15371.
47. Kabsch W (1993) Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. Journal of Applied Crystallography 26: 795–800.
48. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr D Biol Crystallogr 66: 133–144.
49. Collaborative Computational Project n (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50: 760–763.
50. DeLano WL (2002) The PyMOL Molecular Graphics System. San Carlos, CA, USA: DeLano Scientific.
51. Mendel CM, Mendel DB (1985) 'Non-specific' binding. The problem, and a solution. Biochem J 228: 269–272.
52. Rocchia W, Sridharan S, Nicholls A, Alexov E, Chiabrera A, et al. (2002) Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. J Comput Chem 23: 128–137.

ARTICLE

doi:10.1038/nature14009

Structural insight into cap-snatching and RNA synthesis by influenza polymerase

Stefan Reich[1,2]*, Delphine Guilligay[1,2]*, Alexander Pflug[1,2]*, Hélène Malet[1,2], Imre Berger[1,2], Thibaut Crépin[2], Darren Hart[1,2], Thomas Lunardi[1,2], Max Nanao[1,2], Rob W. H. Ruigrok[2] & Stephen Cusack[1,2]

Influenza virus polymerase uses a capped primer, derived by 'cap-snatching' from host pre-messenger RNA, to transcribe its RNA genome into mRNA and a stuttering mechanism to generate the poly(A) tail. By contrast, genome replication is unprimed and generates exact full-length copies of the template. Here we use crystal structures of bat influenza A and human influenza B polymerases (FluA and FluB), bound to the viral RNA promoter, to give mechanistic insight into these distinct processes. In the FluA structure, a loop analogous to the priming loop of flavivirus polymerases suggests that influenza could initiate unprimed template replication by a similar mechanism. Comparing the FluA and FluB structures suggests that cap-snatching involves *in situ* rotation of the PB2 cap-binding domain to direct the capped primer first towards the endonuclease and then into the polymerase active site. The polymerase probably undergoes considerable conformational changes to convert the observed pre-initiation state into the active initiation and elongation states.

The influenza virus genome comprises eight segments of single-stranded RNA (viral RNA, or vRNA), each packaged in separate ribonucleoprotein particles (RNPs). Both conserved 3′ and 5′ ends of the vRNA (the promoter) are bound to the RNA-dependent RNA polymerase, and the rest of the pseudo-circularised vRNA is coated with nucleoprotein. The polymerase is a heterotrimer composed of subunits PA, PB1 and PB2 and, in the context of the RNP, it performs the distinct processes of transcription and replication using the same template vRNA (reviewed in refs 1 and 2). Transcription of viral mRNA occurs through a unique process called cap-snatching, in which short capped oligomers, derived from host pre-mRNA, are bound by the PB2 subunit[3–5], cleaved by an endonuclease in the PA subunit[6,7] and then used to prime mRNA synthesis by the PB1 subunit. Stuttering of the polymerase at an oligo-U stretch near the vRNA 5′ end leads to auto-polyadenylation[8]. Thus, translation-competent viral mRNAs are generated without the need for a viral-encoded capping enzyme nor the host poly-adenylation machinery, which is shut down by viral-encoded NS1 protein[9]. By contrast, replication involves unprimed synthesis of an exact, full-length copy of the vRNA into complementary RNA (cRNA) and subsequently the inverse process back to progeny vRNA. Nascent replicates are immediately packaged with nucleoprotein into new viral RNPs (vRNPs) or complementary RNPs (cRNPs). In contrast, viral mRNA is not so packaged but is treated as host pre-mRNA[10] and further spliced (in the case of NS and M segments) and/or exported to the cytoplasm by host cell machineries. Interestingly, cRNPs do not perform transcription in infected cells and may require a second polymerase to replicate[11,12]. Despite many years of study, the mechanism by which RNPs are able to perform these different functions and what determines the type of RNA synthesis that occurs are still obscure. Here we infer, using complementary information from atomic resolution structures of influenza A and B polymerases in complex with the vRNA promoter together with known structures of other viral RNA polymerases, the mechanism by which the polymerase can perform either cap-dependent transcription or unprimed (*de novo*) RNA synthesis. The structures thus open the way to a detailed description of how the influenza transcription/replication machine works in a context-dependent manner.

Structure of FluB polymerase compared with FluA

Full-length heterotrimeric influenza polymerase from B/Memphis/13/03 (FluB) was obtained by expression in insect cells as a self-cleaving polyprotein[13] (Extended Data Fig. 1). Recombinant FluB polymerase was active in the absence of nucleoprotein in cap-dependent transcription and both ApG-primed and, less efficiently, unprimed replication assays using short model vRNAs (Extended Data Fig. 2). Two different crystal forms of FluB polymerase were obtained with consensus promoter sequences for influenza B[14] (Extended Data Table 1). Both contain nucleotides 1–14 from the vRNA 5′ end (5′-pAGUAGUAACAAGAG-3′OH) and either nucleotides 5–18 (FluB1 form) or nucleotides 1–18 (FluB2 form) from the 3′ end (3′OH-UCGUCUUCGUCUCCAUAU-5′OH). The FluB1 form yielded a fully interpretable experimental map (Extended Data Fig. 3a–c) at 3.4 Å resolution, allowing an almost complete model of FluB polymerase to be built (Fig. 1a). The 2.7-Å resolution FluB2 structure, solved by molecular replacement using the FluB1 structure, is extremely well ordered (Extended Data Fig. 3d, e). Owing to crystal contacts, it has the best defined endonuclease domain, which, however, is in the same position as in all other structures. By contrast, the C-terminal two-thirds of PB2 (PB2-C) completely lacks electron density in the FluB2 form (Fig. 1b), although intact PB2 is present in the crystal.

Sequence alignments with bat FluA, the structure of which is described elsewhere[15], show that influenza B/Memphis/13/03 has 36.0 (48.6), 59.5 (71.0) and 37.0 (50.9) per cent amino acid identity (similarity) for PA, PB1 and PB2, respectively (with higher than average conservation in the functionally important regions) (Supplementary Fig. 1). The FluA and FluB polymerase structures and their mode of binding to the vRNA promoter are remarkably similar (Extended Data Fig. 4). However, a striking difference of 70° in the orientation of the PB2 cap-binding domain (Fig. 1c, d) suggests that this domain can rotate *in situ*. Concerning

[1]European Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, CS 90181, 38042 Grenoble Cedex 9, France. [2]University Grenoble Alpes-Centre National de la Recherche Scientifique-EMBL Unit of Virus Host-Cell Interactions, 71 Avenue des Martyrs, CS 90181, 38042 Grenoble Cedex 9, France.
*These authors contributed equally to this work.

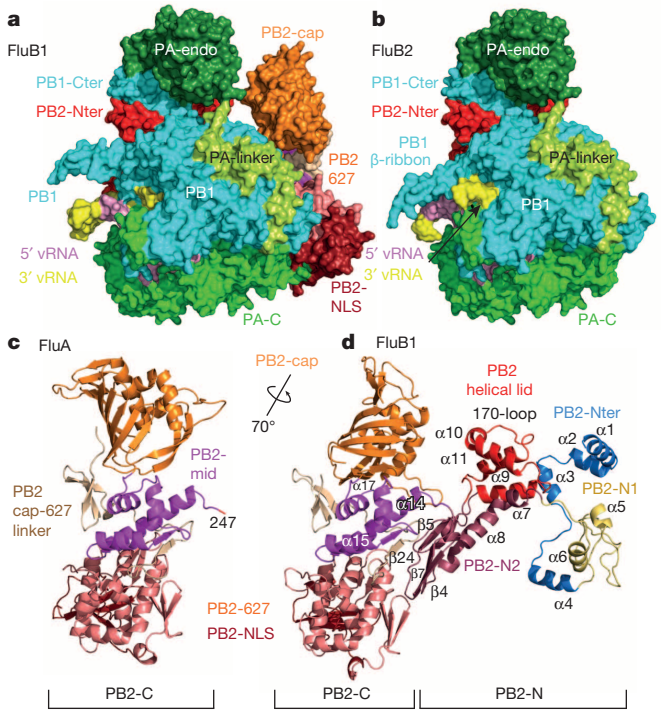


Figure 1 | Structure of influenza B polymerase. **a**, Surface view of FluB1 structure colour-coded according to domain structure (Extended Data Fig. 4) except that PA-C, PB1and PB2-N are uniformly green, cyan and red, respectively. The vRNA 5′ and 3′ ends are pink and yellow, respectively. **b**, Surface view of the polymerase in the FluB2 crystal form that lacks the entire PB2-C domain but includes the full-length 3′ end of the vRNA (black arrow). **c**, Bat FluA PB2-C colour-coded according to domain structure (Extended Data Fig. 4). **d**, The complete PB2 subunit as in the FluB1 crystal form in the same orientation as in **c**, highlighting the 70° difference in orientation of the cap-binding domain.

vRNA binding, all FluA and FluB structures exhibit identical conformations of 5′ nucleotides 1–14 and 3′ nucleotides 5–13 as described elsewhere[15], and most protein–RNA contacts for these regions are conserved between FluA and FluB (Extended Data Table 2). The higher resolution FluB2 structure shows that the protein–RNA interface is highly hydrated with numerous water-mediated protein–RNA interactions. In both the FluB1 and FluB2 structures, the 3′–5′ duplex region of the promoter comprises four base pairs (3′ 10-UCUC-13 with 5′ 11-AG AG-14). In the FluB1 structure, the 5-nucleotide 3′ overhang (14-CA UAU-18) forms a triple-stranded structure at a two-fold crystal contact, including two base triples with the symmetry-related duplex (Extended Data Fig. 5), whereas in the FluB2 structure the RNA does not participate in crystal contacts.

Promoter 3′ end binding

Only in the FluB2 crystal form is the complete 3′ end of the promoter structurally ordered (including nucleotides 1–5, not visible in other structures). The single-stranded 3′ extremity, 1-UCGUCUUCG-9, perhaps unexpectedly, does not enter the polymerase active site but is bound in an alternative location on the surface of the polymerase in an arc conformation, such that U1 is not far from the distal 3′–5′ duplex region (Fig. 2a). Bases 1–4 stack on each other but other bases are bound in individual pockets. Most bases are orientated towards the protein and all except U1 make base-specific RNA–RNA or RNA–protein interactions (Fig. 2a and Extended Data Table 2). All three subunits are involved in binding 3′ nucleotides 6–9, whereas nucleotides 1–5 only interact with PB1. Residues 670–679 of PB1 are involved in binding both extremities of the 3′ end whereas the PB1 β-ribbon interacts with 3′ nucleotides 1–3 (Fig. 2a). The sequence-specific nature of the 3′-end binding and conservation of interacting residues strongly suggests that this binding site is functionally important. This implies that there must be a mechanism for relocating the 3′ end into the PB1 active site during initiation of RNA synthesis. The observed 3′-end conformation is inconsistent with a hook conformation[16], but overall the promoter structure is consistent with that proposed in ref. 17, which suggests that the sequence constraints imposed on the 3′ end by the necessity of almost exact complementary to the 5′ hook would make it appear that the 3′ end would also take a hook conformation.

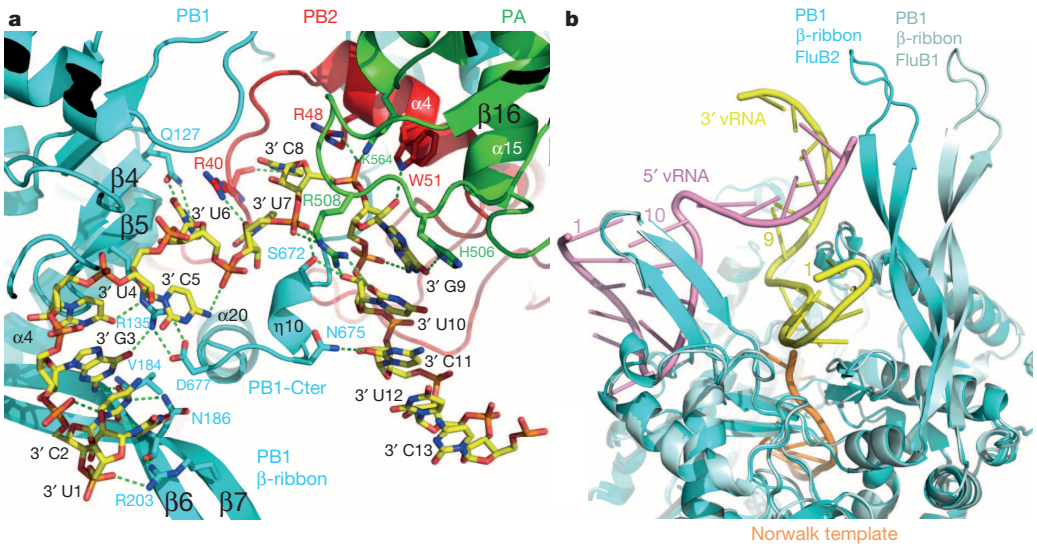


Figure 2 | Promoter 3′-end binding and PB1 β-ribbon flexibility. **a**, Diagram showing RNA–RNA and RNA–protein interactions of the complete 3′ end (nucleotides 1–13, yellow sticks) of the promoter as in the FluB2 structure. For clarity, not all interactions (nor water-mediated interactions) are depicted. All three subunits, PA (green ribbons and residues), PB1 (cyan) and PB2 (red), are involved. Nucleotides 1–9 are single-stranded, and 10–13 form a duplex with the 5′ end (not shown). The PB1 β-ribbon interacts with the proximal part of the 3′ end and PB1-Cter interacts with both proximal and distal nucleotides. Specific RNA–RNA interactions include N2 to OP2 of G9, N4 of C5 to OP2 of U6, O2′ of U1 to OP2 of C2. **b**, Superposition of PB1 from the FluB2 (cyan) and FluB1 (light cyan) structures, showing flexibility of the long β-ribbon. The 3′ (nucleotides 1–13, yellow) and 5′ (nucleotides 1–14, pink) ends of the promoter are as in the FluB2 structure. The 3′ deviates from the path into the PB1 active site that is depicted by the template strand (orange) from the superposed Norwalk template–primer elongation complex (PDB code 3BSO).

In all structures, the unusually long PB1 β-ribbon (residues 177–214) has a role in interacting with the vRNA on the exterior of the polymerase. In the FluB1 structure, the β-ribbon is straight and projects away from the polymerase, its tip (residues 195–196) interacting with crystal symmetry-related RNA (Fig. 2b and Extended Data Fig. 5). In the bat FluA structure, the ribbon is bent towards the polymerase and its central part contacts the duplex region of the promoter, whereas its extremity is disordered (not shown). In the FluB2 structure, the β-ribbon is the most bent and residues 184–186 and influenza-conserved Arg 203 interact with the proximal 3′ end (Fig. 2a, b). These observations show that the PB1 β-ribbon has an affinity for RNA and is flexible. It could therefore have a dynamic role in translocating the RNA into the polymerase from the RNP and/or could mediate interactions with proximal nucleoprotein molecules of the RNP. This hypothesis is supported by fitting of the polymerase–promoter structure to the available electron microscopy map of the mini-RNPs, which predicts the close proximity of the ribbon to nucleoprotein (Extended Data Fig. 6).

Mechanism of replication

Influenza virus polymerase catalyses primer-independent (*de novo*) replication to generate cRNA from vRNA and vice versa. It has been proposed that efficient replication requires nucleoprotein[18] and/or polymerase oligomers[11,12,19]. RNA polymerases that perform *de novo* synthesis generally possess a special 'priming' loop that is thought to stabilize the priming and incoming NTPs in the absence of a priming oligonucleotide. This phenomenon was first structurally characterized for bacteriophage Φ6 polymerase, in which a tyrosine at the extremity of the priming loop stacks on the priming nucleotide[20]. Flavivirus polymerases such as those of hepatitis C virus (HCV) or dengue virus (DENV) also have an aromatic residue (a tyrosine in HCV and a tryptophan for DENV) as putative priming platforms[21]. For PB1, a β-hairpin loop (residues 641–657), structurally analogous to that of HCV, is observed in an ordered configuration in the FluA structure[15] but is disordered in the FluB polymerase structures. The loop tip contains the 648-Ala-His-Gly-Pro motif, conserved in all influenza polymerases. Modelling, on the basis of the Φ6 initiation complex structure, shows that the loop could potentially act as a priming platform to promote correct initiation, with His 649 plausibly interacting with the initial incoming nucleotides (Fig. 3a, b). More details of the active site configuration, which largely involves the canonical polymerase motifs, are given for the FluA structure[15]. A model of the elongation step of influenza polymerase can be obtained by superposing the primer–template complex of poliovirus polymerase[22] on PB1, the high conservation of the polymerase active site ensuring an unambiguous superposition (Fig. 3c, d). The putative priming loop would need to be displaced once elongation starts because it would sterically clash with an emerging template–product duplex (Fig. 3c).

These results lead to the following two observations. As highlighted above, in our structures, the vRNA 3′ extremity does not enter the PB1 active site. However, comparison for instance with the polio template–product complex shows that vRNA 3′ nucleotide 8/9 is at the template tunnel entrance but corresponds to 3′ nucleotide 5/6 in the polio polymerase complex (Fig. 3d). Thus, the 3′ end, on reorientation into the PB1 active site, would have to draw back three nucleotides to initiate at the first position, perhaps concomitantly with breaking of the 3′–5′ duplex region. The mechanism to do this is unclear at present. Interestingly, it has been proposed that vRNA and cRNA initiate replication differently, either synthesizing pppApG at positions 1 and 2 (1-UC) of the 3′ end directly for vRNA, or internally at 4-UC followed by realignment at 1-UC for cRNA[23]. This suggests that the 3′ end is differently positioned in the active site depending on whether it is vRNA or cRNA. This could be because the c3′ end sequence differs at three positions and is one nucleotide longer than the v3′ end before the 3′–5′ duplex region. According to the modelling, it would thus be positioned correctly for internal initiation and the putative priming loop could have a role in this.

The second observation concerns the fact that modelling with the polio template–product elongation complex shows that an extended duplex cannot be accommodated in the cavity of the current structures

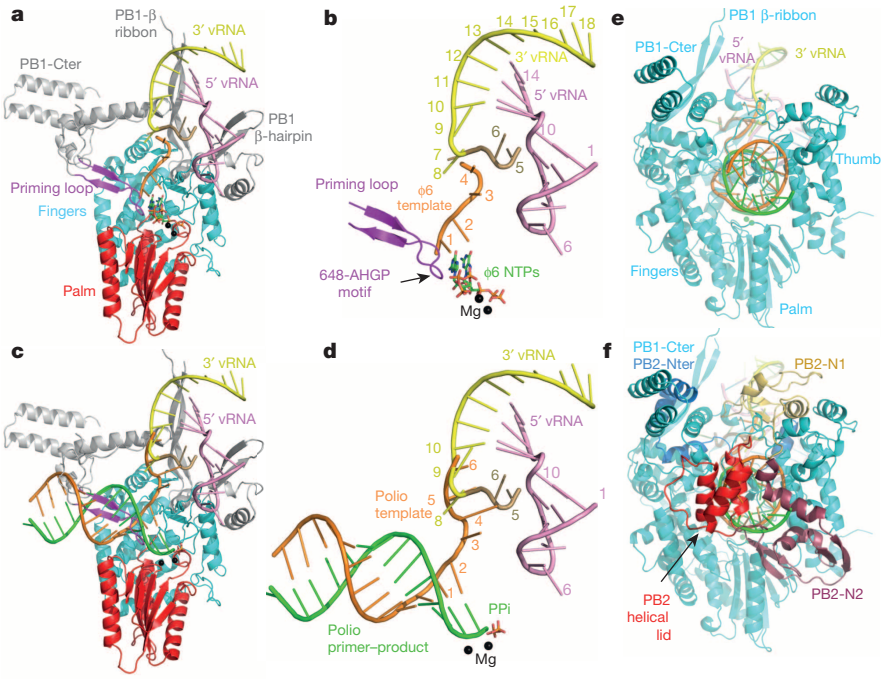


Figure 3 | Model for replication initiation and elongation by influenza polymerase. **a**, FluA PB1 with bound 3′ (nucleotides 5–18, yellow) and 5′ (nucleotides 1–14, pink) vRNA superposed with the Φ6 initiation complex structure (PDB code 1HI0) with template (orange) and two initial incoming NTPs with magnesium (green sticks and black spheres). The PB1 putative priming loop is magenta with the palm (red) and fingers (cyan). The thumb is omitted for clarity. **b**, As in **a** but showing only the PB1 putative priming loop. The influenza conserved 648-Ala-His-Gly-Pro motif at the loop tip could stabilize the initiation complex (the electron density for the His-Gly residues is poor). 3′-end nucleotides 5 and 6 that deviate from the canonical template pathway (orange) are sand coloured. **c**, As in **a** but with the primer (green) and template (orange) RNA from the poliovirus polymerase elongation complex (PDB code 3OL7) after superposition of the polymerase domains. The PB1 putative priming loop clashes with the duplex RNA and therefore must be displaced. **d**, As in **c** but excluding the protein. The influenza vRNA 3′ nucleotide 8/9 is at the template entrance but corresponds to the 3′ nucleotide 5/6 template in the polio virus polymerase complex. **e**, As in **c** but end-on view and with PB1 uniformly coloured cyan. The template–product duplex can be accommodated in the PB1 cavity although the thumb domain is expected to open. **f**, As in **e** but including the PB2-N domain with subdomains coloured as in Extended Data Fig. 4. The product/primer strand (green) can potentially exit/enter (see also Fig. 5) but the template strand is blocked by the PB2 helical lid domain (red).

because of a severe steric clash of the outgoing template strand. This is not primarily due to PB1 (apart from the putative priming loop) (Fig. 3e), but to elements of PB2-N that lie directly on top of the duplex (Fig. 3f). As discussed below, the current structures do provide an open channel into the PB1 active site for a capped primer to initiate transcription but the outgoing template is blocked by helices α8–α10 of the PB2 lid domain. In the case of HCV polymerase, the structure of a product–template duplex complex revealed a 20° rotation of the thumb domain that opened up the product–template binding cavity[24]. In analogy to this, we expect elongation to be accompanied by equivalent conformational changes in which thumb opening could be coupled to displacement of the priming loop and rotation out of the way of the PB2 lid domain. The length of the product–template duplex that is accommodated by influenza polymerase, what causes strand separation (although the PB2 N2 and lid domains are plausibly involved) and which exit path the two strands take are open questions. As discussed below, in the case of cap-dependent transcription, a likely exit pathway for the nascent mRNA is away from the nuclease domain and towards the PB2 627-domain (containing the host-specific amino acid residue 627).

Mechanism of cap-dependent transcription

Cap-snatching is uniquely performed by segmented negative-strand RNA viruses including orthomyxoviruses, bunyaviruses and arenaviruses[25,26]. The PB2 cap-binding and PA endonuclease domains involved in this process were previously characterized structurally and functionally[5–7]. The complete polymerase structures now allow a plausible mechanism for cap-snatching and cap-dependent priming to be proposed. All structures show the PA-Nter endonuclease in the same position and orientation, anchored to the PB1-Cter–PB2-Nter interface (Fig. 1a). By contrast, comparison of the FluA and FluB1 structures after superposition suggests that the PB2 cap-binding domain is able to rotate as a rigid body *in situ*. Whereas PB2 residues before Ile 319 (Ile 321 in FluB) and after Arg 495 (Lys 496 in FluB) align very well, the entire cap-binding domain in between differs in orientation by 70° between the two structures, suggesting that it is flexibly hinged at these anchor points (Fig. 1c, d and Supplementary Video 1). In the FluA structure, the cap-binding site faces the endonuclease active site directly across a solvent channel at a distance of about 50 Å (Fig. 4a, b). This configuration is consistent with a cap-bound host pre-mRNA being cleaved 10–15 nucleotides downstream by the nuclease[27], bearing in mind that the observed cap-binding domain orientation, probably constrained by crystal contacts, is not necessarily optimal for cap-snatching. The observed variability of the primer length[27] would be explained by flexibility in both the cap-binding domain orientation and RNA conformation and possibly the sequence preference of the nuclease cleavage site[28]. Cleaved primers would then be further selected by their efficiency in priming mRNA synthesis, which probably correlates with the complementary to the extremity of the 3′ template[27,29–32]. In the FluB1 structure, the rotated position of the cap-binding domain both shields the bound-capped primer from the endonuclease (Fig. 4c, d) and directs it down into the polymerase RNA catalytic cavity (Fig. 4e). This model is supported by the observation in FluB1 crystals (but not other crystal forms) of residual difference electron density, strongly suggestive of RNA, that descends precisely from the Trp 369–Phe 406 sandwich in the FluB cap-binding site into the throat of the polymerase, which leads to the PB1 active site (Fig. 4e and Extended Data Fig. 7). The nature and origin of this RNA is unclear, making it difficult to fit a precise model. But whatever the RNA origin, its fortuitous occurrence in the FluB1 structure gives a very plausible model of how a capped primer might be configured during transcription initiation. The 424-loop of the cap-binding domain seems to have key roles in channelling the capped primer into the polymerase throat (the integrity of this loop was previously shown to be important for transcription[5]), as well as the projecting amino-terminal end of PB2 lid domain helix α9, and in particular the double prolines 157-Pro-Pro that force the RNA into a ~90° bend (Fig. 4e and Extended Data Fig. 7). The observed RNA density corresponds to about six nucleotides plus

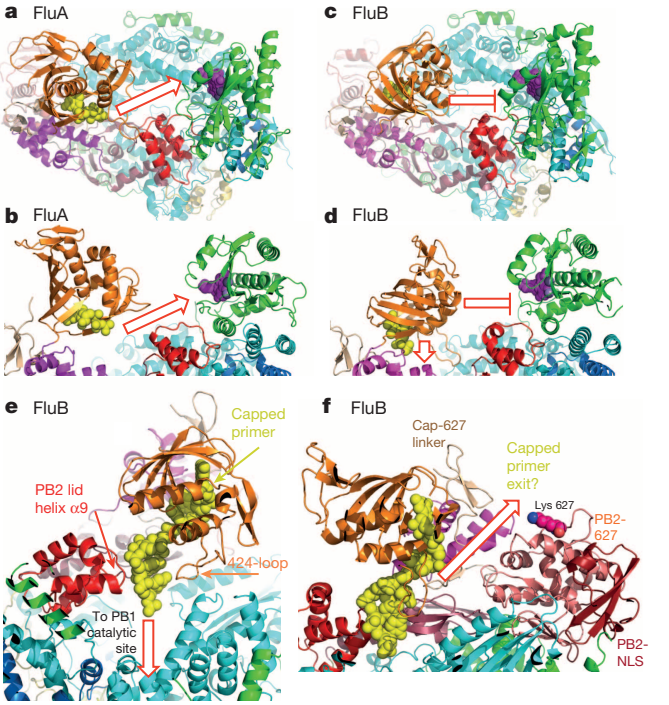


Figure 4 | Cap-snatching and cap-dependent priming of transcription. **a**, Cap-snatching configuration. Top view of the relative orientations in the FluA structure of the cap-binding domain (orange) with bound cap analogue (m^7GTP, yellow spheres, obtained by superposition with PDB code 4CB4) and endonuclease (green) with active site indicated by a bound inhibitor (purple spheres). PA is uniformly green, PB1 cyan and PB2 subdomains coloured as in Extended Data Fig. 4. Cap-bound host pre-mRNA can reach the endonuclease active site unimpeded across a solvent channel (red arrow) at a straight-line distance of around 50 Å (although the cap-binding domain orientation observed is not necessarily optimal for primer cleavage). **b**, As in **a** but side view. **c**, **d**, As in **a** and **b** but for the FluB1 structure. The rotated orientation of the cap-binding domain shields cap-bound RNA from the endonuclease. **e**, The FluB1 cap-binding domain configuration is compatible with cap-dependent priming. Yellow spheres represent model of bound capped primer derived from RNA-like residual electron density (Extended Data Fig. 7). The primer is channeled towards the PB1 active site by the 424-loop of the cap-binding domain and the N-terminal end of PB2 lid domain helix α9. **f**, The putative exit channel for the capped transcript is between the PB2 cap-627 linker and 627-domains towards host-specific residue Lys 627, and away from the nuclease.

the cap, and extends over a straight-line distance of around 26 Å to the bend. The remaining distance to the polymerase active site is around 28 Å, which is compatible with a primer of around 12–14 nucleotides (Extended Data Fig. 7). An overall model of how cap-dependent priming is likely to occur in influenza polymerase is given in Fig. 5.

One can hypothesize about subsequent steps in cap-dependent transcription (Extended Data Fig. 8). Once the capped primer 3′ end engages the vRNA template in the PB1 active site, primer elongation occurs by template-directed nucleotide addition. Further rotation *in situ* of the cap-binding domain could initially accommodate the growing mRNA while still maintaining cap-binding (Extended Data Fig. 8c). However, at some stage the buckling out of the lengthening mRNA would force cap release, which has previously been estimated to be after 11–15 nucleotides[33] (Extended Data Fig. 8d). The transcript would naturally emerge into the basic channel between the cap-binding domain and the cap-627 linker/627-domain in the vicinity of host-specific residue 627 (Fig. 4f), possibly explaining why capped RNA was crosslinked to the 627-domain[34,35]. This exit pathway avoids the endonuclease, consistent with reports that the polymerase protects its own mRNAs from degradation by transiently binding to the conserved AGCAAAGCAGG

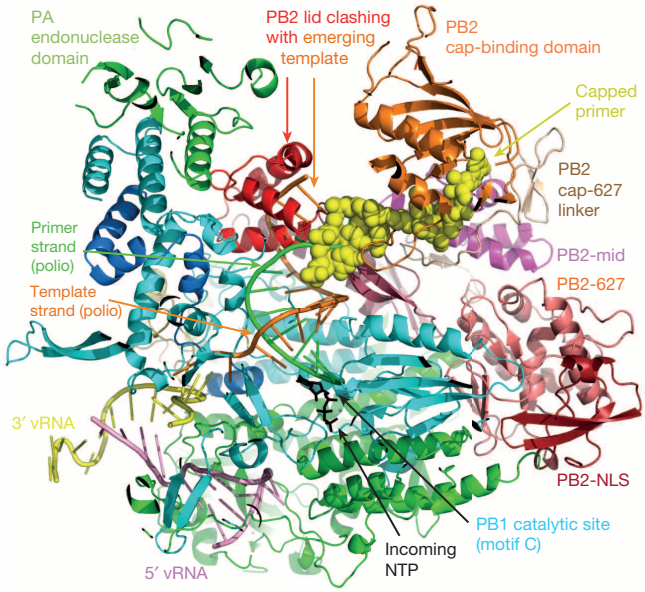


Figure 5 | Model for cap-dependent transcription. The FluB1 structure is superposed with the template–primer (orange/green) duplex and incoming NTP (black sticks) from the poliovirus complex (PDB code 3OL7). PA is uniformly green, PB1 cyan and PB2 subdomains coloured as in Extended Data Fig. 4. The capped RNA primer (yellow spheres) is as in Fig. 4e and connects with the primer strand in the polio complex. The polio template strand connects with the vRNA 3′ end (yellow tube) at the template tunnel entrance. During elongation the emerging template strand would clash with the PB2 helical lid (red), which therefore has to move. When the vRNA template has mostly passed through the polymerase there will be a minimal loop remaining with the tightly bound 5′ hook (pink tube), which will generate the poly(A) tail on the transcript by stuttering on the oligo-U sequence at 17–22 nucleotides from the 5′ end. During elongation, the polymerase will sequester 20–25 nucleotides of the template. In the context of transcription by an RNP, at least this amount of RNA would have to dissociate from nucleoprotein and re-associate after exiting the polymerase.

sequence, which occurs just downstream of the host mRNA-derived primer sequence and is transcribed from the conserved 3′ end of the template[36]. The 627-domain may have a role in this as it has a binding preference for 5′ vRNA-like sequences[37]. When eventually released, the 5′ cap structure itself is bound by the nuclear cap-binding complex and the mRNA subsequently processed by host machinery[10] (Extended Data Fig. 8d). More generally, the same exit pathway could be used for cRNA or vRNA replicates, and the 627-NLS domain (a double domain containing host-specific PB2 residue 627 and the PB2 nuclear localization signal (NLS)) could have a role in their packaging with incoming nucleoprotein into nascent cRNPs or vRNPs[38].

