



**HAL**  
open science

# Détermination des structures de petites protéines par RMN

Beate Bersch

► **To cite this version:**

Beate Bersch. Détermination des structures de petites protéines par RMN. Biologie structurale [q-bio.BM]. Université Joseph Fourier (Grenoble I), 2003. tel-01449090

**HAL Id: tel-01449090**

**<https://hal.univ-grenoble-alpes.fr/tel-01449090>**

Submitted on 30 Jan 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Dossier**

présenté par

**Beate Bersch**

pour obtenir

**L'Habilitation à Diriger des Recherches**

**Biologie**

soutenue le 24 novembre 2003

Composition du jury☐

Guy Lippens

Rapporteur

Alain Milon

Rapporteur

Jean-Marc Lancelin

Rapporteur

Roland Douce

François Penin

## Table des Matières

|       |   |     |
|-------|---|-----|
| 1     | Introduction.....   | 1   |
| 2     | Détermination des structures de petites protéines par RMN□l'apport du marquage isotopique.....  | 2   |
| 2.1   | Le marquage isotopique.....   | 2   |
| 2.2   | Approche classique par la RMN du proton□Introduction .....  | 4   |
| 2.2.1 | Principe d'attribution et d'obtention des données structurales dans le cas d'une protéine non-marquée aux isotopes .....  | 5   |
| 2.2.2 | Etude structurale du cytochrome <i>c551</i> d' <i>Ectothiorhodospira halophila</i> .....  | 7   |
|       | <u>Article 1.</u> <sup>1</sup> H and <sup>13</sup> C NMR assignments and structural aspects of a ferrocyclochrome□-551 from the purple phototrophic bacterium <i>Ectothiorhodospira halophila</i> ..... | 8   |
|       | <u>Article 2.</u> <i>Ectothiorhodospira halophila</i> ferrocyclochrome□ <sub>551</sub> : solution structure and comparison with bacterial cytochromes□ .....  | 19  |
| 2.2.3 | Etude structurale du module de type EGF de la protéase C1r du complément humain. ....   | 36  |
|       | <u>Article 3.</u> Solution structure of the Epidermal Growth Factor-like Module of Human Complement Protease C1r, an Atypical Member of the EGF Family† .....   | 37  |
| 2.3   | Approche pour des protéines doublement marquées de petite taille□Introduction .....   | 48  |
| 2.3.1 | Approche optimisée d'attribution et d'obtention des informations structurales dans le cas d'une petite protéine doublement marquée.....   | 48  |
| 2.3.2 | Détermination optimisée de structure à partir des données RMN non-ambiguës.....   | 51  |
| 2.3.3 | Le domaine N-terminal de la réductase du mercure de <i>Ralstonia metallidurans</i> □ repliement de la chaîne principale et liaison au mercure. ....   | 53  |
|       | <u>Article 4.</u> Biophysical Characterization of the MerP-like Amino-Terminal Extension of the Mercuric Reductase from <i>Ralstonia metallidurans</i> CH34.....  | 54  |
| 2.4   | Conclusion.....   | 64  |
| 3     | Etudes des protéines > 15kDa par RMN.....   | 65  |
| 3.1   | Etude par RMN de la protéine Fur d' <i>Escherichia coli</i> □du monomère au dimère.....   | 66  |
| 3.2   | Le domaine flavodoxine de la sulfite réductase d' <i>Escherichia coli</i> .....   | 72  |
|       | <u>Article 5.</u> Reactivity, Secondary Structure and Molecular Topology of the <i>Escherichia coli</i> Sulfite Reductase Flavodoxin-like Domain.....   | 74  |
| 3.3   | La protéine MsrA d' <i>Erwinia chrysanthemi</i> □les bases structurales du mécanisme réactionnel revisités. ....  | 86  |
|       | <u>Article 6.</u> Direct structure determination using residual dipolar couplings: reaction-site conformation of methionine sulfoxide reductase in solution.....  | 88  |
| 4     | Conclusion.....   | 95  |
| 5     | Références .....  | 96  |
|       | Annexe: Curriculum vitae et production scientifique.....  | 103 |



## 1 Introduction

Depuis mon entrée au CNRS en 1992, mon travail de recherche a porté sur les études de protéines par Résonance Magnétique Nucléaire (RMN). Au sein du Laboratoire de Résonance Magnétique Nucléaire (LRMN) de l'Institut de Biologie Structurale à Grenoble, mon rôle se situe à l'interface entre la biologie et la RMN. Ainsi, il a été possible d'établir des étroites collaborations entre des chercheurs biochimistes et notre équipe de RMN. Les informations structurales qui ont été obtenues à partir des données RMN fournissent des informations précieuses complétant souvent les études fonctionnelles établies dans les laboratoires de mes collaborateurs. Ma motivation principale consiste donc en l'établissement des liens forts entre la biochimie et la biologie structurale.

Avec la cristallographie à rayons X, la RMN est une méthode puissante pour l'obtention de l'information structurale à haute résolution des macromolécules biologiques (protéines, ADN, glucides ...). De plus, la RMN est souvent la technique de choix pour accéder à des informations précises sur la dynamique des biomolécules ainsi sur des interactions intermoléculaires.

Pendant les dernières quinze années, la RMN a connu des avancées importantes, permettant aujourd'hui l'étude de macromolécules des tailles de plus en plus élevées, approchant les poids moléculaires de 80 kDa. Plusieurs facteurs ont contribué à cette évolution

L'instrumentation l'intensité des champs magnétiques est passée de 2.1 Tesla (fréquence  $^1\text{H}$  de 90 MHz) à 21 Tesla (fréquence  $^1\text{H}$  de 900 MHz) ce qui a résulté en une augmentation considérable de la résolution et de la sensibilité de l'expérience RMN. En parallèle, l'utilisation des cryosondes contribue à augmenter le rapport du signal sur bruit d'un facteur 2 à 3, ce qui permet de travailler à des concentrations des échantillons plus faibles et/ou de réduire le temps d'accumulation des expériences.

Les développements méthodologiques vers la fin des années 1980, les techniques de la RMN hétéronucléaire ont été introduites. Dans ces expériences, l'aimantation est transférée du proton vers un ou plusieurs hétéronoyaux ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) pour être à nouveau détecté au niveau du proton. Dans ces expériences, la sensibilité est déterminée par le rapport gyromagnétique important du proton, tandis que le transfert très efficace de l'aimantation est dû au constantes de couplage  $^1\text{J}$  élevées. De plus, la détection d'un deuxième voire d'un troisième noyau permet d'augmenter la résolution de l'expérience par l'ajout d'une ou plusieurs dimensions supplémentaires.

Plus récemment, le développement de la technique TROSY (transverse relaxation optimized spectroscopy), fondée sur l'utilisation constructive de la corrélation croisée entre deux mécanismes de relaxation à champ élevé, permet aujourd'hui des études de macromolécules à haut poids moléculaire (> 50 kDa) par RMN.

En ce qui concerne l'obtention des informations structurales, il faut mentionner l'exploitation des paramètres expérimentaux liés à l'orientation d'une interaction donnée par rapport à un système de référence extérieur. Ceci inclut les interactions dues au paramagnétisme des protéines, à leur alignement partiel dans le champ magnétique ou à la relaxation modulée par l'anisotropie de diffusion rotationnelle. Cette information à longue portée est évidemment complémentaire des contraintes structurales classiques (effet Overhauser nucléaire, couplages J) mais devient extrêmement précieuse dans le cas où l'exploitation des données classiques est limitée par la taille de la molécule (dégénérescence des fréquences proton, nombre réduit des protons dans des molécules deutériées, diffusion de spin) ou par sa forme (molécules allongées ou contenant plusieurs domaines).

La surexpression et le marquage isotopique des protéines. La RMN hétéronucléaire n'aurait pas pu rencontrer autant de succès, si les techniques de la biologie moléculaire n'avaient pas permis le clonage des gènes et la surexpression des protéines correspondantes dans des organismes appropriés. Il est ainsi devenu possible de produire des protéines enrichies jusqu'à 100% en  $^{15}\text{N}$ ,  $^{13}\text{C}$  et  $^2\text{H}$  d'une manière aisée. La deutériation des macromolécules devient surtout utile pour des masses moléculaires élevées pour la RMN (> 15 kDa) où elle permet la diminution significative de la relaxation dipolaire.

Dans ce document, j'aimerais illustrer les différentes facettes de la RMN appliquée à l'étude de protéines par quelques exemples choisis parmi les projets de recherche accomplis durant ces dernières années par moi-même et également par des étudiants que j'ai encadrés lors de leur séjour doctoral au laboratoire. Les sujets que je vais traiter sont variés, même si plusieurs sont reliés aux métalloprotéines et au transfert d'électrons. Ceci est principalement dû au fait qu'à l'heure actuelle, nous n'avons pas les moyens (personnels et matériels) nécessaires au développement d'une recherche biologique propre au sein du LRMN. Mes projets de recherche sont donc majoritairement nés suite à des discussions avec des chercheurs des laboratoires de biologie ou de biochimie.

## **2 Détermination des structures de petites protéines par RMN et l'apport du marquage isotopique**

Comme nous l'avons vu, le marquage isotopique joue un rôle primordial dans la RMN des macromolécules. Pourtant, il y a une dizaine d'années, le prix élevé du carbone-13 faisait que le double-marquage ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ) était surtout envisagé pour les études des protéines de masses moléculaires supérieures à 10 kDa. Depuis que les produits isotopiques sont devenus plus accessibles, ce type de marquage est appliqué d'une manière plus systématique afin de faciliter aussi bien l'étape de l'attribution que l'obtention des données structurales. Dans cette première partie, je vais illustrer l'importance du marquage isotopique, notamment celui au  $^{13}\text{C}$ , pour l'obtention des informations structurales dans le cas de petites protéines (<10 kDa). Je comparerai la détermination de structure «classique», basée sur la RMN du proton à une étude plus récente d'une protéine marquée au  $^{15}\text{N}$  et au  $^{13}\text{C}$ . Cette dernière a été effectuée en étroite collaboration avec Bernhard Brutscher du LRMN avec le but d'optimiser les étapes nécessaires à l'obtention des informations structurales, ce qui sera illustré plus en détail dans la partie 2.3.

### **2.1 Le marquage isotopique**

Le marquage des biomolécules aux isotopes stables ( $^{15}\text{N}$ ,  $^{13}\text{C}$  et  $^2\text{H}$ ) a révolutionné la RMN biomoléculaire (1,2,3). Des expériences multidimensionnelles basées sur la détection de plusieurs noyaux différents ont permis de réduire les superpositions spectrales, facilitant significativement les étapes d'attribution et d'obtention de contraintes structurales (4, 5 et références citées). De la même manière, l'introduction d'hétéronoyaux permet, parmi d'autres, l'utilisation des filtres isotopiques pour l'étude des complexes de protéines (6), les mesures de relaxation pour la caractérisation de la dynamique de la biomolécule (7) ou encore les mesures de couplages dipolaires résiduels, largement utilisés pour la validation ou l'affinement des structures tridimensionnelles (8). Enfin, le marquage au  $^2\text{H}$  est une condition préalable pour l'étude des protéines de masses moléculaires supérieures à 20 kDa par la RMN (9, 10) L'effet principal du remplacement des protons par le deutérium est la réduction de la relaxation des carbones auxquels ils sont directement liés. Cette diminution de la relaxation est

cruciale pour les expériences à triple-résonances dans lesquelles l'aimantation est transférée par le carbone- $\beta$  ainsi que pour toute autre expérience impliquant des étapes de transfert de cohérence longues. Dans ce contexte, il faut également mentionner qu'une deutériation maximale est requise pour l'utilisation des expériences de type TROSY (11), résultant dans un gain en résolution et en sensibilité dans le cas des très grosses protéines étudiées à champ élevée (fréquence du proton  $\geq$  800 MHz).

Le marquage isotopique nécessite le clonage de la séquence génique codant pour la protéine d'intérêt dans un vecteur de surexpression et la production de la protéine dans un hôte permettant l'enrichissement aux isotopes. Ce marquage est généralement effectué par la culture de cet organisme dans un milieu minimum contenant du  $^{15}\text{N}$ -chlorure d'ammonium et du  $^{13}\text{C}$ -glucose comme seules sources d'azote et du carbone. Ceci représente déjà une sévère limitation, car seuls les organismes étant capables de pousser dans un milieu pauvre sont adaptés à la production des protéines marquées. Généralement, les protéines sont produites soit dans la bactérie *Escherichia coli*, soit dans la levure *Pichia pastoris*. Ce choix limité est à l'origine de difficultés supplémentaires surtout pour l'expression des protéines provenant des eucaryotes ou des archaebactéries. Ceci est lié par exemple à une utilisation différente de codons d'ADN notamment chez les archaebactéries, à l'incapacité des bactéries d'introduire certaines modifications post-traductionnelles (ponts disulfures, glycosylation ...) ou, au contraire, dans les levures à une hyper-glycosylation des protéines exprimées. De plus, la culture cellulaire dans un milieu pauvre se traduit très souvent par une diminution importante du nombre des cellules en fin de culture et par conséquent de la quantité de protéine surproduite. Dans ce contexte, il faut mentionner les coûts associés à la production des protéines marquées qui font qu'un bon niveau de surproduction est impératif. A titre indicatif, une bonne surproduction doit permettre d'obtenir au moins un échantillon RMN par litre de culture, correspondant à 500  $\mu\text{l}$  d'une concentration de 1mM (pour une protéine d'une masse de 20kDa cela correspond à 10mg de protéine purifiée à partir d'un litre de culture). Ainsi, la capacité de surproduire la protéine d'intérêt dans des bonnes conditions afin de permettre son marquage isotopique représente souvent la première étape limitante pour une étude par RMN.

Cependant, durant les dernières années, des nouvelles solutions sont apparues afin de remédier à certains de ces problèmes. Citons notamment la commercialisation des milieux riches marqués, basés sur les hydrolysats d'algues (Spectra Stable Isotopes, MD, USA) ou de bactéries (Silantes, München, RFA) et permettant des simple, double ou triple marquages aux  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$ . Au niveau de la biologie moléculaire, il faut mentionner l'évolution des vecteurs d'expression avec l'introduction des nouvelles étiquettes (tags) et protéines de fusion (amélioration de la purification, augmentation du rendement et du taux de protéine soluble), et le développement des souches spécifiques souvent déficientes en protéases, capables de compenser la présence de codons rares ou de permettre la formation de pont disulfures stables (voir par exemple «[pET System Manual](#)» de Novagen). Des tentatives ont été effectuées avec succès afin d'adapter d'autres systèmes d'expression, comme par exemple des cyanobactéries photoautotrophes (12), des cellules ovariées d'hamster (13) et des systèmes «[cell-free](#)», utilisant des extraits d'*E. coli* (14, commercialisé par Roche, Mannheim, RFA et par Silantes, München, RFA) ou de germe de blé (15). Il est clair que les avancements et le succès de la RMN biomoléculaire de la dernière décennie sont étroitement liés à l'évolution rapide dans le domaine de la surexpression des protéines marquées aux isotopes.

J'aimerais ajouter à cet endroit que la biologie moléculaire et le marquage des protéines forment une partie importante de mon travail, qui n'est pas très visible à travers la production scientifique. Pendant deux ans (1997-1999), j'ai partagé mon temps entre le LRMN et le Laboratoire

d'Enzymologie Moléculaire de l'IBS (LEM, dirigé par Gérard Arlaud et Jean Gagnon) où j'ai appris les techniques de la biologie moléculaire (clonage des gènes, expression des protéines dans *P. pastoris* et dans *E. coli*) dans le contexte de l'étude structurale des composants du complément humain (voir 2.2.3). Par la suite, je me suis occupée de la mise au point des cultures bactériennes pour le marquage isotopique et de l'équipement de notre laboratoire de biochimie, et cela conjointement avec Laurence Blanchard (LRMN).

## 2.2 Approche classique par la RMN du proton

Malgré le fait que la plupart des études des protéines par RMN sont actuellement effectuées en utilisant des protéines marquées aux isotopes, ceci n'est pas toujours possible. Je vais présenter deux études utilisant l'approche classique de la RMN du proton. Les deux protéines en question n'ont pas été obtenues par surexpression dans des bactéries mais par purification de la protéine en abondance naturelle à partir de l'hôte physiologique (cas du cytochrome *c551*) ou par synthèse peptidique en phase solide (cas du module EGF de la protéase C1r du complément humain).

La surexpression des protéines contenant les groupements prosthétiques liés de façon covalente, comme l'hème, peut parfois poser des problèmes. La synthèse et l'introduction de ce groupement requièrent des enzymes spécifiques qui doivent être présents dans l'hôte d'expression en nombre suffisant. Ceci rend la surexpression de ces protéines dans un hôte différent de l'organisme d'origine très difficile, puisqu'il faut également produire toutes les protéines accessoires. Cependant, l'utilisation du même organisme ou d'un organisme proche peut être plus prometteuse, même si le taux d'expression est généralement plus faible que dans *E. coli*. Par exemple, deux des cytochromes *c* étudiés au laboratoire dans les années 1990 ont pu être exprimés à partir des plasmides obtenus par génie génétique dans des organismes qui étaient soit identiques, soit très proches des organismes d'origine : le cytochrome *c553* de *Desulfovibrio vulgaris* dans *Desulfovibrio desulfuricans* (16) et le cytochrome *c2* de *Rhodobacter capsulatus* dans *Rhodobacter sphaeroides* ou dans *Rhodobacter capsulatus* (17). Une telle approche n'a pas été tentée pour le cytochrome *c551* d'*Ectothiorhodospira halophila*, dont 53  $\mu$ mol (correspondant à 400 mg) ont été purifiés à partir de 1500 g de cellules obtenues dans un volume de culture de 500 l (18).

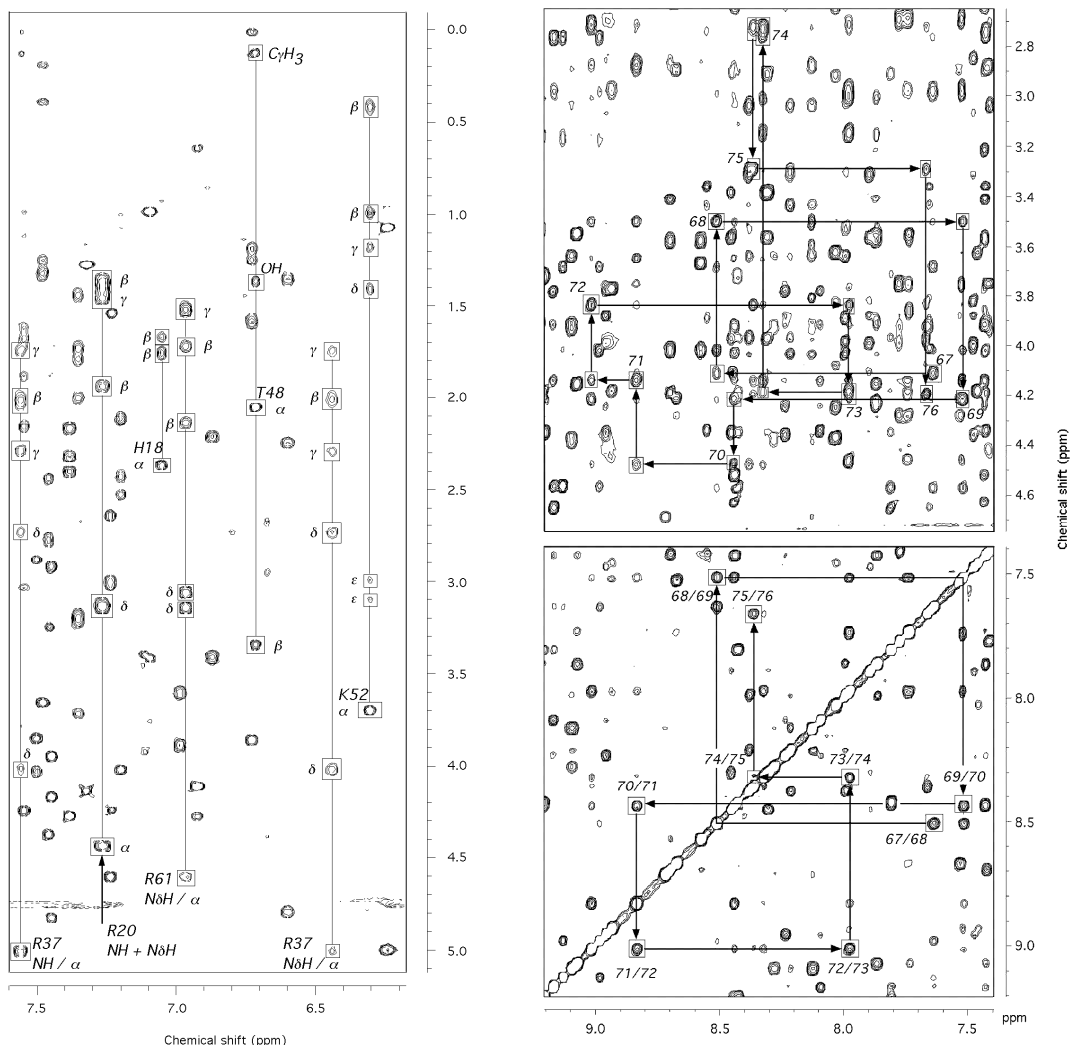
En ce qui concerne les petits peptides ou protéines tel que le module EGF de la composante C1 du complément humain (53 résidus), leur surproduction dans des bactéries n'est pas aisée car ils sont fréquemment dégradés par protéolyse, surtout quand ils ne sont que peu structurés. Par contre, ils peuvent facilement être obtenus par synthèse peptidique, méthode qui a été initialement utilisée pour l'obtention du module EGF étudié (19). Cette technique permet également un marquage isotopique spécifique, en introduisant des acides aminés marqués à des positions données de la séquence. Néanmoins, le coût d'achats d'acides aminés marqués est très élevé, rendant l'application de cette technique très limitée. Cependant, depuis quelques années, des multiples travaux ont été publiés visant la surexpression des petits peptides en vue de leur marquage isotopique (pour un résumé récent, voir 20). Nous avons grâce à cela obtenu des variants du module EGF en fusion avec la thiorédoxine et avec l'ubiquitine (non-publié), ce qui a permis leur marquage isotopique et l'étude dynamique par RMN (21).



## 2.2.1 Principe d'attribution et d'obtention des données structurales dans le cas d'une protéine non-marquée aux isotopes

La RMN ne peut apporter des informations structurales qu'une fois les résonances de noyaux observés attribuées à des atomes de la molécule. Une attribution la plus complète possible est donc une condition préalable à l'interprétation des différents paramètres expérimentaux en termes d'informations structurales.

L'attribution séquentielle des protéines non-marquées a été décrite en détail par K. Wüthrich (22). Ce processus est effectué globalement en deux étapes. (1) Identification des systèmes de spins basée sur les interactions scalaires (couplages  $J$ ) entre les protons à l'intérieur d'un même acide aminé et (2) le positionnement de ces systèmes de spins selon la séquence protéique à l'aide des interactions dipolaires (NOE) séquentielles et à courte distance (voir Figure 1). Une description détaillée est donnée dans l'Article 1 qui résume l'attribution des résonances du cytochrome *c551* d'*Ectothiorhodospira halophila* et l'obtention des premières informations structurales accessibles avant l'analyse complète des effets Overhauser nucléaires (NOE).



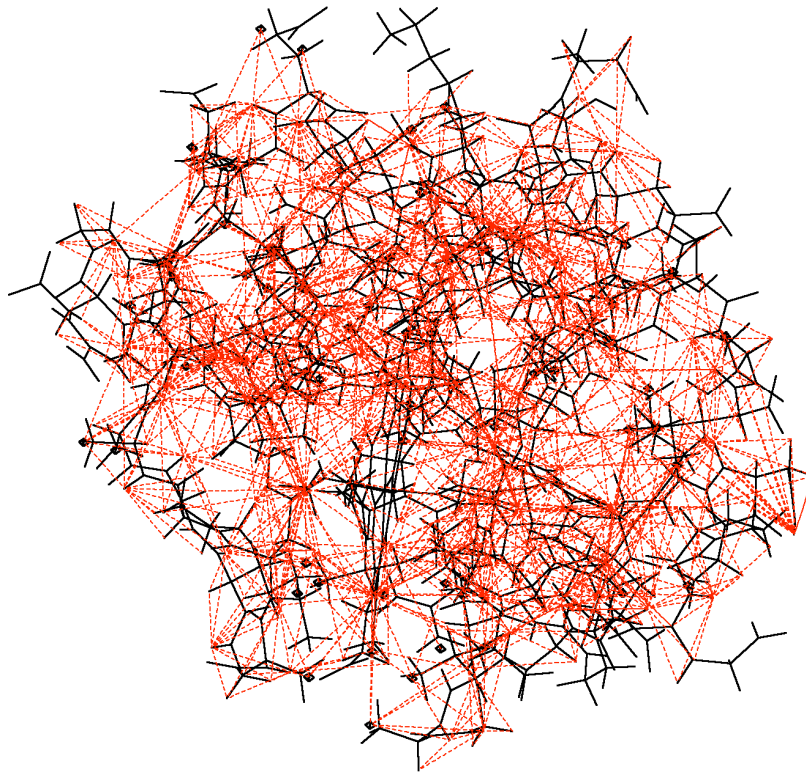
**Figure 1.** Principe de l'attribution séquentielle des protéines non-marquées. A. Attribution des systèmes de spins à l'aide de l'expérience TOCSY. Délimitation de plusieurs systèmes de spin à la fréquence du proton amide ou du proton  $NH$  des chaînes latérales (arginines). B. et C. Attribution séquentielle à l'aide de l'expérience NOESY. Connexions séquentielles  $H_i-NH_{i+1}$  (B) et  $NH_i-NH_{i+1}$  (C). Conditions expérimentales  $[c551]=8mM$ ,  $T=27^\circ C$ ,  $pH=5.8$ , temps de mélange=100ms pour les deux expériences.

Différentes informations sont ensuite utilisées pour le calcul de la structure à une résolution atomique. Les NOE fournissent des informations sur des distances entre protons (23), les mesures des constantes de couplage  $^3J$  permettent d'obtenir des renseignements sur des angles dièdres (24) selon la relation de Karplus (25) et l'observation des protons amides en échange lent avec l'eau indique la présence de liaisons hydrogènes, souvent situées dans des régions de structure secondaire (26). Le NOE est de loin le paramètre le plus important pour les calculs de structures car il est le seul à contenir l'information à longue distance (par rapport à la séquence primaire) qui est indispensable pour le repliement correct de la chaîne peptidique.

Cependant, l'utilisation des NOE comme contrainte structurale nécessite de considérer différents aspects.

La qualité des structures obtenues à partir des NOE dépend, entre autre, de l'attribution correcte de ceux-ci. Pour éviter l'introduction des erreurs, l'attribution de tous les protons des chaînes latérales est cruciale. Il est clair que l'augmentation de la taille de protéine entraîne une multiplication du nombre de protons. Il en résulte que le nombre de NOE qui peuvent être attribués d'une manière non-ambiguë diminue significativement. Une attribution correcte (et de préférence non-ambiguë) est particulièrement importante pour les quelques NOE clés à longue distance, qui déterminent le repliement global de la chaîne peptidique (27, 28, 29). De la même manière, des erreurs dans l'interprétation des NOE en termes de distances réduisent la qualité des structures résultantes (30). Le NOE est proportionnel à la distance entre les deux noyaux en  $1/r^6$ . Néanmoins, des estimations exactes de distances ne sont pas possibles car le NOE est également affecté par la diffusion de spin et par la mobilité locale. De ce fait, les contraintes NOE sont généralement introduites sous forme des gammes de distances ou avec des limites d'erreur assez larges. Enfin, la qualité de l'ensemble structural obtenu dépend également du nombre des contraintes utilisées (31). Plus ce nombre est élevé, meilleure est la définition de l'ensemble structural. Cependant, cette définition est très sensible à l'exactitude de ces informations. Un nombre très élevé des contraintes trop serrées ou la présence de fausses attributions peut introduire des torsions ou des erreurs systématiques dans l'ensemble structural. Ceci est surtout vrai pour l'orientation des vecteurs NH, car la plupart des NOE implique les protons amides. Une illustration de l'ensemble des NOE utilisés pour le calcul de la structure du cytochrome *c551* est donnée dans la Figure 2.

Différentes approches sont utilisées pour le calcul des coordonnées atomiques à partir des paramètres expérimentaux, dont une description plus exhaustive peut être trouvée dans une revue de P. Güntert (32). Dans les deux études que je vais présenter ci-dessous, nous avons utilisé un protocole combinant des étapes de recuit simulé et de dynamique moléculaire sous contraintes. Ce protocole a été développé au laboratoire par M. Blackledge (33).



**Figure 2.** Projection de l'ensemble des distances NOE utilisées pour le calcul de la conformation du cytochrome *c551* sur sa structure tridimensionnelle. Les diamants désignent les pseudoatomes qui remplacent les protons non-attribués individuellement (méthyles, méthylènes dégénérés, aromatiques ...).

### 2.2.2 Étude structurale du cytochrome *c551* d'*Ectothiorhodospira halophila*

Collaboration avec Terry Meyer, University of Tucson, Arizona, USA.

A mon arrivée à l'IBS, un des sujets principaux au laboratoire était la détermination de structures tridimensionnelles des cytochromes *c* bactériens, dans le but de caractériser les propriétés structurales des différentes sous-classes de cette famille de protéines. Les cytochromes *c* jouent un rôle central dans les systèmes biologiques de transport d'électrons et sont présents d'une manière ubiquitaire dans tout organisme. Ce sont des protéines contenant généralement 80 à 120 acides aminés qui sont repliés autour du groupement prosthétique, l'hème. Dans les cytochromes *c* de la classe 1, l'hème est lié d'une manière covalente à une séquence d'acides aminés conservée situé dans la partie N-terminale, CXXCH, et le fer est lié par deux ligands protéiques, une histidine et une méthionine. Malgré un certain nombre d'éléments structuraux communs, les cytochromes *c* sont caractérisés par une variabilité élevée des séquences primaires et du potentiel redox, qui peut varier entre 0 et 450 mV pour les cytochromes de la classe 1. L'étude du cytochrome *c551* d'*E. halophila*, une bactérie pourpre, a été motivée principalement par la relation entre sa structure et son potentiel redox très bas (50 mV) mais aussi par son analogie séquentielle (et structurale) avec des cytochromes *c* des bactéries vertes sulfureuses et les cyanobactéries.

Ce travail a donné suite à deux publications qui sont insérées ci-dessous. La première décrit l'attribution totale du cytochrome *c551* d'*E. halophila*, la détermination des structures secondaires et la proposition d'un modèle structural tandis que la deuxième est consacrée à la détermination de la structure à haute résolution et sa comparaison avec les structures d'autres cytochromes *c* d'origine bactérienne.

# <sup>1</sup>H and <sup>13</sup>C NMR assignments and structural aspects of a ferrocytochrome<sub>c</sub>-551 from the purple phototrophic bacterium *Ectothiorhodospira halophila*

Beate Bersch<sup>1</sup>, Bernhard Brutscher<sup>1</sup>, Terrance E. Meyer<sup>2</sup> and Dominique Marion<sup>1</sup>

<sup>1</sup> Institut de Biologie Structurale - Jean-Pierre Ebel, CNRS-CEA, 41 Avenue des Martyrs, 38027 Grenoble Cedex 1, France.

<sup>2</sup> Department of Biochemistry, University of Arizona, Tucson, Arizona 85721, USA

Two dimensional nuclear magnetic resonance was used to assign the <sup>1</sup>H and <sup>13</sup>C resonances of ferrocytochrome<sub>c</sub>-551 from *Ectothiorhodospira halophila*, a halophilic phototrophic purple bacterium. This 78 residue protein belongs to a small subgroup of class I cytochromes<sub>c</sub> together with the analogous cytochromes<sub>c</sub>-551 from *E. halochloris* and *E. abdelmalekii*. A nearly complete assignment of <sup>13</sup>C resonances was obtained at natural abundance using a gradient enhanced <sup>1</sup>H-detected heteronuclear single quantum coherence experiment (HSQC). This was found to be extremely useful for the unambiguous assignment of side chain protons. The secondary structure of the protein was determined from analyses of short and medium range nuclear Overhauser enhancements (NOE), amide proton exchange and <sup>13</sup>C chemical shifts. Three helices could be identified which are well conserved among the class I cytochromes<sub>c</sub>. There is some evidence for two other regions of less well defined helical structure. From a preliminary analysis of long range NOE it is shown that in the *E. halophila* cytochrome<sub>c</sub>-551 the general cytochrome<sub>c</sub> fold is well conserved, including the three conserved helices (residues 2→8, 41→50, 63→76), the regions around the heme ligands (Cys14-Ser15-Ser16-Cys17-His18, Met55) and the omega loop (residues 18→28). In addition, three variable segments of the protein are discussed in detail, one of those including a *cis*-proline, a feature so far unique in the cytochrome<sub>c</sub> family. Structural alignments of the *E. halophila* cytochrome<sub>c</sub>-551 with two other *Pseudomonas* cytochrome *c*<sub>5</sub> homologs (*Azotobacter vinelandii* cytochrome *c*<sub>5</sub> and *Chlorobium limicola* cytochrome<sub>c</sub>-555) are provided which are based on sequence homologies and secondary structure alignments.

The family *Ectothiorhodospiraceae* belongs to the phototrophic purple bacteria and consists of a single genus, including moderately to extremely halophilic organisms which are metabolically intermediate between the families *Rhodospiraceae* and *Chromatiaceae* (Imhoff, 1989). Two groups can be distinguished among the *Ectothiorhodospiraceae*, those which contain either bacteriochlorophyll *a* or *b*. Only the bacteriochlorophyll *a* containing bacteria are able to grow photoorganotrophically with reduced sulfur compounds as electron donors (Brune, 1989). The extremely halophilic *Ectothiorhodospira halophila*, a chlorophyll *a* containing species, possesses a cyclic electron transfer chain, composed of a photosynthetic

reaction center and a cytochrome *bc*<sub>1</sub> complex (Leguijt & Hellingwerf, 1991; Leguijt et al., 1993). In addition, soluble metallo-redox proteins have been isolated and characterized from *E. halophila* (Meyer, 1985 ; Bartsch, 1991), among those a small cytochrome<sub>c</sub>-551 with a particularly low redox potential of 58 mV (Meyer, 1985). As has been shown for the analogous protein from *E. abdelmalekii*, cytochrome<sub>c</sub>-551 is situated at the external face of the cytoplasmic membrane and seems to be involved in sulfide oxidation (Then & Trüper, 1983). As it is located in the periplasm, it is likely that some adaptation to high salinities has occurred over the course of evolution. The sequence of this 78 residue cytochrome has been published recently (Ambler et al., 1993) together with that of the homologous protein from *E. halochloris*, a bacteriochlorophyll *b* containing species. These two proteins show 63% sequence identity but do not closely resemble other cytochromes<sub>c</sub> from purple photosynthetic bacteria. However, there seem to be some structural homologies to *Azotobacter* and *Pseudomonas* cytochrome *c*<sub>5</sub>, to the cytochromes *c*<sub>6</sub> of cyanobacteria and eukaryotic algal chloroplasts, as well as to the cytochromes<sub>c</sub>-555 from green photosynthetic bacteria (*Chlorobiales*) (Ambler et al., 1993). Therefore, it

-----  
Corresponding author □ Dominique Marion, Institut de Biologie Structurale - Jean-Pierre Ebel, 41 Avenue des Martyrs, 38027 Grenoble Cedex 1, France.

Abbreviations □ BURP, band selective, uniform response, pure-phase; DQF, double quantum filtered; COSY, J-correlated spectroscopy; HSQC, <sup>1</sup>H-detected heteronuclear single quantum coherence; NOESY, nuclear Overhauser spectroscopy; PFG, pulsed B<sub>0</sub> field gradient; TOCSY, total correlation spectroscopy.

has been proposed that the cytochromes $c_5$ -551 of the *Ectothiorhodospiraceae* belong to a small subclass of class I bacterial cytochromes $c$  (Ambler et al., 1993).

The class I cytochromes $c$  are small soluble heme proteins shuttling electrons between various components of the electron-transport chains. These proteins are characterized by their heme binding sequence CXXCH which occurs near the N-terminus, and a methionine as a sixth heme ligand which is located near the C-terminus. Their sizes range from approximately 80 to 120 residues. While class I cytochromes generally show a common folding pattern at least for some segments of the protein chain (Chothia & Lesk, 1985), they are characterized by a great variability in protein sequence and redox potential, the latter varying between 0 and +450 mV (Pettigrew & Moore, 1987). Besides the familiar mitochondrial cytochrome $c$ , at least 16 distinct subclasses of type I cytochromes are now recognized in bacteria (Meyer & Kamen, 1982; Pettigrew & Moore, 1987, Moore & Pettigrew, 1990; Meyer, 1995). Subgroups may contain either many individual species or only a few identified so far, as in the case of the *Ectothiorhodospiraceae* cytochromes $c$ -551. During evolution, unique structures were designed to optimize interactions in particular pathways such as respiration, photosynthesis, and denitrification, amongst others. Crystal structures have been determined for several mitochondrial cytochromes $c$  (horse, tuna, rice, yeast; Bushnell et al., 1990; Louie & Brayer, 1990; Murphy et al., 1992 and references therein), for bacterial cytochromes $c_2$  (*Rhodospirillum rubrum*, *Rhodobacter capsulatus*, *Paracoccus denitrificans*; Bhatia et al. 1984, Benning et al., 1991; 1994 and references therein), and for more divergent bacterial class I cytochromes $c$  *Pseudomonas aeruginosa*  $c$ -551 (Matsuura et al., 1992), *Azotobacter vinelandii*  $c_5$  (Carter et al., 1985), *Desulfovibrio vulgaris*  $c$ -553 (Nakagawa et al., 1990), *Chlorobium limicola*  $c$ -555 (Korszun & Salemme, 1977), *Pseudomonas putida* para cresol methyl hydroxylase (Mathews et al., 1991), and *Paracoccus denitrificans* methylamine dehydrogenase associated cytochrome $c$ -551 (Chen et al., 1994). Nevertheless, representative structures for only 7 out of 16 subgroups have been determined to date.

NMR represents a valuable tool to investigate the structure and dynamics of small proteins in solution. NMR studies of cytochromes $c$  have been reviewed recently (Marion, 1994). To date, solution structures have been obtained for horse cytochrome $c$  (Qi et al., 1994), *Ps. aeruginosa* cytochrome $c$ -551 (Detlefsen, 1991), *Ps. stutzeri* cytochrome $c$ -551 (Cai et al., 1992) and *D. vulgaris* Hildenborough cytochrome $c$ -553 (Blackledge et al., 1994). In this paper we present results obtained from the NMR investigation of the ferrocyclochrome $c$ -551 from *E. halophila*. We were able to completely assign the  $^1\text{H}$  resonances using a single  $\text{H}_2\text{O}$  sample and a great majority of the  $^{13}\text{C}$  resonances of the protein by the combined use of TOCSY (total correlation spectroscopy), NOESY (nuclear Overhauser spectroscopy) and natural abundance  $^{13}\text{C}$ - $^1\text{H}$  correlations at different temperatures.  $^{13}\text{C}$ - $^1\text{H}$  correlation spectroscopy turns out to be indispensable for specifically assigning side-chain  $^1\text{H}$ , as recently reported for *D. vulgaris*

cytochrome $c$ -553 (Medvedeva et al., 1993). In addition, we will discuss the secondary structure of the protein as accessed by the analyses of short range NOE, amide proton exchange, as well as  $^{13}\text{C}$  chemical shift. The general folding of the protein will be discussed by analogy to previously obtained crystal structures of the *A. vinelandii* cytochrome  $c_5$  (Carter et al., 1985) and *Chl. limicola* cytochrome $c$ -555 (Korszun & Salemme, 1977).

## Materials and Methods

**Sample preparation** The cytochrome $c$ -551 was isolated and purified from cultures of *E. halophila* strain BN9626 as described elsewhere (Ambler et al., 1993). The purity of the protein solution was better than 95% as determined by UV/Vis absorption spectroscopy. 2.7 ml of a 15 mM stock solution were washed 4 times with 500  $\mu\text{l}$  of 0.1 M potassium phosphate buffer, pH 6.0, on a centricon 3 microconcentrator (Amicon Inc., Beverly, USA) to remove residual protonated buffer components. Finally, the protein was dissolved in 350  $\mu\text{l}$  0.1 M potassium phosphate buffer in 90%  $\text{H}_2\text{O}$  / 10%  $\text{D}_2\text{O}$ , containing 20  $\mu\text{M}$  chloramphenicol to prevent any bacterial growth. Reduction of the cytochrome $c$ -551 was achieved by adding 150  $\mu\text{l}$  of 0.1 M sodium dithionite dissolved in 0.1 M phosphate buffer (pH 8.0) under argon atmosphere. The NMR tube was sealed to avoid reoxidation of the protein. The final protein concentration was 8 mM at pH 5.8.

For the amide proton exchange experiment, 500  $\mu\text{l}$   $\text{D}_2\text{O}$  were added to the protein which had been previously lyophilized in 0.1 M phosphate buffer pH 6.0. 10  $\mu\text{l}$  of 0.5 M sodium dithionite were added and the NMR tube was sealed immediately. The final protein concentration for this experiment was 1 mM at pH 5.9 (direct pH meter reading).

**NMR experiments:** All NMR experiments were carried out on a Bruker AMX-600 spectrometer operating at  $^1\text{H}$  600 MHz, equipped with a standard 5 mm Bruker probe with internal pulsed  $B_0$  field gradient (PFG) coil. Data were processed using either the UXR software of the Bruker spectrometer or Felix software (version 2.1) kindly provided by Biosym Technologies.

For the sequential  $^1\text{H}$  assignment, a set of three 2D correlation experiments were recorded at 27°C: (1) COSY - DQF (double-quantum filtered correlation spectroscopy, Rance et al., 1983), (2) TOCSY (total correlation spectroscopy, Braunschweiler & Ernst, 1983; Davis & Bax, 1985) and (3) NOESY (nuclear Overhauser spectroscopy, Macura et al., 1981). All spectra were recorded in the phase sensitive mode (States-TPPI, Marion et al., 1989). Suppression of the water resonance was achieved by weak pre-irradiation during the relaxation delay. For a spectral width of 15 ppm, 1024 and 800 complex points were recorded in the  $t_2$  and  $t_1$  dimension, respectively. Prior to Fourier transformation the data were multiplied by a skewed sine bell function and zero filled to final matrices of 2048\*2048 real points. In the TOCSY experiment, isotropic mixing was optimized to balance longitudinal and rotating frame cross relaxation ('Clean TOCSY', Griesinger et al., 1988). The second trim pulse was replaced by a z-filter using a pulsed  $B_0$  field gradient to remove the baseline distortion due to residual water (Simorre, personal

communication). The mixing time for both the NOESY and the TOCSY experiment was set to 100ms.

These transfer delays have been optimized previously using semi-selective 2D-experiments with different mixing-times (40-150ms) as described by Brutscher et al., (1993) using 2.2 ms BURP (band selective, uniform response, pure-phase) pulses (Geen & Freeman, 1991). The incremented dimension was reduced to the amide proton resonances (6 ppm) and consequently, experimental time was shorter (8 h instead of 18 h).

For the assignment of the  $^{13}\text{C}$  resonances and the completion of  $^1\text{H}$  side chain assignment a  $^1\text{H}$ -detected heteronuclear single quantum coherence (HSQC) and a relayed HSQC-TOCSY experiment were acquired at 27°C. The sensitivity of these experiments is greatly enhanced by the suppression of  $^{12}\text{C}$ - $^1\text{H}$  coherences using PFG (Medvedeva et al., 1993). The spectral width in the  $^{13}\text{C}$  dimension was 70 ppm for a 200 $\times$ 1024 complex data set (128 scans for HSQC or 160 for HSQC-TOCSY).

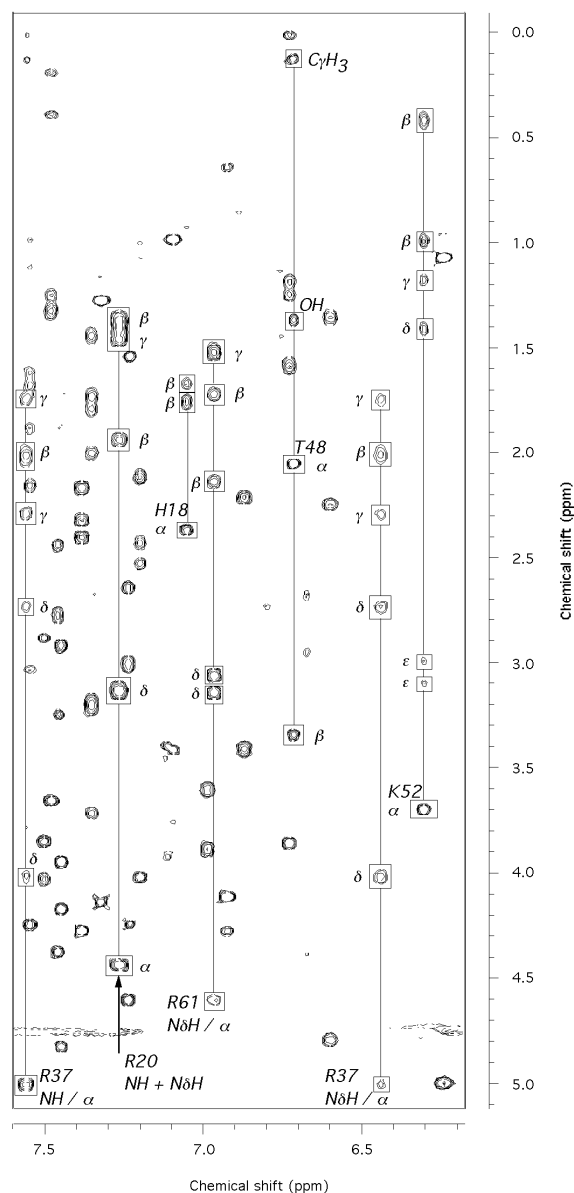
In order to solve various overlap problems, additional  $^1\text{H}$ - $^1\text{H}$  and  $^1\text{H}$ - $^{13}\text{C}$  experiments were recorded at 12°C and 40°C.

The amide proton exchange was measured using semi-selective TOCSY experiments of the fingerprint region (see above). 8 scans were acquired for each of the 160  $t_1$  increments, leading to a spectral resolution of 19 Hz in  $f_1$ . The experimental time was 30 min.

## Results

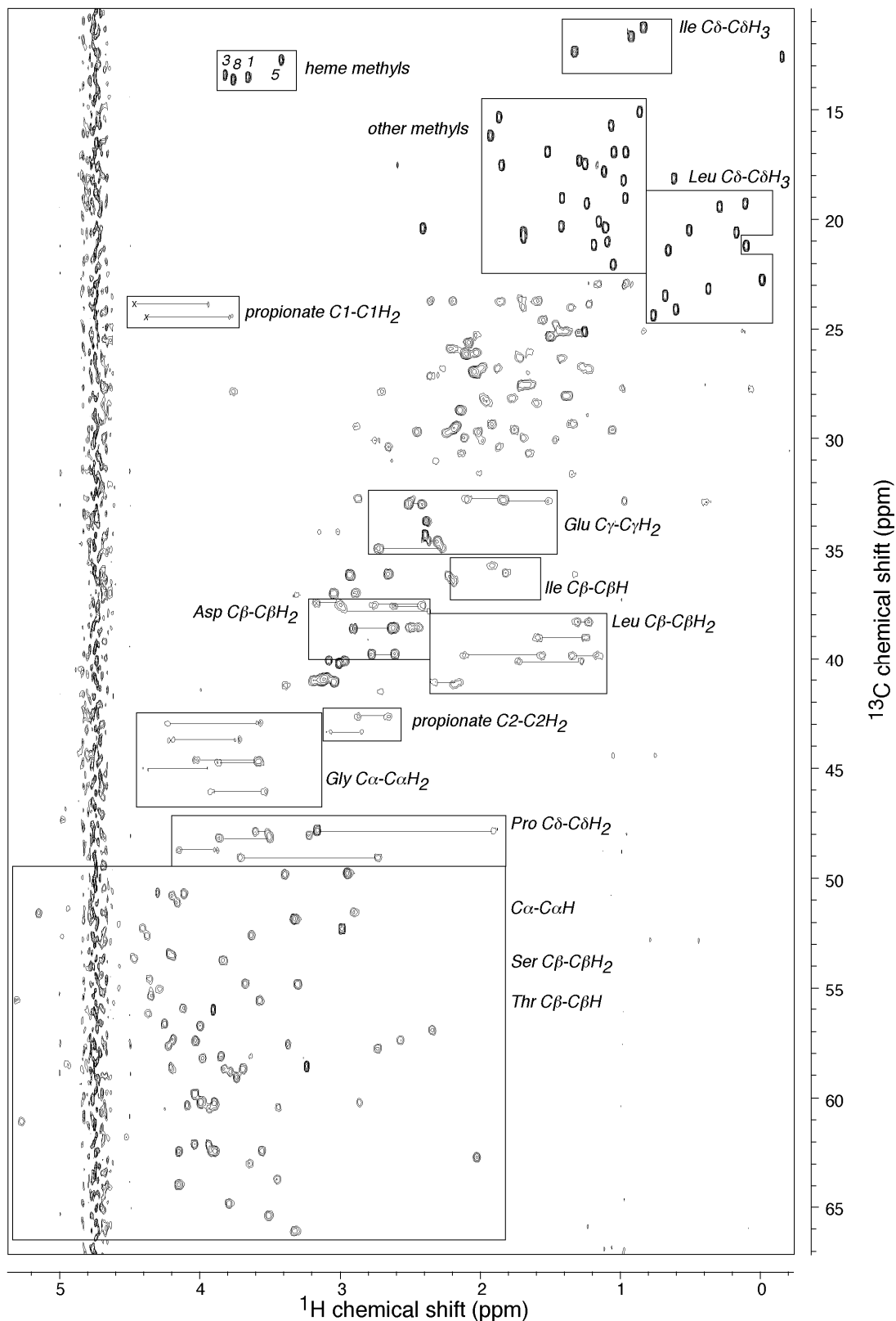
**Sequential assignment** All residues showed very narrow line widths, and their spin systems could be resolved completely from the amide proton frequency in the TOCSY spectra (see Fig. 1). Only one threonine (Thr48) could not be assigned directly due to the unexpected presence of an additional proton, later identified as the OH proton (see below). All four arginine residues gave rise to  $\text{C}\alpha\text{-NH}_2\text{-N}\alpha\text{H}$  correlations and showed twofold repetition of the whole spin-system in the TOCSY spectrum (see Fig. 1). Methionine 55 was recognized by its unusual chemical shifts due to the heme ligation (see Table 1.a). Spin systems were aligned according to the protein sequence using  $\text{NH}_i\text{-NH}_{i+1}$ ,  $\text{C}\alpha\text{-NH}_i\text{-NH}_{i+1}$ ,  $\text{C}\beta\text{-NH}_i\text{-NH}_{i+1}$  and  $\text{C}\gamma\text{-NH}_i\text{-NH}_{i+1}$  NOESY connectivities. In the case of the proline residues which lack the amide proton, the  $\text{C}\alpha\text{-NH}_i\text{-C}\alpha\text{-NH}_{i+1}$  cross-peaks with the preceding residue allowed their unambiguous identification. The unusual spin system of Thr48 was clearly identified by the connectivities with both neighboring amino acids. Proton assignments are given in Table 1.

**Aliphatic side chain assignments** Whereas all protons belonging to a side chain can be identified using TOCSY spectra, one cannot achieve a specific assignment based on relative cross-peak intensities because of multi-spin effects (Cavanagh et al., 1990). Protons connected either by  $^2\text{J}$  or  $^3\text{J}$  couplings can be identified in COSY spectra and a first tentative assignment of the long side chains (Arg, Lys, Glu, Met, Leu) was performed this way. In reality, the unambiguous assignment of long side chains using only the information of the COSY experiment is made difficult by missing peaks due to small J-coupling vs. line-width, non discrimination between  $^2\text{J}$  and  $^3\text{J}$  couplings, and accidental overlaps.



**Fig. 1** NH ( $f_2$ ) - aliphatic ( $f_1$ ) region of the HOHAHA spectrum of *E. halophila* ferrocytochrome-551. The experiment was recorded at 27°C in 90%  $\text{H}_2\text{O}$  / 10%  $\text{D}_2\text{O}$  at pH 5.8 with 100 ms mixing time. Even for long side chains complete spin systems could be detected (Lys52, for example). Spin systems for three of the four arginines are shown aligned from the  $\text{N}\alpha\text{H}$  proton. Note the degeneracy of Arg20-NH and Arg20- $\text{N}\alpha\text{H}$  as well as the unusual spin system in the case of Thr48, containing an OH proton.

These problems can be circumvented by the use of natural abundance  $^{13}\text{C}$ - $^1\text{H}$  correlation spectroscopy as has already been demonstrated for another small, monohemic cytochrome (Medvedeva et al., 1993). Using two experiments (HSQC and HSQC-TOCSY), we were able to specifically assign all the aliphatic side chain protons using two complementary sets of information firstly, methylene protons were correlated pair-wise and secondly, characteristic  $^{13}\text{C}$  chemical shift ranges can be used for  $^{13}\text{C}$  assignment. The principle of this assignment process is illustrated in Fig. 2. The use of the additional HSQC-TOCSY experiment proved to be very useful to complete the assignment, as it greatly helped to identify the spin systems due to the appearance of at least parts of the



**Fig. 2**  $^{13}\text{C}$ - $^1\text{H}$  correlation spectroscopy of ferrocycytochrome c-551.  $^1\text{H}$  resonances appear in  $f_2$ ,  $^{13}\text{C}$  resonances in  $f_1$ . Boxes indicate regions containing one or several classes of  $^{13}\text{C}$ - $^1\text{H}$  pairs. Lines connect methylene protons in  $f_2$ . This spectrum illustrates the advantages of  $^{13}\text{C}$ - $^1\text{H}$  correlated spectroscopy for the side chain assignment (1) carbon atoms can be distinguished by their characteristic chemical shift and (2) methylene protons can be paired as they appear at the same carbon frequency. Peaks which are missing in this plot but which have been detected at a lower contour level are indicated by crosses.

characteristic TOCSY pattern at the corresponding  $^{13}\text{C}$  frequency. For example, we were able to assign all the methyl groups of alanines and valines, which were found in the same spectral region, by the relayed  $^{13}\text{C}$ - $^1\text{H}$  correlations.

(Val) or  $^{13}\text{C}$ - $^1\text{H}$  correlations (Ala). The HSQC-TOCSY experiment also facilitated the completion of the  $^{13}\text{C}$ - $^1\text{H}$  cross-peak assignment which was difficult due to the spectral overlaps. Additional information for the

$^{13}\text{C}$ - $\text{C}\beta\text{H}$  cross-peak assignment was obtained from a second HSQC experiment at 40°C. The  $^{13}\text{C}$  assignments are given in Table 2.

**Aromatic side chain assignment** The two spin systems of the Trp34 side chain were assigned using the COSY connectivities. They were related by the NOE between N1H and C7H. Three tyrosyl residues are present in cytochrome c-551 and the ring resonances were assigned on the basis of the NOE with their respective  $\text{C}\beta\text{H}_2$ . The two tyrosines 58 and 73 gave rise to very broad cross-peaks in the TOCSY spectrum due to an intermediate flipping rate on the NMR timescale (Wagner, 1983). On the other hand, the Tyr6 side chain exhibits four individual resonances, indicating slow ring flipping (in fact the non-symmetry related C3H and C6H overlap at 6.9 ppm). Exchange relayed TOCSY cross-peaks (Feng & Roder, 1988) were observed between the C5H and the C2H proton, which also showed a very weak COSY connectivity. Tyr6-C3H and C5H give sharp resonance lines which are separated by 25 Hz and thus the ring flip rate of Tyr6 is below 55 Hz at 27°C. In the NOESY spectra recorded at 40°C these resonances coalesce which means that the ring flip rate approaches the value of 55 Hz.

#### Observation of non-amide exchangeable protons

As already mentioned above, Thr48 showed a very unusual pattern in the HOHAHA spectrum—four protons were found aligned at the NH frequency. The  $\alpha$ -proton showed COSY cross-peaks to the three other aliphatic resonances; therefore we supposed that the resonance at 1.37 ppm corresponded to the OH-proton. This hypothesis was further confirmed by the absence of a correlation at this  $^1\text{H}$  frequency from the HSQC spectrum as well as by a relatively high chemical shift variation as a function of the temperature.

For two of the three tyrosine side chains (Tyr6 and Tyr58) the OH proton could also be identified. They exhibit no cross-peak in the TOCSY spectrum ( $^4\text{J}$ -coupling) but strong NOE with the H3, H5 protons. Moreover, their interresidue NOE network is similar to that of the other ring protons. The Tyr6-OH disappears at 40°C, probably due to fast solvent exchange at this higher temperature.

In addition, very broad resonances, detected in the downfield region of the NOESY spectrum, were identified by NOE to the  $\text{N}\beta\text{H}$  as the terminal  $\text{NH}_2$  of two arginine residues (Arg37 and Arg64). For Arg37, three resonances could be distinguished, two of them are found at a particularly high resonance frequency and show NOE to both heme propionates. The non-equivalence of these two protons indicates the steric hinderance of this side chain.

In general, the observation of so many exchangeable non-amide protons in a protein is quite rare, even at slightly acidic pH. However, in the  $^1\text{H}$  assignment of horse ferrocycytochrome c, a comparable number of exchangeable protons has been observed (Wand et al., 1989). These protons are probably involved in interresidual hydrogen bonds.

**Heme assignment** The heme protons were assigned according to a procedure originally described by Keller and Wüthrich (Keller & Wüthrich, 1978; see also Detlefsen et al., 1990).  $^{13}\text{C}$  assignments were obtained from the HSQC experiment. Note that the  $^{13}\text{C}$  chemical shift order of the

heme methyls is slightly different from previously reported ones. In our case, methyl-2<sup>1</sup>, -7<sup>1</sup>, -12<sup>1</sup> and -18<sup>1</sup> resonate at 13.5, 13.5, 12.8 and 13.7 ppm respectively, in comparison to 13.9, 13.7, 13.5, and 11.5 (horse ferrocycytochrome c; Gao et al., 1990), 13.3, 12.4, 11.2, and 9.9 ppm (*Rhodobacter capsulatus* ferrocycytochrome c; Caffrey et al., 1994) and to 14.8, 12.8, 12.6 and 12.0 ppm (*Desulfovibrio vulgaris* ferrocycytochrome c-553; Medvedeva et al., 1993). A different chemical shift order of the heme methyl protons may indicate a change in heme ligand geometry, as the chemical shift of these protons is mainly determined by ring current effects in the diamagnetic protein.