Concerning auto-polyadenylation of viral mRNA by the polymerase, the tight binding of the hook at the 5′ end of the template is thought to cause stuttering at the oligo-U stretch typically 17–22 nucleotides from the 5′ end, resulting in the addition of several adenosine residues[8,17]. Because a minimum of ten 5′ nucleotides are required to form the 5′ hook, and, on the basis of the structural alignment with the polio polymerase primer–template complex, a minimum of seven extra nucleotides is required to reach the site of nucleotide addition (Fig. 3e), the crystal structure is fully compatible with the proposed polyadenylation model (Extended Data Fig. 8d).

Conclusions

The FluA and FluB polymerase structures presented here seem to be in an inactive pre-initiation state requiring relocation of the 3′ end into the PB1 active site before RNA synthesis can begin. However, we think the observed 3′-end binding site on the polymerase surface could have functional importance in, for instance, providing an additional docking site for the 3′ end (on the same polymerase or a different polymerase) after it has been copied and exited from the active site. This would be an efficient way to allow several rounds of primary transcription from the same vRNP in the early stages of infection. Alternatively, the 3′ end bound to the surface of one polymerase could translocate into the active site of a second, empty polymerase as has been proposed in some models of replication that imply polymerase oligomerization[11]. It is also clear, as discussed above, that additional conformational changes must occur to allow progression from the initiation to the elongation stages of RNA synthesis. Although it is expected, in analogy to HCV, that the PB1 RNA-binding cavity widens during this step and the lid made by the PB2 N2 domain should open (see above), several other lines of evidence suggest that PB2 as a whole is the most mobile part of the polymerase. First, docking of the polymerase crystal structure into the mini-RNP electron microscopy map[39] shows that the PA–PB1 heterodimer fits well but the extra density assigned to PB2 is detached from the rest and cannot be fitted without a gross conformational change of PB2 (Extended Data Fig. 6). Second, detachment of a large fragment of PB2 is compatible with the polymerase structure in the FluB2 crystal form, in which two-thirds of PB2 (PB2-C) is not visible at all although there is space in the crystal for it. Similarly, in the electron microscopy reconstructions of native RNPs[19], part of the polymerase (the 'arm'), was observed to be detached and flexible. Although this was assigned to PA-C, we think it is most likely to be PB2-C, for the reasons just given and because our structures suggest that the integrity of the PA–PB1 heterodimer is very unlikely to be disrupted at least while both are intimately binding the vRNA 5′ end.

Although there are still very many open questions, our three complementary structures already give considerable new insight into the mechanism of replication and transcription by influenza polymerase. They provide a solid structural framework for future studies aimed at refining understanding of this complex and dynamic molecular machine, not only in isolation but also in the more complicated physiological context of the RNP and host factors.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 August; accepted 29 October 2014.
Published online 19 November 2014.

1. Resa-Infante, P., Jorba, N., Coloma, R. & Ortin, J. The influenza virus RNA synthesis machine: advances in its structure and function. *RNA Biol.* **8,** 207–215 (2011).
2. Fodor, E. The RNA polymerase of influenza a virus: mechanisms of viral transcription and replication. *Acta Virol.* **57,** 113–122 (2013).
3. Ulmanen, I., Broni, B. A. & Krug, R. M. Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m^7GpppNm) on RNAs and in initiating viral RNA transcription. *Proc. Natl Acad. Sci. USA* **78,** 7355–7359 (1981).
4. Blass, D., Patzelt, E. & Kuechler, E. Identification of the cap binding protein of influenza virus. *Nucleic Acids Res.* **10,** 4803–4812 (1982).
5. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature Struct. Mol. Biol.* **15,** 500–506 (2008).
6. Dias, A. *et al.* The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458,** 914–918 (2009).
7. Yuan, P. *et al.* Crystal structure of an avian influenza polymerase PA_N reveals an endonuclease active site. *Nature* **458,** 909–913 (2009).
8. Pritlove, D. C., Poon, L. L., Fodor, E., Sharps, J. & Brownlee, G. G. Polyadenylation of influenza virus mRNA transcribed *in vitro* from model virion RNA templates: requirement for 5′ conserved sequences. *J. Virol.* **72,** 1280–1286 (1998).
9. Nemeroff, M. E., Barabino, S. M., Li, Y., Keller, W. & Krug, R. M. Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3′end formation of cellular pre-mRNAs. *Mol. Cell* **1,** 991–1000 (1998).
10. Bier, K., York, A. & Fodor, E. Cellular cap-binding proteins associate with influenza virus mRNAs. *J. Gen. Virol.* **92,** 1627–1634 (2011).

11. Jorba, N., Coloma, R. & Ortin, J. Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. *PLoS Pathog.* **5,** e1000462 (2009).
12. York, A., Hengrung, N., Vreede, F. T., Huiskonen, J. T. & Fodor, E. Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. *Proc. Natl Acad. Sci. USA* **110,** E4238–E4245 (2013).
13. Nie, Y., Bellon-Echeverria, I., Trowitzsch, S., Bieniossek, C. & Berger, I. Multiprotein complex production in insect cells by using polyproteins. *Methods Mol. Biol.* **1091,** 131–141 (2014).
14. Lee, Y. S. & Seong, B. L. Nucleotides in the panhandle structure of the influenza B virus virion RNA are involved in the specificity between influenza A and B viruses. *J. Gen. Virol.* **79,** 673–681 (1998).
15. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* http://dx.doi.org/10.1038/nature14008 (this issue).
16. Flick, R., Neumann, G., Hoffmann, E., Neumeier, E. & Hobom, G. Promoter elements in the influenza vRNA terminal structure. *RNA* **2,** 1046–1057 (1996).
17. Pritlove, D. C., Poon, L. L., Devenish, L. J., Leahy, M. B. & Brownlee, G. G. A hairpin loop at the 5′ end of influenza A virus virion RNA is required for synthesis of poly(A)$^{+}$ mRNA *in vitro*. *J. Virol.* **73,** 2109–2114 (1999).
18. Newcomb, L. L. *et al.* Interaction of the influenza a virus nucleocapsid protein with the viral RNA polymerase potentiates unprimed viral RNA replication. *J. Virol.* **83,** 29–36 (2009).
19. Moeller, A., Kirchdoerfer, R. N., Potter, C. S., Carragher, B. & Wilson, I. A. Organization of the influenza virus replication machinery. *Science* **338,** 1631–1634 (2012).
20. Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H. & Stuart, D. I. A mechanism for initiating RNA-dependent RNA polymerization. *Nature* **410,** 235–240 (2001).
21. Lescar, J. & Canard, B. RNA-dependent RNA polymerases from flaviviruses and *Picornaviridae*. *Curr. Opin. Struct. Biol.* **19,** 759–767 (2009).
22. Gong, P. & Peersen, O. B. Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proc. Natl Acad. Sci. USA* **107,** 22505–22510 (2010).
23. Deng, T., Vreede, F. T. & Brownlee, G. G. Different de novo initiation strategies are used by influenza virus RNA polymerase on its cRNA and viral RNA promoters during viral RNA replication. *J. Virol.* **80,** 2337–2348 (2006).
24. Mosley, R. T. *et al.* Structure of hepatitis C virus polymerase in complex with primer-template RNA. *J. Virol.* **86,** 6503–6511 (2012).
25. Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. A unique cap(m^{7}GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23,** 847–858 (1981).
26. Reguera, J., Weber, F. & Cusack, S. *Bunyaviridae* RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription. *PLoS Pathog.* **6,** e1001101 (2010).
27. Sikora, D., Rocheleau, L., Brown, E. G. & Pelchat, M. Deep sequencing reveals the eight facets of the influenza A/HongKong/1/1968 (H3N2) virus cap-snatching process. *Sci. Rep.* **4,** 6181 (2014).
28. Datta, K., Wolkerstorfer, A., Szolar, O. H., Cusack, S. & Klumpp, K. Characterization of PA-N terminal domain of Influenza A polymerase reveals sequence specific RNA cleavage. *Nucleic Acids Res.* **441,** 349–353 (2013).
29. Hagen, M., Tiley, L., Chung, T. D. & Krystal, M. The role of template-primer interactions in cleavage and initiation by the influenza virus polymerase. *J. Gen. Virol.* **76,** 603–611 (1995).
30. Rao, P., Yuan, W. & Krug, R. M. Crucial role of CA cleavage sites in the cap-snatching mechanism for initiating viral mRNA synthesis. *EMBO J.* **22,** 1188–1198 (2003).
31. Geerts-Dimitriadou, C., Goldbach, R. & Kormelink, R. Preferential use of RNA leader sequences during influenza A transcription initiation *in vivo*. *Virology* **409,** 27–32 (2011).
32. Geerts-Dimitriadou, C., Zwart, M. P., Goldbach, R. & Kormelink, R. Base-pairing promotes leader selection to prime *in vitro* influenza genome transcription. *Virology* **409,** 17–26 (2011).
33. Braam, J., Ulmanen, I. & Krug, R. M. Molecular model of a eucaryotic transcription complex: functions and movements of influenza P proteins during capped RNA-primed transcription. *Cell* **34,** 609–618 (1983).
34. Li, M. L., Rao, P. & Krug, R. M. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. *EMBO J.* **20,** 2078–2086 (2001).
35. Honda, A., Mizumoto, K. & Ishihama, A. Two separate sequences of PB2 subunit constitute the RNA cap-binding site of influenza virus RNA polymerase. *Genes Cells* **4,** 475–485 (1999).
36. Shih, S. R. & Krug, R. M. Surprising function of the three influenza viral polymerase proteins: selective protection of viral mRNAs against the cap-snatching reaction catalyzed by the same polymerase proteins. *Virology* **226,** 430–435 (1996).
37. Lim, K. *et al.* Biophysical characterization of sites of host adaptive mutation in the influenza A virus RNA polymerase PB2 RNA-binding domain. *Int. J. Biochem. Cell Biol.* **53,** 237–245 (2014).
38. Ng, A. K. *et al.* Influenza polymerase activity correlates with the strength of interaction between nucleoprotein and PB2 through the host-specific residue K/E627. *PLoS ONE* **7,** e36415 (2012).
39. Coloma, R. *et al.* The structure of a biologically active influenza virus ribonucleoprotein complex. *PLoS Pathog.* **5,** e1000491 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the staff of the European Molecular Biology Laboratory (EMBL) eukaryotic expression and high-throughput crystallization facilities within the Partnership for Structural Biology (PSB) and members of the ESRF-EMBL Joint Structural Biology Group for help on European Synchrotron Radiation Facility (ESRF) beamlines. The work was supported by ERC Advanced Grant V-RNA (322586) and EU Grant FLU-PHARM (259751) to S.C. and partially by a Roche Postdoc Fellowship to S.R.

Author Contributions S.R., D.G. and T.L. did protein expression, purification, crystallization and activity assays. A.P. did crystallographic analysis. H.M. did electron microscopy and fitting to the mini-RNP electron microscopy map. M.N. calculated the first interpretable FluB polymerase electron density map. Using the polyprotein vector designed and provided by I.B., and with the help of D.H., S.C. designed the FluB polymerase construct. T.C., D.H., R.R. and S.C. have long-collaborated on studies of influenza polymerase. S.C. supervised the project, collected data, did crystallographic analysis and wrote the paper with input from S.R., D.G., A.P., H.M. and M.N.

Author Information Structure factors and co-ordinates have been deposited in the Protein Data Bank (PDB) under the accessions 4WSA (FluB form 1) and 4WRT (FluB form 2). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.C. (cusack@embl.fr).

METHODS

Construct. The influenza B/Memphis/13/03 polymerase heterotrimer was expressed as self-cleaving polyprotein (Extended Data Fig. 1). A codon-optimized synthetic construct (DNA2.0) with the composition GNH_{BstEII} GSGSENLYFQ$_{TEV}$GSHH HHHHHH$_{8xHis\text{-}tag}$ GSGS-**PA** (GenBank ID AAU94844) GSGSGENLYFQ$_{TEV}$G SGSGSGSG-**PB1** (GenBank ID AAU94857) GSGSGENLYFQ$_{TEV}$G SGSGSGSG-**PB2** (GenBank ID AAU94870) GWSHPQFEK$_{Strep\text{-}tag}$GRSG$_{RsrII}$ was cloned via BstEII and RsrII into the vector pKL-PBac[13], which also contains coding sequences for tobacco etch virus (TEV) protease (5′) and cyan fluorescent protein (CFP) (3′). (TEV cleavage site, His-tag and Strep-tag are underlined.)

Expression and purification. High Five insect cells expressing the target protein complex were resuspended in buffer A (50 mM Tris-HCl, 500 mM NaCl, 10% (v/v) glycerol and 5 mM BME, pH 8) supplemented with protease inhibitors (Roche, complete mini, EDTA-free), lysed by sonication and centrifuged at 30,000 r.p.m. for 30 min at 4 °C (rotor type 45 Ti, Beckman Coulter). Ammonium sulphate was added to the clarified supernatant (0.5 g ml^{-1}), the resulting precipitate collected by centrifugation as above and re-dissolved in buffer A supplemented with 20 mM imidazol. Soluble proteins were loaded on a nickel nitrilotriacetic acid (NTA) column (GE, FF crude) and bound proteins were eluted by 500 mM imidazole in buffer A. The target protein was loaded on a strep-tactin matrix (IBA, Superflow) and bound proteins eluted by 2.5 mM *d*-desthiobiotin in buffer A. Fractions containing the target protein were pooled and diluted with an equal volume of buffer B (50 mM HEPES/NaOH, 10% (v/v) glycerol and 2 mM Tris(2-carboxyethyl)phosphine (TCEP), pH 7.45) before loading on a heparin column (HiTrap Heparin HP, GE Healthcare). Proteins were eluted by a gradient of buffer B supplemented with 1 M NaCl, concentrated (Amicon Ultra, 50 kDa molecular mass cut-off) and further purified by size-exclusion chromatography (S200, GE Healthcare) in buffer C (50 mM HEPES/NaOH, 500 mM NaCl, 5% (v/v) glycerol and 2 mM TCEP, pH 7.45). Homogeneous monomeric FluB polymerase was concentrated as above and stored in aliquots at −80 °C. Protein concentration was determined by measuring the absorbance at 280 nm using the extinction coefficient 287,300 M^{-1} cm^{-1}.

Crystallization. FluB polymerase was concentrated to 9 mg ml^{-1} (37 μM) in a buffer containing 500 mM NaCl, 50 mM HEPES, pH 7.5, 5% glycerol and 2 mM TCEP, and mixed with 40 μM vRNA for crystallization in hanging drops at 4 °C. A trigonal crystal form (FluB1) was obtained by mixing polymerase with nucleotides 5–18 of the 3′ end and 1–14 of the 5′ end of the vRNA (IBA) in a condition containing 0.1 M bicine, pH 9.0, 10% MPD. Large (up to 150 μm) diamond-like crystals grew within a few days and diffracted to around 3.4 Å resolution but were very radiation-sensitive. The structure was solved with data at 6.5 Å resolution from a single heavy metal derivative obtained by soaking native crystals with 1 mM K_2PtCl_4 for 1 h. Selenomethionylated protein crystals were obtained in the same conditions as native ones. Polymerase with nucleotides 1–18 of the 3′ end and 1–14 of the 5′ end of the viral RNA gave thin hexagonal plates (form FluB2) in 1 M LiCl, 10% PEG 6000 and 0.1 M bicine, pH 9.0, that took 3–4 weeks to grow and diffracted to 2.7 Å resolution. All crystals were cryo-protected in mother liquor supplemented with 20% glycerol and flash-frozen in liquid nitrogen. Data was collected at 100 K on beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF), equipped with a Pilatus 6M-F detector, at wavelengths of 0.9730 and 0.9792 Å for FluB1 and FluB2 crystals, respectively. All data were integrated and scaled with XDS[40].

Structure determination. A partial molecular replacement solution (LLG 334) was found with PHASER[41] using the known PA-C–PB1-Nter (PDB codes 2ZN1 and 3CM8) and PB2 627 (PDB code 2VY7) domain structures initially both from FluA. The cap-binding and endonuclease domains could not be located even using the actual FluB domain structures (unpublished data). Nevertheless, ~22% of the complete structure was sufficient to identify around 20 platinum sites by inspection of a model-phased difference anomalous map. Several of the platinum peaks coincided with known positions of methionine residues. After scaling the platinum and native data sets, the platinum substructure was refined in SHARP[42] to 7 Å and treated as SIRAS (single isomorphous replacement with anomalous), using the partial molecular replacement phases in the form of Hendrickson–Lattman coefficients. The final phasing statistics were phasing power $(PP)_{anomalous} = 0.716$, $PP_{iso,centric} = 0.609$, $PP_{iso,acentric} = 0.714$, figure of merit $(FOM)_{centric} = 0.21$, $FOM_{acentric} = 0.36$. Solvent flattening and phase extension to the full resolution of the native data (initially 3.7 Å and subsequently 3.4 Å) was then performed with SOLOMON[43] benefitting from the high solvent content of 73%. The resultant map had an overall correlation on $|E|^2$ of 81.4% and R_{factor} of 23.8%. The exceptionally good continuity of the map (Extended Data Fig. 3a–c) allowed immediate placing of known structures of the cap-binding, endonuclease and PB1–PB2 interface domains and revealed numerous additional secondary structures that could eventually be linked to trace almost the entire chain of each subunit as well as the vRNA. During model building and refinement with REFMAC[44] map sharpening with B_{factor} of −50 Å^2 was used to improve visibility of side chains. Accurate model building was aided by using four high resolution structures of FluB polymerase domains determined during the course of this work (PA endonuclease at 2.1 Å resolution, PA-C/PB1-Nter at 2.4 Å resolution, PB2 cap-binding domain with m^{7}GTP at 1.5 Å resolution, and the PB2 627-domain at 1.05 Å, unpublished data). Sequence assignment was verified by using the methionine positions located using the anomalous differences measured at 4.1 Å resolution from a seleno-methionylated polymerase crystal (Extended Data Fig. 3). The FluB2 crystal structure was determined by molecular replacement using the FluB1 structure. The C-terminal two-thirds of PB2 (PB2-C) is completely absent in the electron density map in this crystal form, although gel analysis shows PB2 to be intact and the crystal packing can accommodate PB2-C. When they became available, the higher resolution bat FluA and FluB2 structures enabled improvement in the quality of the FluB1 model with the help of secondary structure constraints derived using PROSMART[45]. Full crystallographic statistics are given in Extended Data Table 1. Figures were drawn with Pymol[46]. Ramachandran statistics, as calculated by Molprobity[47], are 93.7% (favoured), 0.6% (disallowed) for the FluB1 structure and 97.5% (favoured), 0.1% (disallowed) for the FluB2 structure.

Polymerase activity assays. A T7-transcribed 39-nucleotide mini-panhandle or equimolar mixture of separated synthetic 3′ and 5′ ends were used as vRNA (Extended Data Fig. 2), corresponding to the consensus promoter sequences for influenza B polymerase[14].

For the ApG-primed replication assay, 0.5 μM protein, 0.5 μM vRNA, 0.5 mM ApG, 0.4 mM GTP/CTP, 1 mM ATP, 0.04 mM UTP, ^{32}P-UTP and 0.8 U μl^{-1} Ribolock, in buffer (150 mM NaCl, 50 mM HEPES, pH 7.5, 5 mM $MgCl_2$ and 2 mM TCEP) were mixed and incubated at 30 °C for 2 h.

For the cap-dependent transcription assay, 0.5 μM protein, 0.5 μM vRNA, 0.4 mM GTP/CTP/UTP, 1 mM ATP, ^{32}P-labelled capped RNA in the same buffer (150 mM NaCl, 50 mM HEPES, pH 7.5, 5 mM $MgCl_2$ and 2 mM TCEP) were mixed and incubated at 30 °C for 2 h. For this purpose, a 5′ diphosphate synthetic 20-base RNA, 5′-ppAAUCUAUAAUAGCAUUAUCC-3′ (Chemgenes), was capped by incubating with vaccinia virus capping enzyme (purified in house following ref. 48) and 20 μM SAM, ^{32}P-GTP, 50 mM Tris, pH 8.0, 6 mM KCl, 1.25 mM $MgCl_2$ and 0.8 U μl^{-1} Ribolock.

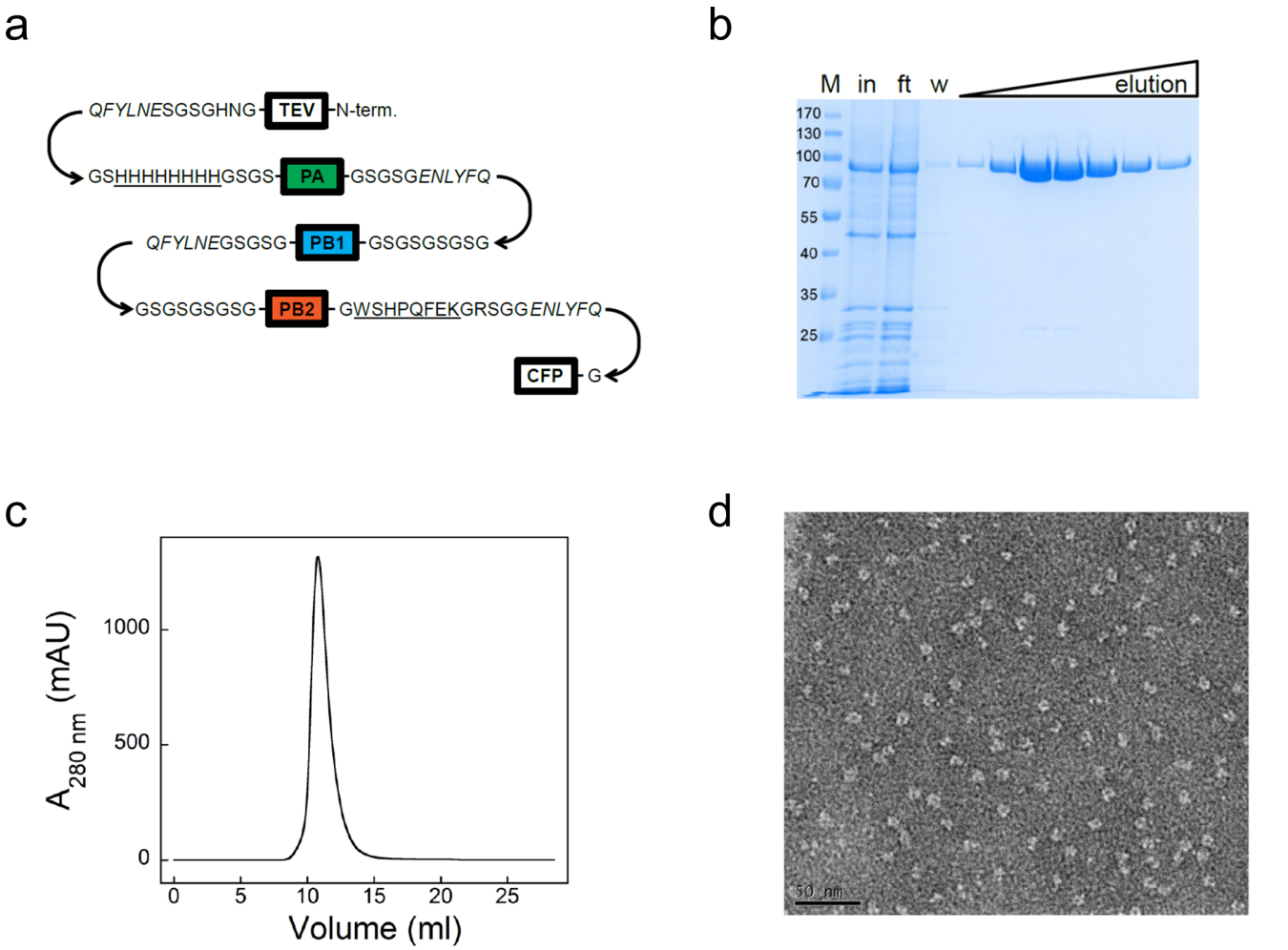
For the endonuclease assay, transcription mix lacking NTPs was incubated at 30 °C for 2 h. Samples were separated on 7 M urea, 20% acrylamide gel in TBE buffer, exposed on a Storage Phosphor screen and read with a Typhoon scanner.

For the time course of unprimed and ApG-primed vRNA replication, 0.5 μM FluB polymerase was mixed with 1 μM 39-nucleotide vRNA mini-panhandle template, NTPs (1 mM ATP, 0.4 mM GTP, 0.4 mM CTP and 0.04 mM UTP) and 0.12 μCi μl^{-1} ^{32}P-UTP, in the absence or presence of 0.5 mM ApG. Reactions were incubated at 30 °C and samples were analysed on a 20% acrylamide, 7 M urea denaturing gel after 0, 2, 5, 10, 15, 20, 30, 40 and 50 min, 1, 2 and 3 h.

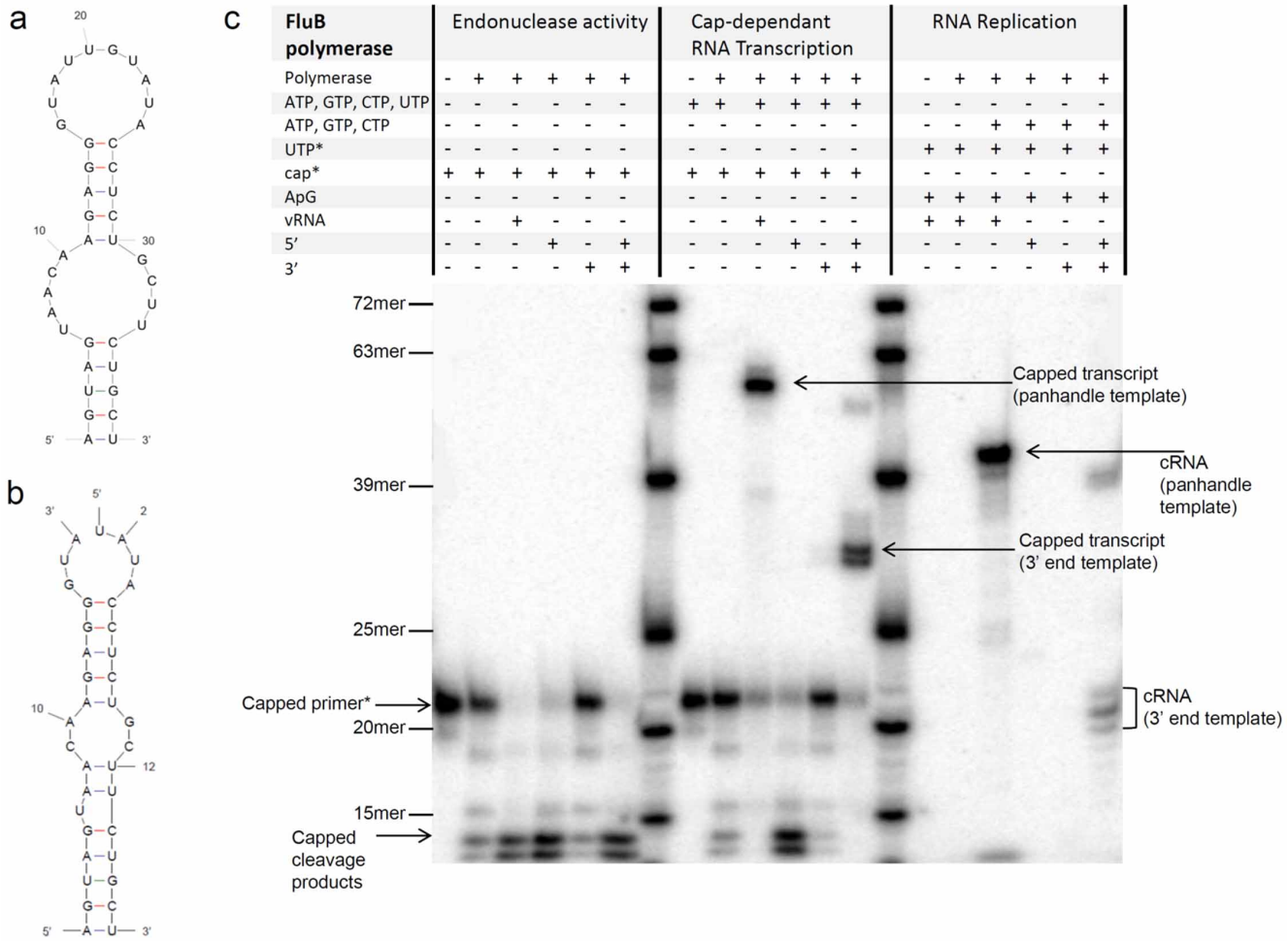
Fitting to the mini-RNP electron microscopy map. Influenza polymerase (FluB2 model) and nine influenza A nucleoproteins (PDB code 2IQH)[49] were rigidly fitted into the 18 Å mini-RNP cryo-EM reconstruction[39] using chimaera fit-in-map module[50] and VEDA[51]. Map scaling was optimized by the cross-correlation between the model and map for different pixel sizes as implemented in VEDA. Down-scaling the electron microscopy map from 2.8 Å/pixel to 2.4 Å/pixel improved the cross-correlation and fit quality considerably. The fitting of the nine nucleoproteins follow the model previously proposed[39], with each nucleoprotein and loop 402–428 of its neighbour being considered as a rigid entity to maintain the nucleoprotein–nucleoprotein interaction mode. For polymerase fitting, different starting positions of the PA–PB1 heterodimer with only 1–32 of PB2 (FluB2 model) were used for rigid body fitting using the chimera fit-in-map module and allowed to identify one preferred rigid fit position. Finally, the model was refined with a simultaneous rigid fit of the polymerase and the nine nucleoproteins using VEDA.

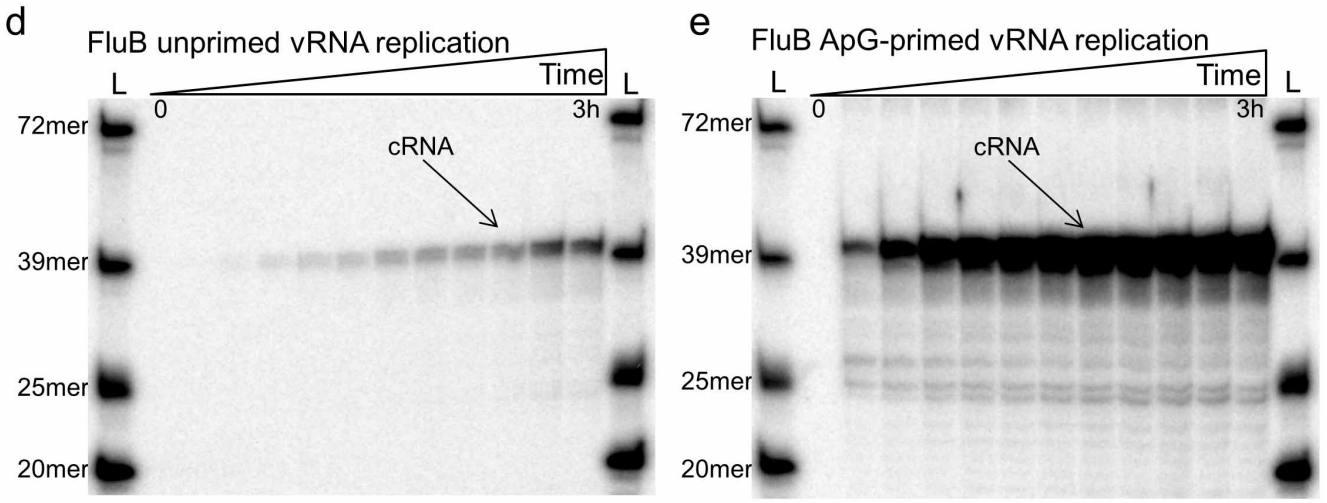
40. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66,** 133–144 (2010).
41. Read, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr. D* **57,** 1373–1382 (2001).
42. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59,** 2023–2030 (2003).
43. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52,** 30–42 (1996).
44. Murshudov, G. N. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
45. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D* **70,** 2487–2499 (2014).
46. DeLano, W. L. The PyMOL Molecular Graphics System; http://www.pymol.sourceforge.net (Schrödinger, LLC, 2002).
47. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
48. De la Peña, M., Kyrieleis, O. J. & Cusack, S. Structural insights into the mechanism and evolution of the vaccinia virus mRNA cap *N*7 methyl-transferase. *EMBO J.* **26,** 4913–4925 (2007).

49. Ye, Q., Krug, R. M. & Tao, Y. J. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444,** 1078–1082 (2006).
50. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comp. Chem.* **25,** 1605–1612 (2004).
51. Siebert, X. & Navaza, J. UROX 2.0: an interactive tool for fitting atomic models into electron-microscopy reconstructions. *Acta Crystallogr. D* **65,** 651–658 (2009).
52. Esnouf, R. M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55,** 938–940 (1999).
53. Area, E. *et al.* 3D structure of the influenza virus polymerase complex: localization of subunit domains. *Proc. Natl Acad. Sci. USA* **101,** 308–313 (2004).
54. Wakai, C., Iwama, M., Mizumoto, K. & Nagata, K. Recognition of cap structure by influenza B virus RNA polymerase is less dependent on the methyl residue than recognition by influenza A virus polymerase. *J. Virol.* **85,** 7504–7512 (2011).

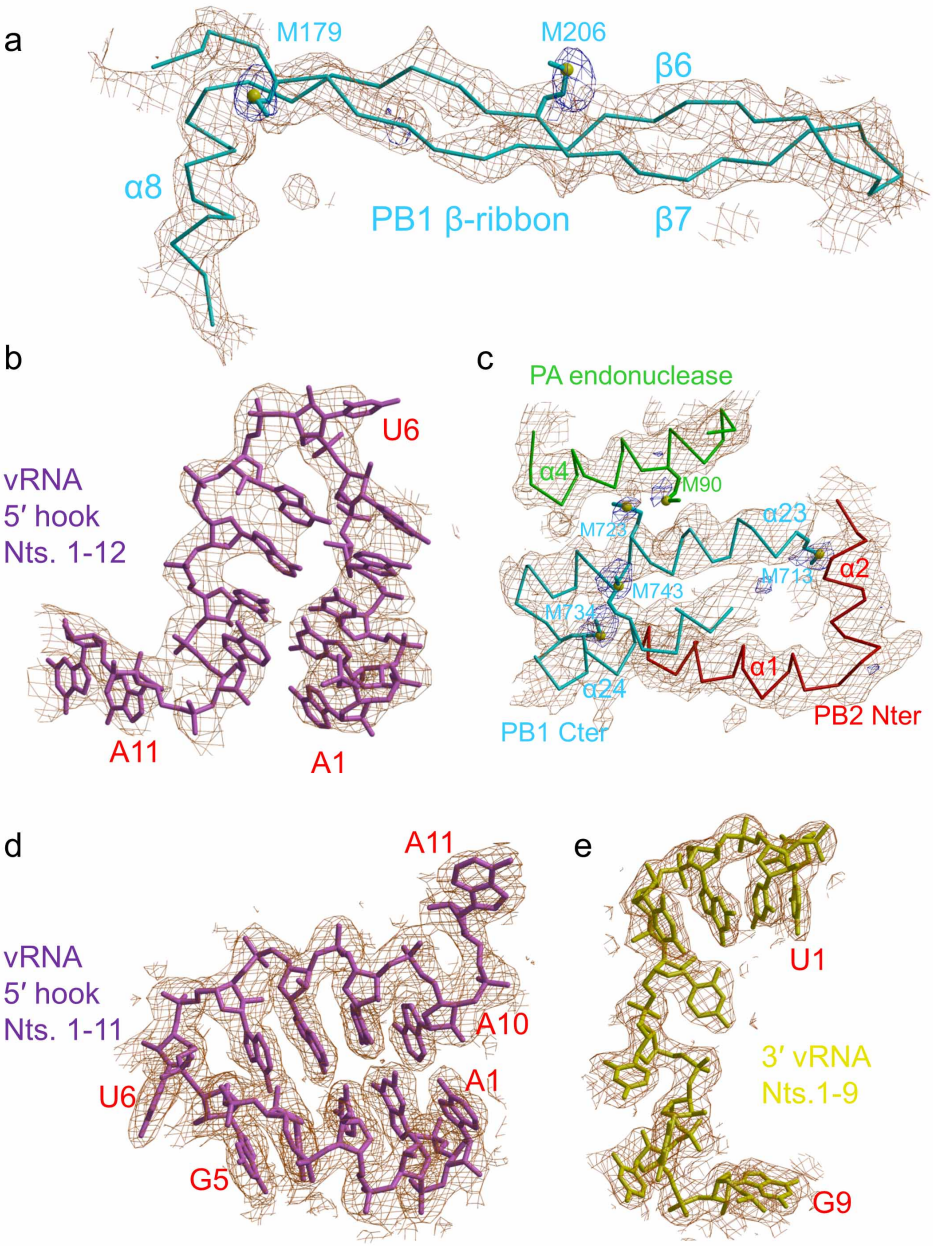


Extended Data Figure 1 | Production and characterization of influenza B polymerase heterotrimer. **a**, Schematic of the self-cleaving polyprotein construct used to express recombinant influenza B heterotrimeric polymerase in insect cells. N-terminally it encodes the tobacco etch virus (TEV) protease that cleaves C-terminal to the amino-acid sequence ENLYFQ (in italics) and releases N-terminally His-tagged PA, PB1, C-terminally Strep-tagged PB2 and cyan fluorescent protein (CFP) for facilitated expression monitoring. Arrows indicate the N-to-C-terminal direction and the termini of each mature protein. The histidine and streptavidin tags are underlined. **b**, Using the PB2 C-terminal strep-tag, most contaminating proteins could be separated from the polymerase as judged by 10% SDS–PAGE followed by Coomassie blue staining. Lanes 'M' contain the protein markers (molecular masses indicated); 'in', 'ft' and 'w' denote the input, flow-through and wash of the engineered streptavidin (strep-tactin) column, respectively, and 'elution' indicates the re-mobilization of bound heterotrimeric polymerase by a sharp gradient of *d*-desthiobiotin. The three subunits, PA (85.7 kDa), PB1 (86.1 kDa) and PB2 (90.8 kDa), run together on the gel. **c**, After ammonium sulphate precipitation, IMAC, strep-tactin affinity and heparin chromatography, the final purification step consists of size-exclusion chromatography. The elution profile (monitored by the absorbance at 280 nm) with a single and nearly symmetric peak suggests a homogeneous and monomeric polymerase complex. **d**, Recombinant influenza B polymerase was analysed by electron microscopy following negative staining with sodium silico-tungstate of 0.02 mg ml^{-1} protein sample. The image demonstrates that the sample is homogeneous and monodisperse with a V- or doughnut-like shape with a central cavity.

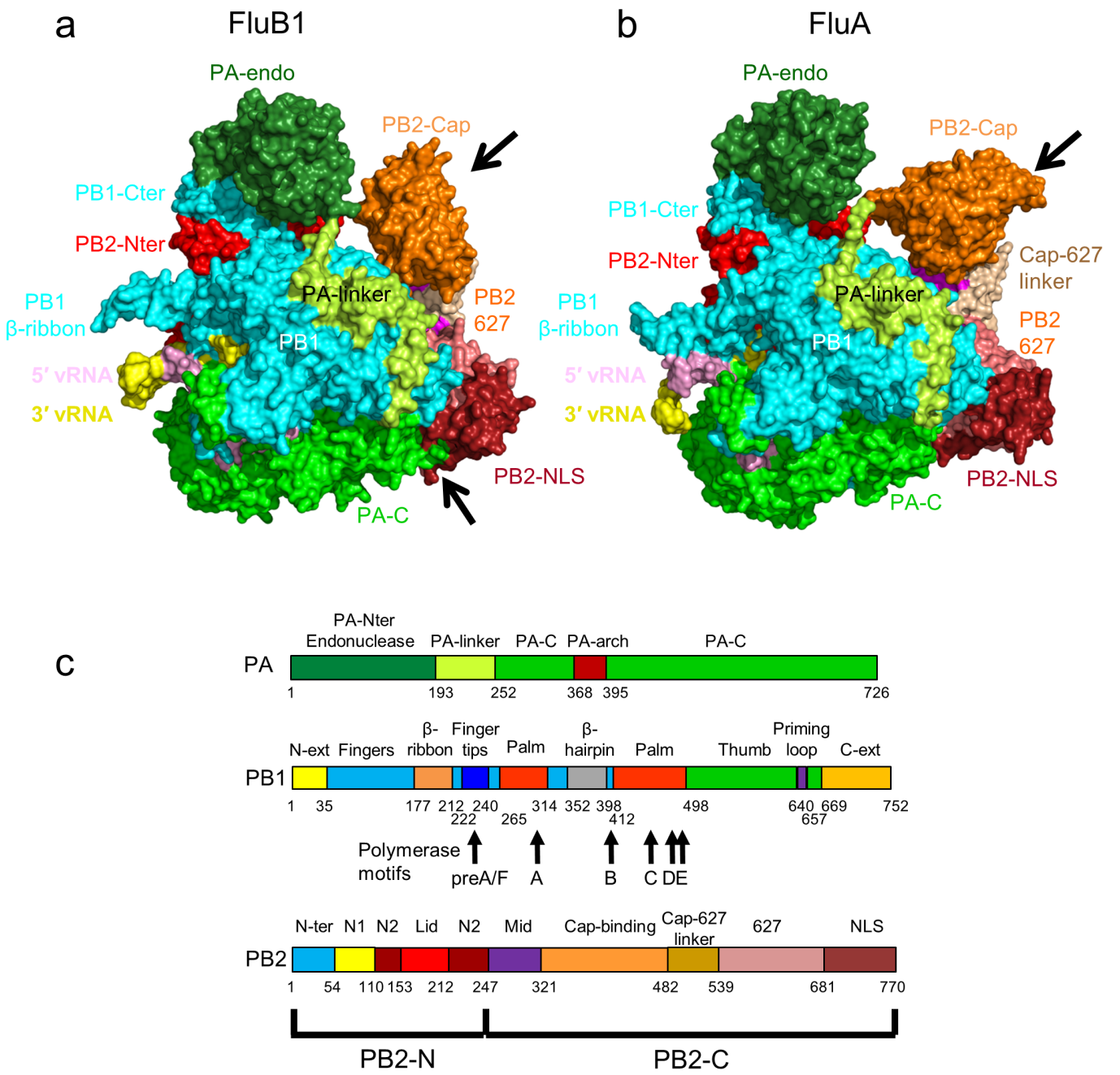




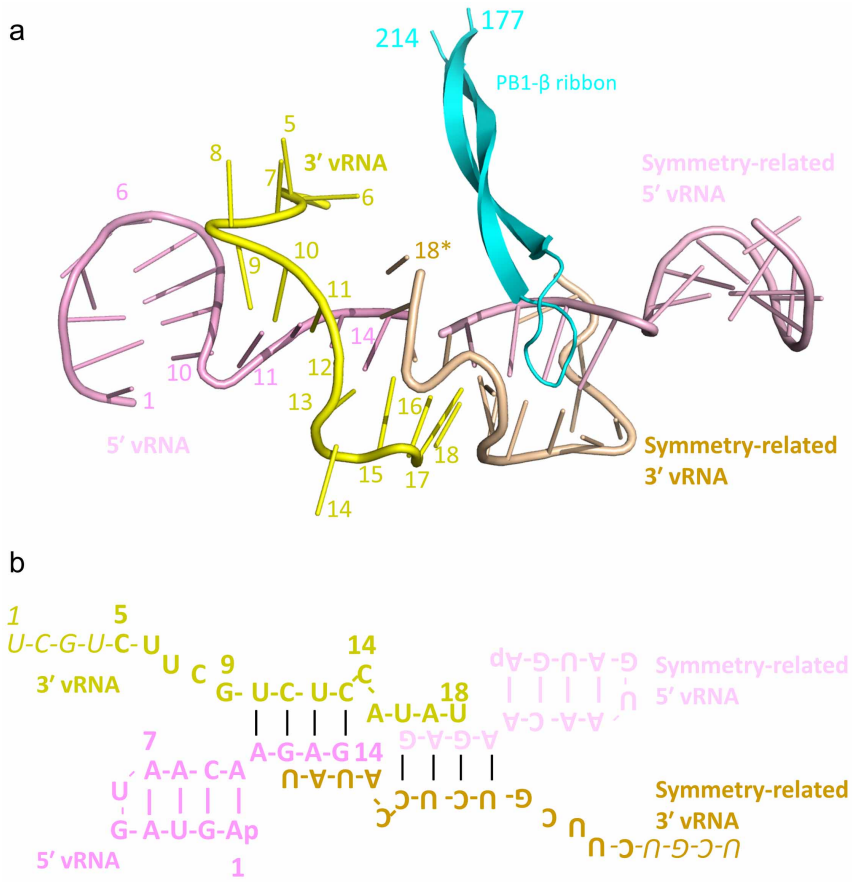
Extended Data Figure 2 | Endonuclease, transcription and replication activities of FluB polymerase. **a**, Schematic of mini-panhandle vRNA: 5′-pppAGUAGUAACAAGAGGGUAUUGUAUACCUCUGCUUCUGCU-3′. **b**, Schematic of separate 5′ and 3′ ends: 5′: 5′-pAGUAGUAACAAGAGGGUA-3′; 3′: 5′-UAUACCUCUGCUUCUGCU-3′. **c**, Endonuclease, cap-dependent transcription and ApG-primed replication assays. Cleavage of the cap donor is visible in lanes 2–6 and enhanced in the presence of the 5′ end, but not the 3′ end. Capped transcripts are visible in lanes 10 (from vRNA panhandle template) and 13 (from separated 5′ and 3′ vRNA ends) as well as cRNA produced in lanes 17 and 20. Markers, with size shown on the left, are RNA ladders labelled with ^{32}P-pCp nucleotide. **d**, **e**, Time course of unprimed (**d**) and ApG-primed (**e**) vRNA replication by influenza B polymerase. The products of replication (cRNA) are indicated with an arrow. Ladders (lanes L) are ^{32}P-pCp nucleotide-labelled RNA oligomers. ApG-primed replication is more efficient than unprimed replication.



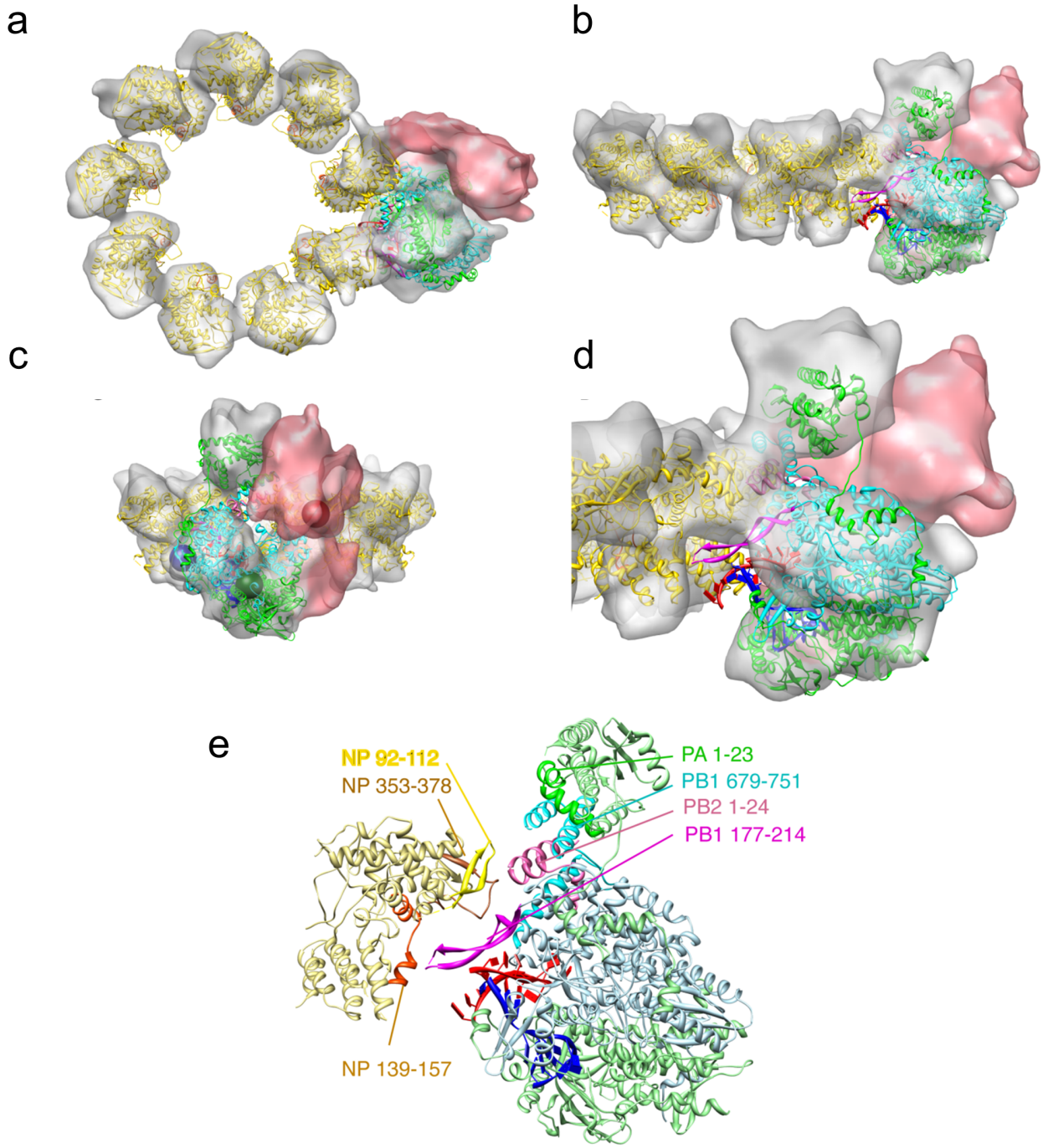
Extended Data Figure 3 | Examples of electron density map for FluB polymerase. **a–c**, Initial platinum SIRAS-phased and phase-extended experimental map at 3.6 Å resolution contoured at 1.1σ (brown) with superposed final model for the FluB1 crystal form. Also shown is the final model-phased selenium anomalous difference map at 4.1 Å resolution contoured at 3.2σ (purple) highlighting methionine positions. **a**, PB1 β-ribbon. **b**, vRNA 5′ hook. **c**, PA–PB1–PB2 helical interface. **d**, **e**, final $2F_o - F_c$ omit map at 2.7 Å resolution for the FluB2 crystal form contoured at 1.1σ. **d**, vRNA 5′ hook nucleotides 1–11. **e**, vRNA 3′ end nucleotides 1–9. Figures drawn with Bobscript[52].



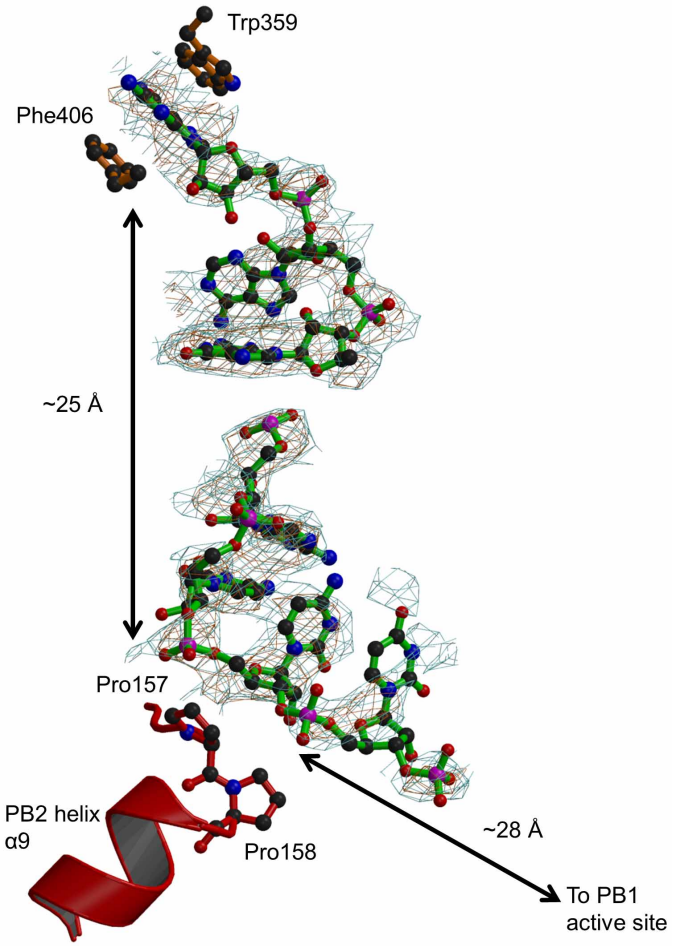
Extended Data Figure 4 | Comparison of FluB and bat FluA polymerase structures. **a**, Surface diagram of FluB1 structure coloured as in **c** except that PA-C, PB1and PB2-N are uniformly green, cyan and red, respectively. The bottom black arrow indicates the extra 12 C-terminal residues of FluB PA that extend the PA C-terminal helix compared to FluA, so that it directly contacts the PB2-NLS domain that is consequently orientated slightly differently from in FluA polymerase. **b**, As in **a** but for bat FluA structure. Arrows highlight the 70° difference in orientation of the cap-binding domain. The structural similarity between FluA and FluB polymerases (LSQMAN, cut-off 3.5 Å) is as follows. PA: 630 Cα atoms aligned, of which 38.6% are identical with root mean squared deviation (r.m.s.d.) 1.34 Å; PB1: 703 Cα atoms aligned, of which 61.3% are identical with r.m.s.d. 1.06 Å; PB2: 428 Cα atoms aligned, of which 40.6% are identical with r.m.s.d. 1.46 Å (excluding the cap-binding domain), and, taking into account the cap-binding domain rotation, 622 Cα atoms aligned, of which 39.0% are identical with r.m.s.d. 1.54 Å). **c**, Subunit domain structure of influenza B polymerase with names and extended colour scheme, showing the positions of the PB1 polymerase motifs. Note that for PB1, the FluB numbering compared to FluA is the same from 1–399 and is thereafter +1. For PB2, FluB is +2 from 1–469 and +1 from 470–628. For PA it is more complicated owing to several short insertions and deletions. See Supplementary Fig. 1.



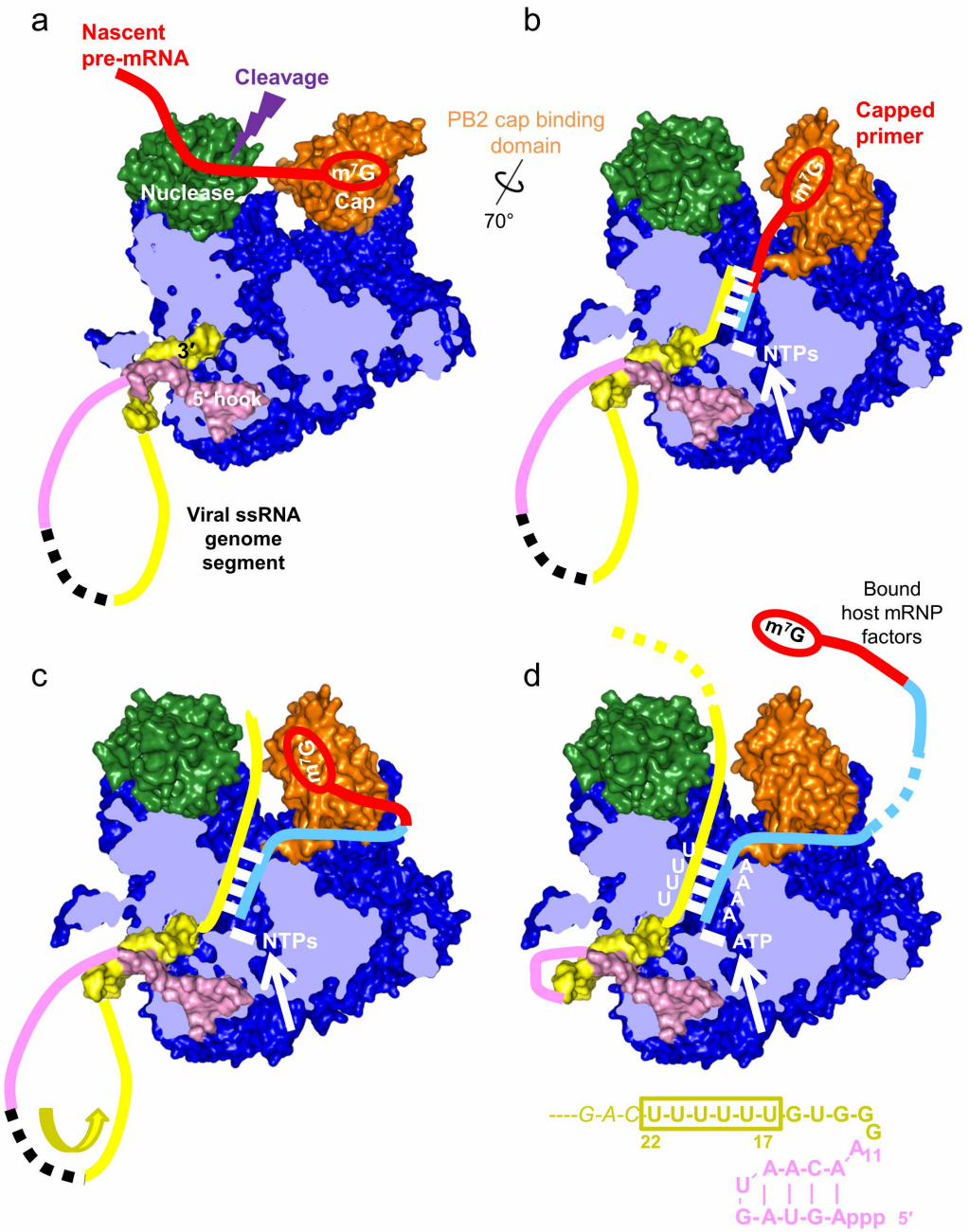
Extended Data Figure 5 | RNA–RNA crystal contact in FluB1 crystal form. **a**, Cartoon of 5′ and 3′ vRNA ends (left, pink and yellow, respectively) interacting with crystallographic two-fold symmetry-related vRNA (right, pale pink and wheat, respectively). The PB1 β-ribbon (cyan) of the left-hand polymerase molecule interacts with the symmetry-related vRNA. **b**, Simplified diagram showing vRNA sequence and secondary structure in the FluB1 crystal form including vRNA-mediated crystal contact.



Extended Data Figure 6 | Polymerase fitting into the mini-RNP electron microscopy map. **a**, **b**, Top (**a**) and side (**b**) view of influenza A mini-RNP pseudo-atomic model with rescaled electron density[39]. PA, PB1 and PB2 (1–32 only) are shown as ribbons and coloured in green, cyan and rose, respectively. Unfilled electron density, likely to contain the rest of PB2, is shown in transparent rose. Nucleoproteins are shown in yellow ribbons, with the nucleoprotein–nucleoprotein interacting loop (residues 402–428) in orange. The vRNA 5′ and 3′ ends are shown in dark blue and red, respectively. **c**, Front view of influenza A mini-RNP pseudo-atomic model. The positions of antibody and tag labelling corresponding to domains of PA, PB1 and PB2 are shown as dark green, dark blue and dark rose spheres, respectively, as localized previously[53]. **d**, Close-up view of **b**. The PB1 β-ribbon (residues 177–214, purple) is located close to one of the proximal nucleoproteins and the vRNA. **e**, Putative interactions between the proximal nucleoprotein and polymerase. Nucleoprotein elements proposed for polymerase interaction are indicated in yellow, brown and orange. Polymerase interacting elements are shown in green, cyan, rose and magenta.