Here again, the HSQC experiment provided the identification of the propionate proton pairs at  $^{13}\text{C}$  frequencies of 24.5 and 42.7 for the propionate-13 and 24.0 and 43.4 for the propionate-17. The corresponding protons were identified as two spin systems in the HOHAHA spectrum. Each proton shows two COSY correlations whereas the fourth COSY cross-peak is missing due to a vanishingly small vicinal coupling constant. Position 1 is more affected by the heme ring current shift than position 2, and the influence of the electronegative group COOH acts in the opposite direction, but to a lesser extent: as a result, position 1 is less screened.

**Secondary structure determination** The classical method of secondary structure prediction consists of a qualitative analysis of short- and medium-range interresidual NOE,  $^3\text{J}_{\text{NH-C}\beta\text{H}}$  coupling constants and NH exchange rates (Wüthrich, 1986). These data are summarized in Fig. 3.a. From these results, three regions of  $\alpha$ -helical structure can be identified—residues Gly2-Asn8 (N-terminal helix), Val41-Ala50 (40s helix) and Asp63-Ser76 (C-terminal helix). Two other regions show NOE patterns which indicate  $\alpha$ -helical structure—the heme binding region around the CXXCH-box (residues Thr13-His18) and another short segment from Ala30 to Asp36. The secondary structure prediction was accomplished by the analysis of chemical shift (Pastore & Saudek, 1990; Spera & Bax, 1991; Wishart & Sykes, 1994). Because  $^{13}\text{C}$  chemical shift range is much larger than that of  $^1\text{H}$ , the contribution of the heme ring current shift to  $^{13}\text{C}$  is relatively moderate. Thus, we can use the  $^{13}\text{C}$  chemical shifts to identify secondary structure elements (Wishart & Sykes, 1994). Standard values for  $^{13}\text{C}$  chemical shifts have been taken from Wishart & Sykes (1994) and were corrected by subtracting 2.81 ppm to account for their calibration using sodium 3-trimethylsilyl-propionate instead of  $\text{Me}_4\text{Si}$ .

The secondary chemical shifts are shown in Fig. 3.b. These results confirm the presence of the N-terminal, the 40s and the C-terminal  $\alpha$ -helices, but not that of the two other small fragments. Therefore we can conclude that there are three well defined  $\alpha$ -helices in the ferrocycytochrome c-551, whereas two further regions might correspond to less well defined or distorted helices.

**Peptide bonds preceding proline residues** No *cis* peptide bonds have been reported so far in the cytochrome c family. As expected, peptide bonds preceding Pro26, Pro31, Pro38 and Pro56 were in the *trans* conformation, as could be shown by the strong  $\text{C}\beta\text{H}_i\text{-C}\beta\text{H}_{i+1}$  NOE. In the case of Pro12 these were absent from the NOESY spectrum.





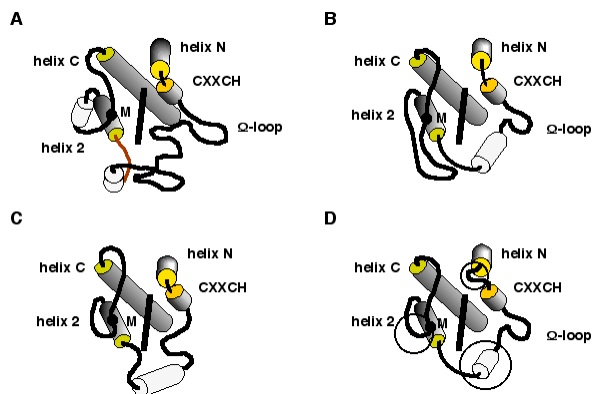
cytochrome $\text{c}_5$ -551 and *R. rubrum* cytochrome $\text{c}_2$  (Chothia & Lesk, 1985). As illustrated in Fig. 4, the common folding pattern includes six segments, these are the N-terminal  $\alpha$ -helix (residues 5 to 11, tuna cytochrome $\text{c}_2$  numbering), the region around the three conserved residues that bind covalently to the heme (CXXCH, residues 14 to 18), a part of the omega loop (residues 24 to 33), a second  $\alpha$ -helix (helix 2, residues 60 to 71), the segment containing the conserved methionine residue that binds to the heme (residues 79 to 81) and the C-terminal  $\alpha$ -helix (residues 87 to 101).

As in all class I cytochrome $\text{c}_2$  structures so far known, the three conserved  $\alpha$ -helices are well defined in the *E. halophila* cytochrome $\text{c}_2$ -551 (compare Figs. 3 and 4). They correspond to the segments Gly2 to Asn8 (N-terminal helix), Val41 to Gly51 (40s helix or helix 2) and Glu65 to Thr77 (C-terminal helix). The N-terminal and the 40s helices very likely pack onto the C-terminal helix as described by Chothia & Lesk (1985). This is demonstrated by several long range NOE between the corresponding helices. In addition, two aromatic residues are generally conserved in the N- and C-terminal helices which are in close contact with each other (Moore & Pettigrew, 1990). In the *E. halophila* cytochrome $\text{c}_2$ -551 they correspond to Tyr6 and Tyr73. NOE between the ring protons of Tyr6 and Tyr73 could be identified at 40°C. Both tyrosines are in slow rotational motion, as has also been observed for the two corresponding aromatic residues in horse heart cytochrome $\text{c}_2$  (Wand et al., 1989) and in *D. vulgaris* Hildenborough cytochrome $\text{c}_2$ -553 (Marion & Guerlesquin, 1992).

In the heme binding segment (Cys14 to His18) the iron coordinating histidine side chain is correctly positioned by a hydrogen bond from the NH1 proton to a carbonyl of a conserved proline residue, as in most class I cytochromes $\text{c}_2$ . This is supported by numerous NOE between the imidazole NH proton and the side chain of Pro26, the slow solvent exchange of the imidazole NH proton and the peculiar chemical shifts of Ala25 and Pro26 (see Table 1.a). The same characteristics have also been observed for the cytochrome $\text{c}_2$  from *Rb. capsulatus* (Gooley et al., 1990; Gooley et al., 1992), the *Ps. aeruginosa* cytochrome $\text{c}_2$ -551 (Chau et al., 1990; Timkovich et al., 1992) and for horse heart cytochrome $\text{c}_2$  (Wand et al., 1989).

There are several indications for the presence of a surface loop extending between the residues His18 and Leu28 in the *E. halophila* cytochrome $\text{c}_2$ -551, which would be analogous to the omega loop. The observed NOE in the region between Arg20 and Asn29 clearly demonstrate the absence of regular secondary structure (see Fig. 3.a), which is characteristic for this structural element (Leszczynski & Rose, 1986). Note also that no slowly exchanging amide protons occur in this region. The absence of NOE between the segment Asp19 to Gly24 with any heme proton agree with its localization on the surface. In addition, several NOE between Gly21 and Glu27 also support the typical folding of the polypeptide chain in an omega loop.

Therefore, without resorting to molecular modelling techniques and using only our NMR data, we can already confirm the conservation of the common cytochrome $\text{c}_2$  fold



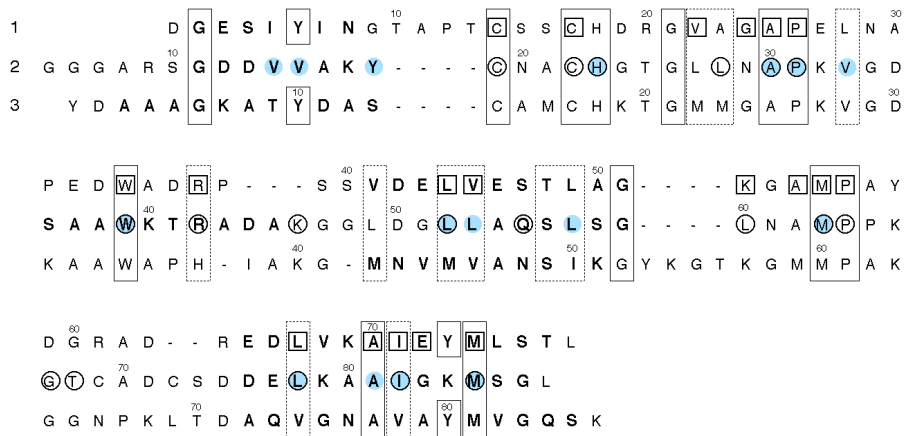
**Fig. 4** Location of conserved and variable segments in different class I cytochromes $\text{c}_2$ . **A.** yeast cytochrome $\text{c}_2$ , **B.** *Ps. aeruginosa* cytochrome $\text{c}_2$ -551, **C.** *A. vinelandii* cytochrome  $\text{c}_5$ , **D.** *E. halophila* cytochrome $\text{c}_2$ -551. The conserved helices are in dark, non-conserved helices in light grey. Heme ligands are also indicated, the CXXCH box is represented as a helix. For simplicity, C- and N-terminal extensions of individual cytochromes are omitted. These schematic drawings illustrate fundamental differences in cytochrome $\text{c}_2$  folding but do not claim to be realistic representations of the three dimensional structures. Molecules **a-c** have been drawn from the corresponding structures deposited in the Brookhaven Protein Data Bank whereas molecule **d** has been derived from protein sequence alignments in combination with our NMR data (see text). Circles mark variable regions which are discussed in more detail.

in the *E. halophila* cytochrome $\text{c}_2$ -551, as is illustrated in Fig. 4.d. techniques and using only our NMR data, we can already confirm the conservation of the common cytochrome $\text{c}_2$  fold in the *E. halophila* cytochrome $\text{c}_2$ -551, as is illustrated in Fig. 4.d.

**Variable segments in the *E. halophila* cytochrome $\text{c}_2$ -551** We will discuss in some detail three variable regions in the *E. halophila* cytochrome $\text{c}_2$ -551 (see Fig. 4.d).

The first segment is the four residue insertion which is found in the *Ectothiorhodospiraceae* between the N-terminal helix and the heme binding site (compare Figs. 4 and 5). We have shown that in the *E. halophila* cytochrome $\text{c}_2$ -551 this segment contains a *cis* peptide bond *E. halophila* and *E. halochloris* (Ambler et al., 1993). An interesting replacement of Gly11 in cytochrome $\text{c}_2$ -551 from *Ps. aeruginosa* to Pro11 in *Ps. stutzeri* has been reported, which induces the formation of a distorted type III turn immediately following the N-terminal helix (Cai et al., 1992). But apparently there is no evidence for a *cis* peptide bond in this protein. In the majority of the class I cytochromes $\text{c}_2$ , the N-terminal helix is directly adjacent to the heme attachment site (Moore & Pettigrew, 1990).

The second variable segment is situated between the omega loop and the second conserved helix (helix 2) and closes the heme binding site at the bottom of the protein. The difference in size between the larger cytochromes $\text{c}_2$  (mitochondrial cytochromes $\text{c}_2$  and bacterial cytochromes $\text{c}_2$ ) and the smaller bacterial cytochromes $\text{c}_2$  is mainly due to the deletion of a large surface loop from this region (residues 41 to 55 in mitochondrial cytochrome $\text{c}_2$ , Moore & Pettigrew, 1990). But even among the small bacterial cytochromes  $\text{c}_2$ ,



**Fig. 5** Structural sequence alignment of three class I bacterial cytochromes  $c_5$ . (1) *E. halophila* cytochrome  $c_5$ -551, (2) *A. vinelandii* cytochrome  $c_5$ , and (3) *Chl. limicola* cytochrome  $c_5$ -555. This alignment is based on X-ray data (Carter et al., 1985; Korszun & Salemme, 1977) and on the NMR data presented in this work. Conserved residues are situated in boxes (dashed lines indicate conservative replacements). Residues given in large bold font are involved in helical structures. The helix involving Met43-Lys51 in *Chl. limicola* cytochrome  $c_5$ -555 has been given according to Moore & Pettigrew, 1990. In the cytochrome  $c_5$  sequence, filled circles indicate residues which have been proposed to form the hydrophobic interior of the protein whereas residues found in van der Waals interaction with the heme are marked by open circles. Open squares in the cytochrome  $c_5$ -551 sequence indicate residues for which NOE were observed between backbone protons and the heme. This sequence alignment has been adapted from that of Ambler et al., 1993, with slight modifications (see text). The sequence of *Chl. limicola* cytochrome  $c_5$ -555 differs from that of Ambler et al. by three residues, Gly58, Met59 and Val82 (compare Fig. 2 in Ambler et al., 1993). The sequence given here is identical to the one deposited in the SwissProt protein data base and in the original sequence paper (Van Beeumen et al., 1976).

the remaining region exhibits substantial structural variability. For example,  $\alpha$ -helices have been determined in the corresponding segments of *Ps. aeruginosa* and *Ps. stutzeri* cytochromes  $c_5$ -551 (residues 27 to 33, Matsuura et al., 1982; Cai et al., 1992) as well as in the *A. vinelandii* cytochrome  $c_5$  (Carter et al., 1985). On the other hand, this helix seems to be absent in *Chl. limicola* cytochrome  $c_5$ -555 (Korszun & Salemme, 1977) and is absent in *D. vulgaris* cytochrome  $c_5$ -553 (Marion & Guerlesquin, 1992). Our NMR data (short range NOE,  $^{13}\text{C}$  chemical shift and amide proton exchange, see Figs. 3.a and 3.b) do not clearly indicate a helical structure in this region, but this might be due to the fact that this segment is too short to show the typical helix characteristics.

The last variable segment we will discuss corresponds to the segment between helix 2 and the methionine ligand (residues 72 to 78 in mitochondrial cytochrome  $c_5$ ). Interestingly, in the *E. halophila* cytochrome  $c_5$ -551 as well as in *A. vinelandii* cytochrome  $c_5$  the conserved  $\alpha$ -helix is adjacent to the sequence around the sixth heme ligand, Met55 or Met63, respectively. In the other small bacterial cytochromes  $c_5$ , such as the *Ps. aeruginosa* cytochrome  $c_5$ -551, *D. vulgaris* cytochrome  $c_5$ -553 and possibly in *Chl. limicola* cytochrome  $c_5$ -555 the protein chain descends after the conserved  $\alpha$ -helix (helix 2 in Fig. 4), closes over the heme at the level of heme propionate-13 and rises again to correctly position the methionine ligand (in *D. vulgaris* cytochrome  $c_5$ -553 this segment even includes an  $\alpha$ -helix, Marion & Guerlesquin, 1992). Roughly, the  $\alpha$ -carbon of the methionine ligand is on a level with the methyl-12<sup>1</sup> (Senn & Wüthrich, 1985). Our NOE data agree with the conserved relative orientation of the methionine ligand, helix 2 and the heme, as given by Chothia and Lesk (1985). Therefore, the deletion of several residues between helix 2 and methionine 55 most likely results in shortening

of the loop. A consequence of the loop deletion observed in *E. halophila* cytochrome  $c_5$ -551 and in *A. vinelandii* cytochrome  $c_5$  might be an increased solvent exposure of heme propionate-13, as has been reported for the crystal structure of *A. vinelandii* cytochrome  $c_5$  (Carter et al., 1985).

**Comparison of *E. halophila* cytochrome  $c_5$ -551, *A. vinelandii* cytochrome  $c_5$  and *Chl. limicola* cytochrome  $c_5$ -555** As already mentioned, the *E. halophila* cytochrome  $c_5$ -551 seems to bridge the green bacterial and cyanobacterial cytochromes  $c_5$  which are all related to the *Pseudomonas* cytochromes  $c_5$ . It has been proposed that the cytochrome  $c_5$  homologs might serve as a molecular marker to relate the phototrophs, which certainly share a common origin (Ambler et al., 1993). Therefore, we will compare the *E. halophila* cytochrome  $c_5$ -551 to two other cytochrome  $c_5$  homologs for which three dimensional structures have already been determined.

Comparing the three cytochrome  $c_5$  sequences, we can identify several similarities. A convincing sequence alignment can be achieved with three major deletion and insertion events, which are all located in variable regions of the cytochrome  $c_5$  sequence. The sequence alignment (see Fig. 5) has been taken from Ambler et al., (1993) but was modified to account for the alignment of structural elements in the region preceding the methionine ligand. Besides the alignment of  $\alpha$ -helical structure, it has been accomplished by the characterization of single residues in contact with the heme or part of the hydrophobic interior of the protein (see Figure caption). This information was taken from the previously published crystal structures of *Chl. limicola* cytochrome  $c_5$ -555 (Korszun & Salemme, 1977) and *A. vinelandii* cytochrome  $c_5$  (Carter et al., 1985). However, these structures have to be regarded with caution. The preliminary low resolution crystal structure of cytochrome  $c_5$ -555 has not been refined and it is not

available from the Brookhaven Protein Data Bank. The cytochrome  $c_5$  structure has been the subject of a critical discussion as the protein used for the structure determination behaves as a monomer in solution but appears to form a dimer in the crystal (Moore & Pettigrew, 1990, p. 181f.). So it is not clear whether the crystal structure really corresponds to the biochemically relevant form. But even with these limitations in mind it becomes obvious that *E. halophila* cytochrome $\alpha$ -551 shares several similarities with these two proteins.

From Fig. 5 it can be seen that nearly all the residues which have been reported to form a hydrophobic pocket in the interior of the cytochrome  $c_5$  (Carter et al., 1985) are conserved or have been subjected to conservative replacements. In addition, comparing *E. halophila* cytochrome $\alpha$ -551 and *A. vinelandii* cytochrome  $c_5$  in a global manner, it can be noted that they seem to possess a nearly identical folding of the protein chain with regard to the heme. Residues which have been found in van der Waals contact with the heme in the crystal structure (Carter et al., 1985) have their equivalents in the *E. halophila* cytochrome $\alpha$ -551, as demonstrated by the NOE between the corresponding residues and the heme. In this context it is interesting to note that two residues for which slowly exchanging OH protons have been observed (Thr48 and Tyr58) have been replaced by residues which are also possible hydrogen bond donors in the cytochromes $\alpha_5$  and  $c$ -555 (Thr48 $\rightarrow$ Ser; Tyr58 $\rightarrow$ Lys; see Fig. 5).

The three deletion and insertion events indicate significant differences in local structure between these three proteins. The first one, the four residue insertion characteristic of *Ectothiorhodospiraceae* cytochromes $\alpha$ -551 has been discussed in detail above. Another deletion in the cytochrome $\alpha$ -551 sequence occurs in a region which is aligned with the end of the non-conserved  $\alpha$ -helix in the cytochrome  $c_5$  sequence. This three residue deletion and the presence of two non-conserved proline residues (Pro31 and Pro38) in the cytochrome $\alpha$ -551 might explain the absence of a clearly defined helical structure compared to cytochrome  $c_5$ . Another significant difference between the three proteins is situated in the region including the surface loop preceding the methionine ligand. The *A. vinelandii* cytochrome  $c_5$  as well as the *E. halophila* cytochrome $\alpha$ -551 show a four residue deletion compared to the *Chl. limicola* cytochrome $\alpha$ -555. In this context it seems interesting to note, that in *Prosthecochloris aestuarii* cytochrome $\alpha$ -555, which shows 53% sequence identity with *Chl. limicola* cytochrome $\alpha$ -555, this loop is further extended by a four residue insertion in the center of the presumed *Chl. limicola* loop (Moore & Pettigrew, 1990).

In most cytochromes $\alpha$ , a tryptophan and a residue which can carry a positive charge (generally Arg or His) are found close to the heme propionates. These residues certainly have analogous functions (hydrophobic interactions, hydrogen-bonding) but originate from different regions of the polypeptide chain (Moore & Pettigrew, 1990). In our case, there is strong evidence that Trp34 and Arg37 correspond to those functionally conserved residues as we can detect NOE between their side chains and heme protons (Arg37 to heme propionates-13 and -17, Trp34 to heme

propionate-17 and methyl-18<sup>1</sup>). Trp34 and Arg37 are homologous to Trp39 and Arg42 in *A. vinelandii* cytochrome  $c_5$  and to Trp34 and His39 in *Chl. limicola* cytochrome $\alpha$ -555. It seems that the arrangement of these two functionally conserved residues in a Trp<sub>1</sub>-Arg(His)<sub>1+3</sub> motif is unique to the *Pseudomonas* cytochrome  $c_5$  homologs. In addition, the orientation of the tryptophan side chain of the cytochrome  $c_5$  analogs seems to differ from that of the large mitochondrial cytochromes $\alpha$  or bacterial cytochromes $\alpha_2$ . Hydrogen bonding between the indole NH proton and heme propionate-17, as shown in the crystal structures of the large class I cytochromes $\alpha$  (Takano & Dickerson 1981a, b; Bathia et al., 1984; Benning et al., 1991, 1994) seems to be excluded in the cytochrome  $c_5$  analogs. This can be seen from the crystal structure of *A. vinelandii* cytochrome  $c_5$  (Carter et al., 1985) and is confirmed by a pattern of NOE observed for the *E. halophila* cytochrome $\alpha$ -551. In these two proteins, the indole NH proton is close to the protein backbone whereas the other side of the tryptophan side chain is oriented towards the heme.

It is interesting to note that, in spite of the fact that these three bacterial cytochromes $\alpha$  belong to three different subgroups of the cytochrome $\alpha$  family (Moore & Pettigrew, 1990; Ambler et al., 1993) they share several common features. For a better characterization of the structural analogies between these three *Pseudomonas* cytochrome  $c_5$  homologs it would be worthwhile to obtain high resolution structures of the corresponding proteins. Our NMR investigation of *E. halophila* cytochrome $\alpha$ -551, which will be completed by a structure calculation using restrained molecular dynamics, is a first step towards the determination of such a three dimensional structure.

### Acknowledgements

This work has been supported in part by the Centre National de la Recherche Scientifique (Interdisciplinary Program for Protein Engineering IMABIO), the Commissariat à l'Energie Atomique (CEA - Department of Life Sciences DSV) and US NIH grant GM 21277. B. Brutscher is the recipient of a Bruker-CEA Ph.D. fellowship.

This is publication n° 251 of the Institut de Biologie Structurale - Jean-Pierre Ebel.

### References

- Ambler, R.P., Meyer, T.E. & Kamen, M.D. (1993) Amino acid sequences of cytochrome $\alpha$ -551 from the halophilic purple phototrophic bacteria *Ectothiorhodospira halophila* and *E. halochloris*, *Arch. Biochem. Biophys.* 306, 83-93.
- Bartsch, R.G. (1991) The distribution of soluble metallo-redox proteins in purple phototrophic bacteria, *Biochim. Biophys. Acta* 1058, 28-30.
- Benning, M., Wesenberg, G., Caffrey, M., Bartsch, R., Meyer, T., Cusanovich, M., Rayment, I. & Holden, H. (1991) Molecular structure of cytochrome $\alpha_2$  isolated from *Rhodobacter capsulatus* determined at 2.5 Å resolution, *J. Mol. Biol.* 220, 673-685.

- Benning, M.M., Meyer, T.E. & Holden, H.M. (1994) X-ray structure of the cytochrome<sub>2</sub> isolated from *Paracoccus denitrificans* refined to 1.7 Å resolution, *Arch. Biochem. Biophys.* 310, 460-466.
- Bhatia, G.E., Finzel, B.C. & Kraut, J. (1984) 2Å resolution X-ray coordinates of *R. rubrum* cytochrome<sub>2</sub>. Deposited with the Brookhaven Protein Data Bank,
- Blackledge, M., Medvedeva, S., Poncin, M., Guerlesquin, F., Bruschi, M. & Marion, D (1994) Structure and dynamics of cytochrome c553 from *Desulfovibrio vulgaris* studied by NMR spectroscopy and restrained molecular dynamics, *J. Mol. Biol.* (in press)
- Braunschweiler, L. & Ernst, R.R. (1983) Coherence transfer by isotropic mixing application to proton correlation spectroscopy, *J. Magn. Res.* 53, 521-528.
- Brune, D.C. (1989) Sulfur oxidation by phototrophic bacteria, *Biochim. Biophys. Acta* 975, 189-221.
- Brutscher, B., Simorre, J.-P. & Marion, D. (1992) Optimization of multidimensional NMR using band-selective pulses, *J. Magn. Res.* 100, 416-424.
- Bushnell, G.W., Louie, G.V. & Brayer, G.D. (1990) High resolution three-dimensional structure of horse heart cytochrome<sub>2</sub>, *J. Mol. Biol.* 214, 585-595.
- Caffrey, M., Brutscher, B., Simorre, J.-P., Fitch, J., Cusanovich, M. & Marion, D. (1994) Assignment of the <sup>13</sup>C and <sup>13</sup>CO resonances for *Rhodobacter capsulatus* ferrocyanochrome<sub>2</sub> using double-resonance and triple-resonance NMR spectroscopy, *Eur. J. Biochem.* 221, 63-75.
- Cai, M., Bradford, E.G. & Timkovich, R. (1992) Investigation of the solution conformation of cytochrome<sub>2</sub>-551 from *Pseudomonas stutzeri*, *Biochemistry* 31, 8603-8612.
- Carter, D.C., Meli, K.A., O'Donnell, S.E., Burgess, B.K., Furey Jr., W.F., Wang, B.-C. & Stout, C.D. (1985) Crystal structure of *Azotobacter* cytochrome c<sub>5</sub> at 2.5 Å resolution, *J. Mol. Biol.* 184, 279-295.
- Cavanagh, J., Chazin, W.J. & Rance, M. (1990) The time dependence of coherence transfer in homonuclear isotropic mixing experiments, *J. Magn. Res.* 87, 110-131.
- Chau, M.-H., Cai, M. & Timkovich, R. (1990) *Biochemistry* 29, 5076-5087. NMR comparison of prokaryotic and eukaryotic cytochromes<sub>2</sub>.
- Chen, L., Durley, R.C.E., Mathews, F.S., Davidson, V.L. (1994) Structure of an electron transfer complex: methylamine dehydrogenase, amicyanin, and cytochrome<sub>2</sub>-551, *Science* 264, 86-90.
- Chothia, C. & Lesk, A.M. (1985) Helix movements and the reconstruction of the haem pocket during the evolution of the cytochrome<sub>2</sub> family, *J. Mol. Biol.* 182, 151-158.
- Davis, D.G. & Bax, A. (1985) Assignment of complex <sup>1</sup>H NMR spectra via two-dimensional homonuclear Hartmann-Hahn spectroscopy, *J. Am. Chem. Soc.* 107, 2820-2821.
- Detlefsen, D.J., Thanabal, V., Percoraro, V.L. & Wagner, G. (1990) Sequential <sup>1</sup>H NMR assignment of Iron(II) cytochrome<sub>2</sub>-551 from *Pseudomonas aeruginosa*, *Biochemistry* 29, 9377-9386.
- Detlefsen, D.J., Thanabal, V., Percoraro, V.L. & Wagner, G. (1991) Solution structure of Fe(II) cytochrome<sub>2</sub>-551 from *Pseudomonas aeruginosa* as determined by two-dimensional <sup>1</sup>H NMR, *Biochemistry* 30, 9040-9046.
- Feng, Y. & Roder, H. (1988) Relayed magnetization transfer by isotropic mixing in exchanging systems, *J. Magn. Res.* 78, 597-602.
- Gao, Y., Boyd, J. & Williams, R. (1990) A systematic approach towards the complete assignment of <sup>13</sup>C resonances for horse ferrocyanochrome<sub>2</sub>, *Eur. J. Biochem.* 194, 355-365.
- Geen, H. & Freeman, R. (1991) Band-selective radiofrequency pulses, *J. Magn. Res.* 93, 93-141.
- Gooley, P.R., Caffrey, M.S., Cusanovich, M.A. & MacKenzie, N.E. (1990) Assignment of the <sup>1</sup>H and <sup>15</sup>N NMR spectra of *Rhodobacter capsulatus* ferrocyanochrome<sub>2</sub>, *Biochemistry* 29, 2278-2290.
- Gooley, P.R., Caffrey, M.S., Cusanovich, M.A. & MacKenzie, N.E. (1992) Mutations Pro→Ala-35 and TYR→Phe-75 of *Rhodobacter capsulatus* ferrocyanochrome<sub>2</sub> affect protein backbone dynamics: measurements of individual amide proton exchange rate constants by <sup>1</sup>H-<sup>15</sup>N HMQC spectroscopy, *Biochemistry* 31, 443-450.
- Griesinger, C., Otting, G., Wüthrich, K. & Ernst, R.R. (1988) Clean TOCSY for <sup>1</sup>H spin system identification in macromolecules, *J. Am. Chem. Soc.* 110, 7870-7872.
- Imhoff, J.F. (1989) In J.T. Staley, M.P. Bryant, N. Pfennig, J.G. Holt (ed.) *Bergey's manual of systematic bacteriology*, Vol. 3, Williams & Wilkins, Baltimore, pp 1654-1658.
- Keller, R.M. & Wüthrich, K. (1978) Evolutionary change of the heme c electronic structure ferricytochrome<sub>2</sub>-551 from *Pseudomonas aeruginosa* and horse heart ferricytochrome<sub>2</sub>, *Biochem. Biophys. Res. Commun.* 83, 1132-1139.
- Korszun, Z.R. & Salemme, F.R. (1977) Structure of cytochrome<sub>2</sub>-555 of *Clorobium thiosulfatophilum* Primitive low potential cytochrome<sub>2</sub>, *Proc. Natl. Acad. Sci. USA* 74, 5244-5247.
- Leguijt, T. & Hellingwerf, K.J. (1991) Characterization of reaction center / antenna complexes from bacteriochlorophyll *a* containing *Ectothiorhodospiraceae*, *Biochim. Biophys. Acta* 1057, 353-360.
- Leguijt, T., Engels, P.W., Crielaard, W., Albracht, S.P.J. & Hellingwerf, K.J. (1993) Abundance, subunit composition, redox properties and catalytic activity of the cytochrome bc<sub>1</sub> complex from alkaliphilic and halophilic, photosynthetic members of the family *Ectothiorhodospiraceae*, *J. Bacteriol.* 1993, 1629-1636.
- Leszczynski, J.F. & Rose, G.D. (1986)  $\beta$  loops in globular proteins: a novel category of secondary structure, *Science* 234, 849-855.
- Louie, G.V. & Brayer, G.D. (1990) High-resolution refinement of yeast iso-1-cytochrome<sub>2</sub> and comparisons with other eukaryotic cytochromes<sub>2</sub>, *J. Mol. Biol.* 214, 527-555.
- Macura, S., Huang, Y., Suter, D. & Ernst, R.R. (1981) Two-dimensional chemical exchange and cross-relaxation

- spectroscopy of coupled nuclear spins, *J. Magn. Res.* 43, 259-281.
- Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989) Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins, *J. Magn. Res.* 85, 393-399.
- Marion, D. & Guerlesquin, F. (1992) Sequential NMR resonance assignment and secondary structure of ferrocyclochrome  $c_{553}$  from *Desulfovibrio vulgaris* Hildenborough, *Biochemistry* 31, 8171-8179.
- Marion, D. (1994) Nuclear magnetic resonance studies of cytochromes in solution, *Biochimie*. 76 (in press)
- Mathews, F.S., Chen, Z., Bellamy, H.D. & McIntire, W.S. (1991) Three-dimensional structure of *p*-cresol methylhydroxylase (flavocytochrome) from *Pseudomonas putida* at 3.0-Å resolution, *Biochemistry* 30, 238-247.
- Matsuura, Y., Takano, T. & Dickerson, R.E. (1982) Structure of cytochrome-551 from *Ps. aeruginosa* refined at 1.6 Å resolution and comparison of the two redox forms, *J. Mol. Biol.* 156, 389-409.
- Medvedeva, S., Simorre, J.-P., Brutscher, B., Guerlesquin, F. & Marion, D. (1993) Extensive  $^1\text{H}$  NMR resonance assignment of proteins using natural abundance gradient-enhanced  $^{13}\text{C}$ - $^1\text{H}$  correlation spectroscopy, *FEBS Lett.* 333, 251-256.
- Meyer, T.E. (1985) Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic photrophic bacterium *Ectothiorhodospira halophila*, *Biochim. Biophys. Acta* 806, 175-183.
- Meyer, T.E. (1995) Evolution and classification of c-type cytochromes, in *Cytochromes. A multidisciplinary approach*, R.A. Scott, A.G. Mauk, eds. University Science, Mill Valley, CA (in press).
- Meyer, T.E. & Kamen, M.D. (1982) New perspectives on c-type cytochromes, *Adv. Protein Chem.* 35, 105-212.
- Moore, G.R. & Pettigrew, G.W. (1990) *Cytochromes - evolutionary, structural and physicochemical aspects*, Springer-Verlag, Berlin Heidelberg.
- Murphy, M.E.P., Nall, B.T. & Brayer, G.D. (1992) Structure determination and analysis of yeast iso-2-cytochrome and a composite mutant protein, *J. Mol. Biol.* 227, 160-176.
- Nakagawa, A., Higuchi, Y., Yasuoka, K., Katsube, Y. & Yagi, T. (1990) S-class cytochromes have a variety of folding patterns. Structure of cytochrome  $c_{553}$  from *Desulfovibrio vulgaris* determined by the multi-wavelength anomalous dispersion method, *J. Biochem.* 108, 701-703.
- Pastore, A. & Saudek, V. (1990) The relationship between chemical shifts and secondary structure in proteins, *J. Magn. Res.* 90, 165-176.
- Pettigrew, G.W. & Moore, G.R. (1987) *Cytochromes - biological aspects*, Springer-Verlag, Berlin Heidelberg.
- Qi, P.X., Di Stefano, D. & Wand A.J. (1994) Solution structure of horse heart ferrocyclochrome determined by high resolution NMR and restrained simulating annealing, *Biochemistry* 33, 6408-6417.
- Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. & Wüthrich, K. (1983) Improved spectral resolution in COSY  $^1\text{H}$  NMR spectra of proteins via double quantum filtering, *Biochem. Biophys. Res. Commun.* 117, 479-485.
- Senn, H. & Wüthrich, K. (1985) Amino acid sequence, haem-iron co-ordination geometry and functional properties of mitochondrial and bacterial c-type cytochromes, *Q. Rev. Biophys.* 18, 111-134.
- Spera, S. & Bax, A. (1991) Empirical correlation between protein backbone conformation and  $\text{C}\alpha$  and  $\text{C}\beta$   $^{13}\text{C}$  nuclear magnetic resonance chemical shifts, *J. Am. Chem. Soc.* 113, 5490-5492.
- Takano, T. & Dickerson, R.E. (1981a) Conformation change of cytochrome ferrocyclochrome refined at 1.5 Å resolution, *J. Mol. Biol.* 153, 79-94.
- Takano, T. & Dickerson, R.E. (1981b) Conformation change of cytochrome ferricytochrome refinement at 1.8Å and comparison with the ferrocyclochrome structure, *J. Mol. Biol.* 153, 95-115.
- Then, J. & Trüper, H.G. (1983) Sulfide oxidation in *Ectothiorhodospira abdelmalekii*. Evidence for the catalytic role of cytochrome-551, *Arch. Microbiol.* 135, 254-258.
- Timkovich, R., Walker II, L.A. & Mengli, C. (1992) Hydrogen exchange in *Pseudomonas* cytochrome-551, *Biochim. Biophys. Acta* 1121, 8-15.
- Van Beeumen, J., Ambler, R.P., Meyer, T.E., Kamen, M.D., Olsen, J.M. & Shaw, E.K. (1976) The amino acid sequences of the cytochromes c-555 from two green sulfur bacteria of the genus *Chlorobium*, *Biochem. J.* 189, 757-774.
- Wagner, G. (1983) Characterization of the distribution of internal motions in the basic pancreatic trypsin inhibitor using a large number of internal NMR probes, *Q. Rev. Biophys.* 16, 1-57.
- Wand, A.J., Di Stefano, D.L., Feng, Y., Roder, H. & Englander, S.W. (1989) Proton resonance assignments of horse ferrocyclochrome, *Biochemistry* 28, 186-194.
- Wishart, D.S. & Sykes, B.D. (1994) The  $^{13}\text{C}$  chemical-shift index. A simple method for the identification of protein secondary structure using  $^{13}\text{C}$  chemical-shift data, *J. Biomol. NMR* 4, 171-180.
- Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, John Wiley, New York.
- Wüthrich, K., Billeter, M. & Braun, W. (1984) Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances, *J. Mol. Biol.* 180, 715-740.

**Two tables, showing the  $^1\text{H}$  and  $^{13}\text{C}$  chemical shift assignments are omitted in the context of this document.**

## *Ectothiorhodospira halophila* ferrocytochrome<sub>551</sub> : solution structure and comparison with bacterial cytochromes<sub>c</sub>

Beate Bersch,\* Martin J. Blackledge, Terry E. Meyer,<sup>\$</sup> and Dominique Marion

Institut de Biologie Structurale - Jean-Pierre Ebel, CEA-CNRS, 38027 Grenoble Cedex 1, France.

<sup>\$</sup> Department of Biochemistry, University of Arizona, Tucson 85721, USA.

\* corresponding author

The solution structure of the *Ectothiorhodospira halophila* ferrocytochrome<sub>551</sub> has been determined. This molecule belongs to a separate class of small bacterial cytochromes *c* for which no 3D structure has been reported so far. It is characterized by a very low redox potential (58 mV) and is isolated from the periplasm of halophilic purple phototrophic bacteria. For the 78 residue protein, 1445 NOE derived distance constraints were used in a combined simulated annealing / restrained molecular dynamics calculation. The final ensemble of 37 structures presents a backbone r.m.s.d. of less than 0.5 Å compared to the mean structure. The physical viability of these structures was investigated by subjecting eight of them to a constraint free molecular dynamics simulation. No systematic conformational change was observed and the average backbone r.m.s.d. compared to the initial structures was less than 1.5 Å. The structure of the *E. halophila* cytochrome<sub>551</sub> shows a striking resemblance with *Azotobacter vinelandii* cytochrome<sub>5</sub>. Significant differences in backbone conformations occur in three small regions which are implicated in solvent protection of the heme propionates and thiomethyl-8<sup>1</sup>. Comparison with *Pseudomonas aeruginosa* cytochrome<sub>551</sub> reveals that only the common cytochrome<sub>c</sub> core, i.e. 3 helices, is conserved. The folding of the protein chain around the heme propionates is very different and results in more efficient solvent protection in *P. aeruginosa*. The electrostatic surface of *E. halophila* cytochrome<sub>551</sub> was found to be significantly different from mitochondrial cytochromes<sub>c</sub> and bacterial cytochromes<sub>c2</sub> but similar to that of *P. aeruginosa* cytochrome<sub>551</sub>.

**Key words :** electron transfer protein, protein structure, NMR, purple bacteria, electrostatic surface.

Class I cytochromes<sub>c</sub> are small, soluble redox proteins implicated in electron transport with nearly ubiquitous occurrence among prokaryotes and eukaryotes. Consequently, they show a great variability in redox potential as well as sequential and structural divergence due to adaptation to the different metabolisms in which they function. These proteins are characterized by covalent attachment of the heme to two cysteines, situated near the N-terminus, and histidine and methionine ligation of the iron. Most often, the class I cytochromes<sub>c</sub> are subdivided into two major categories, the larger eukaryotic cytochromes<sub>c</sub> and bacterial cytochromes<sub>c2</sub>, and the smaller bacterial and algal cytochromes<sub>c</sub>. The size difference is mainly due to the presence or absence of a large loop at the bottom of the heme, i.e. near the propionates (Moore & Pettigrew, 1990). It has been proposed that such a major change could have occurred only once during evolution and so gave rise to two different lines of descent (Pettigrew & Moore, 1987). Therefore, this classification is not only based on phenomenological criteria but seems to reflect a

phylogenetic relationship between mitochondrial cytochromes<sub>c</sub> and bacterial cytochromes<sub>c2</sub>, as shown by Meyer *et al.*, (1986). Additional structural and sequential homologies have been recognized between representatives of these two groups (Chothia & Lesk, 1985; Sogabe & Miki, 1995). The larger class I cytochromes<sub>c</sub> have been well characterized with high resolution X-ray structures for cytochrome<sub>c</sub> from tuna (Takano & Dickerson, 1981b; Takano & Dickerson, 1981a), rice (Ochi *et al.*, 1983), yeast iso-1 and iso-2 (Louie & Brayer, 1990; Murphy *et al.*, 1992) and horse (Bushnell *et al.*, 1990) and cytochromes<sub>c2</sub> from *Rhodospirillum rubrum* (Salemme *et al.*, 1973a; Salemme *et al.*, 1973b; Bhatia, 1981), *Rhodobacter capsulatus* (Benning *et al.*, 1991), *Paracoccus denitrificans* (Benning *et al.*, 1994), *Rhodobacter sphaeroides* (Axelrod *et al.*, 1994), and *Rhodopseudomonas viridis* (Sogabe & Miki, 1995). In addition, the solution structure of the ferrocytochrome<sub>c</sub> from horse and from the cyanide adduct of triply mutated yeast iso-1 cytochrome<sub>c</sub> have been determined by NMR (Qi *et al.*, 1994; Banci *et al.*, 1995). The smaller class I

cytochromes  $c$  show much lower sequence homology reflecting the high diversity of different bacterial and algal metabolisms. There is less structural information available for this category. High resolution structures have been obtained for the cytochromes  $c_6$  of the green algae *Chlamydomonas reinhardtii* (Kerfeld *et al.*, 1995) and *Monoraphidium braunii* (Frazão *et al.*, 1995), the cytochrome  $c, d_1$ -nitrite reductase from *Thiosphaera pantotropha* (Fülöp *et al.*, 1995a), and for the cytochrome  $c_{551}$  from *Pseudomonas aeruginosa* (also called  $c_8$ ), for which both X-ray and NMR structures are available (Matsuura *et al.*, 1982; Detlefsen *et al.*, 1991). NMR structures have also been determined for the closely related cytochromes  $c_{551}$  from *Ps. stutzeri* (Cai *et al.*, 1992) and *Ps. stutzeri* ZoBell (Cai & Timkovich, 1994) and for the cytochrome  $c_{553}$  from *Desulfovibrio vulgaris* Hildenborough (Blackledge *et al.*, 1995b). Lower resolution structures have been reported for *Chlorobium limicola* cytochrome  $c_{555}$  (Korszun & Salemme, 1977), the cytochrome  $c_{54}$  from *Anacystis nidulans*, now called  $c_6$  (Ludwig *et al.*, 1982), *Azotobacter vinelandii* cytochrome  $c_6$  (Carter *et al.*, 1985), *Desulfovibrio vulgaris* Miyazaki  $c_{551}$  (Nakagawa *et al.*, 1990), p-cresol methylhydroxylase, a flavocytochrome  $c$  from *Pseudomonas putida* (Mathews *et al.*, 1991), the di-heme flavocytochrome  $c$  sulfide dehydrogenase from *Chromatium vinosum* (Chen *et al.*, 1994b), and the di-heme cytochrome  $c$  peroxidase from *Ps. aeruginosa* (Fülöp *et al.*, 1995b). In addition, the structure of cytochrome  $c_{551}$  from the denitrifying bacterium *Paracoccus denitrificans* has been resolved in a ternary protein complex with methylamine dehydrogenase and amicyanin (Chen *et al.*, 1994a).

3D structural data may allow a better understanding of electron transfer or redox potential control on an atomic level. During the last few years, much effort has been devoted to the characterization of structurally and functionally important residues. Comparative analyses of cytochrome  $c$  structures isolated from different organisms can give insight into the conservation of individual residues, structural motifs or hydrogen bond networks (Chothia & Lesk, 1985; Moore & Pettigrew, 1990; Sogabe & Miki, 1995; Blackledge *et al.*, 1996). In addition, site-specific mutagenesis has been used to determine the effect of specific mutations on electron transport kinetics, redox potential or protein stability (Mauk, 1991; Caffrey, 1994; Caffrey & Cusanovich, 1994 for reviews). Key residues involved in the intermolecular interaction were recently identified by the structural analysis of two electron transfer complexes, eukaryotic cytochrome  $c$  with cytochrome  $c$  peroxidase (Pelletier & Kraut, 1992), and the ternary complex of methylamine dehydrogenase, amicyanin and cytochrome  $c_{551}$  from *Paracoccus denitrificans* (Chen *et al.*, 1994a).

Here we present the solution structure of the cytochrome  $c_{551}$  from *Ectothiorhodospira halophila* (Eha  $c_{551}$ ). The *Ectothiorhodospiraceae* are phototrophic sulfur bacteria which belong to the  $\beta$ -subgroup of the purple bacteria (Stackebrandt *et al.*, 1984; Woese, 1987). Among the soluble metallo-redox proteins isolated from the extremely halophilic bacterium *E. halophila*, a small

cytochrome  $c_{551}$  was characterized with a particularly low redox potential of 58 mV (Meyer, 1985; Ambler *et al.*, 1993). This cytochrome  $c$  shows no significant sequence homology with cytochromes  $c$  from other purple phototrophic bacteria. Instead, it exhibits sequential resemblance with the *Pseudomonas* and *Azotobacter* cytochromes  $c_5$ , with the cytochromes  $c_{555}$  from the green sulfur bacteria and with the cytochromes  $c_{576}$  from cyanobacteria and eukaryotic algal chloroplasts (Ambler *et al.*, 1993; Bersch *et al.*, 1995).

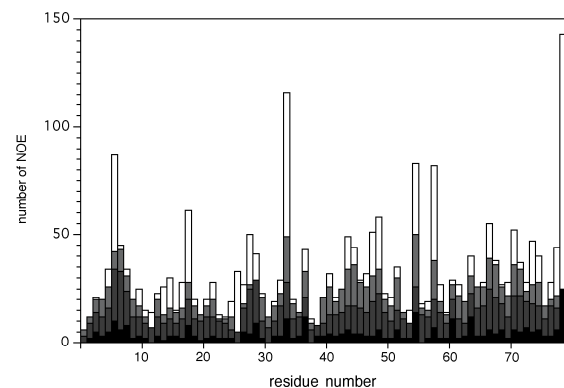
Structure calculations were performed with 1445 NOE distance constraints using a recent three stage simulated annealing (SA) - restrained molecular dynamics (rMD) protocol (Blackledge *et al.*, 1995b). The overall stability and physical viability of the ensemble was investigated by comparing restrained and unrestrained MD calculations.

The structure of Eha  $c_{551}$  is also compared to the crystal structures of *A. vinelandii* cytochrome  $c_5$  (Avi  $c_5$ ) and *P. aeruginosa* cytochrome  $c_{551}$  (Pae  $c_{551}$ ). Avi  $c_5$  shows the greatest sequence homology. The comparison with Eha  $c_{551}$  is also interesting from a phylogenetic point of view, as the cytochrome  $c_5$  homolog has been proposed to be present in very different phototrophic bacteria (Ambler *et al.*, 1993). In contrast, comparison with the unrelated small bacterial Pae  $c_{551}$  provides insight into the conservation of structural properties essential for cytochrome  $c$  function. These should be interpreted in relation to the results of a recent structural comparison between two unrelated bacterial and one eukaryotic cytochrome  $c$  (Blackledge *et al.*, 1996). Finally, comparing the electrostatic surfaces of different cytochromes  $c$  gives important information concerning a possible molecular interaction within an electron transfer complex.

## Results

**Experimental constraints.** 1445 distance constraints (including 118 heme to protein distances) and 10  $\phi_i$  dihedral constraints were used for the 78 residue protein. The sequential distribution of NOESY derived distances according to their range is shown in Figure 1.

**Structure calculations.** The global fold of the protein is determined in a first simulated annealing (SA) stage, starting from randomized coordinates. It allows the



**Figure 1.** Sequential distribution and range of NOE distance constraints used in the calculation of the cytochrome  $c_{551}$  structures. Black, intraresidue NOE; dark grey, sequential NOE ( $i+1$ ); light grey, medium range NOE ( $i \leq 4$ ); white, long range NOE. Residue 79 corresponds to the heme moiety.



**Table 1.** Statistics of cytochrome c551

| A Experimental statistics     |                 |                   |
|-------------------------------|-----------------|-------------------|
| No of experimental violations | SA <sup>a</sup> | E_II <sup>b</sup> |
| > 0.05 Å                      | 18.9 ± 4.1      | 30.2 ± 5.6        |
| > 0.10 Å                      | 5.2 ± 2.3       | 11.8 ± 3.4        |
| B Energetic statistics        |                 |                   |
| energy term <sup>c</sup>      | SA <sup>a</sup> | E_II <sup>b</sup> |
| Bond                          | 13.8 ± 0.8      | 12.1 ± 0.7        |
| Angle                         | 102.7 ± 7.1     | 88.2 ± 4.8        |
| Dihedral                      | 82.2 ± 5.3      | 83.7 ± 4.4        |
| Out of plane                  | 3.0 ± 0.5       | 2.9 ± 0.6         |
| H bond                        | -31.6 ± 1.9     | -37.5 ± 1.6       |
| VDW                           | -240.0 ± 10.3   | -263.1 ± 7.7      |
| Electrostatic                 | -973.11 ± 12.4  | -1053.5 ± 9.1     |
| Total                         | -1042.2 ± 29.1  | -1166.6 ± 19.9    |
| experimental <sup>d</sup>     | 10.9 ± 3.0      | 13.0 ± 4.7        |
| C Structural statistics       |                 |                   |
|                               | SA <sup>a</sup> | E_II <sup>b</sup> |
| BB <sup>e</sup> (1-78)        | 0.57±0.07       | 0.48±0.18         |
| BB (3-77)                     |                 | 0.39±0.12         |
| Heavy (1-78)                  |                 | 0.91±0.16         |
| Heavy (3-77)                  |                 | 0.83±0.12         |

<sup>a</sup> SA = ensemble of 9 structures after SA stage. <sup>b</sup> E\_II = ensemble of 37 structures after SI/SII stage. <sup>c</sup> all values are given in kcal mol<sup>-1</sup>. <sup>d</sup> force constants used : SA, kNOE = 50 kcal mol<sup>-1</sup> Å<sup>-2</sup>, k $\phi$ <sub>I</sub> = 0.0 kcal mol<sup>-1</sup> rad<sup>-2</sup>; E\_II, kNOE = 25 kcal mol<sup>-1</sup> Å<sup>-2</sup>, k $\phi$ <sub>I</sub> = 60 kcal mol<sup>-1</sup> rad<sup>-2</sup>. <sup>e</sup> Rmsd values (in Å) are the average pairwise rmsd relative to the mean of the corresponding ensemble, for the residues and atoms shown; backbone atoms correspond to C, C<sup>α</sup>, N.

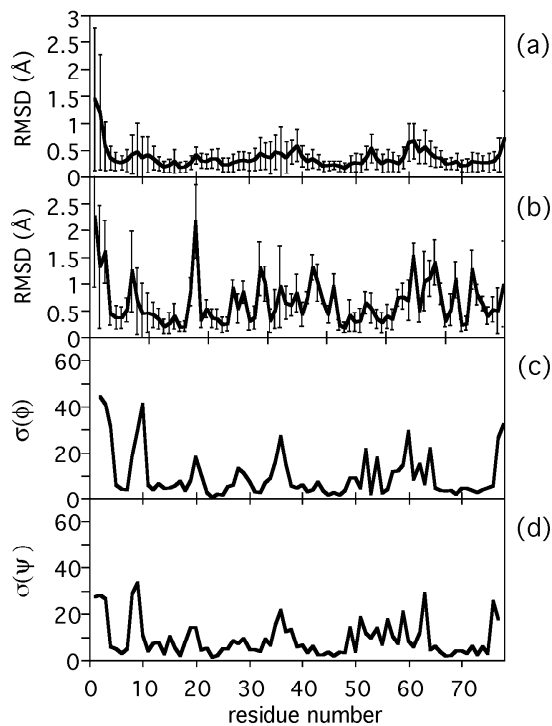
complete investigation of the conformational space as has been demonstrated elsewhere (Blackledge *et al.*, 1995b). Mirror images that occur because the chiral force constants are scaled more slowly than the experimental ones, can easily be distinguished from the correct fold by the experimental and physical energy term. An ensemble of correctly folded structures is then submitted to simulated annealing/restrained molecular dynamics (SA/rMD) calculations for an iterative structure refinement. The global fold of the protein was calculated twice, from the initial and from the final set of experimental constraints. The variation of the constraint list during the refinement did not lead to a significant modification of the global fold of the protein. The ensemble of 9 structures calculated with the final constraint set was used as starting conformations for the final simulated annealing/restrained molecular dynamics (SA/rMD) calculation. Among the 40 energy minimized structures obtained, 3 were eliminated on the basis of experimental violation energy (> 100 kcal mol<sup>-1</sup>); the total physical energy was also about 100 kcal mol<sup>-1</sup> higher than for the selected structures. The resulting ensemble will further be referred to as E\_II.

**Structural analysis.** The energetic and geometric statistics of the E\_II ensemble are given in Table 1 and Figure 2. Among these structures, the backbone was remarkably well defined with an average r.m.s.d. of 0.48 Å for the backbone C, C<sup>α</sup>, N atoms when superimposed on the mean structure (0.91 Å for all heavy atoms). Experimental constraints were also well satisfied. No violation greater than 0.43 Å was observed in E\_II. Despite the lower experimental force constant used in the rMD calculation, the resulting population of conformers was distributed around the same average backbone conformation as the global fold ensemble calculated with the simplified force field (backbone r.m.s.d. of SA[E\_II] 0.53 Å). The average backbone r.m.s.d. between the individual SA and the resulting E\_II structures was 0.9±0.18 Å. There were only a few residues for which the backbone conformation did not converge during the two consecutive SA calculations, which was resolved in most cases during the rMD calculation. The major changes during the refinement step were due to reorientations of sidechains in the rMD calculation, resulting in an energy drop of 100 kcal/mol (compare Blackledge *et al.*, 1995b; Albrand *et al.*, 1995).

**Table 2.** Secondary structure elements in the *E. halophila* cytochrome<sub>551</sub>.

| residue number | secondary structure element | hydrogen bond (NH O) <sup>a</sup> |
|----------------|-----------------------------|-----------------------------------|
| Gly2 - Asn8    | □ helix (helix-N)           | [6→2] 7□ 3 ... 9□ 5               |
| Thr10 - Thr13  | □ turn type VIb             | [13→10]                           |
| Gly21 - Gly24  | □ turn type II              | 24□ 21                            |
| Glu27 - Ala30  | □ turn type II              | [30→27]                           |
| Pro31 - Trp34  | □ turn type III             | 34□ 31                            |
| Trp34 - Arg37  | □ turn type I               | 37□ 34                            |
| Val41 - Ala50  | □ helix (helix-2)           | 44□ 40 ... [46□ 42] ... 51□ 47    |
| Lys52 - Met55  | □ turn type II'             | 55□ 52                            |
| Tyr58 - Arg61  | □ turn type II              | 61□ 58                            |
| Arg64 - Ser76  | □ helix (helix-C)           | 67□ 63 ... 76□ 72                 |

<sup>a</sup> secondary structure stabilizing hydrogen bonds identified in more than 90% of the E<sub>II</sub> ensemble. In the □ helices, only the first and the last hydrogen bonds are specified. Missing hydrogen bonds are given in brackets.

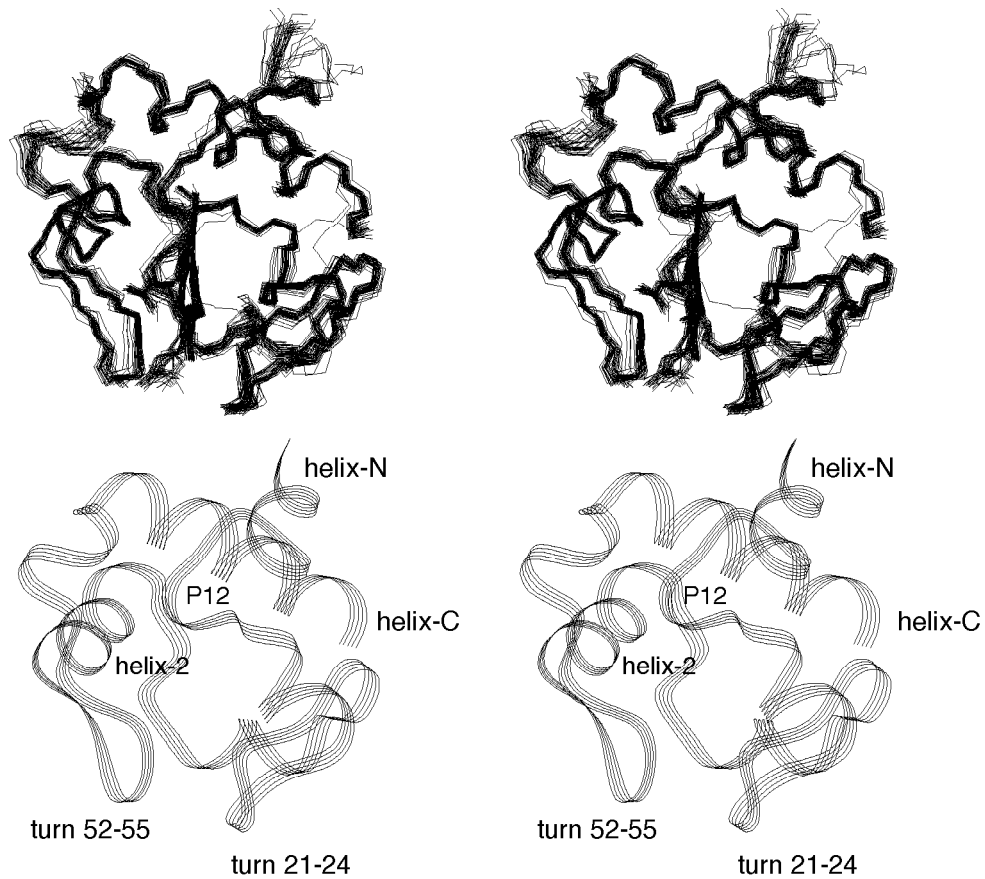


**Figure 2.** Structural analysis of the E<sub>II</sub> ensemble. a, the positional backbone r.m.s.d calculated with respect to the mean structure. b, the heavy atoms r.m.s.d. c, standard deviation of  $\phi$  angles compared with the mean value. d, standard deviation of  $\psi$  angles compared with the mean value.