Extended Data Figure 7 | Residual electron density in the FluB1 crystal form mimicking capped primer binding to the PB2 cap-binding domain. Residual m$2F_o - F_c$ (blue mesh at 0.9σ) and m$F_o - F_c$ (orange mesh at 2.5σ) electron density showing RNA-like density bound in the cap-binding site in the FluB1 crystal form. The low resolution and partial occupancy do not allow identification of the RNA and the discontinuous model shown is for illustrative purposes only. Owing to the rigorous purification procedure it is unlikely to be insect cell-capped RNA that is trapped on the polymerase. More likely it derives from the input vRNA used in crystallization, possibly partially digested by the endonuclease that generates 3′ ends. That this RNA could even be uncapped is explicable by the fact that the FluB cap-binding domain, unlike that of FluA, promiscuously binds both methylated and unmethylated guanosine[54]. Indeed, the density seems to be better fit with a free 3′ end sandwiched between Phe 406 and Trp 359 in the cap-binding site rather than a capped 5′ end. As the primer emerges from the cap-binding site it is initially channelled on one side by the base of the 424-loop, and on the other by residues 518–522 of the cap-627 linker. Further down, the extended 424-loop continues to guide the RNA, as well as, on the other side, the projecting N-terminal end of PB2 helix α9 (155-EMPPDE in FluB), with the double proline forcing the RNA into a ~90° bend. Arg 425 and Arg 438 are well placed to interact with phosphates and one base seems to stack on the Glu 155–Arg 217 salt bridge. Conserved basic residues on PB2 N2 domain strands β7, 144-Arg-Lys-Arg (FluA 142-Arg-Lys-Arg), and β8, 216-Arg-Arg-Arg-Phe (FluA 214-Arg-Thr-Arg-Phe), are also likely to be involved. Straight-line distances from the cap-binding site to the bend and from the bend to the PB1 active site are indicated. See also Fig. 5.



Extended Data Figure 8 | Schematic diagram of steps in cap-dependent transcription by influenza virus polymerase. **a**, Cap-snatching from host pre-mRNA (red). The m^7G cap is bound by the cap-binding domain (orange, orientated as in the FluA structure) and the pre-mRNA cleaved 10–14 nucleotides downstream by the endonuclease (green). The single-stranded vRNA genome is bound by its 5′ (hook, pink) and 3′ (template, yellow) ends to the polymerase (blue, depicted as a cutaway section). **b**, Transcription initiation. The cap-binding domain rotates to the position observed in the FluB1 structure directing the capped primer into the PB1 active site, where it potentially makes limited base pairs with the extremity of the template. Template-directed NTP addition (white) extends the host sequences (red) with virally encoded sequences (cyan). Note that in **b**–**d** additional conformational changes in the polymerase are expected, but not depicted since they are currently unknown. **c**, Transcription elongation. Transcription elongation proceeds, eventually leading to the release of the cap from the cap-binding domain (**d**) and the binding of host mRNP factors. **d**, Polyadenylation by stuttering. After most of the vRNA template has been translocated through the polymerase, only a tight turn connects it to the bound 5′-hook. The nucleotide sequence of this region is given at the bottom. This places the 5′ proximal oligo-U stretch in the PB1 active site allowing poly(A) tail synthesis by a stuttering mechanism in which the template is no longer translocated but the product strand is able to slip.

Extended Data Table 1 | Data collection and refinement statistics for FluB polymerase structures

	FluB1 3′ end: 5-18 5′ end: 1-14	FluB2 3′ end: 1-18 5′ end: 1-14
Data collection		
Space group	$P3_221$	$P6_222$
Cell dimensions		
a, *b*, *c* (Å)	199.70,199.70,252.68	207.37,207.37,345.69
α, β, γ (°)	90.0, 90.0,120.0	90.0, 90.0,120.0
Number of crystals	4	1
Resolution (Å)	50.0-3.40(3.52-3.40)*	50.0-2.70(2.80-2.70)
R_{meas}	12.1 (130.1)	25.2 (142.9)
$I/\sigma I$	12.6 (1.5)	10.3 (1.9)
Completeness (%)	98.5 (89.9)	99.9 (100.0)
Redundancy	11.8 (5.8)	10.1 (10.3)
Refinement		
Resolution (Å)	50.0-3.40(3.49-3.40)	50.0-2.70(2.77-2.70)
No. reflections	79266	120050
R_{work}/ R_{free}	22.9/26.5 (40.2/42.7)	17.3/21.1 (32.3/34.6)
No. atoms		
Protein	17351	13548
RNA	589	615
Water	-	809
B-factors (Å^2)	158.5	43.5
Protein	159.6	43.5
RNA	126.1	52.7
Water	-	38.2
R.m.s deviations		
Bond lengths (Å)	0.003	0.003
Bond angles (°)	0.657	0.683

*Highest resolution shell is shown in parenthesis.

Extended Data Table 2 | Direct polar polymerase–vRNA contacts for the FluB2 structure

Calculated with CONTACT (CCP4i) with 3.5 Å cutoff.

PA (chain B) - 5′ vRNA (chain V)

Residue		Atom		RNA base		Atom		Distance (Å)	
Lys	330B	NZ	...	A	1V	OP1	...	3.21	***
Gly	367B	N	...	A	1V	N1	...	3.44	*
Gly	367B	O	...	A	10V	O2'	...	2.75	***
Gly	369B	N	...	A	11V	OP2	...	2.76	***
Leu	370B	N	...	A	11V	OP1	...	3.23	***
Thr	371B	N	...	A	11V	OP1	...	3.02	***
Thr	371B	OG1	...	A	10V	OP2	...	2.91	***
			...	A	10V	N7	...	3.45	*
Gln	504B	NE2	...	A	11V	O4'	...	3.33	*
			...	A	11V	O5'	...	3.05	***
His	506B	NE2	...	A	11V	N7	...	3.32	*
Val	513B	N	...	C	9V	O2'	...	2.89	***
Thr	515B	OG1	...	A	1V	N1	...	2.88	***
Arg	558B	NH2	...	U	3V	OP1	...	2.98	***
Val	559B	O	...	G	2V	O3'	...	3.41	*
Asn	560B	O	...	G	2V	O2'	...	3.21	***
Gly	561B	O	...	U	3V	O3'	...	3.41	*
			...	U	3V	O2'	...	3.42	*
Gln	566B	NE2	...	A	4V	OP1	...	2.94	***
Asn	692B	ND2	...	G	5V	O6	...	2.90	***

PB1 (chain P) - 5′ vRNA (chain V)

His	32P	NE2	...	G	5V	OP1	...	3.10	***
His	32P	O	...	A	7V	O4'	...	3.43	*
Thr	34P	N	...	A	8V	OP1	...	3.20	***
Tyr	38P	OH	...	U	6V	O5'	...	3.32	*
Lys	365P	NZ	...	C	9V	OP2	...	2.92	***

PA (chain B) - 3′ vRNA (chain T, numbered from 3′ end)

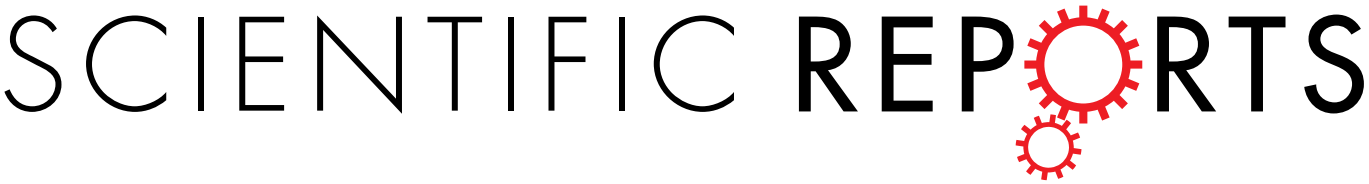
His	506B	ND1	...	G	9T	O6	...	3.01	***
Arg	508B	NE	...	U	7T	OP1	...	3.02	***
Arg	508B	NH1	...	U	10T	O2	...	3.45	*
Arg	508B	NH2	...	U	10T	O2'	...	2.49	***
			...	U	10T	O2	...	3.23	***
Lys	564B	NZ	...	C	8T	OP2	...	2.79	***

PB1 (chain P) - 3′ vRNA (chain T, numbered from 3′ end)

Gln	127P	OE1	...	U	6T	N3	...	2.97	***
			...	U	6T	O2	...	3.33	*
Val	133P	O	...	U	4T	N3	...	3.42	*
			...	U	4T	O4	...	3.07	***
Arg	135P	NE	...	U	4T	O4	...	3.01	***
Arg	135P	NH2	...	U	4T	O4	...	3.16	***
			...	G	3T	O6	...	3.15	***
Asn	136P	O	...	U	6T	N3	...	2.89	***
Val	184P	O	...	G	3T	N2	...	2.83	***
Asn	186P	N	...	C	2T	O2	...	2.71	***
Asn	186P	ND2	...	C	2T	N3	...	3.26	***
Arg	203P	NH2	...	U	1T	OP2	...	2.41	***
Arg	350P	NH2	...	U	4T	O4'	...	2.90	***
Asn	670P	OD1	...	U	7T	OP2	...	3.35	*
Asn	670P	ND2	...	U	7T	OP2	...	2.92	***
Arg	671P	N	...	G	9T	OP1	...	2.88	***
Arg	671P	NH1	...	U	10T	OP1	...	2.67	***
Ser	672P	N	...	U	10T	O3'	...	3.39	*
Ser	672P	OG	...	U	10T	O2'	...	2.82	***
			...	U	7T	OP2	...	2.61	***
Asn	675P	ND1	...	C	11T	O2'	...	2.54	***
			...	C	11T	O2	...	3.17	***

PB2 (chain E) - 3′ vRNA (chain T, numbered from 3′ end)

Thr	38E	O	...	U	7T	N3	...	2.82	***
			...	U	7T	O2	...	3.27	***
Arg	40E	N	...	U	7T	O4	...	3.25	***
Arg	40E	NH1	...	U	7T	O4'	...	2.82	***
Arg	40E	NH2	...	U	6T	O2	...	3.45	*
Arg	40E	O	...	C	8T	N4	...	2.99	***
Glu	42E	OE2	...	C	8T	N3	...	3.47	*
Arg	48E	NH1	...	C	8T	OP2	...	2.71	***
Arg	48E	NH2	...	C	8T	O4'	...	3.26	***
Trp	51E	NE1	...	G	9T	O2'	...	2.74	***



OPEN

Structural characterization of recombinant IAV polymerase reveals a stable complex between viral PA-PB1 heterodimer and host RanBP5

Received: 17 November 2015
Accepted: 04 April 2016
Published: 20 April 2016

Christopher Swale[1,2,3,*], Alexandre Monod[1,2,*], Laura Tengo[1,2,3], Alice Labaronne[1,2,3], Frédéric Garzoni[1,2,4], Jean-Marie Bourhis[1,2,3], Stephen Cusack[1,2,4], Guy Schoehn[1,2,3], Imre Berger[1,2,4,5], Rob WH Ruigrok[1,2,3] & Thibaut Crépin[1,2,3]

The genome of influenza A virus (IAV) comprises eight RNA segments (vRNA) which are transcribed and replicated by the heterotrimeric IAV RNA-dependent RNA-polymerase (RdRp). RdRp consists of three subunits (PA, PB1 and PB2) and binds both the highly conserved 3′- and 5′-ends of the vRNA segment. The IAV RdRp is an important antiviral target, but its structural mechanism has remained largely elusive to date. By applying a polyprotein strategy, we produced RdRp complexes and define a minimal human IAV RdRp core complex. We show that PA-PB1 forms a stable heterodimeric submodule that can strongly interact with 5′-vRNA. In contrast, 3′-vRNA recognition critically depends on the PB2 N-terminal domain. Moreover, we demonstrate that PA-PB1 forms a stable and stoichiometric complex with host nuclear import factor RanBP5 that can be modelled using SAXS and we show that the PA-PB1-RanPB5 complex is no longer capable of 5′-vRNA binding. Our results provide further evidence for a step-wise assembly of IAV structural components, regulated by nuclear transport mechanisms and host factor binding.

The influenza virus is a negative sense RNA virus with a segmented genome belonging to the *Orthomyxoviridae* family. The viral RNA (vRNA) is divided into 8 segments that encode for a total of 10 core proteins and a few additional accessory proteins. Every vRNA segment is coated by numerous nucleoprotein (NP) units and by one RNA-dependent RNA-polymerase (RdRp) complex composed of PA (716 aa for influenza A virus (IAV) strains), PB1 (756 aa for IAV strains) and PB2 (757aa for IAV strains). The RdRp binds to the 5′- and 3′- terminal extremities of the vRNA which acts as a promoter region[1–3]. The macromolecular assembly between the vRNA, NP and the RdRp forms the ribonucleoprotein (RNP). The RdRp performs both replication and transcription of the vRNA genome in the nucleus of the cell. As such, the RdRp can produce either cRNA/vRNA through *de novo* replication or viral mRNA through "cap snatching" transcription[4]. The factors regulating the switch from a transcriptive to a replicative RdRp remain to be characterised.

Recently, the structures of influenza B[5], bat influenza A[6] and influenza C[7] were published, providing tremendous insight into the complex architecture of the RdRp. All three subunits are tightly intertwined and form intricate quaternary structures at the vRNA promoter binding sites. Current models of RdRp assembly suggest that PB2 is imported via the importin-α pathway[8,9] whereas the PA-PB1 heterodimer is imported through direct interaction with RanBP5[10–12].

[1]Université Grenoble Alpes, Unit of Virus Host Cell Interactions, UMI 3265 UJF-EMBL-CNRS, 71 avenue des Martyrs, CS 90181, F-38042 Grenoble Cedex 9, France. [2]CNRS, Unit of Virus Host Cell Interactions, UMI 3265 UJF-EMBL-CNRS, 71 avenue des Martyrs, CS 90181, F-38042 Grenoble Cedex 9, France. [3]Institut de Biologie Structurale (IBS), Univ. Grenoble Alpes, CEA, CNRS, 38044 Grenoble, France. [4]EMBL Grenoble, 71 avenue des Martyrs, CS 90181, F-38042 Grenoble Cedex 9, France. [5]The School of Biochemistry, University of Bristol, Clifton BS8 1TD, United Kingdom. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.C. (email: thibaut.crepin@ibs.fr)

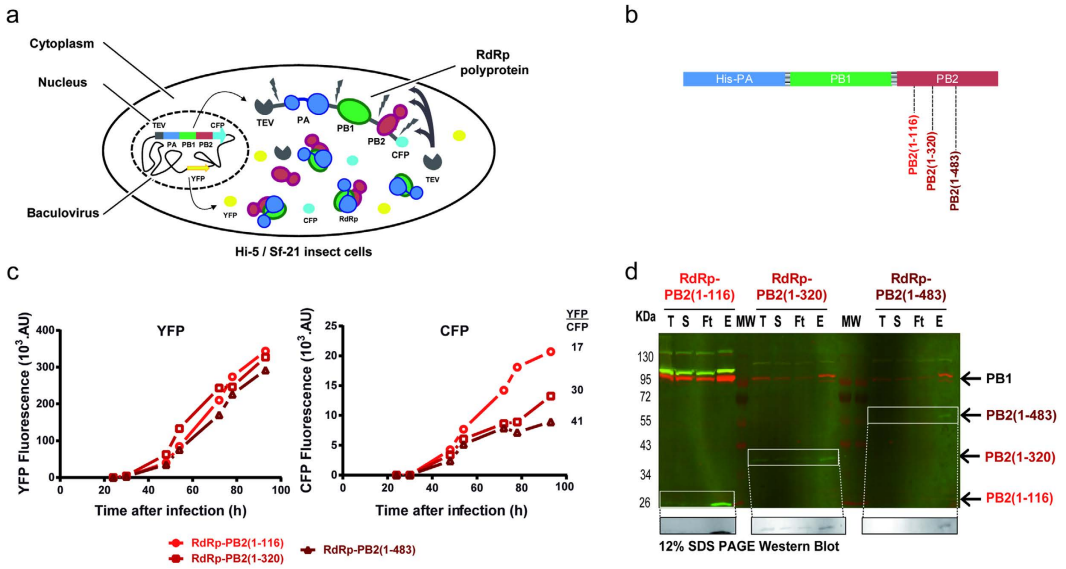


Figure 1. Truncated human-IAV RdRp polyprotein expression. (**a**) Logic of the TEV-PA-PB1-PB2-CFP polyprotein expression in a schematic view. During expression, TEV will process the polyprotein ensuring a stoichiometric assembly of PA, PB1 and PB2. YFP and CFP are produced during the process and monitor respectively baculovirus proliferation and polyprotein translation. (**b**) Truncated RdRp constructs where PB2 is incrementally extended until residues 116, 320 or 483. (**c**) YFP (left) and CFP (right) fluorescence kinetics measured during Hi-5 insect cells culture of truncated RdRp constructs. YFP ($\lambda_{exi} = 488$ nm, $\lambda_{emi} = 525$ nm) and CFP ($\lambda_{exi} = 430$ nm, $\lambda_{emi} = 480$ nm) measurements were performed on cellular extracts prepared by sonicating 1×10^6 cells in PBS (500 µL) follow by centrifugation. Fluorescence intensities are plotted against time after infection. (**d**) Small scale nickel resin purification analysis by western blot. Purifications were performed on the 50 mL of Hi-5 insect cells cultures used for the YFP and CFP fluorescence kinetics (**c**). Deposits feature total lysate after freeze/thaw (T), supernatant after centrifugation (S), resin flow through (Ft) and the primary elution fraction (E). After migration on a 12% SDS-PAGE, proteins were transferred on PVDF membrane. Primary antibodies targeting human-IAV PB2 (rabbit IgG) and human-IAV PB1 (mouse IgG) have been used. Revelation was performed with secondary goat antibodies coupled with Alexa Fluor 532 ($\lambda_{exi} = 632$ nm, $\lambda_{emi} = 647$ nm) and Alexa Fluor 633 ($\lambda_{exi} = 531$ nm, $\lambda_{emi} = 554$ nm) targeting mouse and rabbit H + L domains respectively, using a Typhoon Trio imaging system (GE Healthcare). After integration of the raw data, PB1 and PB2 revelation are visible in red and green respectively. Black and white signal of PB2 is also shown (bottom) to highlight the PB2 truncations. The upper bands appearing in green/yellow, correspond to unprocessed polyproteins.

Structural studies of RdRp from the human infecting influenza A (human-IAV) strains in contrast, have been limited to date, partly due to difficulties to produce recombinant H3N2 or H5N1 polymerases. Nonetheless, a cryo-EM reconstruction was recently obtained of a truncated influenza A/H5N1 polymerase[13]. In the present article, we identify PB2 as the main bottleneck hampering complete recombinant polymerase expression in our insect cells expression system. Moreover, we present the characterisation of several constructs of human-IAV RdRp, including the biologically relevant PA-PB1 heterodimer in solution. The titration measurements against the 5′- and 3′-vRNA promoters show a strong sub-nanomolar affinity of the PA-PB1 heterodimer towards the 5′-vRNA whereas the specific binding of the 3′-vRNA requires the presence of PB2. By applying our co-expression strategy, we show that the previously proposed PA-PB1-RanBP5 import complex can be purified, which we characterise using small angle X-ray scattering (SAXS). Biochemical analysis of this complex reveals a role for RanBP5 in hindering 5′-vRNA binding. Taken together, these data provide evidence for a cellular RdRp assembly pathway following a sequential and conditional process of assembly.

Results

PB2 is a key limiting factor for recombinant expression of the heterotrimer. Polyproteins are naturally used by certain viruses to structure their proteome[14,15]. Recombinant polyproteins mimicking viral polyproteins have recently emerged as a powerful tool to express protein complexes for functional characterisation as well as structural determination (reviewed in[16]). Using this strategy[17], soluble and active heterotrimeric RNA-dependent RNA-polymerases of influenza B virus (IBV) and bat influenza A (bat-IAV) virus could be produced, crystallized and the structure determined at high resolution[5,6]. We have applied a similar strategy to the RdRp of two human-IAV strains, A/Victoria/3/1975(H3N2) and the highly pathogenic A/Viet-Nam/1203/2004(H5N1). The three genes of each heterotrimeric complex were combined with Tobacco Etch Virus (TEV) protease and Cyan Fluorescent Protein (CFP) in a single large open reading frame (ORF). Each gene was separated by a DNA sequence encoding for a peptide segment comprising a short serine/glycine linker and a TEV protease cleavage site (Fig. 1a and supplementary Figure 1). The constructs were optimized for recombinant expression in insect

	Construct number	PA	PB1	PB2	RanBP5	YFP/CFP	Duplicate in H5N1	Solubility
heterodimers	1	1–716	1–686	–	–	6	✓	yes
	2	197–716	1–686	–	–	4	✓	yes
	3	197–716	1–660	–	–	4	✓	yes
	4	197–716	1–560	–	–	4	✓	no
	5	210–716	1–686	–	–	6	✓	no
	6	222–716	1–686	–	–	6	✓	no
	7	231–716	1–686	–	–	6	✓	no
	8	240–716	1–686	–	–	6	✓	no
	9	250–716	1–686	–	–	6	✓	no
	10	263–716	1–686	–	–	6	✓	no
	11	197–263	16–686	–	–	n.d.*	✓	no
heterotrimers	12	1–716	1–757	1–759	–	50	✓	n.d.
	13	1–716	1–757	1–36	–	15	✓	yes
	14	197–716	1–757	1–116	–	7	✓	yes
	15	1–716	1–757	1–116	–	17	✓	yes
	16	1–716	1–757	1–250	–	28	✓	n.d.
	17	1–716	1–757	1–320	–	30	✓	n.d.
	18	1–716	1–757	1–483	–	41	✓	n.d.
	19	1–716	1–757	1–116 – MBP	–	17	✓	yes
	20	1–716	1–757	1–116 – 320–483	–	20	✓	yes
	21	1–716	1–686	–	1–1115	12	–	yes

Table 1. Details of human-IAV RdRp polyproteins. The expressions have been made on both A/Victoria/3/1975(H3N2) and A/Viet-Nam/1203/2004(H5N1) strains. Identical results were obtained with the 2 strains. The table provides the data obtained with the A/Victoria/3/1975(H3N2) strains. The ratio YFP/CFP were calculated with the 2 maxima of the fluorescence spectra measured on the same sample (YFP: $\lambda_{exi} = 488$ nm, $\lambda_{emi} = 525$ nm ; CFP: $\lambda_{exi} = 430$ nm, $\lambda_{emi} = 480$ nm) corresponding to 1×10^6 infected cells. This ratio is proportional to the level of expression of the corresponding constructs. The table contains also data on the solubility of each construct in the classical purification buffer (i.e.: 50 mM Tris-HCl pH 8.5, 300 mM NaCl, 2 mM β-mercaptoethanol, 2–10% glycerol). *This construct has been cloned in pFastBac-HTB without TEV nor CFP. The only fluorescent reporter protein was the YFP. Its signal was in accordance with the YFP values obtained for the other constructs.

cells using the MultiBac system[18,19]. During expression, the TEV protease cleaves the polyprotein co-translationally, resulting in a stoichiometric assembly of the RdRp that can then be selectively purified using an affinity nickel resin. The production of the polyprotein is monitored using the fluorescence of two reporter proteins: CFP encoded within the polyprotein reports directly on recombinant protein yield, while YFP, integrated in the baculovirus genome, monitors virus performance[19]. The ratio of YFP/CFP in our system is a highly useful criterion to determine recombinant polyprotein translation efficacy of the different constructs.

Table 1 summarizes the results obtained with all constructs tested. Several attempts were made to express the H3N2 polymerase heterotrimer (construct number 12, Table 1). Although YFP and CFP can be detected and quantified, CFP values were extremely low compared to YFP with a YFP/CFP ratio of 50, the highest observed for any polyprotein constructs that we expressed. No clear bands were observed on western blot using specific PA, PB1 and PB2 antibodies. Similar results were obtained with the H5N1 RdRp polyprotein, indicating failure of these recombinant expression experiments for these full-length IAV heterotrimeric polymerases. As a comparison, using the same strategy, the YFP/CFP ratio for the expression of IBV or Bat-IAV heterotrimer was close to 20, similar to the ratios we obtained for the expressions of constructs 13 and 15 (i.e. without most of PB2).

In order to understand the difference between the expression of the human-IAV and IBV polymerases, we designed new constructs with a C-terminally truncated PB2 based on available structural information. These were PB2(1–36)[20], PB2(1–116)[21] and based on the cap-binding domain[22], PB2(322–483). Expression and purification experiments were undertaken on constructs 15, 17 and 18, all of which all have full-length PA, PB1 and truncations of PB2 extending to residues 116, 320 and 483 respectively (Fig. 1b). As the PB2 extension increases beyond residue 116, we observed a loss of expression of the polyprotein, thus identifying a critical region in PB2 that limits yield. This can first of all be observed through the monitoring of YFP/CFP reporter genes during expression (Fig. 1c). YFP fluorescence values reach a similar plateau for all constructs whereas CFP values are reduced significantly for constructs 17 and 18. Moreover, after nickel resin pull-down on lysates from an equal amount of cells, we observed by western blot (Fig. 1d) decreased intensity bands of not only PB2 but also PB1, indicative of an overall expression loss. Polyprotein size increase alone could not explain the reduced expression because other control constructs (19 and 20) with additional protein sequences beyond PB2 residue 116 conserve lower YFP/CFP ratios. Furthermore, replacing the PB2(1–116) by another large protein such as RanBP5 (residues 1–1115) generates a construct with a lower YFP/CFP ratio than construct 15. The analysis of the amino

acids sequences of PB2 does not provide more indication on the real nature of this phenomenon (supplementary Figure 2).

To determine whether this "loss-of-expression" phenomenon results from our polyprotein expression strategy, which combines PA-PB1 and PB2 in one ORF, we generated a MultiBac baculovirus co-expressing a PA-PB1 fusion and PB2 from an independent expression cassette[23]. This strategy also resulted in high YFP/CFP ratios ($\geq$35), even though the CFP gene was only fused to PA-PB1. These results identify PB2, especially when extended beyond the residue 116, as the bottleneck for expression of complete human-IAV RdRp.

Dissecting the heterotrimer. In order to identify a minimal active core of the human-IAV polymerase, we have systematically dissected the heterotrimeric complex for further characterisation using the polyprotein strategy. The expression of PA-PB1 increases significantly when PB2 is totally removed (compare constructs 1 and 13/15; Table 1). The same observation is made after removing the endonuclease domain (i.e. PA-Nter; compare constructs 14 and 13/15). After purification, PA-PB1 forms homogeneous monomeric and stable particles (Fig. 2b) whereas the complexes with PB2 (constructs 15 and 16) give rise to dimers in solution (Fig. 2a). The dimers are stable enough to withstand salt concentrations up to 2 M NaCl during purification, but the SEC elution peaks are broad suggesting that the oligomerization process is quite dynamic. By adding vRNA promoter-like molecules (e.g. the IAV panhandle)[24], we found that the dimers dissociated into monomeric RNA:protein complexes. We were able to show that the truncated PA-PB1 (construct 1) and the PA-PB1-PB2 (construct 14) exhibit expected polymerase activities. The endonuclease activity is similar to that of the isolated PA-Nter domain (supplementary Figure 3), can be inhibited by the same point mutation (i.e. PA-E80A) or compound (i.e. DPBA) and shows comparable dependency on manganese[25,26]. The constructs are also functional in RNA synthesis. Upon addition of a mixture of IAV panhandle, NTPs and [α-^{32}P]-UTP, both truncated PA-PB1 and PA-PB1-PB2 are able to generate an 80-nucleotide long product plus other minor products in absence or in presence of ApG (supplementary Figure 4).

We have also shown that further N-terminally truncating PA and/or C-terminally truncating PB1 has no impact on the expression of the corresponding polyproteins (constructs 1 to 11). The heterodimer PA-PB1 can be purified with or without the N-terminal domain of PA (i.e. PA-Nter; compare constructs 1 and 2). However, as soon as the PA-hinge (residues 200–260) is removed, PB1 becomes insoluble. This implies that the PA-hinge linking the Nter- and Cter-domains of PA is crucial for the stability of PB1. However, it is not sufficient in itself since the C-terminal region of PA also interacts with the N-terminal part of PB1[27,28] and cannot be deleted either (compare constructs 2 and 5 to 11). PB1 can be shortened on its C-terminus but when purifying the construct 4 (i.e. PB1 until amino acid 560), only PA-Cter remains soluble. Thus, the minimal construct required to obtain soluble PB1 is construct 3 with PA from 197 to 716 and PB1 from 1 to 660. All the PA-PB1 constructs without the N-terminal domain of PA form dimers in solution (supplementary Figure 5) and adding vRNA-like molecules has no incidence on the dimerization.

Over all the constructs tested, crystals have only been obtained using construct 2 (PA from 197 to 716 and PB1 from 1 to 686). Crystals grew in a few days in low PEG content solutions. They diffract poorly, with diffraction limited to typically $\approx$ 10 Å.

PA-PB1 dimer exhibits high-affinity interaction with 5′-vRNA but requires PB2 for binding 3′-vRNA. All influenza RNA segments have the same organization, a central coding region flanked by 2 un-translated regions containing the highly conserved and complementary 5′- and 3′-ends[29–31]. The viral polymerase specifically interacts with both the 5′- and 3′-ends, as recently visualized in crystal structures[5,6], and uses them as a promoter[1–3]. During their biochemical characterisation, we have seen significant effects of vRNA-like RNA molecules on the stability and/or the oligomeric state of most of the constructs. The initial RNA molecule used was the 80-nt panhandle[24] from which we made the 5′-vRNAp corresponding to the 5′-end (5′-AGUAGAAACAAGGGUA) and it's 3′ equivalent called 3′-vRNAp (AUACCCUGCUUUUGCU-3′). By thermal shift assay experiments, a 10 °C stabilisation was observed when 5′-vRNAp was added to all the constructs, whereas when 3′-vRNAp was added, the effect was less significant and dependent on the construct. Similar data had already been published[32].