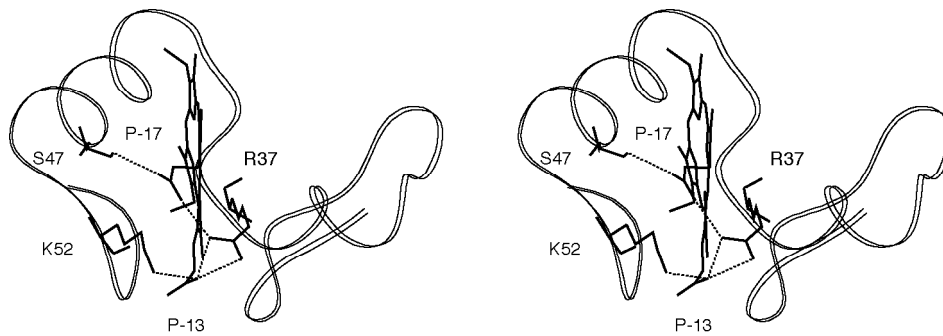
**Structure description.** Figure 3 shows the superposition of the 37 backbone structures of the E<sub>II</sub> ensemble, the secondary structure elements of the protein are listed in Table 2. The Eha<sub>551</sub> shows the typical cytochrome<sub>551</sub> fold with the 3 conserved helices in their characteristic relative orientation (Chothia & Lesk, 1985). The N-terminal helix (helix-N) spans residues 2 to 8, the central helix (helix-2) residues 41 to 50 and the C-terminal helix (helix-C) residues 64 to 77. Residues Ile5 and Tyr6 on the N-terminal helix interact mainly with residues Asp66, Ala70 and Tyr73 from the C-terminal helix, the interaction between the two aromatic amino acids is conserved among most

cytochromes<sub>551</sub>. The flipping of the Tyr6 ring is very slow and NOE could be attributed to the four individual aromatic protons. This resulted in a very well defined orientation of the aromatic ring with standard deviations of  $\phi_1$  and  $\phi_2$  of less than 5°. The segment Asn8 to Thr10 at the end of helix-N is characterized by a  $\phi/\psi$  heterogeneity mainly of Gly9, which is disordered. Following Thr10, the protein backbone forms a type VIb  $\phi$  turn, including a *cis* peptide bond preceding Pro12. This *cis* peptide bond, unique so far in cytochromes<sub>551</sub>, has already been recognized from the typical NOE pattern (Bersch *et al.*, 1995) and is confirmed by the structure calculations, during which no constraints were used to force this torsion ( $\phi = -10.89^\circ \pm 2.9^\circ$ ). This turn is in close contact with the top of the methionine arm on the other side of the heme binding site. The heme binding motif (Cys14 to His18) shows a highly conserved conformation due to steric constraints with the heme. The protein chain then forms a sequence of several non-conserved loops. In Eha<sub>551</sub>, the well characterized  $\phi$  loop of the mitochondrial cytochrome<sub>551</sub>, Pae *c*<sub>551</sub> and *D. vulgaris* *c*<sub>553</sub> (see Blackledge *et al.*, 1996 for comparison) is replaced by a short, non-standard turn between residues His18 and Gly21. Residues 21 to 24 form a type II  $\phi$  turn at the bottom of the heme. The chain rises again to correctly position Pro26 which forms a conserved hydrogen bond with the histidine sidechain. This is another example of a residue which shows functional but not sequential homology among different cytochromes<sub>551</sub>, as has been proposed for amino acids interacting with the heme propionates (Moore & Pettigrew, 1990) and recently for the methionine ligand (Blackledge *et al.*, 1995a). In the eukaryotic cytochromes<sub>551</sub>, the proline homolog is situated on the  $\phi$ -loop, which has nearly disappeared in the Eha<sub>551</sub>. The small, helix like fragment comprising residues Ala31 to Arg37 can be identified as a sequence of two consecutive  $\phi$  turns (see Table 2), followed by helix-2. The methionine arm of Eha<sub>551</sub> is remarkably short and comprises only 4 residues, Lys52 to Met55 which form a type II'  $\phi$  turn.

**Sidechain conformation.** As can be expected from the heavy atom r.m.s.d. of 0.9 Å, the sidechain conformations are well defined. Only 6 out of 59 residues show a mixture



**Figure 3.** Superposition of the 37 structures of the E\_II ensemble on their mean (top) and the schematic ribbon representation (bottom) in stereo representation. For clarity, only the backbone C $\alpha$ , C and N atoms and the heme heavy atoms are displayed.



**Figure 4.** Hydrogen bonds in *E. halophila* cytochrome c<sub>551</sub>. Stereo representation of the hydrogen bonding of the heme propionates.

of the three possible  $\phi_1$  rotamers (Glu3, Arg20, Glu32, Asp42, Asp63 and Met74). Especially the sidechains of the amino acids located in helix-2 and helix-C as well as those implicated in hydrogen bonds (see below) were very well defined. 30 out of 49 unconstrained  $\phi_1$  angles adopt the same rotamer in more than 80% of the structures and 12 of them populate a single rotamer in all structures.

In this context it is also interesting to address the question of the methionine ligand which can adopt a number of sidechain conformations with either R or S chirality at sulfur for the iron coordination (Senn & Wüthrich, 1985). Because the chirality of the Eha c<sub>551</sub> methionine sulfur was not known, it was held unconstrained during the initial calculations. The resulting SA and E\_II ensembles included

structures with R or S chirality. These two conformers could be distinguished by comparing the relative intensities of the NOESY cross-peaks of the Met55 C $\beta$ H<sub>3</sub> with either the heme 15- and 20-*meso* protons or the Met55 C $\beta$ H<sub>2</sub> and C $\beta$ H<sub>2</sub> sidechain protons. In the structures with S chirality, the distance between the methionine methyl and the 15-*meso* proton was about 1.0 to 1.5 Å shorter than the corresponding distance to the 20-*meso* proton. Therefore, in the final structure calculation, the chirality of the methionine sulfur was constrained to S. 33 E\_II structures showed the same, very well-defined sidechain conformation with  $\phi_1$ ,  $\phi_2$  and  $\phi_3$  angles of  $-175^\circ \pm 3^\circ$ ,  $158^\circ \pm 3^\circ$  and  $60^\circ \pm 5^\circ$ . Only 4 structures had a different Met55 sidechain conformation, but several distances to heme, Tyr48 and Leu49 protons did not

correspond to their relative intensities in the NOESY spectra.

**Hydrogen bonds.** The helix-stabilizing hydrogen bonds ( $\text{HN}_{i+4}\text{-CO}_i$ ) of the E\_II structures are in good agreement with the NMR determination of slowly exchanging amide protons (Bersch *et al.*, 1995), with the exception of Ile7, Asn8, Ser47 and Met74. These amino acids are situated in helices but their amide protons were found to be exchanged after 12 hr in  $\text{D}_2\text{O}$ . It should be noted that no H-bond constraints were used in the structure calculation. Good agreement with the NMR data was also found for residues Trp34, Arg37 and Met55 corresponding to the hydrogen bond donors in three of the six reverse turns (compare Table 2). It should be borne in mind that hydrogen exchange within 12 hr does not exclude hydrogen bonding, but might indicate a somewhat weaker hydrogen bond. On the other hand, non-exchanged hydrogens should be found implicated in a hydrogen bond, as hydrogen bonding provides the predominant solvent shielding mechanism in a small protein. Two amide protons (Leu28 and Ala30) did not exchange but could not be identified as hydrogen bond donors.

Apart from the hydrogen bonds occurring in the secondary structure elements, there are several long range hydrogen bond motifs which seem to be important for the stabilization of the 3D structure of the protein. The following interactions have been recognized in more than 9 out of 37 E\_II structures:

(1) Asp1 interacts with a residue in the C-terminal helix. In fact, the N-terminus is highly disordered and two possible hydrogen bonds, each with a population below 25% in the set of E\_II structures were identified: Asp1- $\text{NH}_3^+$  with Lys69-O or Lys69- $\text{N}^+\text{H}_3$  with Asp1-O

(2) The heme ligation site is stabilized by two H-bonds between Ser15 HN and Tyr6 CO as well as between the Ser15 OH and the Ile7 CO. In addition, the well characterized hydrogen bond between the Pro26 carbonyl and the histidine sidechain is present in all 37 structures.

(3) In 25% of the E\_II structures, the Tyr6 ring forms a hydrogen bond with the Asp19 sidechain. This interaction stabilizes the relative orientation of helix-N and the  $\beta$ -loop like structure. The low frequency of this hydrogen bond throughout the E\_II ensemble is somewhat surprising because the ring OH proton is one of the exchangeable non-amide protons visible in the NMR spectra and should therefore belong to a stable hydrogen bond (Bersch *et al.*, 1995). Inspection of the E\_II ensemble revealed that an Asp19  $\phi_1$  value of  $-60^\circ$  is required for H-bond formation whereas  $\phi_1=180^\circ$  is observed in most of the E\_II structures. The observed  $\text{C}^{\beta}\text{H}/\text{C}^{\alpha}\text{H}$  coupling and the  $\text{NH}-\text{C}^{\beta}\text{H}_{1,2}$  NOE strongly suggest the unique population of the  $-60^\circ$  conformer, but this dihedral angle has not been constrained during the structure calculation as different rotamers were present in the global fold ensemble (see "Materials and Methods" for the use of  $\phi_1$  constraints). This  $\phi_1$  angle would also be in agreement with two weak NOE between the Tyr6 ring OH and the two Asp19  $\text{C}^{\beta}\text{H}$  protons present in the NOESY spectra.

(4) The type II  $\beta$  turn Glu27 to Ala30 is hydrogen bonded to the C-terminus (Asn29 HN with Thr77 OH) and

to the following two consecutive  $\beta$  turns (Glu27 amide with the Asp33 sidechain carboxyl, and from the Trp34 sidechain to the Glu27 carbonyl). The previously predicted orientation of the Trp34 sidechain is confirmed, which excludes hydrogen bonding to the heme propionates (Bersch *et al.*, 1995). Therefore, Trp34 is not analogous to Trp56 of Pae $\beta_{551}$  or Trp59 of mitochondrial cytochrome $\beta$ .

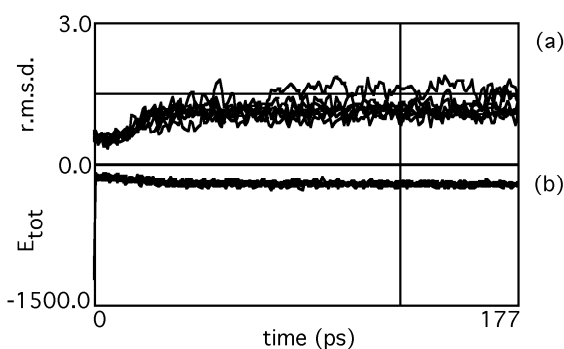
(5) Helix-2 and helix-C interact *via* a hydrogen bond between the Arg64 sidechain ( $\text{N}^+\text{H}_2$ ) and the Asp42 and Glu46 sidechains. These sidechains are situated at the molecular surface. The Arg64- $\text{N}^+\text{H}_2$  belong to the slowly exchanging protons which were detected in the NMR spectra (Bersch *et al.*, 1995).

(6) The top of the methionine ligand arm is stabilized by hydrogen bonds from Tyr58 HN and Asp59 HN, neither of which exchanges during 12 hr in  $\text{D}_2\text{O}$ , to Thr48 CO and Leu49 CO.

(7) Hydrogen bonds from the Tyr58 ring to the Pro12 CO and in some cases from the Arg61 sidechain to the Ala11 CO connect the two halves of the molecule on top of the heme. Note that the Tyr58 ring OH proton was also visible in the NMR spectra.

(8) Figure 4 shows the hydrogen bonding to the heme propionates. In all calculated structures, Arg37 interacts with the two propionates simultaneously, by means of the slowly exchangeable  $\text{N}^+\text{H}_2$ . Due to sidechain disorder, the Lys52 amine forms hydrogen bonds with either propionate, although propionate-13 is preferred. In addition, the Ser47 sidechain hydrogen bonds to propionate-17.

**Constraint-free molecular dynamics.** To further validate the structural ensemble obtained, 8 arbitrarily chosen E\_II structures were subjected to a constraint-free MD simulation in the full AMBER force field. The total energy and the positional backbone r.m.s.d during the calculation are shown in Figure 5. All 8 structures were remarkably stable during the calculation with an average backbone r.m.s.d of less than 1.5 Å compared to the initial E\_II structures (see Figure 6a). The geometric statistics of the cfMD ensemble are listed in Table 3. The positional backbone r.m.s.d comparison of the conformational means of the cfMD and the rMD ensemble is given in Figure 6a and the two ensembles are shown in Figure 7. There is no



**Figure 5.** Analysis of the constraint free dynamic for the cfMD ensemble. a, backbone r.m.s.d (in Å) during the eight calculations compared to the initial structures, the horizontal line indicates an r.m.s.d of 1.5 Å and the vertical line the beginning of the sampling period. b, total energy during the calculations (in  $\text{kcal mol}^{-1}$ ).

**Table 3.** Statistics of the constraint free molecular dynamics calculations.

| A. Experimental statistics                     |                   |
|--|-------------------|
| Number of experimental violations <sup>a</sup> | cfMD <sup>b</sup> |
| > 0.5 Å  | 8.8 ± 2.5         |
| > 1.0 Å  | 2.5 ± 1.7         |
| B. Structural statistics                       |                   |
|  | cfMD <sup>b</sup> |
| cfMD BB (3-77) <sup>c</sup>                    | 0.81±0.2          |
| cfMD: rMD BB (3-77) <sup>d</sup>               | 0.84              |

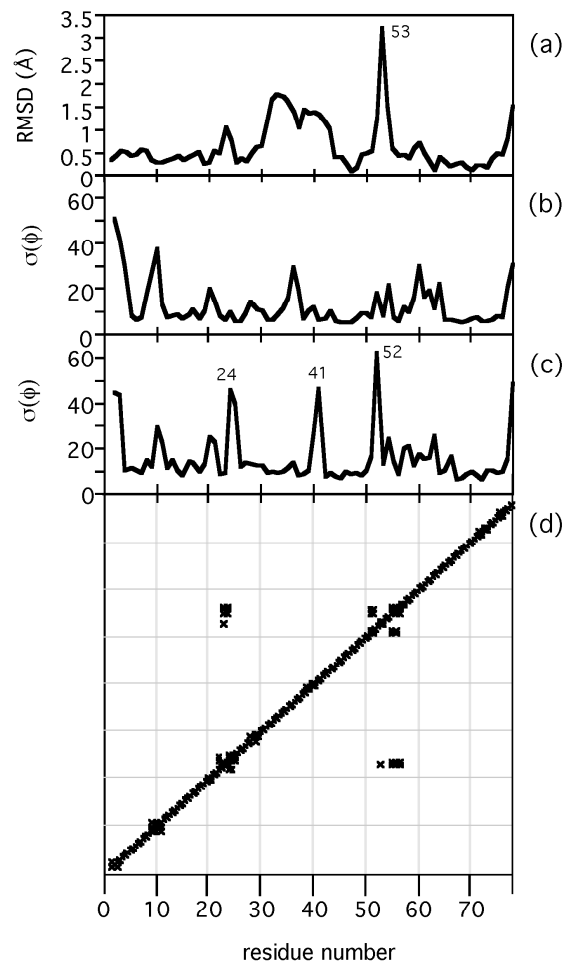
<sup>a</sup> Only backbone-backbone distances are considered (415 distances).

<sup>b</sup> cfMD = ensemble of 40 structures resulting of eight unconstrained molecular dynamic calculations.

<sup>c</sup> R.m.s.d values (in Å) are the average pairwise r.m.s.d relative to the mean of the corresponding ensemble, for the residues and atoms shown; backbone atoms correspond to C, C, N.

<sup>d</sup> R.m.s.d between the conformational means of the specified ensembles.

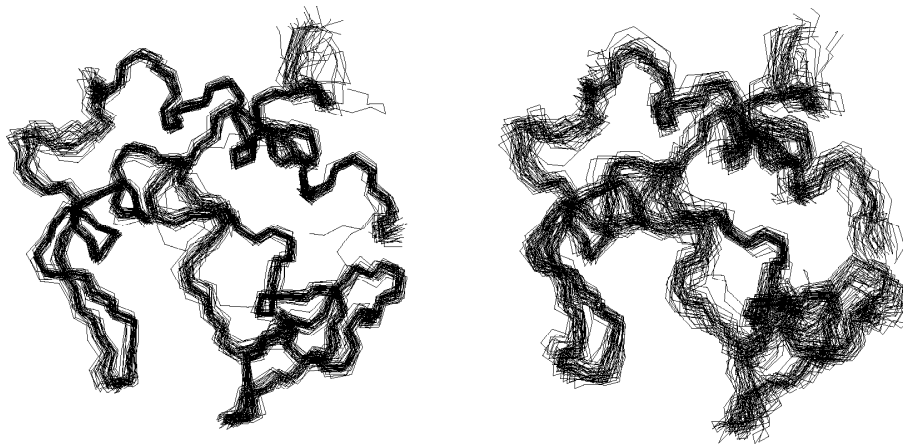
significant structural reorientation during the unconstrained MD simulation and all long range hydrogen bonding networks described above are conserved (data not shown). This means that the conformations calculated with the experimental constraints correspond to a well defined energy minimum with respect to the forcefield used in the absence of experimental constraints. However, the backbone r.m.s.d between the cfMD and the rMD means indicates a slight reorientation of residues 31 to 41. In the unconstrained MD, residues 31 to 36 form an  $\alpha$  helix instead of the two consecutive turns, as indicated by the backbone dihedral angles and the hydrogen bond network (data not shown). Interestingly, as shown by the  $\sigma(\phi)$  deviation, these residues are less flexible during the unconstrained than during the constrained MD (compare Figure 6b and 6c). Thus, the helical conformation seems to be preferred but is in contradiction with two experimental constraints which are violated in more than 75 % of the cfMD ensemble. On the other hand, three pairs of residues can be identified for which the flexibility is enhanced in the absence of experimental constraints : Gly24-Ala25, Ser40-Val41 and residues Lys52-Gly53. Again, the positional r.m.s.d indicates a slight reorientation of the backbone. Residues 24, 25 and 52 and 53 are localized in the two tight turns which close the bottom of the molecule. These two turns are in a concerted movement as can be seen from the normalized crosscorrelation of backbone dihedral variances over the cfMD ensemble shown in Figure 6d. In fact, conformational disorder and/or systematic reorientation in the corresponding regions have also been observed during the MD simulations of the *D. vulgaris* cytochrome $c_{553}$  (Blackledge *et al.*, 1995b). These residues are always close to the heme propionates, which we consider to be deprotonated, as their *pK* is not known. Therefore, a reduced negative charge was attributed to simulate the effects of solvent, implying their solvent exposure (compare Table 5 and Blackledge *et al.*, 1996). It might be that the systematic reorientation we observe is due to an oversimplification in the treatment of heme propionates.



**Figure 6.** Structural analysis of the cfMD ensemble and comparison with the rMD calculation. a, positional backbone r.m.s.d in Å between the conformational means of the cfMD and the rMD ensemble. b, standard deviation of  $\phi$  angles in comparison with their mean in the cfMD ensemble. c, standard deviations of  $\phi$  angles in comparison with their mean in the rMD ensemble. d, covariances of  $\phi\phi$  angles throughout the cfMD ensemble. Only values with  $c_{ij} > 0.75$  are shown.

## Discussion

**Comparison of the Eha $c_{551}$  with other bacterial cytochromes $c$ .** For this purpose, we measured structural homology by superposing small well-fitting regions and trying to expand them. A common motif was retained if the peptide segment comprised at least 4 residues with a trace r.m.s.d of less than 0.4 Å between the two respective structures (compare Blackledge *et al.*, 1996). The results of the comparison of the Eha $c_{551}$  with Avi $c_5$  and Pae $c_{551}$  are summarized in Table 4. Eha $c_{551}$  and Avi $c_5$  (Ambler *et al.*, 1993), show a significant structural homology with 48 C $\alpha$  (out of 78) at similar positions. Even with the less related Pae $c_{551}$ , 40 structurally analogous residues can still be identified, which represents roughly 50% of the Eha $c_{551}$ . A sequence alignment based on the identification of the homologous structural elements is given in Figure 8. Note that most of the conserved residues lie in regions of structural homology which are highlighted in Figure 8. Figure 9 shows the backbone superposition of these bacterial cytochromes $c$ . Hydrogen bond networks connecting different secondary structure elements are



**Figure 7.** Comparison of constrained and unconstrained molecular dynamics. Superposition of 37 rMD structures (left) and the 40 cfMD structures (right) on their means. The heme has been omitted for clarity.

summarized in Table 5 and the heme solvent accessibility for the proteins is given in Table 6. After a more detailed comparison of these bacterial cytochromes<sub>c<sub>5</sub></sub>, the electrostatic properties of the Eha<sub>c<sub>5</sub></sub> will be described at the end of the discussion.

**The common core structure of cytochromes<sub>c<sub>5</sub></sub>** It can be seen that the homologous segments defined by Chothia and Lesk (helix-N, helix-2, helix-C, CXXCH heme binding site,  $\square$ -loop and methionine ligand, Chothia & Lesk, 1985) are conserved with the exception of the  $\square$ -loop, which is reduced to a turn-like structure involving only four residues in the Eha<sub>c<sub>5</sub></sub> and in the Avi<sub>c<sub>5</sub></sub>. Interestingly, in the three small bacterial cytochromes<sub>c<sub>5</sub></sub> there is a strong homology of the methionine ligand motif, extending to at least five residues in the top of the methionine arm, including the methionine ligand. In the Pae<sub>c<sub>5</sub></sub> a six-residue insertion is found at the bottom of the methionine arm, which covers the heme at the level of heme propionate-13 (compare Figure 9).

***E. halophila* cytochrome<sub>c<sub>5</sub></sub> and *A. vinelandii* cytochrome<sub>c<sub>5</sub></sub> : two *Pseudomonas* cytochrome<sub>c<sub>5</sub></sub> analogs.**

The amino acid sequence and even more significantly the backbone conformation of these two proteins show a surprising resemblance, despite their large difference in redox potential (58 mV and 312 mV, respectively) and their function in two metabolically very different organisms. The *Ectothiorhodospiraceae* are anoxygenic phototrophs whereas *Azotobacter* belongs to the aerobic and facultatively denitrifying bacteria, as does *Pseudomonas*. In addition, these two cytochromes<sub>c<sub>5</sub></sub> might differ in their oligomerization state. As judged by the linewidth of the NMR resonances (Bersch *et al.*, 1995) the Eha<sub>c<sub>5</sub></sub> must be a monomer, at least under our experimental conditions, whereas the Avi<sub>c<sub>5</sub></sub> has been isolated both as a monomer and a dimer (Meyer & Kamen, 1982). Therefore, a comparative analysis of the corresponding structures is very interesting, and to our knowledge, represents one rare case where resolved molecular structures are available for two closely resembling cytochromes<sub>c<sub>5</sub></sub> exhibiting very different redox properties.

**Table 4.** Homologous structural motifs of bacterial cytochromes *c<sub>5</sub>*.

|                         | Eha <sub>c<sub>5</sub></sub> /Avi <i>c<sub>5</sub></i> | Eha <sub>c<sub>5</sub></sub> /Pae <i>c<sub>5</sub></i> |
|-------------------------|--|--|
| Helix-N <sup>a</sup>    | 0.210<br>(3-8)/(12-17)                                 | 0.385<br>(2-7)/(3-8)                                   |
| CXXCH                   | 0.384<br>(14-19)/(19-24)                               | 0.105<br>(14-18)/(12-16)                               |
| $\square$ turn type II  | 0.281<br>(21-27)/(26-32)                               | -  |
| $\square$ turn type I / | 0.359<br>(31-35)/(36-40)                               | -  |
| helix H'                | 0.315<br>(45-50)/(53-58)                               | 0.388<br>(41-50)/(41-50)                               |
| Helix-2                 | 0.329<br>(51-56)/(59-64)                               | 0.223<br>(51,52,55-58)/<br>(51,52,61-64)               |
| Methionine arm          | 0.373<br>(64-75)/(74-85)                               | 0.284<br>(65-77)/(69-81)                               |
| Helix-C                 | 1.14<br>(48 residues)                                  | 1.30<br>(40 residues)                                  |

<sup>a</sup> r.m.s.d values in Å for the superposition of C $\square$  atoms of the residues given in parentheses below.

Comparing the backbone conformations, only four short well-defined regions with significantly different folding can be identified : (1) the N-terminus, which is extended in the Avi<sub>c<sub>5</sub></sub> and often ragged (Meyer & Kamen, 1982; Carter *et al.*, 1985), (2) the four-residue insertion forming a type VIb  $\square$  turn in the Eha<sub>c<sub>5</sub></sub>, (3) the helix H' which is replaced by two consecutive turns in the Eha<sub>c<sub>5</sub></sub> and (4) the loop connecting the methionine arm and helix-C which is in close contact with the segment connecting helix-N and the heme binding site (the  $\square$  turn type VIb in the Eha<sub>c<sub>5</sub></sub>).

The four-residue insertion between helix-N and the heme binding site is unique in the *Ectothiorhodospiraceae* (Ambler *et al.*, 1993). The *cis*-proline (Pro12) interacts *via* a hydrogen bond with Tyr58 which is situated on the other half of the protein in the loop connecting the methionine arm and helix-C (Figure 10 a). These two bulky sidechains partially protect the cysteine edge of the heme from the solvent. In the Avi<sub>c<sub>5</sub></sub> the two halves of the molecule are connected by an interaction between Tyr18 and the Cys69-Cys72 disulfide bond present only in cytochromes<sub>c<sub>5</sub></sub>. It has

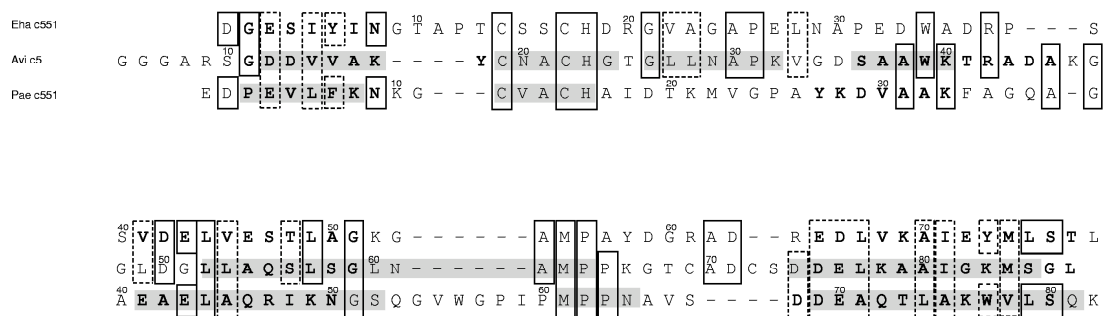
**Table 5.** Conserved hydrogen bond networks in three small bacterial cytochromes *c*

| Characterization                        | Eha $\Gamma_{551}$                           | Avi $c_5$                         | Pae $c_{551}$                    |
|---|--|-----------------------------------|----------------------------------|
| First residue - helix-C                 | D1:sc - K69:sc <sup>a</sup><br>D1:HN - K69:O | G5:HN - A79:O                     | E1:sc - E70:sc                   |
| Helix-N - CXXCH                         | Y6:O - S15:HN                                | V15:O - N20:HN                    | F7:O - V13:HN                    |
| Histidine ligand - proline              | H18:sc - P26:O                               | H23:sc - P31:O                    | H16:sc - P25:O                   |
| Stabilization of turn region / helix H' | N29:HN - T77:sc<br>W34:sc - E27:O            | G34:HN - M84:O<br>W39:HE1 - K32:O | -                                |
| Helix-C - helix-2                       | R64:sc - D42,E46:sc                          | - <sup>b</sup>                    | Q72:sc - E41:sc                  |
| Top of methionine arm                   | Y58:HN - T48,L49:O<br>D59:HN - L49:O         | K66:HN - S56:O<br>G67:HN - S56:O  | K49:sc - V66:O<br>N64:sc - I48:O |
| Connection above the heme               | Y58:sc - Pro12:O                             | - <sup>c</sup>                    | K10:sc - A65:O                   |
| Heme propionate-13                      | R37:sc                                       | R42:sc                            | -                                |
| Heme propionate-17                      | S47:sc                                       | Gln55:sc <sup>d</sup>             | R47:sc                           |

<sup>a</sup> sc, sidechain. <sup>b</sup> replaced by hydrogen bond between helix-C and helix H'. <sup>c</sup> replaced by interaction between Y18 sidechain and C69-C72 disulfide (see text). <sup>d</sup> hydrogen bond geometry not ideal (see text)

been shown that proteins are stabilized by weakly polar interactions between sulfur atoms and properly oriented aromatic rings (Reid *et al.*, 1985; Burley & Petsko, 1988). In the cytochrome $c_5$ , the criteria for a favourable interaction, as given by Reid *et al.* (1985) are fulfilled (S69: D=3.86Å,  $\Delta_1=51^\circ$ ,  $\Delta_2=89^\circ$ ; S72: D=5.77Å,  $\Delta_1=57^\circ$ ,  $\Delta_2=51^\circ$ , with D the distance between the sulfur atom and ring centroid and  $\Delta_1$ ,  $\Delta_2$  determining the relative orientation of sulfur and ring plane). Solvent protection of the cysteine edge is achieved by the Gly67 backbone which occupies a position analogous to the Tyr58 sidechain in the Eha $\Gamma_{551}$ . This results in a slightly more significant heme exposure in the case of cytochrome $\Gamma_5$ . Much interest has been devoted to the dielectric environment and to the solvent accessibility of cytochrome $\Gamma_5$  heme propionates, which have been proposed to be important factors in the modulation of the redox potential (Moore *et al.*, 1986; Moore & Pettigrew, 1990). Interestingly, the bottom of the molecule (cf Figure 3), where the propionates are situated, shows much more structural and sequential variability than the top of the molecule, with the conserved helices and the heme binding site. In what follows, we will discuss residues implicated in H-bonding and solvent protection of the heme propionates. In the Eha $\Gamma_{551}$ , H-bond interactions are found with the sidechains of three residues, Arg37, Ser47 and Lys52 (see Figure 4 and 10b). Arg37 is one of the functionally conserved residues in the cytochrome $c_5$  homologs (Ambler *et al.*, 1993). It is situated in the two non-conserved loops

and forms H-bonds with both propionates. *A. vinelandii* Arg42 belongs to helix H'. Its sidechain occupies nearly the same position as *E. halophila* Arg37 but only forms a H-bond with propionate-17 (Figure 10b). *E. halophila* Ser47, which interacts with heme propionate-17, is analogous to *A. vinelandii* Gln55. Both residues are situated in the last turn of the conserved helix-2. Even if the geometry for hydrogen bonding is not ideal, a possible hydrogen bond interaction between Gln55 and propionate-17 has been proposed (Carter *et al.*, 1985). The H-bonds formed between the *E. halophila* Lys52 sidechain and both propionates do not exist in the cytochrome $\Gamma_5$ , as the analogous position is occupied by Leu60. *E. halophila* Lys52 is more efficient in protecting heme propionate-17 from the solvent than the much shorter Leu60, resulting in a partial exposure of *A. vinelandii* heme propionate-17 in this region (compare Figure 10b). On the other hand, the significant solvent accessibility of Eha $\Gamma_{551}$  belongs to a region covered by Gln55 and Lys47 in Avi $\Gamma_5$ . These two sidechains are in close spatial proximity. As mentioned above, Gln55 is analogous to Ser47, which is less efficient due to its small size. *A. vinelandii* Lys47 belongs to the third helix H' turn which has been deleted in Eha $\Gamma_{551}$  (Figure 10b). It points towards heme propionate-17, but no H-bond is formed (Carter *et al.*, 1985). This lysine is the only residue in the additional turn of helix H' which interacts with the heme. The lysine and the glutamine are conserved in



**Figure 8.** Sequential homology of *E. halophila* cytochrome $\Gamma_{551}$ , *A. vinelandii* cytochrome $\Gamma_5$  and *P. aeruginosa* cytochrome $\Gamma_{551}$ . Structural sequence alignment. **□**-helices are indicated by bold font, homologous residues are indicated by boxes (dashed lines : conservative replacements) and regions structurally homologous to the Eha $\Gamma_{551}$  are highlighted.