We used fluorescence anisotropy to measure the interaction between our truncated polymerase constructs and the conserved vRNA ends by following the fluorescence polarization increase of a fluorescently labelled RNA when it binds the polymerase. For this purpose, we used 5′-vRNAp and 3′-vRNAp, both labelled with fluorescein amidite (FAM) at the opposite extremity of the putative interaction, i.e. at the 3′-end for the 5′-vRNAp and *vice versa* (Fig. 3 and Table 2). At 300 mM NaCl, the Kd for the 5′-vRNAp are similar (sub-nanomolar) for both constructs 1 and 14, whereas for the 3′-vRNAp, a Kd of 36 nM for the construct with PB2 is obtained (Fig. 3a) and no binding was observed for the construct without PB2 (Fig. 3c). As the estimated Kd for the 5′-end was in a range below the working concentration and beyond the sensitivity limit of the fluorescence anisotropy measurement, another method was required to determine the 5′-vRNAp affinity. Filter binding assays (FBA) were then therefore used for precise Kd determination (Fig. 3b). With the corresponding radioactive probes, we found values of 0.2–0.4 nM for the 5′-end. For the 3′-end, the Kd was 100-fold higher and critically dependent on the presence of PB2. Salt concentration is also an important parameter. We show different binding characteristics at 150 and 300 mM NaCl of the PA-PB1 towards different RNAs. The Kd value of PA-PB1 for the 5′-end remains sub-nanomolar in the range of 150 to 300 mM NaCl. In contrast, at 150 mM NaCl, 3′-vRNAp binds to PA-PB1 with the same affinity than for a poly-UC (Fig. 3c), and so is non-specific. All relevant titrations were performed at 300 mM NaCl to prevent non-specific affinity measurements. Recently published affinity data[33] working at 500 mM NaCl likewise consistent with very high affinity of the 5′-vRNAp towards IAV polymerase but this study failed to measure 3′-vRNAp binding within the nanomolar range, indicating that 3′-vRNAp binding is also tightly influenced by salt concentration.

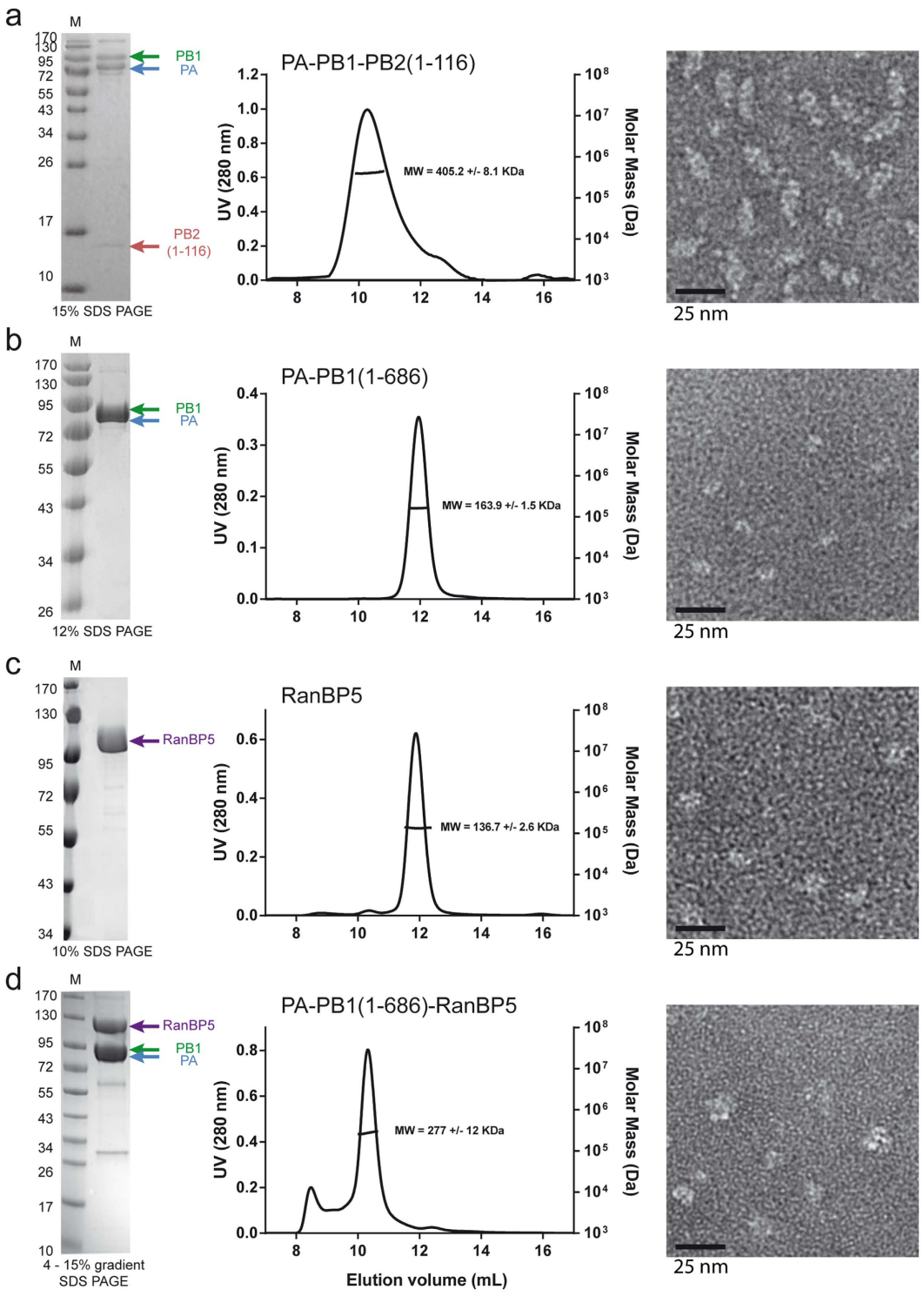


Figure 2. Homogeneous sample analysis of recombinant polymerases, RanBP5 and polymerase-RanBP5. Purified product analysis are horizontally grouped for (**a**) PA-PB1-PB2(1-116), (**b**) PA-PB1(1-686), (**c**) RanBP5 and (**d**) PA-PB1(1-686)-RanBP5. On the left are the Coomassie blue colored SDS PAGE gel of the purified sample with corresponding subunits bands indicated by colored arrows (PA in blue, PB1 in green and RanBP5 in purple). In the middle are the SEC-MALLS chromatograms with the UV signal as a backtrace and molecular weight estimate below the peak. Estimated average molecular weight for each sample is also detailed. SEC-MALLS-RI runs of PA-PB1(1-686), RanBP5 and PA-PB1(1-686)-RanBP5 were performed in the same buffer (50 mM Tris-HCl pH 8.0 and 150 mM NaCl) whereas the SEC-MALLS-RI run of PA-PB1-PB2(1-116) was performed with 50 mM Tris-HCl pH 8.0, 300 mM NaCl and 10% glycerol. (right) corresponding electron microscopy images.

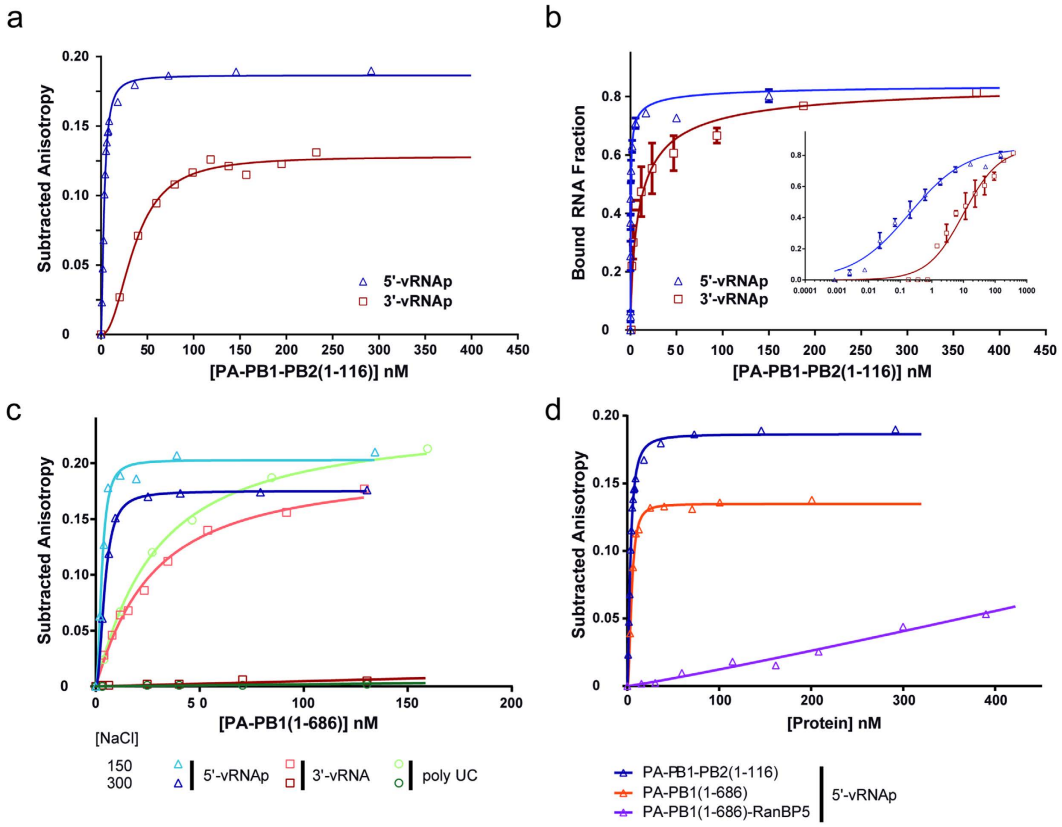


Figure 3. vRNA binding and specificity. (**a**) Binding titration of the truncated trimer PA-PB1-PB2(1-116) towards the 5′-vRNAp (blue triangle) and 3′-vRNAp (red square) sequences using fluorescence anisotropy at 300 mM NaCl. (**b**) Binding titration performed by filter binding assay against P^{32} labelled 5′-vRNAp (blue triangle) and 3′-vRNAp (red square) using 300 mM NaCl. Bound RNA fraction is plotted as a function of polymerase concentration. (**c**) Binding titration of the truncated dimer PA-PB1(1-686) performed at 150 and 300 mM NaCl against the 5-′vRNAp (dark and light blue triangles), 3′-vRNAp (orange and dark red squares) and polyUC RNA (light and dark green circles) by fluorescence anisotropy. (**d**) Binding titration of different polymerases and polymerase-RanBP5 constructs against the 5′-vRNAp at 300 mM NaCl by fluorescence anisotropy. PA-PB1-PB2(116) and PA-PB1(1-686) are depicted by blue and orange triangles respectively, PA-PB1(1-686)-RanBP5 is depicted with purple triangles. For all anisotropy titrations (**a**,**c**,**d**) subtracted anisotropy is plotted as a function of protein concentration.

Construct	NaCl (mM)	RNA	Kd (nM)	std error (nM)	R^2
construct **14**:PA-PB1-PB2(1-116)		5′-vRNAp	0.83 (0.19)	0.08 (0.02)	0.99 (0.97)
	300	3′-vRNAp	36 (17)	2 (2)	0.99 (0.95)
		polyUC	≥1000	n.d.	n.d.
construct **1**:PA-PB1(1-686)		5′-vRNAp	0.87 (0.38)	0.1 (0.05)	0.99 (0.99)
	300	3′-vRNAp	≥1000	n.d.	n.d.
		polyUC	≥1000	n.d.	n.d.
		5′-vRNAp	0.53	0.12	0.99
	150	3′-vRNAp	22.1	1.0	0.99
		polyUC	22.8	0.7	0.99
construct **21**:PA-PB1(1-686) RanBP5	300	5′-vRNAp	≥1000	n.d.	n.d.
NP	300	polyUC	29	1.2	0.94

Table 2. Titration measurements against vRNA promoters. Values in parentheses correspond to the values obtained by filter-binding assay experiments. All the experiments have been made in triplicate.

RanBP5 interacts tightly with PA-PB1(1-686) when co-expressed. During the viral cycle, the assembly of influenza RdRp follows a multi-factorial pathway involving many host partners proteins. After transcription in the nucleus, viral mRNAs are exported to the cytoplasm to be translated by the cellular machinery. The components of the replication machinery (i.e. PA, PB1, PB2 and NP) then must be imported into the nucleus.

	construct 1:PA-PB1(1-686)	construct 21:PA-PB1 (1-686)-RanBP5	construct 22 :RanBP5
Data collection parameters			
Instrument	ESRF - BM29		
Beam size at sample (μm)	700 × 700		
Wavelength (Å)	0.9919		
q range ($Å^{-1}$)	0.25–50		
Detector	Pilatus 1 M		
Detector distance (m)	2.867		
Exposure (s per image)	1		
Column	S200inc 5/150 GL		
Flow rate ($mL.min^{-1}$)	0.5	0.4	0.5
Injected sample concentrations ($mg.mL^{-1}$)	3.4	4.3	8.5
Injection volume (μL)	50		
Temperature (K)	293		
Structural parameters			
R_g (Å) [from $P(r)$]	37.9 ± 0.1	52.2 ± 0.1	39.4 ± 0.2
R_g (Å) [from Guinier]	36.2 ± 0.4	51.8 ± 0.5	38.8 ± 0.8
D_{max} (Å)	128	181	136
Porod volume estimate ($Å^3$)	254 630	577 070	215 890
Molecular-mass determination			
Molecular mass M_r (Da) [from Rambo]	146 493	323 311	143 827
Calculated M_r (Da) from sequence	165 915	291 983	125 892
Software employed			
Primary data reduction	PRIMUS		
Data processing	PRIMUS		
Ab initio analysis	DAMMIF		
Validation and averaging	DAMAVER & DAMMIN/MONSA		
Computation of model intensities	CRYSOL		
3D graphics representations	CHIMERA		

Table 3. SAXS data-collection and scattering-derived parameters.

Whereas PB2 and NP use importins-α[9,34–37], PA and PB1 are conjointly imported as a preformed heterodimeric submodule by the importin-β RanBP5[10,12,38]. Having confirmed the existence of a stable PA-PB1 complex we sought to demonstrate its interaction with RanPB5. The human *IPO5* gene which encodes for RanBP5, was expressed in insect cells. After purification, RanBP5 forms a homogeneous monomeric sample (Fig. 2c) that has been used in mixing experiments with freshly purified PA-PB1(1-686) followed by SEC-MALLS-RI experiments. The two samples were mixed in stoichiometric ratio and incubated before size exclusion chromatography (SEC). Both RanBP5 and PA-PB1(1-686) were eluted in the same single peak, but MALLS-RI indicated a molecular weight corresponding of a mixture rather than a stable ternary complex (supplementary Figure 6). We then attempted to produce a PA-PB1(1-686)-RanBP5 heterotrimer by the self-processing polyprotein strategy. A long ORF encoding for the trimeric complex was created by inserting the *IPO5* gene downstream of the PB1(1-686) coding sequence (Table 1, construct 21). Remarkably, from this polyprotein construct, PA-PB1 co-purifies with RanBP5, forming a homogeneous and stable heterotrimer (Fig. 2d).

Online SEC-SAXS was used for the characterisation of PA-PB1(1-686), RanBP5 and PA-PB1(1-686)-RanBP5, respectively in solution. Using the Vc determination method[39] on the diffusion data of PA-PB1(1-686) shows a Mw estimate of 146 kDa (Table 3 and supplementary Figure 7). The calculated Mw is 166 kDa. The Guinier transform measures a hydrodynamic radius (Rg) of 36.2 Å and GNOM produces the pair distribution function fit with a Dmax of 128 Å. Further analysis of the SAXS curve (Fig. 4a) shows a visually adequate correlation with the CRYSOL curve which uses the bat-IAV polymerase structure (PDB id: 4WSB) as a model. To evaluate the statistical similarity between experimental intensities and those computed from a model has been derived. The high chi^2 values are observed for this dataset, due to a noisy SAXS curve. A slight deviation of the fit is observed above a q range of 0.8 nm^{-1}, suggesting conformational differences between the scattering curve and the crystal structure coordinates. *Ab-initio* modelling was performed using 15 DAMMIF models which were averaged by DAMAVER. Averaged model correlation with the diffusion curve was then checked using the damstart file as a starting envelope for DAMMIN. The polymerase DAMAVER envelope (Fig. 4a) appears as a pear shaped structure in which the homologous model can be fitted. Additional envelope volume is visible close to the endonuclease domain, implying that it adopts multiple conformations in solution.

Using the same methodology, we determine RanBP5 to be a monomer in solution with a measured Mw of 144 kDa close to the theoretical Mw of 126 kDa (Table 3 and supplementary Figure 7). RanBP5 displays an Rg of 38.8 Å and a Dmax of 136 Å. With a Mw lower than that of PA-PB1(1-686), RanBP5 displays a larger Dmax and Rg indicating an extended structure. We have looked over the *Protein Data Base* (PDB) and found yeast

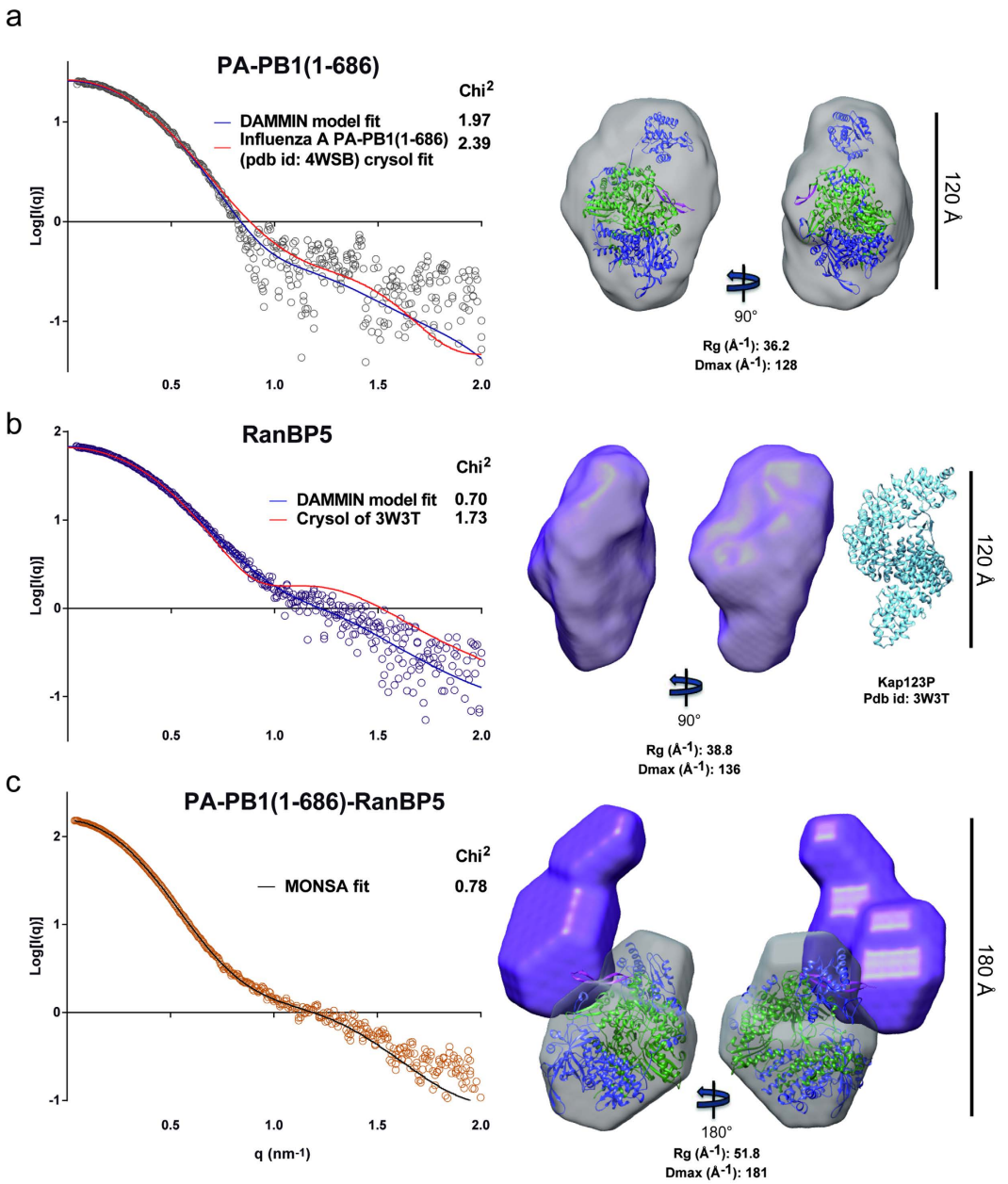


Figure 4. Online SAXS analysis of PA-PB1, RanBP5 and PA-PB1-RanBP5 complex in solution. Results are shown for (**a**) PA-PB1(1-686), (**b**) RanBP5 and (**c**) PA-PB1-RanBP5. On the left are the experimental Ln[I(q)] curves with the *ab-initio* DAMMIN curve fit (blue) using the Damstart (DAMAVER) as an initial constraint and the CRYSOL fit (red) of the closest homologous X-ray structure. The MONSA fit is also shown for PA-PB1-RanBP5 in black. The Chi^2, to evaluate the statistical similarity between experimental intensities and those computed from a model, are also shown. On the right, the homologous PDB structure of PA-PB1(1-686) from Bat-IAV polymerase is depicted in cartoon and fitted in both the DAMAVER and the MONSA generated envelops of PA-PB1(1-686) which are depicted in grey. PA is coloured in blue, PB1 in green and the NLS of PB1 in magenta. The DAMAVER and MONSA envelop of RanBp5 is coloured in purple. As a comparison, the atomic structure of Kap123P (pdb is: 3W3T) is depicted in cartoon with a sky blue colour alongside the DAMAVER envelop of RanBP5. All envelopes and models are adjusted at the same size scale.

karyopherin Kap121[40] to be the closest homologue available for RanBP5 (30% sequence identity over 1027 residues). CRYSOL curve fitting using the apo form of Kap121 (PDB id: 3W3T) against the diffusion curve reveals important structural and conformational differences between the two structures (Fig. 4b). However, envelope modelling confirms an elongated structure of RanBP5 in solution, comparable in size with yeast Kap121 in the crystal structure.

Analysis of PA-PB1(1-686)-RanBP5 describes a significantly larger complex in solution with an Rg of 51.8 Å and a Dmax of 181 Å. The Mw estimate of 323 KDa is close to the expected Mw of 292 KDa, confirming the

presence of a stoichiometric 1:1:1 PA-PB1(1-686)-RanBP5 complex in solution. MONSA builds *ab initio* shapes using the scattering curves for complexes together with those of their individual components[41]. MONSA was used to perform the *ab-initio* modelling of the complex PA-PB1(1-686)-RanBP5 (supplementary Figure 8). It separately uses the PA-PB1(1-686) and RanBP5 diffusion curves (Fig. 4a,b) in combination with that of PA-PB1(1-686)-RanBP5 to propose a consensus dual envelope (Fig. 4c). Model fitting was performed against the three experimental curves and matches all 3 curves with similar quality. With this *ab-initio* modelling we also find a comparable envelope for PA-PB1(1-686) like the one produced with DAMMIF/DAMAVER. RanBP5 on the other hand displays a different envelope suggesting conformational differences between the unbound and bound forms of RanBP5. The model suggests a close interaction between the two proteins, RanBP5 (violet envelope) interacts with PA-PB1(1-686) (grey envelope), suggesting molecular contacts well beyond those of the proposed PB1 nuclear localization signal (NLS) containing domain, which the crystal structure shows is a mobile and solvent exposed β-ribbon[6]. This model further supports the hypothesis that RanBP5 may play a chaperone function for PA-PB1 prior to assembly with PB2 in the nucleus to complete the polymerase.

RanBP5 regulates the binding of the 5′-vRNAp. The PA-PB1 complex not only exists as a stable submodule but can moreover recognise the 5′-vRNAp with a high affinity and acts as a replicase and endonuclease on its own, without PB2 present. The current model predicts the NLS of PA-PB1 to be present within PB1 (residues 187–211)[12], at the end of a beta-strand located parallel to the vRNAp binding sites, indicating a potential link between the binding to RanBP5 and/or the binding to vRNAp.

Subsequently to measuring the sub-nanomolar Kd of PA-PB1(1-686) and PA-PB1-PB2(1-116) to the 5′-vRNAp, an identical experiment was performed with the PA-PB1(1-686)-RanBP5 complex (Fig. 3d). The association curve clearly shows a complete loss of binding affinity towards the 5′-vRNAp when RanBP5 is present, both at 300 mM and 150 mM NaCl. Using SEC-MALLS-RI experiments, we have shown that an excess of 5′-vRNAp does not induce any complex dissociation of the PA-PB1(1-686)-RanBP5 trimer (supplementary Figure 9). Taken together, these experiments suggest that RanBP5 binding obscures the 5′-vRNAp binding site, thus blocking specific 5′-vRNAp binding.

Discussion

Expression of full-length IBV and bat-IAV polymerases has been successfully achieved using a polyprotein approach, resulting in atomic resolution structures by X-ray crystallography[5,6]. In marked contrast and unexpectedly, efficient expression of active full-length human or avian IAV polymerase at yields high enough for structural studies has not been reported to date, irrespective of the method used. Our studies identify PB2, and notably residues 120 to 250 as responsible for this poor expression. Interestingly, attempts by a different research group to reconstitute a full IAV polymerase (e.g. full IAV polymerase expression/purification for structural studies) likewise failed beyond the N-terminal PB1 interacting domain of PB2[13]. Furthermore, the recent bat-IAV, IBV and ICV RdRp crystal structures have shown that PB2 can adopt a plethora of conformations, demonstrating how dynamic PB2 is within the heterotrimer[5–7,42,43]. The reason why PB2 hinders polyprotein expression is not clear at the moment. We speculate that the extended fold of the PB2 region 1–250 is unstable unless bound to one side of PB1[5,6] and thus may require chaperones to maintain its stability prior to assembly onto PA-PB1. Of note, Hsp90 has been shown to stimulate influenza virus RNA synthesis[44] and the nuclear import of the subunits[35]. Moreover, Hsp90 inhibitors reduce influenza virus replication in cell culture[45]. We have shown previously that provision of co-factors in *trans* can ameliorate the quality and quantity of target proteins expressed in the baculovirus/insect cell system[46,47]. Provision of Hsp90 and the other chaperones, within the polyprotein or in *trans* on the baculovirus backbone, may provide a powerful handle to overcome the bottleneck of human-AIV polymerase production. Furthermore, although the cellular context between viral infection and the expression of recombinant proteins is different, it cannot be excluded that the polyprotein interferes with the cellular innate immunity. IAV RdRp, and PB2 in particular, has been shown to regulate host anti-viral response through the binding to IPS-1 and the inhibition of interferon β production[48].

PA-PB1 can be produced as a stable submodule and forms a discrete heterodimer in solution. Dimerization of PA-PB1 heterodimers can appear when either the endonuclease is removed or the N-terminal of PB2 is added which may indicate that the superstructure formation could be artificially induced by truncations of the full-length polymerase. PA-PB1 can specifically bind to the 5′-vRNAp with sub-nanomolar affinity in solution. In contrast, 3′-vRNAp binding requires the presence of PB2 and is within the low nanomolar range, indicating a much stronger preference of the polymerase towards the 5′-vRNAp. These results are consistent with the high-resolution crystal structures in which the 5′-vRNAp binding site is located between the PA and PB1 subunits[5,6]. The 3′-vRNA, on the other hand, interacts with all three subunits suggesting a sequential vRNAp binding mechanism.

Our results corroborate key steps of influenza polymerase assembly[10,38,49]. Purification of stoichiometric complexes using the polyprotein strategy confirms that PA-PB1 can exist as a preformed and stable submodule, independently of additional host factors, with PA acting as a putative "chaperone" for PB1. Our data also shows that PA-PB1 is functional in RNA synthesis, which has interesting physiological implications. Moreover, we demonstrate here that the PA-PB1(1-686) heterodimer can assemble with RanBP5 into a stable, heterotrimeric complex that can be purified *in vitro*. This likely corresponds to the import complex for these two subunits, although other chaperones such as HSP90 may be co-imported[35]. In contrast to PA-PB1 with RanBP5, we failed to reconstitute a binary PB2-importin-α complex using the polyprotein strategy even though functional and structural data show that PB2 nuclear import depends on the importin-α pathway, through a direct interaction between the PB2 C-terminal domain that contains the NLS[8,9]. This is probably partly due to an inherent instability of isolated IAV PB2 in insect cells but could also reflect the lack of sufficient chaperones, co-chaperones or other unidentified factors that stabilise PB2. Furthermore, cytoplasmic expression of PB2 appears to destabilise

many cellular processes[50,51]. We also still do not know the reason for the difference in the import pathway usage between the PA-PB1 submodule and PB2. The lack of high expression levels of fully assembled PA-PB1-PB2 heterotrimers unfortunately currently limits our efforts to elucidate the structure of the complete human IAV polymerase at atomic resolution. Nevertheless, docking of bat-IAV and IBV crystal structures into the PA-PB1(1-686)-RanBP5 SAXS envelope suggests that assembly of PB2 onto the PA-PB1-RanBP5 complex might be possible, based on steric considerations. Indeed RanBP5 could even facilitate trimer assembly perhaps concomitantly with Ran-dependent dissociation from PA-PB1. In contrast, RanBP5 profoundly affects the ability of PA-PB1 to bind vRNA. We show that whereas the 5′-vRNA is specifically recognized by PA-PB1 with high affinity, this is abolished in the ternary complex with RanBP5. This again suggests that RanBP5 disassociation may be coupled to assembly of the polymerase with the promoter RNA. In any case, only when PB2 associates is the polymerase able to associate with the 3′-vRNA end, consistent with the crystal structures[5,6]. But surprisingly, the affinity of the heterotrimer for 3′-vRNA is in the same order of magnitude than the affinity of NP for RNA substrates (Table 2), revealing a potential competition between the complete polymerase heterotrimer and the nucleoprotein.

We demonstrate here for the first time that the polyprotein expression strategy can be used not only to co-express subunits that form a stable complex, but also to successfully reconstitute interactions with downstream partner proteins *in vivo* in the expression host for direct purification and functional characterisation. Functional studies on IAV RdRp reveal that vRNAp promoter binding is completely abolished when PA-PB1(1-686) is bound to RanBP5, suggesting a possible role for RanBP5 in maintaining PA-PB1 in an inactive form during nuclear import. This discovery holds a promise for influenza research as it provides a novel point of therapeutic intervention for targeting and inhibiting influenza virus during assembly of its active components. We anticipate that compounds which block polymerase/RanBP5 binding or release may prove to be highly efficient, broad spectrum antiviral inhibitors in the treatment of influenza, in particular for patient groups where classical vaccination strategies fail.