**Table 6.** Heme solvent accessibility

|                               | Eha $\square_{551}$         | Avi $\square_5$ | Pae $c_{551}$ |  |
|-------------------------------|-----------------------------|-----------------|---------------|--|
| Methyl-7 <sup>1</sup>         | 6.54 $\pm$ 2.2 <sup>a</sup> | 7.74            | 20.65         |  |
| S-methyl-8 <sup>2</sup>       | 21.14 $\pm$ 1.3             | 23.02           | 23.88         |  |
| Methyl-12 <sup>1</sup>        | 7.47 $\pm$ 3.3              | 3.28            | 5.83          |  |
| Propionate-13: O1,O2          | 13.25 $\pm$ 2.0             | 15.59           | 4.48          |  |
| Propionate-17: O1,O2          | 11.66 $\pm$ 3.8             | 8.31            | 0.0           |  |
| heme surface area exposed (%) | 15.8 $\pm$ 1.5              | 15.9            | 12.6          |  |

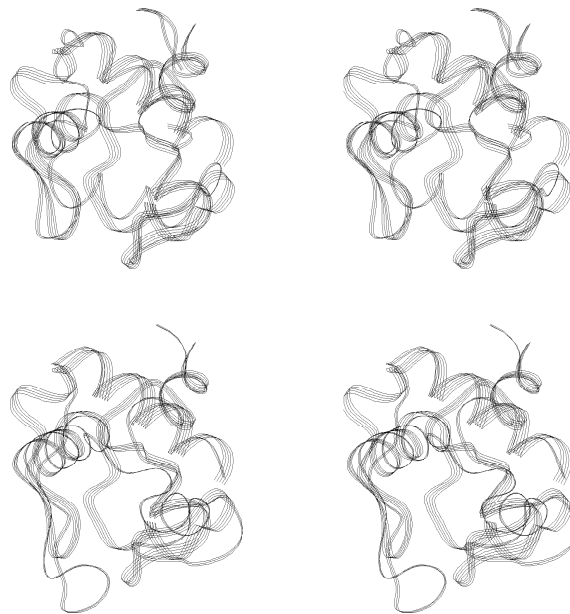
<sup>a</sup> all values are in Å<sup>2</sup>.

cytochromes $\square_5$  and cytochromes $\square_{555}$  (Lys46 and Gln55 in Avi $\square_5$ , Gln46 and Lys55 in *P. mendocina* cytochrome $\square_5$ , Lys40 and Asn48 in *C. limicola* cytochrome $\square_{555}$  and Gln40 and Lys48 in *Prosthecochloris aestuarii* cytochrome $\square_{555}$  (Moore & Pettigrew, 1990), but their relative positions are inverted in *P. mendocina* and *P. aestuarii*. These observations underline once more the proposed relationship between cytochromes $\square_5$ , cytochromes $\square_{555}$  and the Eha $\square_{551}$  (Ambler *et al.*, 1993). We can conclude that the solvent accessibility of Eha $\square_{551}$  and Avi $\square_5$  heme propionate-17 have the same magnitude which is much greater than that of most other cytochromes $\square$  (compare Kerfeld *et al.*, 1995; Blackledge *et al.*, 1996 and Table 6). However, different sidechains are involved in hydrogen bonding and solvent protection. The solvent protection of heme propionate-13 is comparable for the two proteins. In *E. halophila* it is achieved by the Arg37 and Lys52 sidechains and by the residues adjacent to the two turns at the bottom of the heme (compare Figure 4). *A. vinelandii* Leu60 is as efficient as Lys52 in protecting heme propionate-13 from the solvent.

**Comparison of *E. halophila* cytochrome  $c_{551}$  and *P. aeruginosa* cytochrome $\square_{551}$ .**

Comparing their backbone conformations it becomes obvious that structural homology is limited to the common core of cytochromes $\square$  described above (see Table 3 and Figure 9). The main differences between the two proteins are situated above the heme and again involve the type VIb  $\square$  turn of the Eha $\square_{551}$  and at the bottom of the molecule, where the polypeptide chain folds in very different conformations.

The solvent accessibility of the heme cysteine edge is significantly greater for the Pae $\square_{551}$  (compare table 6). This is partially due to the absence of a residue analogous to *E. halophila* Pro12 which interacts with Tyr58, protecting the heme edge from the top. The sidechain of *P. aeruginosa* Asn64 has the same orientation as *E. halophila* Tyr58, but hydrogen bonds to the methionine sulfur and to the backbone of Ile48 in helix-2 (Matsuura *et al.*, 1982). The H-bond between the sidechain of Lys10 and the Ala65 backbone closes the upper part of the heme crevice in the Pae $\square_{551}$  (Matsuura *et al.*, 1982). At the bottom of the exposed heme edge, the sidechain of *E. halophila* Ala54 points towards the heme, whereas the sidechain of *P. aeruginosa* Pro60 is directed to the protein surface, resulting

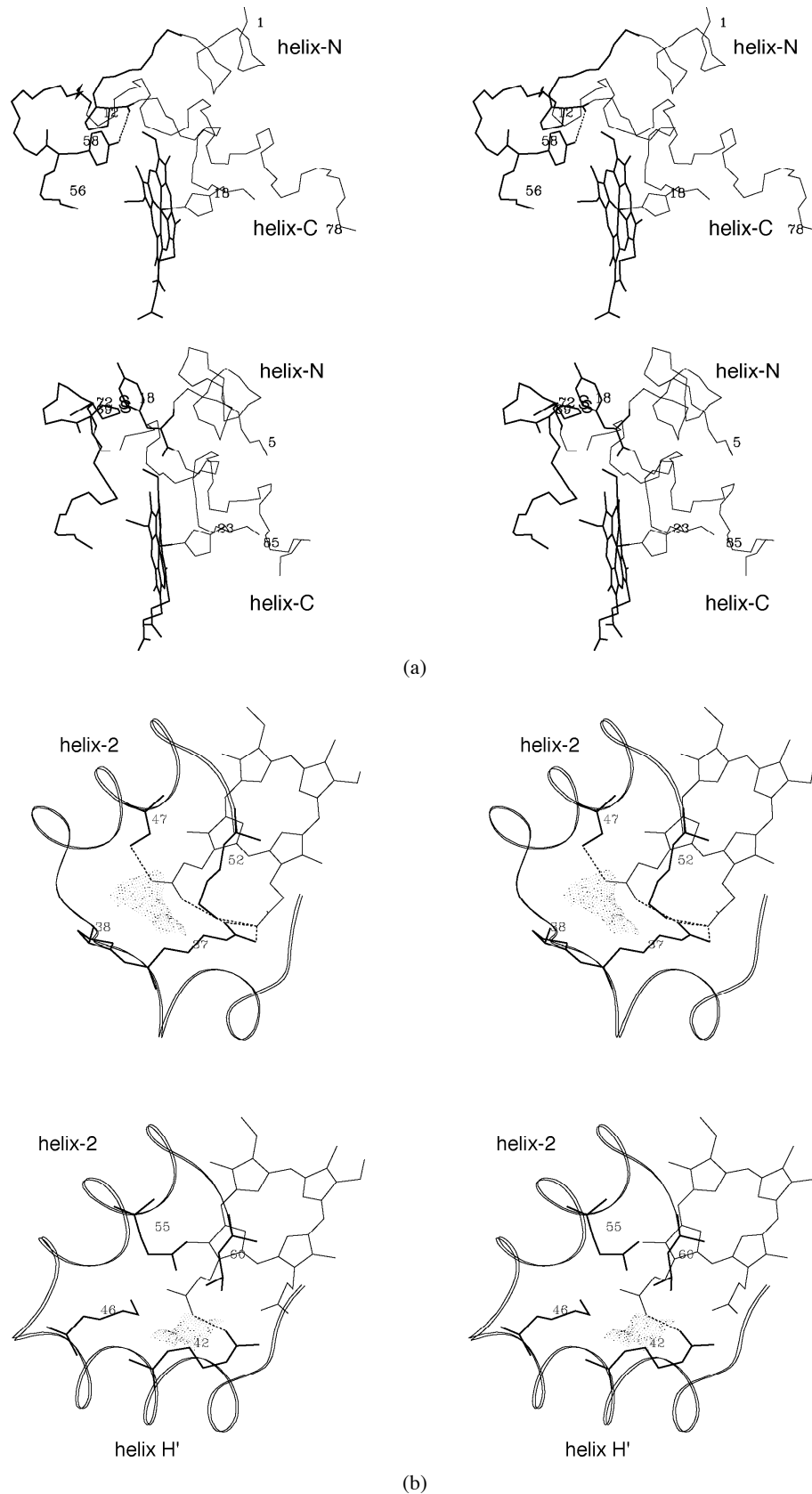


**Figure 9.** Structural homology of *E. halophila* cytochrome $\square_{551}$ , *A. vinelandii* cytochrome $\square_5$  and *P. aeruginosa* cytochrome $\square_{551}$ . Backbone superposition of Eha $\square_{551}$  (broad ribbon) and Avi $\square_5$  (narrow ribbon, top) and Pae $\square_{551}$  (narrow, bottom) in stereo representation. The heme is omitted for clarity.

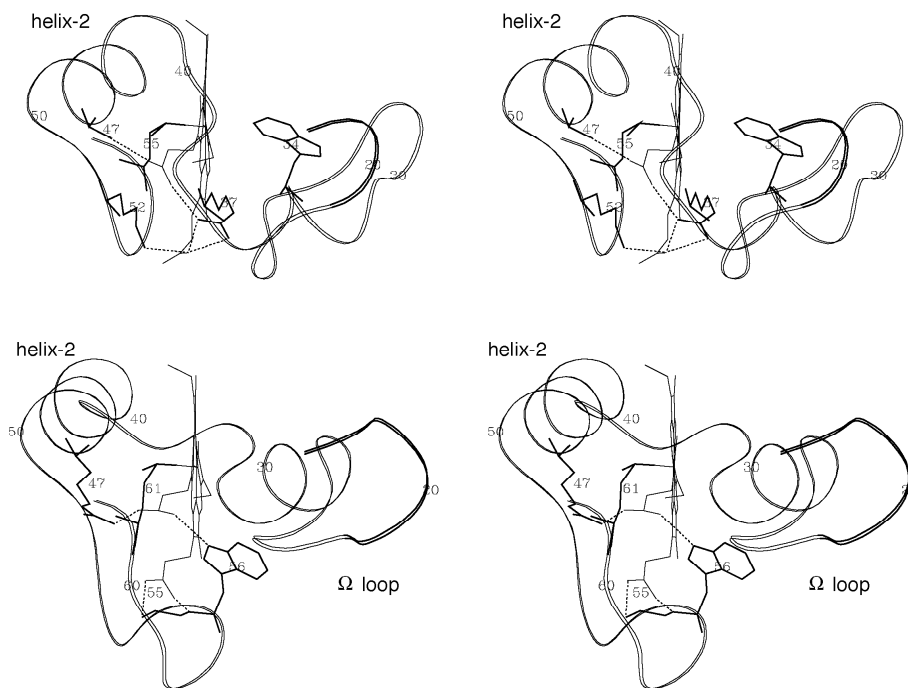
in the increased solvent accessibility of the *P. aeruginosa* heme.

The most significant difference between these two proteins is the folding of the protein backbone at the bottom of the molecule (Figure 11). As the two hemes but not the protein core have been superposed, the relative orientation of the two molecules differs from that shown in Figure 9. In the Pae $\square_{551}$ , the  $\square$ -loop (residues 16 to 22) is followed by a turn and the non-conserved helix H', whereas in the Eha $\square_{551}$  the polypeptide chain forms 4 consecutive turns. The type II  $\square$  turn (residues 21-24) does not play the same role as the  $\square$  turn between the *P. aeruginosa*  $\square$ -loop and helix H' as it closes the bottom of the protein at the level of propionate-13 and seems to compensate for the shorter methionine arm on the other side of the heme (see also Blackledge *et al.*, 1996). Note also that the *P. aeruginosa* helix H' does not show the same orientation as the two turns in the *E. halophila* protein which correspond to the *A. vinelandii* helix H', and thus these two non-conserved helices cannot be regarded as analogous, despite their similar location in the protein sequence (compare Figure 8). There is clearly no analogy between Trp34, situated at the end of the  $\square$  turn type III in the Eha $\square_{551}$  and the *P. aeruginosa* Trp56. Trp56 belongs to the six residue insertion in the methionine arm which is implicated in solvent protection of the heme propionates. The change in the backbone conformation induces different H-bond networks with the heme propionates in the two cytochromes as shown in Figure 11. The only conserved interaction is the one between the Ser47 sidechain and heme propionate-17 in the Eha $\square_{551}$  (Arg47 in Pae $\square_{551}$ ), but the shorter sidechain results in a less efficient solvent protection for the Eha $\square_{551}$ . Solvent protection of the *P. aeruginosa* heme propionates is mainly achieved by the





**Figure 10.** Structural comparison of *E. halophila* cytochrome<sub>b551</sub> and *A. vinelandii* cytochrome<sub>b5</sub> in stereo representation. a, interaction between the type VIb  $\beta$  turn and Tyr58 in the Eha<sub>b551</sub> (top) and the analogous interaction between Tyr18 and the disulfide in Avi<sub>b5</sub> (bottom). Non-homologous regions are given by thick lines, the hydrogen bond between Tyr58 and Pro12 is dashed. b, local environment of heme propionates in the Eha<sub>b551</sub> (top) and the Avi<sub>b5</sub> (bottom). The backbone is shown in ribbon representation, sidechains discussed in the text are drawn in thick lines and hydrogen bonds are dashed. The solvent accessible surface of propionate-17 is indicated by dots.



**Figure 11.** Structural comparison of *E. halophila* cytochrome $c_{551}$  (top) and *P. aeruginosa* cytochrome $c_{551}$  (bottom) in stereo representation. The folding of the peptide chain around the bottom of the heme crevice is shown in ribbon representation. The  $\Omega$ -loop and its analog as well as the sidechains discussed in the text are drawn in thick lines, hydrogen bonds are dashed.

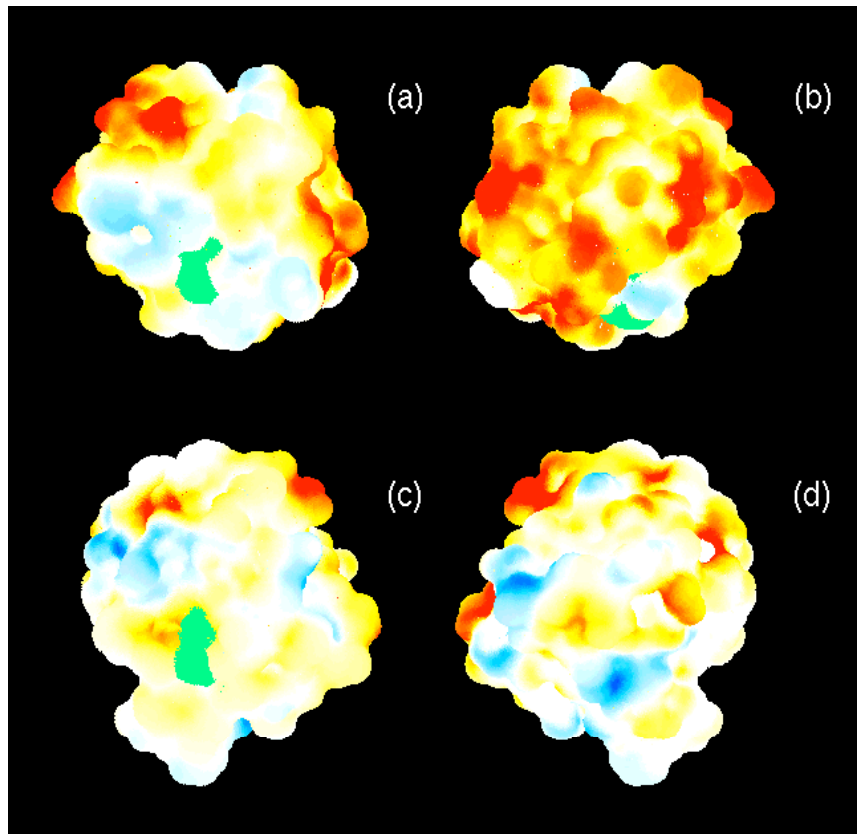
three bulky sidechains of Val55, Trp56 and Arg47 missing in the Eha $c_{551}$  (see Table 6 for the heme solvent accessibility). The direct environment of the heme propionates in the Eha $c_{551}$  is characterized by a higher polarity due to the increased solvent exposure and the interaction with two charged and a polar sidechain, whereas in the Pae $c_{551}$ , several of the heme propionate stabilising H-bonds are formed with hydrophobic residues.

**The electrostatic surface of cytochrome $c_{551}$ .** Based on experimental results at varying ionic strength and on the knowledge of 3D structures, electrostatic interactions have been proposed to stabilize complexes between various cytochromes $c$  and cytochrome $c$  peroxidase (Pelletier & Kraut, 1992), bacterial reaction centers (Tiede *et al.*, 1993; Meyer *et al.*, 1993a), plastocyanin (Meyer *et al.*, 1993b) and a soluble cytochrome $c$  isolated from *Paracoccus denitrificans* (Davidson & Jones, 1995). In the large class I cytochromes $c$ , a ring of conserved lysine residues around the exposed heme edge has been proposed to form a positively charged encounter surface able to interact with a negatively charged surface of their redox partners (Pelletier & Kraut, 1992; Allen *et al.*, 1987; Tiede *et al.*, 1993). A similar pattern of positive charges has been determined for cytochrome $c_{553}$  from *Desulfovibrio vulgaris* (Blackledge *et al.*, 1996). However, Pae $c_{551}$  has been shown to behave completely different in its interaction with bacterial reaction centers (Tiede *et al.*, 1993; Meyer *et al.*, 1993a) and plastocyanin (Meyer *et al.*, 1993b), due to its slightly negative surface charge at the heme edge. According to the sequence, the overall charge of the *E. halophila* ferro cytochrome $c_{551}$  is -11, assuming the two propionates to be deprotonated. As this relatively high negative charge is characteristic for proteins isolated from halophilic

organisms (Lanyi, 1971) we expected a quite homogenous negative surface for this cytochrome $c$ . Surprisingly, the Eha $c_{551}$  exhibits a significant polarization with the front face bearing weak positive and negative charges while most of the negative surface charges are clustered at the back of the molecule (Figure 12 a and b). The front face of Eha $c_{551}$  closely resembles that of Pae $c_{551}$  (Figure 12 a and c), even if their net charges are significantly different (-11 and -3, respectively). The very low charge density on this face compared to the backside of the Eha $c_{551}$  might indicate that electrostatic interactions do not have the same importance for complex stabilization than for the large class I cytochromes $c$ , if the same encounter surface is used. This hypothesis is also underlined by the fact that cytochrome $c_{551}$  functions in the periplasm of an extremely halophilic organism, i.e. at ionic strengths of greater than 3 M, at which electrostatic interactions are strongly shielded and non-specifically bound ions compete with complex formation.

Due to the strong polarization of the Eha $c_{551}$  this cytochrome $c$  could provide a good tool for the differentiation of possible binding sites by the measurement of electron transport rates at varying ionic strengths. However, the physiological redox partners of Eha $c_{551}$  as well as Pae $c_{551}$  remain to be determined.

**Conclusion.** We have determined the 3D solution structure of the Eha $c_{551}$  and compared it to two other small bacterial cytochromes $c$ . Comparison to Pae $c_{551}$  reveals poor structural homology with only the elements of the common cytochrome $c$  core being conserved. On the other hand, significant structural resemblance exists between Eha $c_{551}$  and Avi $c_5$ . The backbone fold is nearly identical with the exception of three short segments, of which two are



**Figure 12.** Electrostatic potential maps of bacterial cytochromes $\text{L}$ . a, Eha $\text{L}_{551}$  front side, b, Eha $\text{L}_{551}$  back side, c, Pae $\text{L}_{551}$  front side and d, Pae $\text{L}_{551}$  back side. The molecules in a and c are shown in the same orientation as in Figure 3, the exposed heme surface is given in green. Potential values between 0 and 3 kT, and 0 and -3 kT are scaled between white, cyan and blue and white, yellow and red, respectively.

interacting. They both have the same H-bond networks stabilizing the respective secondary structure elements, including two of the three protein to heme H-bonds of Eha $\text{L}_{551}$ . The amino acid sequences have been aligned with respect to the structural homology between the two cytochromes $\text{L}$ . More than 40% of the *E. halophila* residues are conserved or have been submitted to conservative replacements. These include the functionally conserved Pro12, Trp34 and Arg37 which have also been found in *Pseudomonas mendocina* cytochrome $\text{L}_5$  and in green bacterial cytochromes $\text{L}_{555}$  (Arg37 is replaced by His37 but is also in H-bond distance to the heme, Korszun & Salemme, 1977). In fact, from previous sequence alignments it has been proposed that *Pseudomonas* cytochrome $\text{L}_5$  analogs are present in cyanobacteria, green sulfur and purple bacteria (Ambler *et al.*, 1993). The detailed comparison of the respective 3D structures may provide additional information for the establishment of the phylogenetic relationship between phototrophic bacteria. The heme environment of the two proteins was compared in detail. The more than 250 mV difference in redox potential is not reflected in a different solvent exposure of the heme propionates, which is similar in the Eha $\text{L}_{551}$  and the monomeric Avi $\text{L}_5$  but higher than in most other cytochromes $\text{L}$ . Dimerization of cytochrome $\text{L}_5$  as observed in the crystal structure does not seem to alter the solvent accessibility of the propionates, but results in protection of the exposed heme cystein edge (Carter *et al.*, 1985). The observed variation of the backbone conformation occurs in

proximity of solvent exposed heme regions: the interacting Tyr58 - Pro12 (Tyr18 - disulfide in cytochrome $\text{L}_5$ ) are close to the exposed cysteine edge in the encounter surface whereas the consecutive tight turns (helix H' in cytochrome $\text{L}_5$ ) are close to heme propionate-17 on the opposite face. These observations suggest that the control and the variation of heme solvent accessibility are important factors during evolution.

The electrostatic surface potential was calculated for Eha $\text{L}_{551}$  and compared with Pae $\text{L}_{551}$ . The front face shows a striking resemblance in Eha $\text{L}_{551}$  and Pae $\text{L}_{551}$ . It is only poorly charged and therefore differs from that of the larger class I cytochromes $\text{L}$ . It is tempting to suggest an identical mechanism for the stabilization of electron transfer complexes for these two proteins in which electrostatic interactions do not play the same role as in large cytochromes $\text{L}$ .

We hope that the present structural study will contribute to the overall understanding of cytochrome $\text{L}$  function and electron transfer complex formation and will give additional hints for the classification of phototrophic bacteria and of the *Ectothiorhodospiraceae* in particular.

#### Materials and Methods

**Sample.** The sample was prepared as previously described (Bersch *et al.*, 1995). Protein concentration was 8 mM in 100 mM phosphate buffer, pH 5.8.

**Experimental constraints.** Distance constraints were obtained from NOESY experiments acquired at three different temperatures (12°C, 27°C, 40°C) with 50 and 100

ms mixing time on a Bruker-AMX600 spectrometer as described elsewhere (Bersch *et al.*, 1995). Cross-peaks from the amide-amide and amide-aliphatic regions were assigned by hand and classified in four categories with distance bounds corresponding to 3.0, 4.4, 5.75 and 6.9 Å. The use of the rather long distances is justified by the excellent quality of the NMR spectra in terms of resolution and signal/noise. Pseudo-atom corrections were added for degenerate methylene, methyl and aromatic protons as proposed by Wüthrich (1986). Resolved but non-stereospecifically assigned protons or methyl groups were allowed to float between the two prochiral positions during the SA protocol. In addition to the distance constraints, ten  $\phi$  dihedral constraints were added after determination of the global fold. One of the three possible rotamers was chosen on the basis of the relative NH/C $\beta$ H<sub>1,2</sub> peak intensities in a 50 ms NOESY and the NH/C $\beta$ H cross-peak shapes at 40°C. From the latter, strong coupling constants were identified with the C $\beta$ H/C $\beta$ H coupling in the detection dimension. The dihedral constraints were added in the constraint list only if the corresponding rotamer came out in more than 90% of the globular fold structures to give priority to the distance constraints. No hydrogen bonding distance constraints were used for the helical regions.

**Structure calculations.** Structure calculations were performed using DISCOVER (Version 2.9.7, 1993) interfaced to the INSIGHT II program (Version 2.3.0, 1993) for visualization and result analysis (Biosym Technologies Inc.). The AMBER4 force field was used for all calculations except for the SA protocols, in which a simple quartic non-bond term was employed. All  $\psi$  dihedrals were forced to *trans* except the one preceding Pro12, as there was strong evidence from the NOE data that this peptide bond was in the *cis* conformation (Bersch *et al.*, 1995). For more detailed information on the three-stage protocol employed, which will be briefly described below, the reader is referred to a previous publication (Blackledge *et al.*, 1995b).

1. **Determination of the global protein fold.** The global fold of the protein was determined using a 60 ps SA protocol starting from randomized initial coordinates containing 10 ps of slow cooling to 100 K. The exploratory period was calculated at a nominal temperature of 1000 K. Experimental data were scaled to a maximum of 50 kcal mol<sup>-1</sup>Å<sup>-2</sup> during the first 30 ps. The molecule was then minimized in the complete AMBER force field. This protocol was employed twice: in the very beginning with an initial set of distance constraints, and once the final list of experimental constraints has been obtained after the iterative structure refinement.