Methods

Molecular biology. To avoid expression of alternative proteins, all the genes used for this work are synthetic genes, codon optimised for insect cells expression. The DNA coding sequences of A/Victoria/3/1975(H3N2), the highly pathogenic A/Viet-Nam/1203/2004(H5N1) polymerase subunits and human *IPO5* have been ordered to GeneArt (ThermoFisher Scientific), optimized for the expression in insect cells. Cloning has been achieved following the supplier procedures (New England Biolabs). The pPBAC plasmid was used for the polyprotein constructs[16,18]. Few constructs have been expressed using pFastBac-HTB (Life Technologies). vRNA-like molecules (i.e. IAV panhandle)[24] have been produced and purified using classical *in vitro* transcription protocols[52,53].

Expression and purification. Large scale suspension cultures expressing polymerase fusion constructs were prepared using High Five insect cells grown in Express Five media (Life Technologies) at 0.5×10^6 cells/mL infected at 0.2% (V/V) with the baculovirus mother solution. Cultures were maintained at $0.5–1 \times 10^6$ cells/mL until proliferation arrest (24–48 h after infection). Following the proliferation arrest YFP and CFP measurements were performed every 12 h until a fluorescence plateau was reached (72–96 h after infection). Cultures were then spun down at 800 g for 10 min and cell pellets were stored at −80 °C.

Cell pellets were resuspended in 50 mL of lysis buffer (50 mM Tris-HCl pH 8.5, 300 mM NaCl and 2 mM β-mercaptoethanol) per 500×10^6 cells in the presence of EDTA-free anti-protease cocktail (complete from Roche). Lysis was performed with two cycles of freezing (−180 °C)/thawing (26 °C) after which 10% of glycerol were added to the lysate before centrifugation (45 min, 40 000 g, 4 °C). After retrieval of the clarified lysate, 30 mM of imidazole pH 8.0 were added before loading on Ni-NTa superpose resin (Quiagen). After flowing the lysate through the resin, three wash steps with 10 column volumes (CV) of buffer A (lysis buffer with 30 mM imidazole), 10 CV of wash buffer B (50 mM Tris-HCl pH 8.5, 1 M NaCl, 10% glycerol and 2 mM β-mercaptoethanol) and 10 CV of wash buffer A were performed. Elution of the bound complex was performed with 300 mM imidazole. Elution fractions containing polymerase were then pooled together and directly injected on a 5 mL Hitrap heparin resin (GE healthcare) which had previously been equilibrated with 5 CV of buffer A. After binding to the resin, a 5 CV wash was performed with buffer A before eluting with a 40 mL salt gradient on a FPLC system. The elution peak corresponding to stoichiometric polymerase assemblies was then pooled and injected on an S200 (GE healthcare) size exclusion chromatography equilibrated in 50 mM Tris-HCl pH 8.5, 300 mM NaCl, 5 mM β-mercaptoethanol. Peak fractions were pooled and concentrated to the desired concentration using a 100 kDa concentrator. Once concentrated the protein prep could be stored by addition of 20% glycerol and flash freezing in liquid nitrogen.

NP was purified using the protocol described in[54,55].

SEC-MALLS-RI analysis. All MALLS runs were performed using a S200 increase SEC column (10/300 GL, GE Healthcare). Sample injection and buffer flow was controlled by a Hitachi L2130 pump, following the SEC column was a L-2400 UV detector (Hitachi), Optilab T-rEX refractometer (Wyatt technologies) and a DAWN HELEOS-II multi angle light scattering detector (Wyatt technologies). Prior to injection, columns and systems were equilibrated in 5 to 10 column volumes of running buffer. 50 μL injections were performed using protein samples concentrated at a minimum of 2 $mg.mL^{-1}$, constant flow rate of 0.5 $mL.min^{-1}$ was used. Accurate MALLS mass prediction was performed with the Astra software (Wyatt Technologies). Curves were represented with Graphpad (Prism).

Electron microscopy. Samples (0.1 $mg.mL^{-1}$) were applied to the clean side of carbon on mica. The carbon was then floated on sodium silicotungstate (2% v/w) or uranyl acetate (2% v/w) and a grid placed on the top of

it. After air-drying, the samples were observed in a T12 FEI electron microscope and images were taken using an Orius SC1000 CCD camera (Gtan Inc., Pleasanton, CA).

SAXS analysis. All datasets were collected on BM29 (ESRF). The initial trials using direct sample measurement showed a slight concentration dependency of the Guinier estimated Rg. Online SEC-SAXS was therefore used for all experiments in order to limit the contribution of partially aggregated, oligomeric or dissociated protein to the experimental curve. The experimental setup consists of a High Pressure Liquid Chromatography (HPLC) system connected to an analytical S200 increase column (5/150 GL, GE Healthcare) followed downstream by the SAXS sample capillary. SAXS measurements were performed every second with a Pilatus 1 M detector at distance of 2.87 m allowing a q range of 0.03 to 4.5 nm with a wavelength of 0.01 nm.

Following data collection, experimental curves were subtracted and analysed using Primus (ATSAS) or Scatter (Bioisis) programs suits[56]. To verify the molecular mass, the Rambo and Tainer method was used[39,57]. Rg predictions using the Guinier extrapolation were plotted against the elution volume to select the most monodisperse part of the protein elution peak. SAXS datasets within this zone were then scaled and averaged to produce one unique I(q) curve. GNOM[58] was then used to produce a P(r) function which was subsequently used by DAMMIF[59] to generate 15 *ab-initio* models. DAMAVER[60] was then used to generate an average model. Consistency of the DAMAVER averaging with the original experimental curve was assessed by using DAMMIN[41] and using the damstart.pdb envelope as a starting model. Homologue PDB structure comparison was assed using Crysol[61]. Multiple diffusion curve *ab-initio* modelling was performed using Bunch[62]. Homologue structure fitting within the DAMAVER envelope was performed with Chimera[63] and curve representations using Graphpad (Prism).

Fluorescence anisotropy assay. Equilibrium RNA binding experiments were performed with the following vRNA set synthetized by Integrated DNA Technologies (IDT): 5′-vRNAp (5′-pAGUAGAAACAAGG GUA-FAM3′), 3′-vRNAp (5′FAM-AUACCCUGCUUUUGCU-3′) and polyUC (5′-pUCUCUCUCUCU CUCUCUC-FAM3′). 5′-vRNA sequences were synthetized with a 5′-mono phosphate and labeled with 6-FAM in 3′OH. 3′ vRNA were labeled with FAM in 5′. Polymerase constructs was titrated into fluorescently labeled RNA using the buffer (50 mM HEPES pH 7.5, 5 mM β-mercaptoethanol plus 150 mM or 300 mM NaCl) at room temperature. Initial reactions were performed with 2–4 nM of RNA with an initial volume of 600 µL to limit the dilution effect of protein addition (<5%). Anisotropy measurements were undertaken using a PTI fluorometer equipped with automated polarizers. The excitation and emission wavelengths were 494 and 521 nm, respectively. 50 anisotropy measurements with a 1 second integration time were performed per titration point.

Filter Binding Assay. Equilibrium RNA binding experiments were undertaken with the same vRNA set synthetized by IDT. vRNA were labeled with ^{32}P in 5′ using the T4 polynucleotide kinase (New England Biolabs). Double filter binding was used following the previously established protocol[64]. Protein concentration was titrated against constant concentrations of vRNA (≤0,01 nM) within a final volume of 200 µL using the standard protein buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 5 mM β-mercaptoethanol). 180 µL were then filtered on two membranes using a 96 well Whatman Minifold dot-blot apparatus. The first membrane was Protran BA 85 membrane (Whatman) to retain the polymerase and the second one is a nylon Hybond-N + membrane (Amersham Bioscience) to retain unbound nucleic acid. Both membranes were pre-incubated 1 h at room temperature within the protein buffer prior to use. After blotting, another 180 µL of protein buffer were run through the membranes to wash any unbound RNA before letting the vacuum dry the dot-blots for 30 seconds. Both membranes were then dismounted from the dot-blot apparatus and exposed to a storage phosphor screen (BAS storage phosphor screen, GE Healthcare) overnight. Revelation of the phosphor screen was then performed with a Typhoon Trio imaging system (GE Healthcare). Dot blot phosphorescence intensity was integrated using the Image J software. A derived bound ratio was then calculated using the following formula (1):

$$Br = \frac{Ibound}{Ibound + Ifree} \tag{1}$$

Where I_{bound} is the integrated intensity of protein/RNA complex which is retained by the protran membrane and I_{free} is the unbound RNA retained by the nylon Hybond-N + membrane.

Binding assay data analysis. RNA binding curves were plotted with the subtracted anisotropy or Bound RNA ratio as a function of the protein concentration. Sigmoidal binding curves were fitted to the data using GraphPad (Prism) with the following two equations:

(1) For the anisotropy titrations, [vRNA] ≈ Kd. Assuming a 1/1 stoichiometry the following equation (2) was used to estimate the Kd

$$\Delta A = \left\{ \frac{(R_T + E_T + K_d) - [(R_T + E_T + K_d)^2 - 4 \times R_T \times E_T]^{1/2}}{2 \times R_T} \right\} \times \Delta A_T \tag{2}$$

Where ΔA is the change in subtracted anisotropy, ΔA_T the total change of subtracted anisotropy. R_T is the total vRNA concentration, E_T is the total polymerase concentration at each given point and Kd is the dissociation constant.

(2) In the case of filter binding assay, a single site isotherm was used as we consider that [vRNA] ≪ Kd. This equation (3) therefor assumes an effective equality between the free polymerase concentration and the total polymerase concentration.

$$\Delta Br = \left\{ \frac{\Delta Br_T(E_T/K_d)}{(1 + E_T/K_d)} \right\} \quad (3)$$

Where ΔBr is the change in binding ratio, ΔBr_T is the total change in binding ratio. E_T is the total polymerase concentration at every point in titration and Kd is the dissociation constant.

Standard error calculations for the apparent Kd within a given condition were calculated through triplicate data measurement.

References

1. Cianci, C., Tiley, L. & Krystal, M. Differential activation of the influenza virus polymerase via template RNA binding. *J Virol* **69,** 3995–3999 (1995).
2. Hagen, M., Chung, T. D., Butcher, J. A. & Krystal, M. Recombinant influenza virus polymerase: requirement of both 5′ and 3′ viral ends for endonuclease activity. *J Virol* **68,** 1509–1515 (1994).
3. Tiley, L. S., Hagen, M., Matthews, J. T. & Krystal, M. Sequence-specific binding of the influenza virus RNA polymerase to sequences located at the 5′ ends of the viral RNAs. *J Virol* **68,** 5108–5116 (1994).
4. Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23,** 847–858 (1981).
5. Reich, S. *et al.* Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516,** 361–366, doi: 10.1038/nature14009 (2014).
6. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516,** 355–360, doi: 10.1038/nature14008 (2014).
7. Hengrung, N. *et al.* Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. *Nature* **527,** 114–117, doi: 10.1038/nature15525 (2015).
8. Mukaigawa, J. & Nayak, D. P. Two signals mediate nuclear localization of influenza virus (A/WSN/33) polymerase basic protein 2. *J Virol* **65,** 245–253 (1991).
9. Tarendeau, F. *et al.* Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat Struct Mol Biol* **14,** 229–233 (2007).
10. Deng, T. *et al.* Role of ran binding protein 5 in nuclear import and assembly of the influenza virus RNA polymerase complex. *J Virol* **80,** 11911–11919 (2006).
11. Huet, S. *et al.* Nuclear import and assembly of influenza A virus RNA polymerase studied in live cells by fluorescence cross-correlation spectroscopy. *J Virol* **84,** 1254–1264 (2010).
12. Hutchinson, E. C., Orr, O. E., Man Liu, S., Engelhardt, O. G. & Fodor, E. Characterization of the interaction between the influenza A virus polymerase subunit PB1 and the host nuclear import factor Ran-binding protein 5. *J Gen Virol* **92,** 1859–1869 (2011).
13. Chang, S. *et al.* Cryo-EM structure of influenza virus RNA polymerase complex at 4.3 A resolution. *Mol Cell* **57,** 925–935, doi: 10.1016/j.molcel.2014.12.031 (2015).
14. Al-Tawfiq, J. A., Zumla, A. & Memish, Z. A. Coronaviruses: severe acute respiratory syndrome coronavirus and Middle East respiratory syndrome coronavirus in travelers. *Current opinion in infectious diseases* **27,** 411–417, doi: 10.1097/QCO.0000000000000089 (2014).
15. Lucas, S. & Nelson, A. M. HIV and the spectrum of human disease. *The Journal of pathology* **235,** 229–241, doi: 10.1002/path.4449 (2015).
16. Crepin, T. *et al.* Polyproteins in structural biology. *Curr Opin Struct Biol* **32,** 139–146, doi: 10.1016/j.sbi.2015.04.007 (2015).
17. Nie, Y., Bellon-Echeverria, I., Trowitzsch, S., Bieniossek, C. & Berger, I. Multiprotein complex production in insect cells by using polyproteins. *Methods in molecular biology* **1091,** 131–141, doi: 10.1007/978-1-62703-691-7_8 (2014).
18. Bieniossek, C., Imasaki, T., Takagi, Y. & Berger, I. MultiBac: expanding the research toolbox for multiprotein complexes. *Trends Biochem Sci* **37,** 49–57, doi: 10.1016/j.tibs.2011.10.005 (2012).
19. Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F. & Berger, I. New baculovirus expression tools for recombinant protein complex production. *J Struct Biol* **172,** 45–54 (2010).
20. Sugiyama, K. *et al.* Structural insight into the essential PB1-PB2 subunit contact of the influenza virus RNA polymerase. *Embo J* **28,** 1803–1811 (2009).
21. An, Y., Meresse, P., Mas, P. J. & Hart, D. J. CoESPRIT: a library-based construct screening method for identification and expression of soluble protein complexes. *PLoS One* **6,** e16261, doi: 10.1371/journal.pone.0016261 (2011).
22. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol* **15,** 500–506 (2008).
23. Bieniossek, C., Richmond, T. J. & Berger, I. MultiBac: multigene baculovirus-based eukaryotic protein complex production. *Current protocols in protein science / editorial board, John E. Coligan …*[*et al.*] Chapter 5, Unit 5 **20,** doi: 10.1002/0471140864.ps0520s51 (2008).
24. Baudin, F., Bach, C., Cusack, S. & Ruigrok, R. W. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *Embo J* **13,** 3158–3165 (1994).
25. Crepin, T. *et al.* Mutational and metal binding analysis of the endonuclease domain of the influenza virus polymerase PA subunit. *J Virol* **84,** 9096–9104 (2010).
26. Dias, A. *et al.* The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458,** 914–918 (2009).
27. He, X. *et al.* Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature* **454,** 1123–1126 (2008).
28. Obayashi, E. *et al.* The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* **454,** 1127–1131 (2008).
29. Desselberger, U., Racaniello, V. R., Zazra, J. J. & Palese, P. The 3′ and 5′-terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. *Gene* **8,** 315–328 (1980).
30. Robertson, J. S. 5′ and 3′ terminal nucleotide sequences of the RNA genome segments of influenza virus. *Nucleic Acids Res* **6,** 3745–3757 (1979).
31. Skehel, J. J. & Hay, A. J. Nucleotide sequences at the 5′ termini of influenza virus RNAs and their transcripts. *Nucleic Acids Res* **5,** 1207–1219 (1978).
32. Brownlee, G. G. & Sharps, J. L. The RNA polymerase of influenza a virus is stabilized by interaction with its viral RNA promoter. *J Virol* **76,** 7103–7113 (2002).

33. Tomescu, A. I., Robb, N. C., Hengrung, N., Fodor, E. & Kapanidis, A. N. Single-molecule FRET reveals a corkscrew RNA structure for the polymerase-bound influenza virus promoter. *Proc Natl Acad Sci USA* **111,** E3335–3342, doi: 10.1073/pnas.1406056111 (2014).
34. Melen, K. *et al.* Importin alpha nuclear localization signal binding sites for STAT1, STAT2, and influenza A virus nucleoprotein. *J Biol Chem* **278,** 28193–28200 (2003).
35. Naito, T., Momose, F., Kawaguchi, A. & Nagata, K. Involvement of Hsp90 in assembly and nuclear import of influenza virus RNA polymerase subunits. *J Virol* **81,** 1339–1349, doi: 10.1128/JVI.01917-06 (2007).
36. Portela, A. & Digard, P. The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication. *J Gen Virol* **83,** 723–734 (2002).
37. Weber, F., Kochs, G., Gruber, S. & Haller, O. A classical bipartite nuclear localization signal on Thogoto and influenza A virus nucleoproteins. *Virology* **250,** 9–18, doi: 10.1006/viro.1998.9329 (1998).
38. Fodor, E. & Smith, M. The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex. *J Virol* **78,** 9144–9153, doi: 10.1128/JVI.78.17.9144-9153.2004 (2004).
39. Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496,** 477–481, doi: 10.1038/nature12070 (2013).
40. Kobayashi, J. & Matsuura, Y. Structural basis for cell-cycle-dependent nuclear import mediated by the karyopherin Kap121p. *J Mol Biol* **425,** 1852–1868, doi: 10.1016/j.jmb.2013.02.035 (2013).
41. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76,** 2879–2886 (1999).
42. Delaforge, E. *et al.* Large-Scale Conformational Dynamics Control H5N1 Influenza Polymerase PB2 Binding to Importin alpha. *Journal of the American Chemical Society* **137,** 15122–15134, doi: 10.1021/jacs.5b07765 (2015).
43. Thierry, E. *et al.* Influenza Polymerase Can Adopt an Alternative Configuration Involving a Radical Repacking of PB2 Domains. *Mol Cell* **61,** 125–137, doi: 10.1016/j.molcel.2015.11.016 (2016).
44. Momose, F. *et al.* Identification of Hsp90 as a stimulatory host factor involved in influenza virus RNA synthesis. *J Biol Chem* **277,** 45306–45314, doi: 10.1074/jbc.M206822200 (2002).
45. Chase, G. *et al.* Hsp90 inhibitors reduce influenza virus replication in cell culture. *Virology* **377,** 431–439, doi: 10.1016/j.virol.2008.04.040 (2008).
46. Fitzgerald, D. J. *et al.* Multiprotein expression strategy for structural biology of eukaryotic complexes. *Structure* **15,** 275–279, doi: 10.1016/j.str.2007.01.016 (2007).
47. Palmberger, D. & Rendic, D. SweetBac: Applying MultiBac Technology Towards Flexible Modification of Insect Cell Glycosylation. *Methods in molecular biology* **1321,** 153–169, doi: 10.1007/978-1-4939-2760-9_11 (2015).
48. Iwai, A. *et al.* Influenza A virus polymerase inhibits type I interferon induction by binding to interferon beta promoter stimulator 1. *J Biol Chem* **285,** 32064–32074, doi: 10.1074/jbc.M110.112458 (2010).
49. Deng, T., Sharps, J., Fodor, E. & Brownlee, G. G. *In vitro* assembly of PB2 with a PB1-PA dimer supports a new model of assembly of influenza A virus polymerase subunits into a functional trimeric complex. *J Virol* **79,** 8669–8674 (2005).
50. Graef, K. M. *et al.* The PB2 subunit of the influenza virus RNA polymerase affects virulence by interacting with the mitochondrial antiviral signaling protein and inhibiting expression of beta interferon. *J Virol* **84,** 8433–8445 (2010).
51. Patel, D., Schultz, L. W. & Umland, T. C. Influenza A polymerase subunit PB2 possesses overlapping binding sites for polymerase subunit PB1 and human MAVS proteins. *Virus Res* **172,** 75–80, doi: 10.1016/j.virusres.2012.12.003 (2013).
52. Degut, C., Monod, A., Brachet, F., Crepin, T. & Tisne, C. *In Vitro*/*In Vivo* Production of tRNA for X-Ray Studies. *Methods in molecular biology* **1320,** 37–57, doi: 10.1007/978-1-4939-2763-0_4 (2016).
53. Price, S. R., Ito, N., Oubridge, C., Avis, J. M. & Nagai, K. Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J Mol Biol* **249,** 398–408 (1995).
54. Boulo, S. *et al.* Human importin alpha and RNA do not compete for binding to influenza A virus nucleoprotein. *Virology* **409,** 84–90 (2011).
55. Chenavas, S. *et al.* Monomeric nucleoprotein of influenza A virus. *PLoS Pathog* **9,** e1003275, doi: 10.1371/journal.ppat.1003275 (2013).
56. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. ATSAS 2.1 – towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.* **40,** s223–s228 (2007).
57. Rambo, R. P. & Tainer, J. A. Super-resolution in solution X-ray scattering and its applications to structural systems biology. *Annual review of biophysics* **42,** 415–441, doi: 10.1146/annurev-biophys-083012-130301 (2013).
58. Semenyuk, A. V. & Svergun, D. I. GNOM – a program package for small-angle scattering data processing. *J. Appl. Crystallogr.* **24,** 537–540 (1991).
59. Franke, D. & Svergun, D. I. DAMMIF , a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42,** 342–346 (2009).
60. Volkov, V. V. & Svergun, D. I. Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36,** 860–864 (2003).
61. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28,** 768–773 (1995).
62. Petoukhov, M. V. & Svergun, D. I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical journal* **89,** 1237–1250, doi: 10.1529/biophysj.105.064154 (2005).
63. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612, doi: 10.1002/jcc.20084 (2004).
64. Wong, I. & Lohman, T. M. A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc Natl Acad Sci USA* **90,** 5428–5432 (1993).

Acknowledgements

We are very grateful to Daphna Fenel, Francine Gérard, Raphaelle Germi, Laurence Grossi, Darren J Hart, Marc Jamin, Thomas Lunardi, Hélène Malet, Stefan Reich and Adam Round and for their help. We thank Kausiki Datta, Klaus Klumpp and Christoph Bieniossek for discussion. We thank Juan Ortin for the gift of the RdRp antibodies. CS was funded by the Hofmann-La Roche pRED external collaboration program, AM by the Région Rhône-Alpes (ARC 1 Santé) and AL through the Labex GRAL. This work was supported by the French agency for Research through the ANR Flunucleovir (ANR-2010-Blanc-1307-1301), the ANR RNAP-IAV (ANR-14-CE09-0017) and the European Commission projects Flupharm (contract number 259751) and ComplexINC (contact number 279039). This study used the platforms of the Grenoble Instruct centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

Author Contributions

T.C., I.B., S.C. and R.W.H.R. conceived the experiments. C.S., A.M., L.T., F.G., A.L. and T.C. performed the biochemistry. A.L. and G.S. performed EM. C.S., J.M.B. and T.C. performed SAXS experiments. C.S., A.M., S.C. and T.C. performed X-ray crystallography. C.S., A.M., A.L., R.W.H.R. and T.C. analysed the Data. C.S., G.S., J.M.B. and T.C. prepared the figures. C.S., A.M., S.C., G.S., J.M.B., I.B., R.W.H.R. and T.C. wrote the Paper.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Swale, C. *et al.* Structural characterization of recombinant IAV polymerase reveals a stable complex between viral PA-PB1 heterodimer and host RanBP5. *Sci. Rep.* **6**, 24727; doi: 10.1038/srep24727 (2016).

Structure of influenza D virus nucleoprotein and functions of NP_{TAIL}

Alice Labaronne[a], Myriam Miloudi[a], Laura Tengo[a], Caroline Mas[b], Guy Schoehn[a], Rob WH Ruigrok[a] & Thibaut Crépin[a]#

Institut de Biologie Structurale (IBS), Univ. Grenoble Alpes, CEA, CNRS, 38044 Grenoble, France[a]. Integrated Structural Biology Grenoble (ISBG) - UMS 3518 (CNRS-CEA-UJF-EMBL), 38044 Grenoble, France[b]

Running Head: RNA binding of IDV NP is regulated by NP_{TAIL}

#Address correspondence to Thibaut Crépin, thibaut.crepin@ibs.fr

Submitted to Journal of Virology, 2017/01/31

Abstract

This paper focuses on the nucleoprotein (NP) of the newly identified member of the *Orthomyxoviridae* family, influenza D virus. To date several X-ray structures of NP of influenza A (A/NP) and B (B/NP) viruses and the infectious salmon anemia (ISA/NP) virus were solved, highlighting a conserved overall core for the nucleoprotein. Here we purified, characterized and solved the X-Ray structure of the tetrameric D/NP at 2.35 Å resolution. The crystal structure of its core is globally similar to NP of other influenza viruses. However, unlike A/NP and B/NP which possess a flexible amino-terminal tail containing nuclear localization signals (NLS) for their nuclear import, D/NP possesses a carboxy terminal tail (D/NP_{TAIL}). We show that D/NP_{TAIL} harbors a bipartite NLS and designed two C-terminal truncated mutants (D/NP-511 and D/NP-529) to demonstrate the role of D/NP_{TAIL} in the interaction with human importin-α7. We also determine the affinity of D/NP for RNA using different RNAs from 5 to 32 nucleotides in length and we show that D/NP_{TAIL} controls the RNA binding of D/NP.

Importance

The RNAs of influenza viruses are covered by a viral protein; the nucleoprotein. The sequence and the structure of influenza D virus nucleoprotein was compared with the nucleoproteins of influenza A, B and C and 3 other nucleoproteins from animal Orthomyxoviruses. All of these proteins share a core starting with an α-helix and ending with 3 consecutive aromatic residues that plug a hydrophobic hole. Both influenza A and B nucleoproteins possess N-terminal disordered tails before the core that contain a NLS sequence. Influenza C and D viruses have C-terminal tails after the core. Here we show that the tail of influenza D nucleoprotein containing a bipartite NLS, regulates also the binding of the RNA to the protein.

Introduction

Influenza viruses are segmented negative strand RNA viruses. Their genome is made by the ribonucleoproteins (RNPs) corresponding to a set of each RNA segments coated with multiple copies of the nucleoprotein (NP) and associated to the viral heterotrimeric polymerase. RNPs are competent for both transcription and replication. The number of vRNA segments is specific to each type of influenza viruses and related to the number of glycoproteins at the surface of the vial particle. Influenza D virus, isolated for the first time in 2011 from swine, is a novel member of the *Orthomyxiviridae* family (15). Its genome is divided into 7 vRNA segments, similar to influenza C virus.

Orthomyxoviridae family includes 7 different *genera*, influenza A, B, C and D viruses, the infectious salmon anemia (ISA) virus, Thogoto virus, and Quaranfilivirus. To date, several X-ray structures of NP have been published, each of them without RNA. There are three structures of influenza A nucleoprotein (A/NP): Two with NP assembled as a trimer (29, 48) and one for the monomeric R416A mutant (5). The X-ray structure of the tetrameric influenza B nucleoprotein (B/NP) is also known (28), whereas the structure of isavirus NP (ISA/NP) was solved as a dimer (49). The overall folds of A/NP and B/NP are very similar with a root-mean-square deviation (rmsd) of 1.6 Å. They share less structure similarity with ISA/NP which presents an additional N-terminal folded domain (Fig. 1A). Figure 1B shows a schematic representation of the overall architecture of NP from Orthomyxoviruses. A/NP, B/NP and WfB/NP (NP of Wellfleet bay virus) are made by a core with an amino-terminal extension (NP_{TAIL}). The core contains the large and shallow positively charged surface which might bind RNA without sequence specificity (2, 9, 29, 48). For NP of influenza C (C/NP) and D (D/NP) viruses, their cores possess a short very version of the N-terminal tail (only 7 residues), which seems to be compensated by the presence of a carboxy-terminal NP_{TAIL}. This carboxy-

terminal NP_{TAIL} is also present in ISA/NP in parallel of its additional N-terminal folded domain. The length of NP_{TAIL} is variable and seems to be specific for each genus.

It is known that influenza viruses replicate in the nucleus of the infected cells, meaning that all their replication machinery needs to interact with the nuclear transport system of the cell. For influenza A and B viruses, several studies have shown that the nuclear localization signals (NLS) of NP recognized by the cellular importins-α for their transport from the cytoplasm to the nucleus, are located within the flexible N-terminal tail (8, 21, 30, 34, 42, 43, 47) Recently a crystal structure between the non-classical NLS of A/NP and the importin-α1 has been solved (27).

In this paper, we characterized D/NP and solved the structure of its tetramer. The C-terminal D/NP_{TAIL} harbouring a classical bipartite nuclear localization signal (NLS) was not visible in the structure and we designed two C-terminal truncated mutants (D/NP-511 and D/NP-529) to study the interaction of D/NP with importin-α7. Our experiments clearly demonstrate that D/NP_{TAIL} is involved in the interaction with importins-α. Moreover we have determined affinity of D/NP for RNA using different RNAs from 5 to 24 nucleotides in length and demonstrated the implication of D/NP_{TAIL} in the RNA binding of D/NP, whereas NP_{TAIL} alone does not bind the RNA.

Materials and Methods

Molecular biology

The DNA coding sequence of D/bovine/France/2986/2012 NP was bought at GeneArt (ThermoFisher Scientific), optimized for bacterial expression. The different constructs have been PCR amplified and cloned following the supplier procedures (New England Biolabs). D/NP, D/NP-511 and D/NP-529 DNA coding sequences have been cloned in pETM13 (EMBL) to express C-terminal His-tagged proteins whereas D/NP$_{TAIL}$ (505-D/NP-552) DNA coding sequence has been cloned in pETM11 (EMBL) to express an N-terminal His-tagged fragment. The DNA coding sequence of the human importin-α7 (KPNA6; Uniprot O60684) was cloned without its IBB domain (amino-acids 58 to 536) in pET9a (Novagen) to express an N-terminal His-tagged protein (4). Sequencing has been done by Eurofins.