2. **Structure refinement.** For the refinement of the globally refolded structures, a two-stage SA - rMD (SI/SII) protocol was used (Blackledge *et al.*, 1995b). During the SA calculation (SI), the local conformational space available from the experimental data was investigated. This calculation still allowed a large sampling of the conformational space (the backbone r.m.s.d has a radius of greater than 5.0 Å), but was less time consuming than the global fold protocol as it started from already folded structures (Blackledge *et al.*, 1995b). In this step, the nominal temperature was held at 2000 K and the NOE force

constants were scaled up progressively to 50 kcal mol<sup>-1</sup>Å<sup>-2</sup>. Short range distances which define the local conformation were introduced before long range distances. After the scaling period, the molecule was allowed to sample the conformational space for 2 ps before the system is smoothly cooled to 100 K over a period of 11 ps. After 2 ps of dynamics, the molecule was subjected to energy minimization as described above. In the following rMD calculation (SII), the use of the complete force field including Coulombic, Lennard-Jones and hydrogen bond interactions led to a physically more viable structure. In this calculation, enough energy was given to the molecule to overcome local energy barriers. The AMBER force field, with the heme parametrization according to Giammona (1984) was used. The effects of solvent were simulated implicitly by the use of a distance-dependent dielectric constant (Brooks *et al.*, 1983; Weiner *et al.*, 1984) and reduced charges on the charged sidechains (Singh & Kollman, 1983; Blackledge *et al.*, 1995b). The weight of the experimental distance constraints was reduced to 25 kcal mol<sup>-1</sup>Å<sup>-2</sup>. After equilibration, the temperature was regulated using weak coupling to a thermal bath (Berendsen *et al.*, 1984). The molecule was allowed to evolve at 750 K for 10 ps, then it was slowly cooled to 300 K over a period of 5 ps, where dynamics was continued for a further 12 ps. Finally the structure was minimized using a conjugate gradient algorithm.

**Structural analysis.** The final structural ensemble was selected on the basis of the experimental energy function, including contributions from distance,  $\phi$  and  $\psi$  dihedral as well as chiral constraints. The ensemble means were calculated by superposing the heavy atoms of the individual structures. The rms deviation was calculated over the specified atoms in comparison to the mean structure, using INSIGHT II.

For each individual structure, two atoms were considered to form a hydrogen bond if the distance between hydrogen and hetero atom was less than 2.5 Å and the hydrogen bond angle greater than 120°. The solvent accessible surfaces were calculated using the Conolly algorithm provided with the INSIGHT II software. The surfaces were calculated in the absence of protons using the united atom radii proposed by Conolly (Conolly, 1983) and a water molecule radius of 1.4 Å.

**Calculation of electrostatic surfaces.** As the mobility of solvent exposed charged sidechains can significantly alter the electrostatic surface, the charge distribution has been determined for series of structures. After calculating an electrostatic grid for the individual structures, the mean of each grid value is taken to represent the conformational average, whose surface gives a much more realistic representation of the effective encounter surface of the molecule than any individual conformer. Electrostatic surfaces were calculated for Eha $\alpha_{551}$  (37 structures, ensemble E\_II) and have already been calculated for Pae c<sub>551</sub> (Blackledge *et al.*, 1996).

The electrostatic grids were calculated with the program DelPhi (Gilson *et al.*, 1987), implemented in the Biosym package, as described elsewhere (Blackledge *et al.*, 1996). Dielectric constants of 80 and 4 were used to represent the

solvent and the protein interior, respectively. An ionic strength of 0.145 M was assumed for the solvent and the Stern layer was taken to 2.0 Å. Charges were as defined in the AMBER4 forcefield. After subsequent focussing calculations of the 61x65x55 point matrix, the final resolution was 0.6 Å per grid point. The charged exterior of the protein was visualized by projecting the grid values onto a Connolly solvent accessibility surface.

**Constraint-free molecular dynamics calculations.** The physical viability of the calculated NMR structures has been investigated by analyzing the evolution of eight structures from the ensemble in an *in vacuo* MD calculation in the absence of experimental constraints (Blackledge *et al.*, 1995b). Parameters and algorithms used were identical to the SII stage. During the calculation, the temperature was maintained at 300 K and the experimental force constraints were reduced to zero over a period of 40(Δ = 500) fs steps after an equilibration period of 5 ps. The molecule was then allowed to evolve over 170 ps of which the last 50 ps were analyzed. The last 50 ps of the eight MD calculations were sampled every 10 ps to give the final ensemble cfMD (40 structures) which were compared to an ensemble rMD, generated by sampling the last 6 ps of each rMD calculation (111 structures). The presence of concerted or related motions in the constraint free dynamics was investigated by calculating the normalized cross-correlation of the dihedral angle variance. This was determined for all  $\phi_i, \phi_j$  as follows :

$$c_{ij} = \frac{\left| \frac{\langle \phi_i \phi_j \rangle}{\langle \phi_i^2 \rangle^{1/2} \langle \phi_j^2 \rangle^{1/2}} \right|}{}$$

with  $\phi_i, \phi_j$  of all residues (156 points for 78 residues) (Ichiye & Karplus, 1991; Philippopoulos *et al.*, 1996). Note that the normalized cross-correlation is independent of the amplitude of the backbone reorientation; for this reason we have also calculated the individual standard deviations for each dihedral angle during restrained and constraint-free MD.

The coordinates of the 37 E\_II structures of Eha c<sub>551</sub> as well as a list of experimental constraints have been deposited in the Brookhaven Protein Databank (accession code 1gks). Those of Avi<sub>5</sub> and Pae<sub>551</sub> were taken from the same databank (pdb codes 1cc5 and 5cyt, respectively).

### Acknowledgements

This work has been supported by the Centre National de la Recherche Scientifique, the Commissariat à l'Energie Atomique and Biosym Technologies, Inc. This is publication no. 349 of the Institut de Biologie Structurale - Jean Pierre Ebel.

### References

Albrand, J.-P., Blackledge, M.J., Pascaud, F., Hollecker, M. & Marion, D. (1995). NMR and restrained molecular dynamics study of the three-dimensional solution structure of toxin FS2, a specific blocker of the L-type calcium channel, isolated from Black Mamba venom. *Biochemistry* **34**, 5923-5937.

Allen, J.P., Feher, G., Yeates, T.O., Komiya, H. & Rees, D.C. (1987). Structure of the reaction center from

*Rhodobacter sphaeroides* R-26: the protein subunits. *Proc. Natl. Acad. Sci. USA* **84**, 6126-6166.

Ambler, R.P., Meyer, T.E. & Kamen, M.D. (1993). Amino acid sequences of cytochrome c<sub>551</sub> from the halophilic purple phototrophic bacteria *Ectothiorhodospira halophila* and *E. halochloris*. *Arch. Biochem. Biophys.* **306**, 83-93.

Axelrod, H., Feher, G. & Allen, J.P. (1994). Crystallization and X-ray structure determination of cytochrome c<sub>2</sub> from *Rhodobacter sphaeroides* in three crystal forms. *Acta Cryst.* **D50**, 596-602.

Banci, L., Bertini, I., Bren, K.L., Gray, H.B., Somponpiscut, P. & Turano, P. (1995). Three-dimensional solution structure of the cyanide adduct of a Met80Ala variant of *Sacharomyces cerevisiae* iso-1-cytochrome c. Identification of ligand-residue interactions in the distal heme cavity. *Biochemistry* **34**, 11385-11398.

Benning, M., Wesenberg, G., Caffrey, M., Bartsch, R., Meyer, T., Cusanovich, M., Rayment, I. & Holden, H. (1991). Molecular structure of cytochrome c<sub>2</sub> isolated from *Rhodobacter capsulatus* determined at 2.5 Å resolution. *J. Mol. Biol.* **220**, 673-685.

Benning, M.M., Meyer, T.E. & Holden, H.M. (1994). X-ray structure of the cytochrome c<sub>2</sub> isolated from *Paracoccus denitrificans* refined to 1.7 Å resolution. *Arch. Biochem. Biophys.* **310**, 460-466.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. & Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690.

Bersch, B., Brutscher, B., Meyer, T.E. & Marion, D. (1995). <sup>1</sup>H and <sup>13</sup>C NMR assignments and structural aspects of a ferrocycytochrome c<sub>551</sub> from the purple phototrophic bacterium *Ectothiorhodospira halophila*. *Eur. J. Biochem.* **227**, 249-260.

Bhatia, G.E. (1981). *Ph.D. thesis* University of California, San Diego.

Blackledge, M.J., Guerlesquin, F. & Marion, D. (1995a). Novel methionine ligand position in *Desulfovibrio vulgaris* cytochrome c-553 : implications for primary sequence alignment of cytochromes c. *Nature Struct. Biol.* **2**, 532-535.

Blackledge, M.J., Medvedeva, S., Poncin, M., Guerlesquin, F., Bruschi, M. & Marion, D. (1995b). Structure and dynamics of ferrocycytochrome c-553 from *Desulfovibrio vulgaris* studied by NMR spectroscopy and restrained molecular dynamics. *J. Mol. Biol.* **245**, 661-681.

Blackledge, M.J., Guerlesquin, F. & Marion, D. (1996). Comparison of low oxido-reduction potential cytochrome c<sub>553</sub> from *Desulfovibrio vulgaris* with the class I cytochrome c family. *Proteins: Struct. Funct. Genet.* **24**, 178-194.

Brooks, B.R., Broccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M.J. (1983). CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comp. Chem.* **4**, 187-217.

Burley, S.K. & Petsko, G.A. (1988). Weakly polar interactions in proteins. *Adv. Protein Chem.* **39**, 125-189.

- Bushnell, G.W., Louie, G.V. & Brayer, G.D. (1990). High resolution three-dimensional structure of horse heart cytochrome c. *J. Mol. Biol.* **214**, 585-595.
- Caffrey, M. (1994). Strategies for the study of cytochrome c structure and function by site-directed mutagenesis. *Biochimie* **72**, 622-630.
- Caffrey, M. & Cusanovich, M. (1994). Site-specific mutagenesis studies of cytochromes c. *Biochim. Biophys. Acta* **1187**, 277-288.
- Cai, M., Bradford, E.G. & Timkovich, R. (1992). Investigation of the solution conformation of cytochrome c-551 from *Pseudomonas stutzeri*. *Biochemistry* **31**, 8603-8612.
- Cai, M. & Timkovich, R. (1994). Solution conformation of cytochrome c-551 from *Pseudomonas stutzeri* ZöBell determined by NMR. *Biophys. J.* **67**, 1207-1215.
- Carter, D.C., Meli, K.A., O'Donnell, S.E., Burgess, B.K., Furey, W.F., Jr., Wang, B.C. & Stout, C.D. (1985). Crystal structure of *Azotobacter* cytochrome c, at 2.5 Å resolution. *J. Mol. Biol.* **184**, 279-295.
- Chen, L., Durley, R.C.E., Mathews, F.S. & Davidson, V.L. (1994a). Structure of an electron transfer complex: methylamine dehydrogenase, amicyanin, and cytochrome c-551. *Science* **264**, 86-90.
- Chen, Z., Koh, M., Van Driessche, G., Van Beeumen, J.J., Bartsch, R.G., Meyer, T.E., Cusanovich, M.A. & Mathews, F.S. (1994b). The structure of flavocytochrome c sulfide dehydrogenase from a purple phototrophic bacterium. *Science* **266**, 430-432.
- Chothia, C. & Lesk, A.M. (1985). Helix movements and the reconstruction of the haem pocket during the evolution of the cytochrome c family. *J. Mol. Biol.* **182**, 151-158.
- Conolly, M.L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**, 709-713.
- Davidson, V.L. & Jones, L.H. (1995). Roles of dipolar effects and local charge in the ionic strength dependence of redox reactions between c-type cytochromes. *Biochemistry* **34**, 1238-1243.
- Detlefsen, D.J., Thanabal, V., Percoraro, V.L. & Wagner, G. (1991). Solution structure of Fe(II) cytochrome c-551 from *Pseudomonas aeruginosa* as determined by two-dimensional <sup>1</sup>H NMR. *Biochemistry* **30**, 9040-9046.
- Frazaõ, C., Soares, C.M., Carrondo, M.A., Pohl, E., Dauter, Z., Wilson, K.S., Hervás, M., Navarro, J.A., De le Rosa, M.A. & Sheldrick, G.M. (1995). *Ab initio* determination of the crystal structure of cytochrome c6 and comparison with plastocyanin. *Structure* **3**, 1159-1169.
- Fülöp, V., Moir, J.W.B., Ferguson, S.J. & Hajdu, J. (1995a). The anatomy of a bifunctional enzyme: structural basis for reduction of oxygen to water and synthesis of nitric oxide by cytochrome *cd*<sub>1</sub>. *Cell* **81**, 369-377.
- Fülöp, V., Ridout, C.J., Greenwood, C. & Hajdu, J. (1995b). Crystal structure of the di-haem cytochrome c peroxidase from *Pseudomonas aeruginosa*. *Structure* **3**, 1225-1233.
- Giammona, A. (1984). *PhD thesis* Davis, University of California,
- Gilson, M.K., Sharp, K.A. & Honig, B.H. (1987). Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comp. Chem.* **9**, 327-335.
- Ichiye, T. & Karplus, M. (1991). Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Struct. Funct. Genet.* **11**, 205-217.
- Kerfeld, C.A., Anwar, H.P., Interrante, R., Merchant, S. & Yeates, T.O. (1995). The structure of chloroplast cytochrome c6 at 1.9 Å resolution : evidence for functional oligomerization. *J. Mol. Biol.* **250**, 627-647.
- Korszun, Z.R. & Salemme, F.R. (1977). Structure of cytochrome c-555 of *Clorobium thiosulfatophilum* Primitivo low potential cytochrome c. *Proc. Natl. Acad. Sci. USA* **74**, 5244-5247.
- Lanyi, J.K. (1971). Salt-dependent properties from extremely halophilic bacteria. *Bacteriol. Rev.* **38**, 272-290.
- Louie, G.V. & Brayer, G.D. (1990). High-resolution refinement of yeast iso-1-cytochrome c and comparisons with other eukaryotic cytochromes c. *J. Mol. Biol.* **214**, 527-555.
- Ludwig, M.L., Patridge, K.A., Powers, T.B., Dickerson, R.E. & Takano, T. (1982). Structure analysis of a ferricytochrome c from cyanobacterium, *Anacystis nidulans*. In *Electron transport and oxygen utilization, Vol 1* (Chien Ho, ed.), pp 27-32, Elsevier - North Holland Inc., Amsterdam.
- Mathews, F.S., Chen, Z., Bellamy, H.D. & McIntire, W.S. (1991). Three-dimensional structure of p-cresol methylhydroxylase (flavocytochrome c) from *Pseudomonas putida* at 3.0-Å resolution. *Biochemistry* **30**, 238-247.
- Matsuura, Y., Takano, T. & Dickerson, R.E. (1982). Structure of cytochrome c-551 from *Ps. aeruginosa* refined at 1.6 Å resolution and comparison of the two redox forms. *J. Mol. Biol.* **156**, 389-409.
- Mauk, A.G. (1991). Electron transfer in genetically engineered proteins. The cytochrome c paradigm. *Struct. Bonding* **75**, 131-157.
- Meyer, T.E. & Kamen, M.D. (1982). New perspectives on c-type cytochromes. *Adv. Protein Chem.* **35**, 105-212.
- Meyer, T.E. (1985). Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium *Ectothiorhodospira halophila*. *Biochim. Biophys. Acta* **806**, 175-183.
- Meyer, T.E., Cusanovich, M.A. & Kamen, M.D. (1986). Evidence against use of bacterial amino acid sequence data for construction of all-inclusive phylogenetic trees. *Proc. Natl. Acad. Sci. USA* **83**, 217-220.
- Meyer, T.E., Bartsch, R.G., Cusanovich, M.A. & Tollin, G. (1993a). Kinetics of the photooxidation of soluble cytochromes, HiPIP, and azurin by the photosynthetic reaction center of the purple phototrophic bacterium *Rhodospseudomonas viridis*. *Biochemistry* **32**, 4719-4726.
- Meyer, T.E., Zhao, Z.G., Cusanovich, M.A. & Tollin, G. (1993b). Transient kinetics of electron transfer from a variety of c-type cytochromes to plastocyanin. *Biochemistry* **32**, 4552-4559.

- Moore, G.R., Pettigrew, G.W. & Rogers, N.K. (1986). Factors influencing redox potentials of electron transfer proteins. *Proc. Natl. Acad. Sci. USA* **83**, 4998-4999.
- Moore, G.R. & Pettigrew, G.W. (1990). *Cytochromes<sub>c</sub> - evolutionary, structural and physicochemical aspects* Springer-Verlag, Berlin Heidelberg.
- Murphy, M.E.P., Nall, B.T. & Brayer, G.D. (1992). Structure determination and analysis of yeast iso-2-cytochrome<sub>c</sub> and a composite mutant protein. *J. Mol. Biol.* **227**, 160-176.
- Nakagawa, A., Higuchi, Y., Yasuoka, K., Katsube, Y. & Yagi, T. (1990). S-class cytochromes<sub>c</sub> have a variety of folding patterns. Structure of cytochrome<sub>c</sub>-553 from *Desulfovibrio vulgaris* determined by the multi-wavelength anomalous dispersion method. *J. Biochem.* **108**, 701-703.
- Ochi, H., Hata, Y., Tanaka, N., Kakudo, M., Sakuri, T., Aihara, S. & Morita, Y. (1983). Structure of rice ferricytochrome c at 2.0 Å resolution. *J. Mol. Biol.* **116**, 407-418.
- Pelletier, H. & Kraut, J. (1992). Crystal structure of a complex between electron transfer partners. *Science* **258**, 1748-1755.
- Pettigrew, G.W. & Moore, G.R. (1987). *Cytochromes<sub>c</sub> - biological aspects* Springer-Verlag, Berlin Heidelberg.
- Philippopoulos, M., Xiang, Y. & Lim, C. (1996). Identifying the mechanism of protein loop closure: a molecular dynamics simulation of the *Bacillus stearothermophilus* LDH loop in solution. *Prot. Engineering* **8**, 565-573.
- Qi, P.X., Di Stefano, D. & Wand, A.J. (1994). Solution structure of horse heart ferrocyanochrome<sub>c</sub> determined by high resolution NMR and restrained simulating annealing. *Biochemistry* **33**, 6408-6417.
- Reid, K.S.C., Lindley, P.F. & Thornton, J.M. (1985). Sulphur aromatic interactions in proteins. *FEBS Lett.* **190**, 209-213.
- Salemme, F.R., Freer, S.T., Xuong, Ng.H., Alden, R.A. & Kraut, J. (1973a). The structure of oxidized cytochrome c<sub>2</sub> of *Rhodospirillum rubrum*. *J. Biol. Chem.* **248**, 3910-3921.
- Salemme, F.R., Kraut, J. & Kamen, M.D. (1973b). Structural bases for function in cytochrome c. *J. Biol. Chem.* **248**, 7701-7716.
- Senn, H. & Wüthrich, K. (1985). Amino acid sequence, haem-iron co-ordination geometry and functional properties of mitochondrial and bacterial c-type cytochromes. *Q. Rev. Biophys.* **18**, 111-134.
- Singh, U.C. & Kollman, P.A. (1983). An approach to computing electrostatic charges for molecules. *J. Comp. Chem.* **5**, 129-145.
- Sogabe, S. & Miki, K. (1995). Refined crystal structure of ferrocyanochrome c<sub>2</sub> from *Rhodospseudomonas viridis* at 1.6 Å resolution. *J. Mol. Biol.* **252**, 235-247.
- Stackebrandt, E., Fowler, V.J., Schubert, W. & Imhoff, J.F. (1984). Towards a phylogeny of phototrophic purple sulfur bacteria - the genus *Ectothiorhodospira*. *Arch. Microbiol.* **137**, 366-370.
- Takano, T. & Dickerson, R.E. (1981a). Conformation change of cytochrome<sub>c</sub> ferrocyanochrome<sub>c</sub> refinement at 1.8Å and comparison with the ferrocyanochrome structure. *J. Mol. Biol.* **153**, 95-115.
- Takano, T. & Dickerson, R.E. (1981b). Conformation change of cytochrome<sub>c</sub> ferrocyanochrome<sub>c</sub> refined at 1.5 Å resolution. *J. Mol. Biol.* **153**, 79-94.
- Tiede, D.M., Vashishta, A.-C. & Gunner, M.R. (1993). Electron-transfer kinetics and electrostatic properties of the *Rhodobacter sphaeroides* reaction center and soluble c-cytochromes. *Biochemistry* **32**, 4515-4531.
- Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984). A new force field for molecular mechanical representation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765-784.
- Woese, C.R. (1987). Bacterial evolution. *Microbiol. Rev.* **51**, 221-271.
- Wüthrich, K. (1986). *NMR of proteins and nucleic acids* John Wiley, New York.

### 2.2.3 Etude structurale du module de type EGF de la protéase C1r du complément humain.

Collaboration avec Gérard Arlaud et Nicole Thielens au Laboratoire d'Enzymologie Moléculaire à l'IBS.

L'activation du système du complément humain, un système de défense non-spécifique, est déclenchée par la composante C1, un complexe multimoléculaire. Ce complexe est formé par deux entités différentes□la composante C1q, protéine de reconnaissance et l'unité catalytique C1s-C1r-C1r-C1s, un tétramère calcium-dépendant (34, 35 pour des revues récentes). Au sein de ce tétramère, les deux protéases C1r et C1s sont respectivement responsables de l'auto-activation et de l'activité catalytique. C1r et C1s sont des protéines modulaires avec une composition homologue□CUB-EGF-CUB-CCP-CCP-protéase à sérine analogue à la chymotrypsine (CUB□ C1r/C1s, uEGF, bone morphogenetic protein module, EGF□ epidermal growth factor-like module, CCP complement control protein module). Dans le but d'une meilleure compréhension de la relation structure/fonction de ce complexe, Gérard Arlaud et son équipe (Laboratoire d'Enzymologie Moléculaire à l'IBS) ont entrepris une stratégie de dissection moléculaire afin de caractériser la structure des composantes individuelles. C'est dans ce contexte que j'ai étudié la structure tridimensionnelle du domaine EGF de la protéase C1r par RMN (36, Article 3). Par la suite, j'ai continué cette collaboration en essayant de produire des modules CUB ou des paires de modules CUB-EGF de C1r, C1s et de deux protéases analogues, MASP-1 et MASP-2 (MASP□ manan-binding lectin associated protease), en vue de leur marquage isotopique pour l'étude structurale par RMN. Sabine Béraud, étudiante d'abord en DEA (DEA National de Cristallographie et RMN Biologiques), ensuite en thèse au LRMN sous ma direction (codirection avec Dr. D. Marion), a également été impliquée dans cette thématique. Malheureusement, tous nos efforts n'ont pas permis d'obtenir des quantités suffisantes de protéine pour envisager une étude par RMN.

Depuis, les structures de multiples domaines impliqués dans ce complexe ont été résolues par cristallographie depuis□ le domaine catalytique de C1s avec le deuxième module CCP (37), le domaine catalytique de C1r dans une forme proenzyme (38,39) et active (39), et finalement le domaine d'interaction CUB-EGF de C1s (40). L'ensemble de ces résultats a permis de mieux comprendre le mécanisme d'auto-activation de C1r (39) et de proposer un modèle pour l'assemblage de C1q avec l'unité catalytique (40).



## Solution structure of the Epidermal Growth Factor-like Module of Human Complement Protease C1r, an Atypical Member of the EGF Family<sup>†</sup>

Beate Bersch<sup>\*‡</sup>, Jean-François Hernandez<sup>§</sup>, Dominique Marion<sup>‡</sup>, and Gérard J. Arlaud<sup>§</sup>

<sup>‡</sup>Laboratoire de Résonance Magnétique Nucléaire and <sup>§</sup>Laboratoire d'Enzymologie Moléculaire, Institut de Biologie Structurale Jean-Pierre Ebel, CNRS-CEA, 41 avenue des Martyrs, 38027 Grenoble, Cedex 1, France.

The calcium-dependent interaction between C1r and C1s, the two homologous serine proteases of the first component of human complement C1, is mediated by their N-terminal regions. The latter comprise an epidermal growth factor (EGF)-like module exhibiting the consensus sequence characteristic of Ca<sup>2+</sup>-binding EGF modules, surrounded by two CUB modules. Due to its Ca<sup>2+</sup> binding ability, the C1r EGF-like module (C1r-EGF) is supposed to participate in the C1r-C1s interaction. An additional interesting feature of C1r-EGF is the unusually large loop connecting the first two conserved cysteine residues. The solution structure of synthetic C1r-EGF (residues 123-175) has been determined using nuclear magnetic resonance and combined simulated annealing-restrained molecular dynamics calculations. The resulting family of 19 structures is characterized by a well ordered C-terminal part (residues Cys144 to Ala174) with a backbone r.m.s.d. of 0.7 Å, and a disordered N-terminal, including the large loop between the first two cysteines (Cys129 and Cys144). This loop is known to be surface exposed and may be expected to participate in domain-domain or protein-protein interactions. In its C-terminal part, C1r-EGF possesses the characteristic EGF fold with a major and a minor  $\beta$ -sheet. The latter comprises a  $\beta$ -bulge, and comparison with other EGF-like modules reveals the existence of two distinct structural and sequential motifs in the bulged part. Additional experiments in the presence of 80 mM Ca<sup>2+</sup> did not show significant structural variation of C1r-EGF, in keeping with previous observations on blood-clotting factors IX and X.

Human C1 is the multimolecular protease that triggers the classical pathway of complement, a system that participates in innate immunity against various bacteria, parasites, and retroviruses (1). Its activation and enzymic activity are mediated by two serine proteases, respectively C1r and C1s<sup>1</sup>, which share the same type of modular organization, each comprising two CUB modules surrounding a single EGF-like module, a pair of

complement control protein (CCP) modules, and a serine protease domain [see reviews by Schumaker et al. (2), and Volanakis and Arlaud (3)]. A fundamental feature of C1r and C1s is that they exert their catalytic activities within a Ca<sup>2+</sup>-dependent tetramer C1s-C1r-C1r-C1s, which itself associates to a third, non-enzymic protein C1q to form the C1 complex (4). The structural determinants responsible for the Ca<sup>2+</sup>-dependent C1r-C1s interactions involved in the assembly of the C1s-C1r-C1r-C1s tetramer are located in the N-terminal  $\beta$  regions of each protein (5). The corresponding fragments (C1r $\beta$ , C1s $\beta$ ) obtained by limited proteolysis, comprising the first CUB module, the EGF-like module, and the N-terminal disulfide loop of the second CUB module, each contain one high-affinity Ca<sup>2+</sup>-binding site and retain the ability to mediate Ca<sup>2+</sup>-dependent protein-protein interaction (6). Association of the C1s-C1r-C1r-C1s tetramer with C1q also involves the C1r $\beta$  and C1s $\beta$  regions (7,8), which therefore represent key elements of the architecture of the C1 complex. In addition to this structural role, the  $\beta$  region of C1r has been proposed to participate in the control of the autoactivation of the serine protease domain of the molecule (7).

In addition to the usual EGF consensus sequence containing six cysteines forming three disulfide bonds, the EGF-like modules of C1r and C1s exhibit the particular consensus pattern Asp/Asn, Asp/Asn, Asp\*/Asn\*, Tyr/Phe (where \* indicates a  $\beta$ -hydroxylated residue) that is characteristic of a subset of EGF modules shown to be

\*To whom correspondence should be addressed.

<sup>1</sup>Abbreviations: CUB, Complement subcomponents C1r/C1s, Uegf, Bone morphogenetic protein-1; EGF, epidermal growth factor, cbEGF, Ca<sup>2+</sup> binding epidermal growth factor; DQF-COSY, double quantum filtered correlation spectroscopy; TOCSY, total correlation spectroscopy; HSQC, <sup>1</sup>H-detected heteronuclear single quantum coherence; NOESY, nuclear Overhauser spectroscopy; SA, simulated annealing; rMD, restrained molecular dynamics; r.m.s.d., root mean square deviation; t-PA, tissue type plasminogen activator; FacIX\_1, FacIX\_2, FacX\_1, FacX\_2, first and second EGF-like modules of coagulation factors IX and X; hEGF, human epidermal growth factor peptide hormone; TGF- $\beta$ , transforming growth factor- $\beta$ ; Her- $\beta$ , heregulin- $\beta$ ; Fib\_32, Fib\_33, 32nd and 33rd EGF-like module from fibrillin-1; MASP, manan binding protein-associated serine protease. The nomenclature of complement proteins is that recommended by the World Health Organization. The nomenclature of protein modules is that defined by Bork P., and Bairoch, A. (1995) *Trends Biochem. Sci.* 20 (Suppl. March), C03.

involved in  $\text{Ca}^{2+}$ -binding (9). The hydroxylated residue is an asparagine in both C1r (Asn150) and C1s (Asn134), hydroxylation being complete in C1r and only partial in C1s (10,11). Given its  $\text{Ca}^{2+}$ -binding characteristics, it was initially assumed that the EGF module of C1r (C1r-EGF) was responsible for the observed  $\text{Ca}^{2+}$ -binding and  $\text{Ca}^{2+}$ -dependent protein-protein interaction properties of the protein. However, deletion of the first CUB module from C1r abolishes its  $\text{Ca}^{2+}$ -dependent interaction properties (12). In the same way, the isolated C1r-EGF binds  $\text{Ca}^{2+}$  with a  $K_d$  of 10 mM, that is about 300 times weaker than that measured for fragment C1r $\square$  (13), indicating the involvement of residues located outside the EGF module.

A very unusual feature of C1r-EGF is the large size of the loop between the two first cysteines, which comprises 14 residues, as opposed to 2-7 residues in other EGF-like modules (14). This highly charged loop also contains the single polymorphic site (Ser135/Leu) identified in human C1r (15,16).