Expression and purification

Escherichia coli BL21 RIL (DE3) cells (Life Technology) were transformed with the resulting plasmids. Cultures were induced 12 h by adding 0.3 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 18 °C and collected by centrifugation. The pellets were resuspended and sonicated in 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 M NDSB-201 (Sigma), 2 mM β-mercaptoethanol (β-ME) and complete protease inhibitor cocktail (Roche) for D/NP constructs and 50 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM β-ME and complete protease inhibitor cocktail for importin-α7. All purifications were performed at room temperature. All D/NP constructs were purified by Nickel affinity chromatography (Ni-NTA, Qiagen) followed, in the case of D/NP, D/NP-511 and D/NP-529, by a HiTrap^{TM} Heparin HP column (GE-Healthcare) on a NGC system (BioRad). Elution fractions of D/NP, D/NP-511, D/NP-529 and D/NP$_{TAIL}$ were dialyzed against 20 mM Tris-HCl pH 7.5 at 50 mM or 300 mM NaCl, 5 mM β-ME. The last purification step was a size-exclusion chromatography using a Hiload^{TM} 16/600 S200 column (GE-Healthcare) for D/NP,

D/NP-511 and D/NP-529 or a S75 10/300 GL column (GE-Healthcare) for D/NP_{TAIL}. Importin-α7 was purified by Nickel affinity chromatography and elution fraction were dialyzed with TEV (1/100) against 20 mM Tris-HCl pH 7.5 at 150 mM, 5 mM β-ME and 20 mM imidazol. Finally a size-exclusion chromatography using Superdex^{TM} 200 increase 10/300 GL column (GE-Healthcare) was performed in 20 mM Tris-HCl pH 7.5 at 150 mM, 5 mM β-ME. Peak fractions were concentrated using a 10 kDa Amicon concentrator. Protein concentrations were determined using the extinction coefficients at 280 nm, ε = 44 537 $M^{-1}.cm^{-1}$ for D/NP, ε = 41 558 $M^{-1}.cm^{-1}$ for D/NP-529 and D/NP-511, ε = 2980 $M^{-1}.cm^{-1}$ for D/NP_{TAIL}, and ε = 46 785 $M^{-1}.cm^{-1}$ for importin-α7.

SEC-MALLS-RI analysis

SEC was performed with a column (Superdex^{TM} 200 increase 10/300 GL or Superdex 75 10/300 GL) equilibrated with 20 mM Tris-HCl pH 7.5, 150 mM NaCl and 5 mM β-ME. Analytical runs were performed at 20 °C with a flow rate of 0.5 $mL.min^{-1}$. MALLS detection was performed with a DAWN-HELEOS II detector (Wyatt Technology) using a laser emitting at 690 nm and protein concentration was measured on-line with the use of differential refractive-index measurements, with an Optilab T-rEX detector (Wyatt Technology) and a refractive-index increment, dn/dc, of 0.185 $mL.g^{-1}$. Weight-average molar masses (Mw) were calculated with ASTRA (Wyatt Technology) as previously described (13).

Electron microscopy

Samples (concentrations around 0.05 $mg.mL^{-1}$) were applied between a carbon and a mica layer. The carbon was then floated on the top of a 2% (w/v) sodium silicotungstate, pH 7.0 solution. The carbon film was covered by a copper grid. Both were fished using a small piece of journal paper and air dried before insertion in the electron microscope (12). Charge-coupled Device (CCD) frames were taken with a FEI T12 microscope operating at 120 kV and

a nominal magnification of 45 000 times. The dilutions for EM were performed with the size-exclusion buffer right before preparing the grid.

Fluorescence anisotropy measurements

Fluorescence anisotropy assays were performed on a Clariostar (BMG Labtech) microplate reader, fitted with polarization filters to measure fluorescence anisotropy. The binding assays were done in 384-well plates at room temperature (60 µL reaction volume) in 20 mM Tris-HCl pH 7.5, 150 or 300 mM NaCl and 5 mM β-ME. 0 to 3 µM of D/NP were titrated in 5 nM $^{5'}$Phosphate-RNA labelled in 3' with fluorescein amidite (FAM; Integrated DNA Technologies). 5 to 36 nucleotides polyUC RNA (Integrated DNA Technologies) were used. After substracting the polarization values obtained for RNA alone, the mean value of three independent experiments were fitted to the standard binding equation y = ((Bmax x xh)/(Kdh = xh)). All fluorescence anisotropy experiments were done in triplicate.

Thermal shift assays

Thermofluor experiments were performed following the established protocols (31). Sample were diluted at 0.25 mg.mL^{-1} in 50 mM Tris-HCl pH 7.5, 300 mM NaCl and 5 mM β-ME and 5X SYPRO orange dye (Invitrogen) in a final volume of 40 µL. The thermal stability was measured using a real time PCR machine (Mx3005P Q-PCR, Stratagene). The dye was excited at 488 nm and the emission light was recorded at 585 nm while the temperature was increased by increments of 1 °C per minute from 25 to 95 °C. The relative fluorescence emission was then plotted against its corresponding temperature to produce the thermal shift profile curve. The melting temperatures were estimated from the curve as the midpoint temperature between the baseline and maximum value of the curve.

Circular dichroism

A JASCO J-810 CD spectropolarimeter equipped with a temperature-controller (Peltier system) was used to record the far-UV CD spectrum of D/NP_{TAIL} at 20 µM. The spectrum was collected at 20°C with ten accumulations using a cuvette with a path-length of 1 mm. After blank signal subtraction, the CD signal (in mdeg) was converted to mean molar residue ellipticity (in $deg.cm^2\ dmol^{-1}$).

Crystallisation and structure determination

D/NP was crystallised by vapor diffusion using the sitting drop method. The crystals were obtained in 2.4 M sodium malonate pH 4 with a protein concentration at 1.5 $mg.mL^{-1}$. The crystals were directly flash-frozen without any cryoprotectant. Data were collected at the ESRF (beamline ID30A-3) and processed with the XDS package (18). The structure was solved by molecular replacement using the influenza A virus R416A monomeric nucleoprotein structure (PDB ID code 3ZDP) without the oligomerization loop (from residues 391 to 439) as a model. Model building and refinement were performed using CCP4i suite program for crystallography (PHASER, ARP/wARP, REFMAC5, COOT) (6, 10, 23, 26, 46). The final refinement was done using BUSTER (36). The coordinates have been deposited in the Protein Data Bank under PDB ID code XXX. The protein structure figures were drawn using PyMOL (33).

Results

Recombinant D/NP forms a tetramer in solution

The DNA coding sequence of D/NP was cloned in a bacterial expression plasmid and over-expressed in *Escherichia coli* as a C-terminal His-tag recombinant protein. D/NP was purified in three steps; a Ni^{2+} resin followed by a heparin column and a final gel filtration. Figure 2A shows a typical gel filtration elution profile using the absorbance signals at 260 and 280 nm. The protein can be then concentrated between 2 and 6 $mg.mL^{-1}$. The sample was pure (Fig. 2B) without nucleic acid contamination (ratio 280/260 > 1.75). By electron microscopy (negative staining), we show that D/NP forms mainly tetramers in solution (Fig. 2C). A SEC-MALLS-RI experiment has confirmed the homogeneity and the molecular weight of the recombinant tetrameric D/NP (supplementary Fig. 1). It has been previously shown that the oligomerization of recombinant A/NP and B/NP can be modulated by the NaCl concentration of the purification buffers (5, 20, 39). Starting from purified and stable oligomeric samples (trimers for A/NP and tetramers for B/NP), monomeric proteins can be obtained by decreasing stepwise the NaCl concentration. After purification at 300 mM NaCl followed by its dialysis at 150 mM NaCl, D/NP is eluted from gel filtration (with a 150 mM NaCl running buffer) at the same volume, meaning that a smooth reduction of the salt concentration does not change its oligomeric state. However, a more drastic decrease induces an irreversible and total precipitation of D/NP, even with a stepwise reduction at 150 mM NaCl. For the rest of the experiments on D/NP and its mutants, all the samples were purified at 300 mM NaCl.

D/NP structure

Full length D/NP was crystallized in sodium malonate as small fine needles that diffract up to 2.35 Å resolution (Table 1). The structure was solved by molecular replacement using a starting model made by one monomer of the A/NP R416A mutant (5), in which the parts

197 involved in the oligomerization are removed. The automatic search has provided the position
198 for three molecules and a forth one was then fitted manually after the analysis of the electron
199 density. Retrospectively, the molecular replacement using the tetrameric B/NP model did not
200 provide any acceptable solution. The asymmetric unit contains four molecules arranged as a
201 tetramer (Fig. 3A and refinement statistics in Table 1), coherent with the observations made
202 by electron microscopy. Based on the electron density, 87 % of the model could be built, from
203 residues 8 to 514 (Fig. 3B). Several internal loops are missing and the NP_{TAIL} was disordered.
204 The D/NP monomer was compared to the already known structures (Fig. 3D with the statistics
205 included in Fig. 1A). The closest model is B/NP (r.m.s.d of 1.9 Å for 365 C_α aligned), the
206 highest deviation between the two models comes from the part of the protein involved in the
207 oligomerization process; the oligomerization-loop of one monomer plugged into a cavity of its
208 neighbour. The loop is then stabilized by several kind of contacts, mainly hydrophobic
209 because of several aromatic residues lying in the pocket of the neighbouring monomer but
210 also a salt bridge between an arginine from one monomer (Arg-425) interacting with a
211 glutamate of the neighbour (Glu-352; Fig. 3C). These two residues, strictly conserved in all
212 influenza NPs, are well known for modulating their oligomeric state. Once mutated in alanine,
213 the corresponding single mutants form monomers, unable to self-assemble (5, 9, 39, 40, 48).
214 Including ISA/NP, a common architecture for all influenza-like NPs can be defined based on
215 the X-ray structures. It explains the limits we have included on the figure 1A, with the NP_{CORE}
216 starting with an α-helix and finishing with an unfolded part, anchored into the surface of the
217 protein. Figure 3E shows the X-ray structures of A/NP, B/NP, D/NP and ISA/NP with a zoom
218 on the C-termini, where a superimposed patch made by 3 consecutive aromatic residues is
219 found in influenza NPs ($_{487}$YFF$_{489}$ for A/NP, $_{545}$FFF$_{547}$ for B/NP and $_{499}$FFF$_{501}$ for D/NP). A
220 similar patch is also presents in the ISA/NP model ($_{580}$GLF$_{582}$) even if the upstream unfolded

part is not as much superimposed. The sequences analysis shows such patch is present in C/NP ($_{501}$FFF$_{503}$) and could be present in Tho/NP ($_{451}$YLF$_{453}$) and Wfb/NP ($_{532}$VIY$_{534}$). Because D/NP differs from A/NP and B/NP in its global architecture, we have decided to further analyze the C-terminal extension: D/NP$_{TAIL}$.

D/NP$_{TAIL}$ behaves as an intrinsically disordered protein

Based on the crystal structure and the disorder prediction (Fig. 4A), we have designed a construct (from residues 505 to 552) for the expression in *E. coli* as a N-terminal His-tag recombinant version. Based on the structure, even if we consider that the core of D/NP ends after residue Phe-501, we have chosen to start the construct at Gly-505, in order to avoid the hydrophobicity of Phe-503. D/NP$_{TAIL}$ was eluted from gel filtration as a 40 kDa-protein but with a normal migration pattern on SDS-PAGE (Fig. 4B). As intrinsically disordered proteins are known to be eluted from gel filtration with an aberrant volume (13, 41). A SEC-MALLS-RI experiment (Fig. 4C) confirmed the monodispersity of the purified sample with a Mw of 8.2 kDa for D/NP$_{TAIL}$ (with the His-tag). The CD analysis confirms that D/NP$_{TAIL}$ does not contain any significant stable secondary structure in solution (Fig. 4D).

D/NP$_{TAIL}$ interacts with importin-α

The N-terminal tails of A/NP and B/NP are known to be involved in their nuclear import by interacting with importins-α (8, 21, 27, 34, 38, 43). The sequence analysis of D/NP shows no import signals in its N-terminal part but suggests the presence of a bipartite NLS within D/NP$_{TAIL}$ ($_{514}$KRR-X$_{14}$-KKRGR$_{545}$; Fig. 4A). We have tested the bipartite NLS with different constructs. First, we could show that the human importin-α7 is co-eluted with the His-tagged D/NP$_{TAIL}$ from a nickel-affinity resin (Fig. 5A). We then have confirmed the interaction by thermal-shift experiments where importin-α7 appears to be more stable in presence of

D/NP$_{TAIL}$ (Fig. 5B), and could purified the D/NP$_{TAIL}$:importin-α7 complex by gel filtration on a S75 column (sup Fig. 2). We have confirmed the interaction between the two partners using the full-length D/NP. By gel filtration, we have shown that the D/NP:importin-α7 complex can be eluted as a single peak. A molecular weight of 474 kDa has been measured by MALLS for the complex (Fig. 5C) corresponding to the interaction of 4 importins-α7 (Mw = 55 kDa) per tetramer of D/NP (Mw = 260 kDa). To confirm the role of D/NP$_{TAIL}$ in the interaction with importin-α, we have designed two C-terminal truncated constructs of D/NP, D/NP-511 and D/NP-529, ending respectively at residues 511 and 529. D/NP-511 is deprived of the whole bipartite NLS whereas contains D/NP-529 the first part. They both behave as the wild-type protein during the purification. Using pull-down experiments on nickel-affinity resin, we show that D/NP-529 still slightly retains importin-α7 whereas the interaction between D/NP-511 and importin-α7 is totally abolished (Fig.5D).

D/NP$_{TAIL}$ regulates the binding of RNA

By using $^{5'}$phosphate-(UC)$_x$-FAM$^{3'}$ in fluorescence anisotropy experiments, we have measured the affinities of D/NP for RNA molecules, ranging from 5 to 36 nucleotides in length (Fig. 6 and Table 2). We show that for small RNA molecules, there is a clear effect of the NaCl concentration on the affinity, with higher affinities obtained at “low” salt concentration. For longer RNA molecules (> 16 nt), the affinities are stronger, with no significant effect of the ionic strength neither of the length of RNA. The tipping point seems to be around 12 nucleotides, similar to what has been observed for A/NP and B/NP (20) but also for ISA/NP (49).

By focusing on the 10-16 nt fluorescent RNA molecules, we have performed the same experiment using the D/NP-511 and D/NP-529 truncated constructs. We have chosen the 300 mM NaCl condition to get a better tuning of the results. For the 10 nt and 16 nt RNA

molecules, there is no significant difference in the affinities between the two mutants and the *wt* protein. However, using the 12 nt fluorescent RNA, we observed that the binding of D/NP-511 and -529 is more efficient than for D/NP (Fig. 6B and Table 2), suggesting that D/NP_{TAIL} controls the binding of RNA onto the core of the nucleoprotein. We have also confirmed that there is no detectable change in the anisotropy when mixing D/NP_{TAIL} and the corresponding RNAs.

Discussion

In 2011, a new virus has been isolated from pigs exhibiting influenza-like illness. Later, cattle has been found to be the primary reservoir and subsequent electron microscopic, biochemical and genetic studies identified an Orthomyxovirus with seven RNA segments sharing 50 % overall amino acids identity with the human influenza C virus (15). First considered as a subtype of the influenza C virus, it has since been officially named influenza D virus (IDV) by the International Committee of Taxonomy of Viruses in 2016. Since its first isolation, IDV appears to be spread all around the world in many mammals, with serological evidence for its transmission to humans (45). The recent structure of IDV Hemagglutinin-Esterase-Fusion glycoprotein (HEF) has described a rather open receptor-binding cavity capable of accommodating diverse extended glycan moieties that could be one reason for its broad cell tropism (37).

With this paper, we provide a detailed characterization of IDV nucleoprotein. The comparison of the nucleoproteins of influenza A, B, D and of infectious salmon anemia virus (ISAV) structures shows that of these proteins share a core (NP_{CORE}) that start with the first helix in the structure, up to three aromatic residues that plug a hydrophobic hole in the protein. The only structure that differs is the ISAV/NP that possesses a small folded domain before the first helix of the NP_{CORE}. Looking to the sequences of all known Orthomyxoviruses nucleoproteins, one can observe that the NP of Thogoto virus has only the NP_{CORE} without any else. The NPs of influenza A and B viruses have an N-terminal NP_{TAIL} whereas the NPs of influenza C, D and ISAV have a C-terminal NP_{TAIL}. The NP_{TAIL} of A is only 20 residues whereas the tails of the other influenza viruses are much longer; 71 residues for B, 62 for C, 51 for D. We demonstrate that D/NP_{TAIL} presents all the characteristics of an intrinsically disordered protein (IDP) and that is involved in the nuclear import through its interaction with cellular importins-α.

This role is not a big surprise as it had already been shown the presence of the NLS motifs within the flexible NP_{TAILs} of viruses A and B (8, 21, 27, 34, 38, 43). Whereas Tho/NP has been shown to be imported in the nucleus (44), this NP does not a IDP-like additional peptide with an NLS.

Like influenza A and B viruses nucleoproteins, D/NP binds single strand RNA. We found that the 12-16 nt long RNAs bind with a better affinity than smaller RNAs. For A/NP and B/NP we found that a 12-mer RNA binds strongly, as if there is strong binding site for 12 nt, and that longer RNAs do not bind better. For the ISA/NP similar findings were found (49). Using a 12-mer RNA and 300 mM salt, we found that the D/NP binds with a K_d of 108 nM whereas the shorter proteins D/NP-529 and -511 bind the same RNA with K_ds of 46 and 28, close to the K_d obtained when using longer RNAs. This suggests that D/NP_{TAIL} could overlap with the RNA binding site. It is known that IDPs contain many charged residues, more than structured parts of proteins (7). D/NP_{TAIL} has 9 negative charges for 8 positive (7 in the NLSs), suggesting that the charges alone will not influence the RNA binding. We find indeed that D/NP_{TAIL} alone doesn't bind RNA, suggesting that the presence of the D/NP_{TAIL} can control the kinetics of the RNA binding.

The preservation of such IDP-like peptide with a conserved function highlights the amplitude of its importance. The name for N_{TAIL} was first used for the C-terminal disordered tails of Sendai and measles viruses nucleoproteins (22), which are about 125 residues long. These tails are disordered but they can adopt a short helix structure (residue D484-A502) for the binding to the phosphoproteins of these viruses that themselves bind to the corresponding RNA polymerase (3). For these viruses, the tails keep the nucleocapsids in a flexible conformation. When cleaved off with trypsin, the helical nucleocapsids become rigid (17, 32). However, the tails do not change the affinity of N for the RNA but the assembly of the capsids with a slower kinetics (24). The RNPs of influenza A and B viruses are very flexible (1, 11, 19,

25) which makes it difficult to handle for high resolution structures. The EM structure of the measles nucleocapsid with a resolution of 3.6 Å (14) was done on the nucleocapsids without the N_{TAIL}. Maybe it would be interesting to make influenza RNPs with NP without NP_{TAIL} for making rigid RNPs that would be easier for making a structure for seeing the RNA on the nucleoprotein.

Acknowledgements

We are highly grateful to Jean-Marie Bourhis, Daphna Fenel and Joanna Timmins for their help. We thank Darren J Hart for the plasmid to express the human importin-α7. We thank Mariette Ducatez and Anny Slama-Schwok for discussion. We thank the staff of the ESRF-EMBL Joint Structural Biology Group for access to ESRF beamlines. AL and MM were funded through the Labex GRAL (ANR-10-LABX-49-01). This work was supported by the French agency for Research through the ANR RNAP-IAV (ANR-14-CE09-0017). This study used the platforms of the Grenoble Instruct centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL, within the Grenoble Partnership for Structural Biology (PSB).

Author contributions

TC conceived the experiments. AL, MM, LT, CM and TC performed the experiments. AL, GS, RWHR and TC analysed the data. AL and TC wrote the first version of the paper. AL, RWHR, and TC wrote the final version of the paper

Competing interest statement

The authors declare no competing financial interests.

References

1. **Arranz, R., R. Coloma, F. J. Chichon, J. J. Conesa, J. L. Carrascosa, J. M. Valpuesta, J. Ortin, and J. Martin-Benito.** 2012. The structure of native influenza virion ribonucleoproteins. Science **338:**1634-1637.
2. **Baudin, F., C. Bach, S. Cusack, and R. W. Ruigrok.** 1994. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. The EMBO journal **13:**3158-3165.
3. **Bloyet, L. M., J. Brunel, M. Dosnon, V. Hamon, J. Erales, A. Gruet, C. Lazert, C. Bignon, P. Roche, S. Longhi, and D. Gerlier.** 2016. Modulation of Re-initiation of Measles Virus Transcription at Intergenic Regions by PXD to NTAIL Binding Strength. PLoS pathogens **12:**e1006058.
4. **Boivin, S., and D. J. Hart.** 2011. Interaction of the influenza A virus polymerase PB2 C-terminal region with importin alpha isoforms provides insights into host adaptation and polymerase assembly. The Journal of biological chemistry **286:**10439-10448.
5. **Chenavas, S., L. F. Estrozi, A. Slama-Schwok, B. Delmas, C. Di Primo, F. Baudin, X. Li, T. Crepin, and R. W. Ruigrok.** 2013. Monomeric nucleoprotein of influenza A virus. PLoS pathogens **9:**e1003275.
6. **Cohen, S. X., M. Ben Jelloul, F. Long, A. Vagin, P. Knipscheer, J. Lebbink, T. K. Sixma, V. S. Lamzin, G. N. Murshudov, and A. Perrakis.** 2008. ARP/wARP and molecular replacement: the next generation. Acta crystallographica. Section D, Biological crystallography **64:**49-60.
7. **Communie, G., R. W. Ruigrok, M. R. Jensen, and M. Blackledge.** 2014. Intrinsically disordered proteins implicated in paramyxoviral replication machinery. Current opinion in virology **5:**72-81.

8. **Cros, J. F., A. Garcia-Sastre, and P. Palese.** 2005. An unconventional NLS is critical for the nuclear import of the influenza A virus nucleoprotein and ribonucleoprotein. Traffic **6:**205-213.
9. **Elton, D., L. Medcalf, K. Bishop, D. Harrison, and P. Digard.** 1999. Identification of amino acid residues of influenza virus nucleoprotein essential for RNA binding. Journal of virology **73:**7357-7367.
10. **Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan.** 2010. Features and development of Coot. Acta crystallographica. Section D, Biological crystallography **66:**486-501.
11. **Gallagher, J. R., U. Torian, D. M. McCraw, and A. K. Harris.** 2016. Structural studies of influenza virus RNPs by electron microscopy indicate molecular contortions within NP supra-structures. Journal of structural biology.
12. **Gaudin, Y., R. W. Ruigrok, C. Tuffereau, M. Knossow, and A. Flamand.** 1992. Rabies virus glycoprotein is a trimer. Virology **187:**627-632.
13. **Gerard, F. C., A. Ribeiro Ede, Jr., C. Leyrat, I. Ivanov, D. Blondel, S. Longhi, R. W. Ruigrok, and M. Jamin.** 2009. Modular organization of rabies virus phosphoprotein. Journal of molecular biology **388:**978-996.
14. **Gutsche, I., A. Desfosses, G. Effantin, W. L. Ling, M. Haupt, R. W. Ruigrok, C. Sachse, and G. Schoehn.** 2015. Structural virology. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. Science **348:**704-707.
15. **Hause, B. M., M. Ducatez, E. A. Collin, Z. Ran, R. Liu, Z. Sheng, A. Armien, B. Kaplan, S. Chakravarty, A. D. Hoppe, R. J. Webby, R. R. Simonson, and F. Li.** 2013. Isolation of a novel swine influenza virus from Oklahoma in 2011 which is distantly related to human influenza C viruses. PLoS pathogens **9:**e1003176.

16. **Holm, L., and P. Rosenstrom.** 2010. Dali server: conservation mapping in 3D. Nucleic acids research **38:**W545-549.
17. **Jensen, M. R., G. Communie, E. A. Ribeiro, Jr., N. Martinez, A. Desfosses, L. Salmon, L. Mollica, F. Gabel, M. Jamin, S. Longhi, R. W. Ruigrok, and M. Blackledge.** 2011. Intrinsic disorder in measles virus nucleocapsids. Proceedings of the National Academy of Sciences of the United States of America **108:**9839-9844.
18. **Kabsch, W.** 2010. Xds. Acta crystallographica. Section D, Biological crystallography **66:**125-132.
19. **Klumpp, K., R. W. Ruigrok, and F. Baudin.** 1997. Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. The EMBO journal **16:**1248-1257.
20. **Labaronne, A., C. Swale, A. Monod, G. Schoehn, T. Crepin, and R. W. Ruigrok.** 2016. Binding of RNA by the Nucleoproteins of Influenza Viruses A and B. Viruses **8**.
21. **Liu, J., Y. Cong, R. Yin, C. Ding, S. Yu, X. Liu, C. Wang, and Z. Ding.** 2014. The deletion of an extra six nucleotides in the 5' -untranslated region of the nucleoprotein gene of Newcastle disease virus NA-1 decreases virulence. BMC veterinary research **10:**964.
22. **Longhi, S., V. Receveur-Brechot, D. Karlin, K. Johansson, H. Darbon, D. Bhella, R. Yeo, S. Finet, and B. Canard.** 2003. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. The Journal of biological chemistry **278:**18638-18648.
23. **McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read.** 2007. Phaser crystallographic software. Journal of applied crystallography **40:**658-674.

24. **Milles, S., M. R. Jensen, G. Communie, D. Maurin, G. Schoehn, R. W. Ruigrok, and M. Blackledge.** 2016. Self-Assembly of Measles Virus Nucleocapsid-like Particles: Kinetics and RNA Sequence Dependence. Angewandte Chemie **55:**9356-9360.
25. **Moeller, A., R. N. Kirchdoerfer, C. S. Potter, B. Carragher, and I. A. Wilson.** 2012. Organization of the influenza virus replication machinery. Science **338:**1631-1634.
26. **Murshudov, G. N., A. A. Vagin, and E. J. Dodson.** 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta crystallographica. Section D, Biological crystallography **53:**240-255.
27. **Nakada, R., H. Hirano, and Y. Matsuura.** 2015. Structure of importin-alpha bound to a non-classical nuclear localization signal of the influenza A virus nucleoprotein. Scientific reports **5:**15055.
28. **Ng, A. K., M. K. Lam, H. Zhang, J. Liu, S. W. Au, P. K. Chan, J. Wang, and P. C. Shaw.** 2012. Structural basis for RNA binding and homo-oligomer formation by influenza B virus nucleoprotein. Journal of virology **86:**6758-6767.
29. **Ng, A. K., H. Zhang, K. Tan, Z. Li, J. H. Liu, P. K. Chan, S. M. Li, W. Y. Chan, S. W. Au, A. Joachimiak, T. Walz, J. H. Wang, and P. C. Shaw.** 2008. Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. FASEB journal : official publication of the Federation of American Societies for Experimental Biology **22:**3638-3647.
30. **Ozawa, M., K. Fujii, Y. Muramoto, S. Yamada, S. Yamayoshi, A. Takada, H. Goto, T. Horimoto, and Y. Kawaoka.** 2007. Contributions of two nuclear localization signals of influenza A virus nucleoprotein to viral replication. Journal of virology **81:**30-41.
31. **Pantoliano, M. W., E. C. Petrella, J. D. Kwasnoski, V. S. Lobanov, J. Myslik, E. Graf, T. Carver, E. Asel, B. A. Springer, P. Lane, and F. R. Salemme.** 2001. High-density

446 miniaturized thermal shift assays as a general strategy for drug discovery. Journal of
447 biomolecular screening **6:**429-440.
448 32. **Schoehn, G., M. Mavrakis, A. Albertini, R. Wade, A. Hoenger, and R. W. Ruigrok.**
449 2004. The 12 A structure of trypsin-treated measles virus N-RNA. Journal of molecular
450 biology **339:**301-312.
451 33. **Schrodinger, LLC.** 2015. The PyMOL Molecular Graphics System, Version 1.8.
452 34. **Sherry, L., M. Smith, S. Davidson, and D. Jackson.** 2014. The N terminus of the
453 influenza B virus nucleoprotein is essential for virus viability, nuclear localization, and
454 optimal transcription and replication of the viral genome. Journal of virology **88:**12326-
455 12338.
456 35. **Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H.**
457 **McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins.** 2011. Fast,
458 scalable generation of high-quality protein multiple sequence alignments using Clustal
459 Omega. Molecular systems biology **7:**539.
460 36. **Smart, O. S., T. O. Womack, C. Flensburg, P. Keller, W. Paciorek, A. Sharff, C.**
461 **Vonrhein, and G. Bricogne.** 2012. Exploiting structure similarity in refinement:
462 automated NCS and target-structure restraints in BUSTER. Acta crystallographica.
463 Section D, Biological crystallography **68:**368-380.
464 37. **Song, H., J. Qi, Z. Khedri, S. Diaz, H. Yu, X. Chen, A. Varki, Y. Shi, and G. F. Gao.**
465 2016. An Open Receptor-Binding Cavity of Hemagglutinin-Esterase-Fusion Glycoprotein
466 from Newly-Identified Influenza D Virus: Basis for Its Broad Cell Tropism. PLoS
467 pathogens **12:**e1005411.
468 38. **Stevens, M. P., and W. S. Barclay.** 1998. The N-terminal extension of the influenza B
469 virus nucleoprotein is not required for nuclear accumulation or the expression and
470 replication of a model RNA. Journal of virology **72:**5307-5312.

39. **Tarus, B., O. Bakowiez, S. Chenavas, L. Duchemin, L. F. Estrozi, C. Bourdieu, N. Lejal, J. Bernard, M. Moudjou, C. Chevalier, B. Delmas, R. W. Ruigrok, C. Di Primo, and A. Slama-Schwok.** 2012. Oligomerization paths of the nucleoprotein of influenza A virus. Biochimie **94:**776-785.

40. **Turrell, L., J. W. Lyall, L. S. Tiley, E. Fodor, and F. T. Vreede.** 2013. The role and assembly mechanism of nucleoprotein in influenza A virus ribonucleoprotein complexes. Nature communications **4:**1591.

41. **Uversky, V. N.** 2012. Size-exclusion chromatography in structural analysis of intrinsically disordered proteins. Methods in molecular biology **896:**179-194.

42. **Wang, P., P. Palese, and R. E. O'Neill.** 1997. The NPI-1/NPI-3 (karyopherin alpha) binding site on the influenza a virus nucleoprotein NP is a nonconventional nuclear localization signal. Journal of virology **71:**1850-1856.

43. **Wanitchang, A., J. Narkpuk, and A. Jongkaewwattana.** 2013. Nuclear import of influenza B virus nucleoprotein: involvement of an N-terminal nuclear localization signal and a cleavage-protection motif. Virology **443:**59-68.

44. **Weber, F., G. Kochs, S. Gruber, and O. Haller.** 1998. A classical bipartite nuclear localization signal on Thogoto and influenza A virus nucleoproteins. Virology **250:**9-18.

45. **White, S. K., W. Ma, C. J. McDaniel, G. C. Gray, and J. A. Lednicky.** 2016. Serologic evidence of exposure to influenza D virus among persons with occupational contact with cattle. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology **81:**31-33.

46. **Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson.**

2011. Overview of the CCP4 suite and current developments. Acta crystallographica. Section D, Biological crystallography **67:**235-242.

47. **Wu, W. W., Y. H. Sun, and N. Pante.** 2007. Nuclear import of influenza A viral ribonucleoprotein complexes is mediated by two nuclear localization sequences on viral nucleoprotein. Virology journal **4:**49.

48. **Ye, Q., R. M. Krug, and Y. J. Tao.** 2006. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. Nature **444:**1078-1082.

49. **Zheng, W., J. Olson, V. Vakharia, and Y. J. Tao.** 2013. The crystal structure and RNA-binding of an orthomyxovirus nucleoprotein. PLoS pathogens **9:**e1003624.

Figures legends

FIG. 1. Nucleoproteins of Orthomyxoviruses. (A) Pairwise root-mean-square deviation (rmsd) of Cα positions and sequence identity between NP of the members of the *Orthomyxoviridae* family. The sequence identity and the rmsd values have been calculated using respectively Clustal Omega (35) and DaLi (16). The identity between Tho/NP and WfB/NP is 16 %. (B) Schematic representation of the nucleoprotein based on the sequence and structure analysis of the protein from representative members of the *Orthomyxoviridae* family. The schema respects the size of the proteins. A/NP (influenza A virus, strain A/WSN/1933(H1N1); Uniprot id: B4URF1), B/NP (influenza B virus, strain B/Memphis/13/2003; Uniprot id: Q5V913), C/NP (influenza C virus, strain C/AnnArbor/1/1950; Uniprot id: Q6I7C0), D/NP (influenza D virus, strain D/bovine/France/2986/2012; Uniprot id: A0A0E3VZU8), ISA/NP (infectious salmon anemia virus, isolate salmon/Norway/810/9/99; Uniprot id: Q8V3T7), Tho/NP (Thogoto virus; Uniprot id: A0A0B6VKB5) and WfB/NP (Wellfleet Bay virus; Uniprot id: A0A0A1E9N5). The flexible tails are represented with simple lines whereas the folded parts and the cores are represented with filled boxes.

FIG. 2. Purification of influenza D nucleoprotein. (A) Size exclusion chromatography profile of wild-type D/NP. The sample was loaded on a Hiload^{TM} 16/600 S200 column equilibrated with the running buffer 20 mM Tris-HCl pH 7.5, 300 mM NaCl and 5 mM β-mercaptoethanol. (B) Coomassie blue-stained SDS-PAGE (4-20 % gradient polyacrylamide) showing the purified wild-type D/NP and the two C-terminal truncated mutants (D/NP-529 and D/NP-511). (C) Electron microscopy images of the elution peak of D/NP. Samples show different oligomeric states although most oligomers are tetramers.

529 FIG. 3. Structure of influenza D nucleoprotein. (A) Structure of the tetrameric D/NP, with 3
530 monomers shown in surface (respectively in green, pink and cyan) and the fourth in cartoon
531 (deep purple). (B) Ribbon diagram of one monomer of D/NP with the α-helixes in deep purple
532 and β-strands in yellow. (C) Detail of the interactions between two monomers of the
533 tetrameric D/NP as shown in (A). The conserved R425 of one monomer (shown as deep
534 purple cartoon) stabilized the position of the oligomerization loop at the surface of the
535 neighbouring protomer, through its conserved E352. (D) Comparison of D/NP with other
536 nucleoproteins negative sRNA viruses. The structure of one monomer of D/NP (deep purple)
537 has been superimposed from left to right, with one monomer of A/NP (blue; PDB id: 2IQH),
538 B/BP (forest; PDB id: 3TJ0) and ISA/NP (light orange; PDB id: 4EWC). The rmsd values are
539 in Figure 1A. (E) Anchoring of the C-terminus on NP_{CORE} by a patch of 3 consecutive aromatic
540 residues. The panel corresponds to a zoom of the superimposed structures shown in panel C
541 with NP_{CORE} represented in surface.

542 FIG. 4. Biophysical characterization of D/NP_{TAIL}. (A) D-score (score for disorder as a function
543 of residue) of D/NP with a zoom (below the graph) on the last 50 residues. The prediction is
544 based on 22 predictor web servers. The disordered regions have a D-score < 0.50 (13). The
545 yellow boxes correspond to the putative NLS motifs. (B) Coomassie blue-stained SDS-PAGE
546 (Tris-Tricine; 15% polyacrylamide) of the purified D/NP_{TAIL}. It migrates at a higher molecular
547 weight (17 kDa approximately) than expected (8 kDa). (C) SEC-MALLS analysis of D/NP_{TAIL}
548 loaded on S75 10/300 GL column. For this experiment, we have chosen to keep the His-tag
549 encoded with the pETM11 plasmid, for an optimal detection of D/NP_{TAIL}. The experimental
550 molecular weight is coherent with the expected mass. (D) Circular dichroïsm of D/NP_{TAIL}.

551 FIG. 5. Interaction with importin-α7. (A) Pull-down assays of human importin-α7 by His-
552 tagged D/NP$_{TAIL}$. The mixtures (200 μL; molar ratio 1 D/NP$_{TAIL}$ for 1 importin-α7) were
553 incubated 1 hour at room temperature and then loaded (L) on 200 μL of Ni-NTA resin. The
554 flow throughs (FT) were recovered by gravity and the resins were washed with 50 volumes of
555 the equilibration buffer. The final 200 μL of the washing step were recovered (W). The
556 proteins were then eluted with the buffer 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 800 mM
557 imidazol and 5 mM β-mercaptoethanol (100 μL fractions). The figure shows the coomassie
558 blue-stained SDS-PAGE (15 % polyacrylamide) with the L, FT, W and the second fractions of
559 the pull-down (E2). The importin-α7 used for this experiment presents a small degradation.
560 (B) Thermal stability assay of importin-α7 in absence (green) or in presence (purple) of
561 D/NP$_{TAIL}$ using Thermofluor. In presence of D/NP$_{TAIL}$, the melting temperature of importin-α7
562 is 6°C higher. (C) SEC-MALLS analysis of D/NP in complex with importin-α7. The mixture (50
563 μL; molar ratio 1 D/NP monomer for 1.2 importin-α7) was incubated 1 hour at room
564 temperature and then loaded on a Superdex^{TM} 200 increase 10/300 GL. The experimental
565 molecular weight (below the peak) is coherent with the expected mass of four importins-α7
566 bound per D/NP tetramer. (D) Pull-down assays of human importin-α7 by D/NP and the two
567 C-terminal truncated mutants (D/NP-529 and D/NP-511). The his-tags are on D/NP. The
568 mixtures (200 μL; molar ratio 1 D/NP monomer for 1.2 importin-α7) were incubated 1 hour
569 and the experience was done as described in panel (A). The figure shows the coomassie
570 blue-stained SDS-PAGE (12 % polyacrylamide) with the L, FT, W and the second fractions
571 (E2).

572 FIG. 6. RNA interaction. (A) Titrations of NP were done against $^{5'}$phosphate-polyUC-FAM$^{3'}$
573 ranging from 8 to 36 nucleotides. The mixtures were incubated 5 min at room temperature.

574 RNA binding titration curves of D/NP at 300 mM. (B) Comparison of the titration curves
575 obtained for the wild-type D/NP and the two truncated mutants (D/NP-529 and D/NP-511) in
576 presence $^{5'}$phosphate-poly(UC)$_6$-FAM$^{3'}$. The K_d obtained for these experiments are listed in
577 Table 2.
578

579

Table 1: Crystallographic data collection and refinement statistics. Values in parentheses are for highest resolution shell.

	D/NP
Data collection	
X-ray source	ID30A3
Wavelength (Å)	0.9677
Space group	P1
a (Å)	75.16
b (Å)	85.16
c (Å)	103.38
α (º)	91.203
β (º)	101.940
γ (º)	101.017
Resolution range (Å)	40 - 2.35 (2.48 - 2.35)
Completeness (%)	91.4 (93.3)
Rsym (I) [(a)] (%)	9.8 (90.0)
I/σI	7.36 (0.98)
Total reflections	204 313 (31 120)
Unique reflections	93 545 (14 226)
Multiplicity	2.2 (2.2)
Refinement statistics	
molecules/AU	4
R-factor (%)	20.4
R_{free} (%)	26.3
R.m.s deviations from ideal values	
Bond length (Å)	0.009
Bond angle (°)	1.19

580

Table 2: K_d values of the D/NP:RNA interactions obtained from the titration curves. The fluorescence anisotropy measurements have been performed at 150 and 300 mM NaCl using $^{5'}$phosphate-polyUC-FAM$^{3'}$.

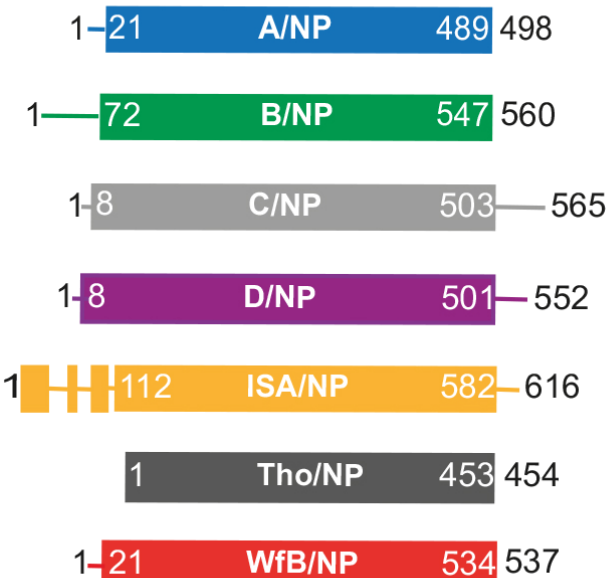
	D/NP				D/NP-529		D/NP-511	
[NaCl]	150 mM		300 mM		300 mM		300 mM	
RNA	K_d (nM)	R^2	K_d (nM)	R^2	K_d (nM)	R^2	K_d (nM)	R^2
5 nt	352 ± 14	0.997	nd	nd	nd	nd	nd	nd
8 nt	108 ± 3.3	0.998	1064 ± 75	0.998	nd	nd	nd	nd
10 nt	48 ± 1.2	0.997	229 ± 9	0.997	221 ± 7.4	0.998	260 ± 19	0.997
12 nt	13 ± 0.3	0.993	108 ± 4.2	0.989	46.1 ± 1.7	0.996	28 ± 1.4	0.991
16 nt	14 ± 0.3	0.988	14 ± 0.3	0.990	11 ± 0.3	0.993	13 ± 0.3	0.987
24 nt	14 ± 0.3	0.991	10 ± 0.2	0.995	nd	nd	nd	nd
32 nt	14 ± 0.2	0.994	16 ± 0.4	0.990	nd	nd	nd	nd

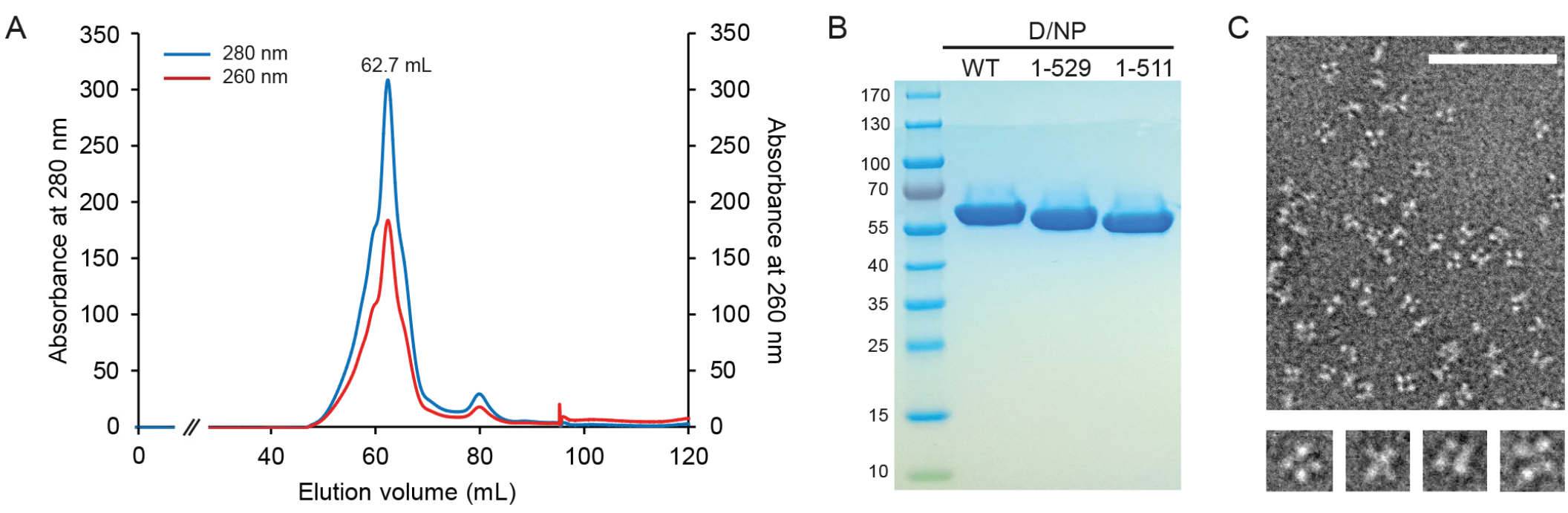
584

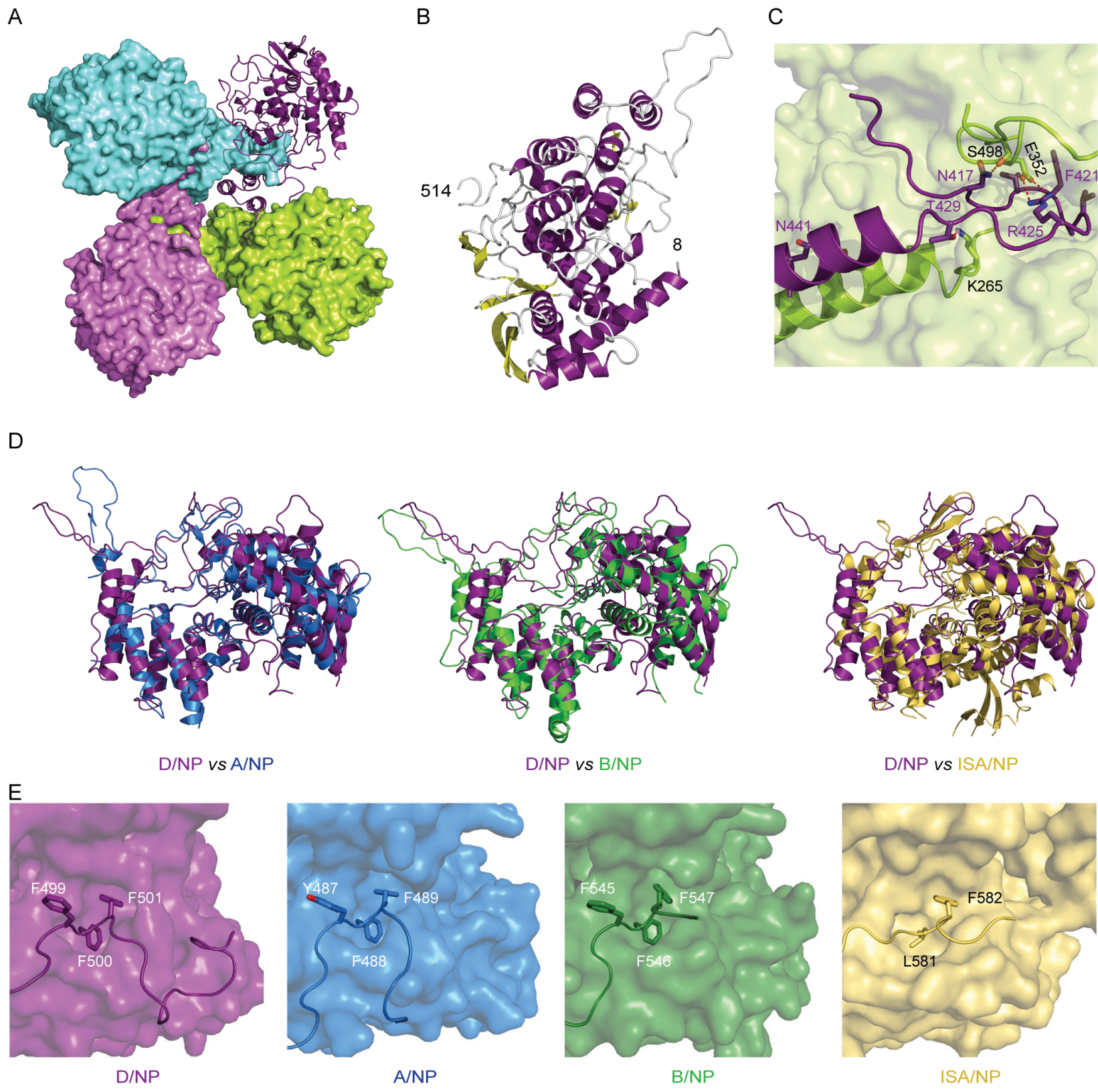
A

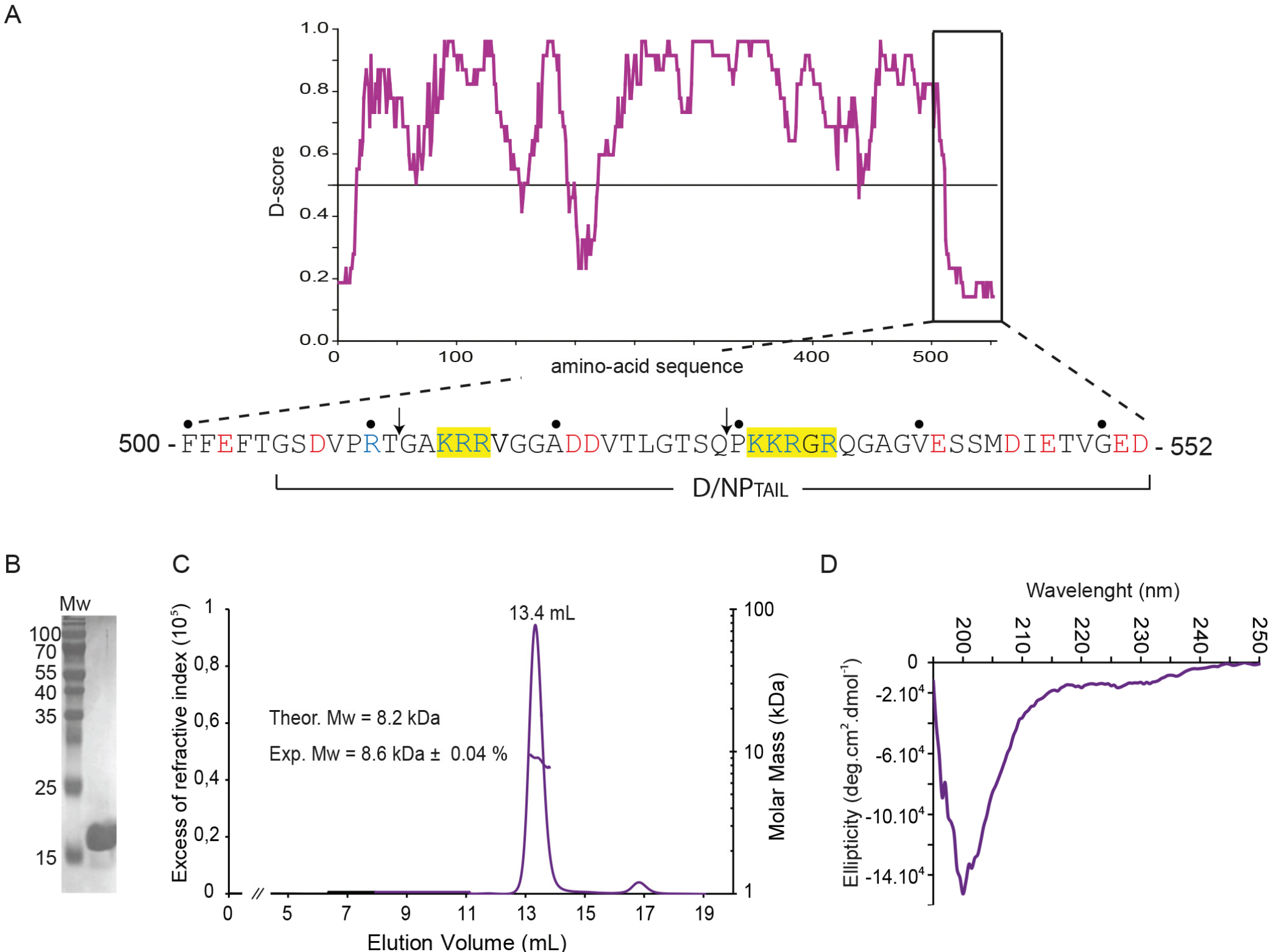
	A/NP	B/NP	C/NP	D/NP	ISA/NP
A/NP (2IQH)		1.6 Å (383 Cα)	-	2.1 Å (356 Cα)	2.9 Å (340 Cα)
B/NP (3TJ0)	37 %		-	1.9 Å (365 Cα)	2.8 Å (340 Cα)
C/NP	18 %	21 %		-	-
D/NP	17 %	21 %	38 %		2.9 Å (330 Cα)
ISA/NP (4EWC)	12 %	12 %	14 %	14 %	
Tho/NP	12 %	15 %	13 %	11 %	13 %
WfB/NP	12 %	12 %	14 %	13 %	12 %

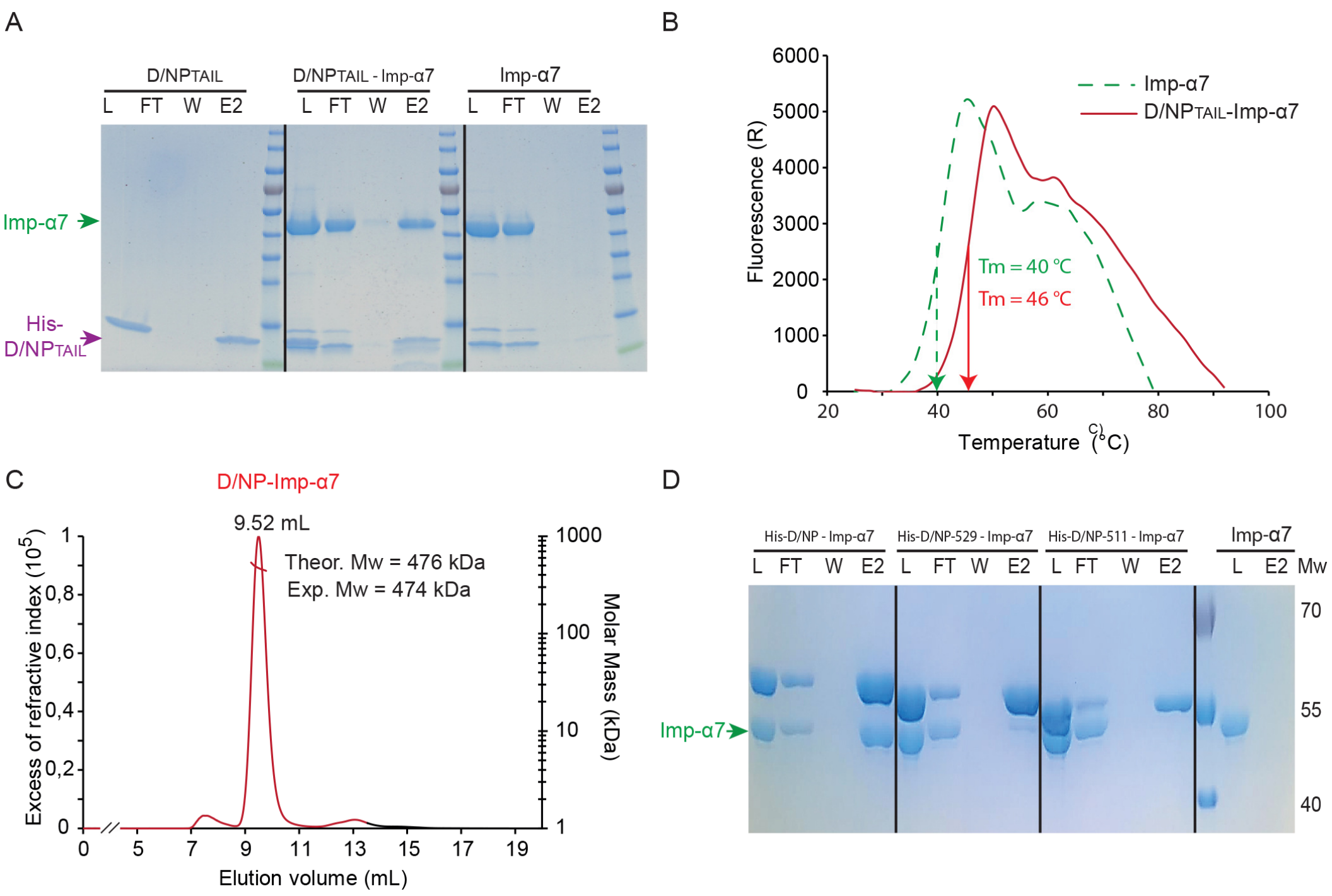
B

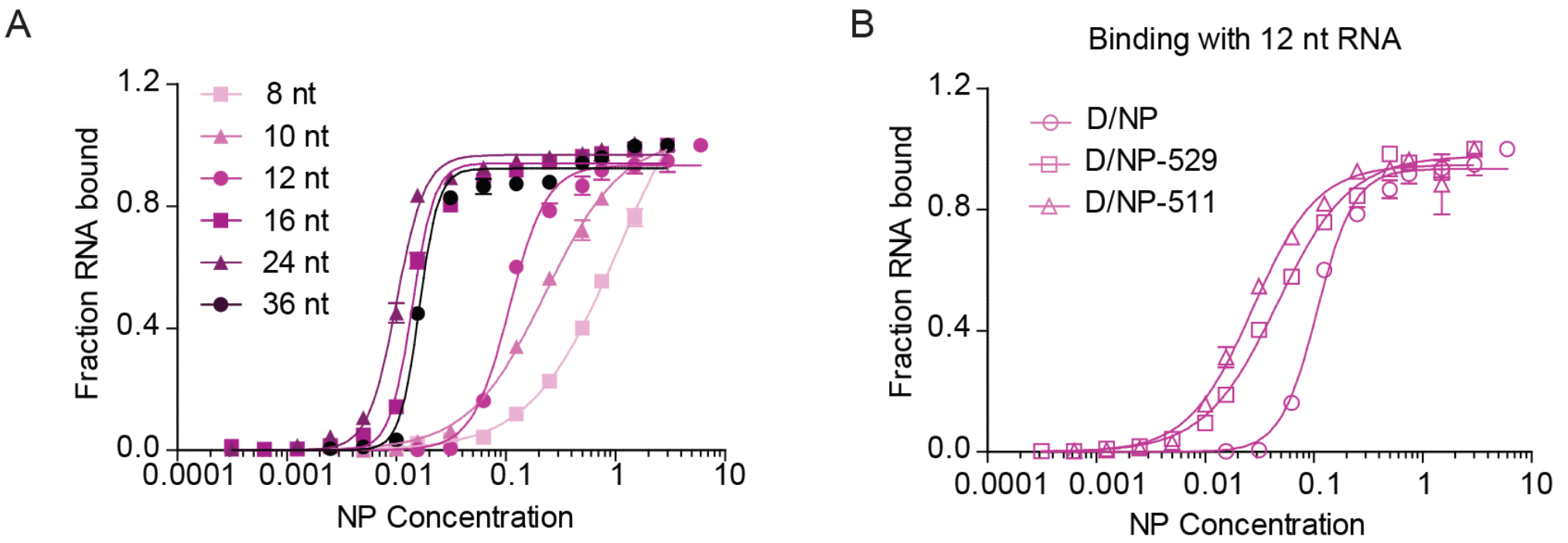



A
280 nm
260 nm
62.7 mL
Absorbance at 280 nm
Absorbance at 260 nm
Elution volume (mL)
0
40
60
80
100
120
50
150
200
250
300
350
B
D/NP
WT
1-529
1-511
170
130
100
70
55
40
35
25
15
10
C


A
B
514
8
C
S498
E352
N417
F421
T429
N441
R425
K265
D
D/NP vs A/NP
D/NP vs B/NP
D/NP vs ISA/NP
E
F499
F501
F500
Y487
F489
F488
F545
F547
F546
F582
L581
D/NP
A/NP
B/NP
ISA/NP


A
D-score
amino-acid sequence
500 - FFEFTGSDVPRTGAKRRVGGADDVTLGTSQPKKRGRQGAGVESSMDIETVGED - 552
D/NPTAIL
B
Mw
100
70
55
40
35
25
15
C
13.4 mL
Theor. Mw = 8.2 kDa
Exp. Mw = 8.6 kDa ± 0.04 %
Excess of refractive index (10^5)
Molar Mass (kDa)
Elution Volume (mL)
D
Wavelenght (nm)
Ellipticity (deg.cm^2.dmol^-1)


A
D/NPTAIL
D/NPTAIL - Imp-α7
Imp-α7
L FT W E2
L FT W E2
L FT W E2
Imp-α7
His-D/NPTAIL
B
Imp-α7
D/NPTAIL-Imp-α7
Fluorescence (R)
6000
5000
4000
3000
2000
1000
0
Tm = 40 °C
Tm = 46 °C
20
40
60
80
100
Temperature (°C)
C
D/NP-Imp-α7
9.52 mL
Theor. Mw = 476 kDa
Exp. Mw = 474 kDa
Excess of refractive index (10^5)
1
0,8
0,6
0,4
0,2
0
1000
100
10
1
Molar Mass (kDa)
0 5 7 9 11 13 15 17 19
Elution volume (mL)
D
His-D/NP - Imp-α7
His-D/NP-529 - Imp-α7
His-D/NP-511 - Imp-α7
Imp-α7
L FT W E2
L FT W E2
L FT W E2
L E2
Mw
Imp-α7
70
55
40


A
1.2
0.8
0.4
0.0
Fraction RNA bound
8 nt
10 nt
12 nt
16 nt
24 nt
36 nt
0.0001
0.001
0.01
0.1
1
10
NP Concentration
B
Binding with 12 nt RNA
1.2
0.8
0.4
0.0
Fraction RNA bound
D/NP
D/NP-529
D/NP-511
0.0001
0.001
0.01
0.1
1
10
NP Concentration