Like other protein modules, EGF modules are found in a wide variety of mostly extracellular proteins, where they occur in different settings. Determination of different high-resolution structures for a specific module-type will hopefully lead to a better identification of its characteristic structural and functional properties. Only very few structural information has been obtained so far on modules from proteins of the complement system. The particular localization of C1r-EGF between two CUB modules, its unusually large loop as well as its  $\text{Ca}^{2+}$  binding properties make it an interesting subject for structural studies. With this aim the 53-residue EGF module of human C1r was synthesized chemically and its three-dimensional structure was determined by NMR spectroscopy. Our study shows that C1r-EGF has a fold similar to that of other EGF-like modules, with the exception of the loop between the first two cysteines which is disordered. Structural comparison with other EGF-like modules leads us to propose a classification of EGF-like modules with regard to structural and sequential characteristics of a  $\square$ -bulge situated in the minor  $\square$ -sheet.

## METHODS

### Sample Preparation

The C1r EGF-like module (residues 123-175) was synthesized chemically by the solid-phase method on an Applied Biosystem 430A synthesizer using the Boc (*tert*-butyloxycarbonyl) chemistry and standard methodology as described elsewhere (13). A serine residue was placed at position 135, where a serine/leucine polymorphism occurs in native C1r, and an asparagine residue was placed at position 150, instead of the  $\square$ -hydroxyasparagine normally found at this position (10,15-17). The latter post-translational modification is not critical for  $\text{Ca}^{2+}$  binding as deduced from studies on recombinant human C1s (18) and blood clotting factors IX and X (19). The peptide was efficiently folded using reduced/oxidized glutathione, yielding a single major species (13). The correct formation of the three disulfide bridges (1-3, 2-4, 5-6) characteristic of EGF-like modules was verified by using mass spectrometry and N-terminal sequence analyses of thermolytic fragments (13).

For the NMR experiments, approximately 5.4 mg of the lyophilized peptide were dissolved in 500  $\square$ l of 95%  $\text{H}_2\text{O}$ /5%  $\text{D}_2\text{O}$  to a final concentration of about 1.8 mM. The pH was adjusted to 6.7 by subsequent addition of small amounts of 0.1 M NaOH. For experiments in  $\text{D}_2\text{O}$ , the peptide was lyophilized and redissolved in  $\text{D}_2\text{O}$ . Additional experiments were recorded in the presence of 80 mM  $\text{CaCl}_2$ , corresponding to a 90 % saturation of the  $\text{Ca}^{2+}$  binding site, as determined previously by 1-D NMR (13).

### NMR Experiments and Experimental Constraints

All NMR experiments were performed on a Bruker AMX-600 spectrometer operating at  $^1\text{H}$  = 600 MHz, equipped with a standard 5-mm Bruker probe with an internal pulsed  $B_0$  field gradient (PFG) coil. Data were processed using the Bruker XWIN-NMR software (version 1.1.1.).

The sequential  $^1\text{H}$  assignment was performed using the following experiments: DQF-COSY (20); clean TOCSY (21,22); NOESY (23) and  $^{13}\text{C}$ -HSQC at natural abundance (24). The latter was used for unambiguous assignment of side-chain protons using two complementary sets of information: methylene protons were correlated pairwise and characteristic  $^{13}\text{C}$  chemical shift ranges were used for  $^{13}\text{C}$  assignment (24). All spectra were recorded in the phase-sensitive mode [States-TPPI, (25)]. Suppression of the water resonance was achieved either by a weak pre-irradiation during the relaxation delay, or by applying the WATERGATE pulse sequence in the case of the NOESY experiments (26). Mixing times for TOCSY and NOESY spectra were 90 ms and 200 ms, respectively. Spectra were recorded at 15°C (NOESY), 20°C (TOCSY, NOESY and  $^{13}\text{C}$ -HSQC) and 30°C (DQF-COSY, TOCSY).

Distance constraints for the structure calculations were obtained from the NOESY experiments acquired in the absence of  $\text{Ca}^{2+}$  at 15°C and 20°C with 1536 and 800 complex points and spectral widths of 8475 and 6495 Hz in the  $t_2$  and  $t_1$  dimensions, respectively. An identical NOESY experiment was performed in  $\text{D}_2\text{O}$  at 15°C in order to discriminate between amide and aromatic resonances. Cross-peaks were assigned by hand and classified in four categories corresponding to upper distance bounds of 2.6, 3.4, 4.5 and 6.0 Å. Pseudo-atom corrections were added for degenerate methylene, methyl and aromatic protons as proposed by Wüthrich (27). In addition to the distance constraints, seven  $\square_1$  dihedral constraints were added on the basis of the relative  $\text{NH}/\text{C}^\square\text{H}_{1,2}$  NOESY peak intensities and the  $\text{C}^\square\text{H}/\text{C}^\square\text{H}_{1,2}$  cross-peaks in the COSY experiment. No hydrogen bonding distance constraint was used in the structure calculations.

### Structure Calculations

Structure calculations were performed using DISCOVER (Version 2.9.7.) interfaced to the INSIGHT II program (Version 95.0) for structure visualization and result analysis (Biosym Technologies Inc.). A three-stage simulated annealing-restrained molecular dynamics (SA-rMD) protocol was employed, as described in detail elsewhere (28). The AMBER4 force field was used in all calculations except the SA protocols, in which a simple quartic non-bond term was employed. All  $\square$  dihedrals were

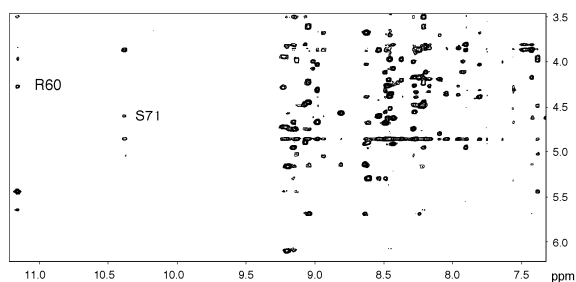


Figure 1: Fingerprint region of the NOESY spectrum. This spectrum was recorded at 15°C in the absence of  $\text{Ca}^{2+}$  with a sample concentration of 1.8 mM. The  $\text{H}^{\text{N}}\text{-C}^{\text{H}}$  cross-peaks of Arg160 and Ser171 are indicated. Note the very crowded region about an  $\text{H}^{\text{N}}$  frequency of 8.4 ppm (see text).

forced to *trans* since all three prolines showed the characteristic strong  $\text{C}^{\text{H}}_2(i)\text{-C}^{\text{H}}(i-1)$  NOESY cross-peaks. The three disulfide bonds, whose localization has been determined independently (13), were introduced as covalent constraints.

First, the global fold of the peptide was determined using a 60 ps SA protocol starting from randomized coordinates. The exploratory period of 50 ps was calculated at a nominal temperature of 1000 K. The NOE force constant was scaled to  $50 \text{ kcal mol}^{-1}\text{\AA}^{-2}$  during the first 30 ps. After slow cooling to 100 K, the molecule was minimized in the complete AMBER force field. Resolved, but non-stereospecifically assigned protons or methyl groups were allowed to float between the two prochiral positions.

Structures were then refined using a two-stage SA-rMD protocol (28). During the SA step, the nominal temperature was set to 2000 K and the NOE force constants were progressively scaled up to  $50 \text{ kcal mol}^{-1}\text{\AA}^{-2}$ . After an exploratory period of 2 ps and subsequent smooth cooling, the molecule was subjected to energy minimization in the full AMBER force field.

In the following rMD calculation the use of the full force field and an implicit solvent simulation led to physically more viable structures (28). After equilibration, the temperature was held to 600 K with the experimental force constant set to  $25 \text{ kcal mol}^{-1}\text{\AA}^{-2}$ . The molecule was allowed to evolve for a period of 10 ps, then it was slowly cooled to 300 K where dynamics was continued for 12 ps. Finally the molecule was minimized using a conjugate gradient algorithm.

As the NOESY spectra acquired in the presence of 80 mM  $\text{CaCl}_2$  showed no significantly different cross-peak patterns, additional structure calculations were performed in order to test whether the conformations obtained in the absence of  $\text{Ca}^{2+}$  were compatible with  $\text{Ca}^{2+}$  binding. In these calculations, a  $\text{Ca}^{2+}$  ion was added whose ligands were taken by analogy with the structures of the first EGF module of factor IX (29) and the EGF module pair from fibrillin-1 (30). Five distance constraints with force constants of  $200 \text{ kcal mol}^{-1}\text{\AA}^{-2}$ , allowing the Ca-O distance to lie between 2.2 and 2.6 Å, were added to the original constraints determined in the absence of  $\text{Ca}^{2+}$ , and the three-step SA-rMD protocol was employed as described above.

The final structural ensembles were selected on the basis of an experimental energy term, including contributions from NOE distance,  $\phi_1$  and  $\phi_2$  dihedral as well as chiral

constraints. The mean structures of the ensembles were calculated using INSIGHT II by superposing the N,  $\text{C}^{\alpha}$  and  $\text{C}^{\beta}$  atoms of residues Gln145 to Glu175. R.m.s.d. values were calculated for the specified atoms in comparison with the mean structure. Hydrogen bonds were considered as identified if the distance between hydrogen and heteroatom or between two heteroatoms was less than 2.5 Å or 3.0 Å, respectively, and provided that the hydrogen bond angle was greater than 135°.

For structural comparisons, the following coordinates for EGF modules were taken from the Brookhaven protein data bank: E-selectin, crystal structure, 1esl (31); human tissue type plasminogen activator (t-PA), minimized average NMR structure, 1tpg (32); porcine blood-clotting factor IX (FacIX\_1, FacIX\_2), crystal structure, 1pfx (33); human FacIX\_1,  $\text{Ca}^{2+}$ -loaded, crystal structure, 1edm (29), human des(1-45) factor Xa (FacX\_2), crystal structure, 1hcg (34), heregulin- $\alpha$  (Her- $\alpha$ ), minimized average NMR structure, 1haf (35). The coordinates for the minimized average NMR structures for human EGF peptide hormone [hEGF, (36)], human transforming growth factor- $\beta$  [TGF- $\beta$ , (37)] and a pair of  $\text{Ca}^{2+}$  binding EGF-like modules from human fibrillin-1 [Fib\_32, Fib\_33, (30)] were taken from Pr. I. D. Campbell's home page (<http://nmra.ocms.ox.ac.uk/idc>). In the case of C1r-EGF, the structure that was the closest to the mean structure in its ordered part was chosen. The coordinates of the structural ensemble have been submitted to the Brookhaven protein data bank (accession code 1apq).

For the search of sequential motifs the PROSITE tool was used [<http://expasy.hcuge.ch/sprot/prosite.html> (38)].

## RESULTS

### Assignment of Proton Resonances

Using standard techniques, the sequential assignment of the C1r-EGF module was straightforward except for a single residue, Cys144, whose amide proton was absent from the TOCSY spectra. Interestingly, the same phenomenon was recently observed for the homologous Cys9 of the P-selectin EGF module (39). Figure 1 shows the fingerprint region of the NOESY spectra. Note the significant downfield shift of the Arg160 and Ser171 amide proton resonances. The same characteristic downfield shift has been observed for a glutamic acid residue in the NMR spectra of several EGF modules [factor IX (40), factor X (41), TGF- $\beta$  (42), urokinase-type plasminogen activator (u-PA) (43) and P-selectin (39)]. The amide proton resonances of most of the residues in the N-terminal part of the peptide (Ala123-Gln143) are clustered in a region centered around 8.4 ppm, suggesting its conformational flexibility. No line broadening is observed as it would occur in the case of slow chemical exchange. A table listing the proton chemical shifts is available as supplementary material. Careful analysis of the NOESY spectra recorded in  $\text{D}_2\text{O}$  revealed cross-peaks between the  $\text{C}^{\text{H}}_2$  protons of cysteines linked by a disulfide bond, providing further confirmation of their correct pairing.

### Structure Calculations and Structure Description

543 distance constraints (94 intraresidual, 203 sequential, 66 medium range and 180 long range) as well as 6  $\phi_1$  dihedral constraints were used for the calculation of the

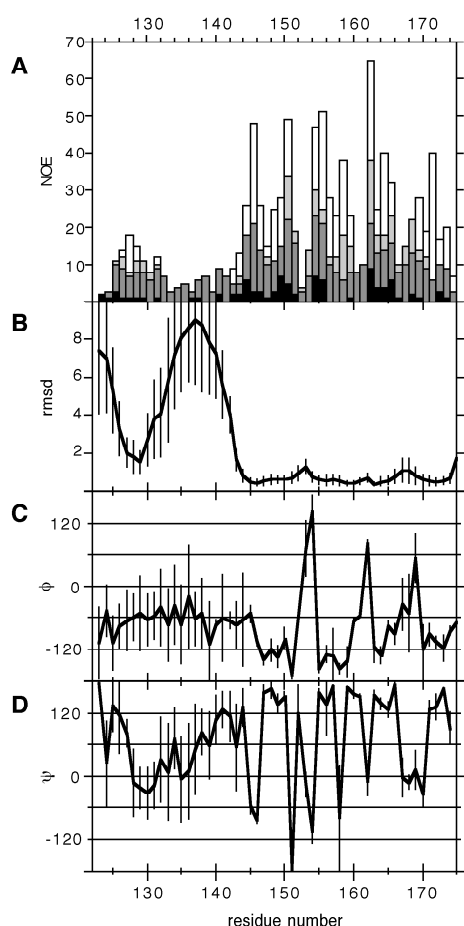


Figure 2: Experimental statistics. (A) Sequential distribution and range of NOE constraints. Black, intraresidual; dark grey, sequential; light grey, medium range ( $i \leq 4$ ); white, long range constraints. (B) The positional backbone r.m.s.d. (N, C $\alpha$ , C $\beta$ ) calculated with respect to the mean structure. (C) Mean values of  $\phi$  angles. (D) Mean values of  $\psi$  angles. The standard deviation is given by error bars.

Ca<sup>2+</sup>-free C1r-EGF module structures. From the sequential distribution of NOESY-derived distances shown in Figure 2(a), it became obvious that the N-terminal part of the molecule, particularly the tip and the segment comprised between residues 134 and 142, lacks a well-defined structure, in agreement with the chemical shift values.

After completion of the three-step SA-rMD calculation, an ensemble of 25 structures was obtained, of which 19 conformations were chosen on the basis of their experimental energy violation. The energetic and geometric statistics of the resulting ensemble are summarized in Table 1 and Figure 2 and the structural ensemble is shown in Figure 3. No structure had NOE violations greater than 0.32 Å. Hydrogen bonds occurring in more than 25% of the final structures are listed in Table S2 (supporting information). The N-terminal part (Ala123 to Gln143) of the molecule is highly disordered, with the exception of residues Asp127-Ala130 which are fixed by the Cys129-Cys148 disulfide bond to the C-terminal part of the module. It is also noteworthy that residues Glu128 to Ser131 show a slight preference for an  $\alpha$ -helical conformation, as can be seen from the  $\phi/\psi$  dihedral means. A similar observation has been made on the corresponding part of human EGF (36).

Table 1. Statistics of C1r-EGF<sup>(a)</sup>

(a) Experimental statistics.

|                               |          |
|-------------------------------|----------|
| No of experimental violations |          |
| > 0.05 Å                      | 26.2±3.5 |
| > 0.10 Å                      | 11.0±2.5 |

(b) Energetic statistics.

|                             |             |
|-----------------------------|-------------|
| energy term <sup>(b)</sup>  |             |
| Bond                        | 8.7±0.6     |
| Angle                       | 73.4±5.4    |
| Dihedral                    | 67.7±7.0    |
| Out of plane                | 2.3±0.5     |
| H bond                      | -26.4±2.1   |
| VDW                         | -146.6±8.2  |
| Electrostatic               | -651.0±19.7 |
| Total                       | -670.7±20.5 |
| experimental <sup>(c)</sup> | 8.9±3.0     |

(c) Structural statistics.

|                             |           |
|-----------------------------|-----------|
| BB <sup>(d)</sup> (145-174) | 0.70±0.16 |
| Heavy (145-174)             | 1.39±0.25 |

<sup>(a)</sup> final ensemble of 19 structures after SA-rMD calculations. <sup>(b)</sup>all values are given in kcal mol<sup>-1</sup>. <sup>(c)</sup>force constants used :  $k_{\text{NOE}} = 25 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ ,  $k_{\phi/\psi} = 60 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ . <sup>(d)</sup>R.m.s.d values (in Å) are the average pairwise r.m.s.d relative to the ensemble mean for the residues and atoms shown; backbone atoms correspond to C<sup>α</sup>, C<sup>β</sup>, N.

Residues Cys144 to Ala174 are structurally well-defined, with a backbone r.m.s.d. of 0.7 Å (Table 1), and exhibit a fold characteristic of the EGF modules, consisting of two double-stranded antiparallel  $\beta$ -sheets. The major sheet is formed by residues Leu147 to Tyr151 and Gly154 to Ser158. It is stabilized by a regular hydrogen bond network as shown in Figure 4. The two strands of this sheet are linked by a type-II  $\beta$ -turn formed by residues Tyr151 to Gly154. The hydrogen bond between Tyr151-O and Gly154-HN is observed in 14 of the 19 structures. A type-II  $\beta$ -turn between residues Arg160 and Tyr163 connects this major  $\beta$ -sheet to the second, minor  $\beta$ -sheet. In that case, the Tyr163-HN-Arg160-O hydrogen bond is found in 5 of the 19 structures. The minor sheet is formed by residues Glu164 to Gln166 and Ser171 to Gln173. A poorly defined tight turn consisting of residues Gln166 to Arg169 is followed by a  $\beta$ -bulge, involving three residues (Arg169, His170 and Ser171). In this particular area, the normal hydrogen bond pattern of a  $\beta$ -sheet is interrupted by a hydrogen bond between Gln166-O and Arg169-H<sup>N</sup>. This bulge can therefore be classified as S3 according to Chan *et al.* (44). A more detailed description of this region is given in the Discussion.

No unique geometry could be determined for the three disulfide bonds, even though  $\phi/\psi$  dihedral constraints were used for Cys148, Cys157, Cys159 and Cys172. According to the  $\phi/\psi$  dihedral, all three disulfides are mixtures of left- and right-handed conformations.

#### Calcium Binding

In order to determine whether Ca<sup>2+</sup> binding induces structural changes, particularly in the disordered N-terminal part of the molecule, a set of NMR spectra was acquired in

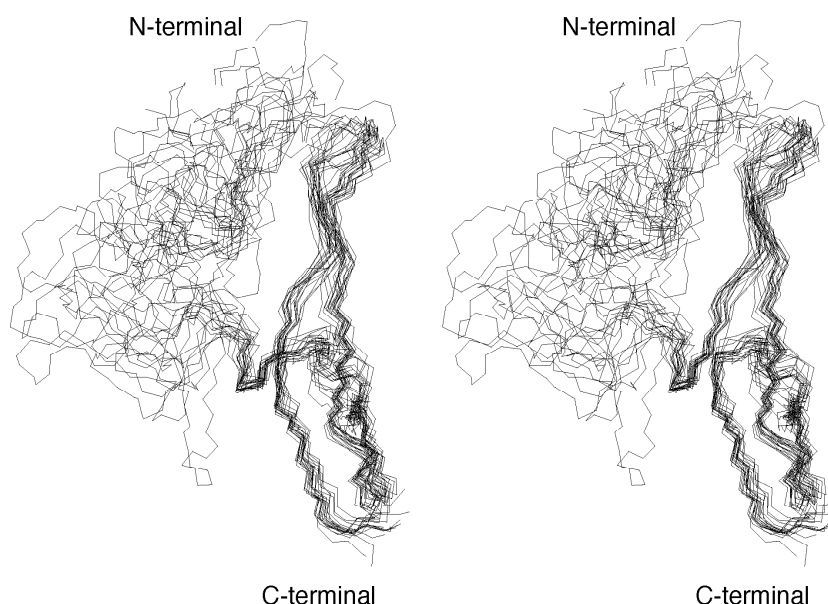


Figure 3: Backbone superposition of 19 C1r-EGF structures in stereo representation. The backbone atoms (N,C $\alpha$ ,C') of residues Cys144 to Ala174 were superposed on the mean structure.

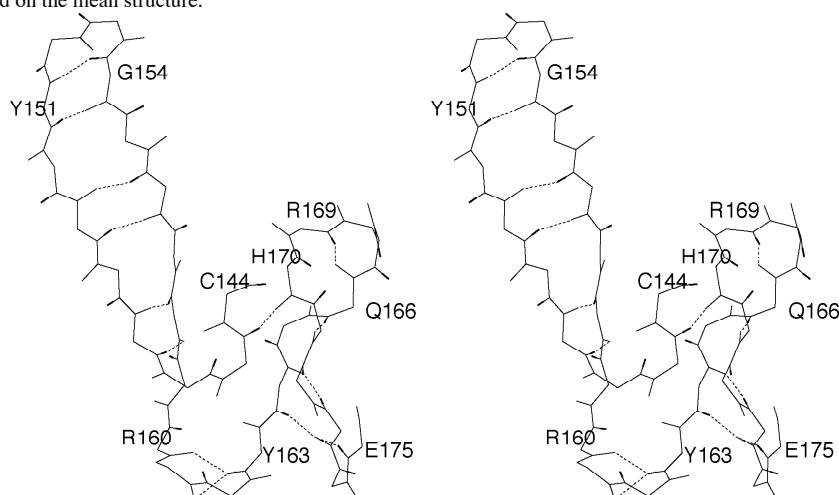


Figure 4: Secondary structure and hydrogen bond network of residues Cys144 to Glu175. Hydrogen bonds are given by dotted lines, the amide proton is represented by a broad tick. Side chains have been omitted for clarity.

in the presence of 80 mM CaCl<sub>2</sub> [i. e. a concentration corresponding to 90% saturation of the binding site (13)], under otherwise unchanged experimental conditions. The complete peak-assignment was performed as in the case of the Ca<sup>2+</sup>-free sample. Chemical shift changes induced by Ca<sup>2+</sup> ligation are summarized in Figure 5 for HN, C<sup>α</sup>H and C<sup>β</sup>H protons. Significant chemical shift changes occur in two regions, comprising residues Asp125 to Ala130 and Tyr151 to Tyr155. Note that the Tyr155-C<sup>β</sup>H chemical shift variation has previously been used to determine the apparent K<sub>d</sub> for Ca<sup>2+</sup> binding by C1r-EGF (13). Both of the regions exhibiting chemical shift changes include or are contiguous to those residues identified as Ca<sup>2+</sup>-binding ligands in other EGF modules (Asp125, Leu126, Glu128, Asn150, and Tyr151 in C1r) (9,29). Therefore, based on sequence comparison and on the chemical shift variations observed upon Ca<sup>2+</sup> binding, it may be postulated that C1r-EGF binds Ca<sup>2+</sup> in a similar way as observed for EGF modules from blood-clotting factors IX (29) and X (45) as well as from human fibrillin-1 (30).

Comparison of the NOESY spectra acquired under identical experimental conditions either in the absence or presence of Ca<sup>2+</sup> did not reveal significant change in the cross-peak patterns. However, the quality of the NOESY spectra decreased with added Ca<sup>2+</sup>, probably because of the dielectric losses in the NMR probe due to increased ionic strength and exchange effects due to incomplete calcium saturation. No additional NOE between the two amino acid stretches involved in Ca<sup>2+</sup> binding were observed. These two regions are already in contact in the *apo* form as deduced from NOE between residues Asp127, Glu128 and Cys129 and the Asn150 side chain and are linked by the Cys129-Cys148 disulfide bond. From these NOE data, we could not deduce significant Ca<sup>2+</sup>-induced structural stabilization of the N-terminal residues involved in Ca<sup>2+</sup> binding, or of the disordered loop between Cys129 and Cys144. This confirms previous observations made on the first EGF module of factor X that Ca<sup>2+</sup> binding does not induce significant structural changes (41,45).

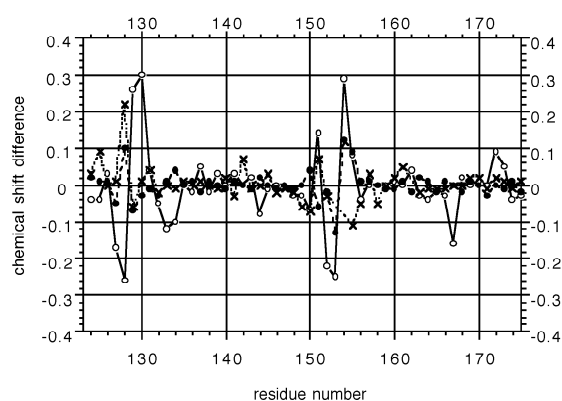


Figure 5: Chemical shift variation induced by  $\text{Ca}^{2+}$  binding. The chemical shift difference ( $\Delta\delta_{\text{Ca}} - \Delta\delta_{\text{apo}}$ ) is given in ppm.  $\text{H}^{\text{N}}$ , open circles, continuous line;  $\text{C}^{\text{D}}\text{H}$ , closed circles, broken line;  $\text{C}^{\text{D}}\text{H}_2$ , crosses and dotted line. In the case of two different values ( $\text{C}^{\text{D}}\text{H}_2$ ,  $\text{C}^{\text{D}}\text{H}_2$ ) the most important one has been taken.

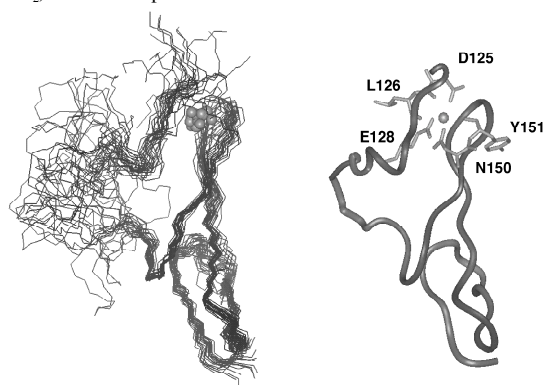


Figure 6: Model of the  $\text{Ca}^{2+}$ -bound form of C1r-EGF. (A) Backbone representation of the 23 selected structures shown in the same orientation as in Figure 3. The backbone atoms of residues Cys144 to Ala174 were superposed on the mean structure. (B) Ribbon representation of the mean structure. The heavy atoms of the  $\text{Ca}^{2+}$ -binding residues are shown for one structure.  $\text{Ca}^{2+}$  is represented by a sphere.

This prompted us to test whether our experimental data set obtained in the absence of  $\text{Ca}^{2+}$  was compatible with  $\text{Ca}^{2+}$  binding by analogy with the crystal structure of the  $\text{Ca}^{2+}$ -bound first EGF module of human factor IX (29). Five additional distance constraints (from  $\text{Ca}^{2+}$  to Asp125- $\text{O}^{\ominus}$ , Glu128- $\text{O}^{\ominus}$ , Asn150- $\text{O}^{\ominus}$ , Leu126-O and Tyr151-O) were added according to this structure (29). A sixth ligand is probably provided by the Glu128 side chain, as factor IX Asp64 is replaced by Asn150 in C1r-EGF [(9), see also the sequence alignment presented in Figure 7]. Addition of  $\text{Ca}^{2+}$ -distance constraints did not lead to systematic violations of NOE distance constraints and both the experimental and physical energy terms were found almost identical to those obtained in the absence of  $\text{Ca}^{2+}$  (23.5 kcal mol $^{-1}$  and -701 kcal mol $^{-1}$ , respectively). A final ensemble of 26 conformations selected on the basis of their experimental energy was obtained after the three-stage SA/rMD protocol, of which three were further excluded because the orientation of the large loop was in contradiction with the experimental constraints. Figure 6(a) shows the resulting 23 structures in the same orientation as the *apo* form (see Figure 3), and Figure 6(b) shows the  $\text{Ca}^{2+}$

Table 2. Comparison of atomic coordinates for different EGF-like modules<sup>(a)</sup>

|              | C1r-EGF (148-151, 154-163) <sup>(b)</sup> | hEGF (20-23, 28-37) <sup>(b)</sup> | TGF- $\beta$ (21-24, 29-38) <sup>(b)</sup> |
|--------------|---|------------------------------------|--|
| C1r-EGF      | -   | 1.21                               | 1.62                                       |
| FacIX_2      | 0.76                                      | 1.16                               | 1.68                                       |
| FaxX_2       | 0.85                                      | 0.95                               | 1.53                                       |
| FacIX_1      | 1.71                                      | 0.97                               | 1.29                                       |
| E-sel        | 1.55                                      | 0.79                               | 0.85                                       |
| hEGF         | 1.21                                      | -                                  | 1.27                                       |
| TGF- $\beta$ | 1.62                                      | 1.27                               | -  |
| t-PA         | 1.85                                      | 1.25                               | 1.59                                       |
| Her- $\beta$ | 1.96                                      | 1.02                               | 1.31                                       |

<sup>(a)</sup> r.m.s.d. calculated in  $\text{\AA}$  over the trace atoms of the reduced EGF core as defined in the text; <sup>(b)</sup> residue numbers corresponding to the reduced EGF core.

binding ligands on a ribbon representation. The ordered C-terminal part of the molecule adopts the same conformation as in the *apo*-form as deduced from the r.m.s.d. between the two ensembles (0.42  $\text{\AA}$  for the  $\text{C}^{\text{D}}$ , N and C' atoms of residues 145-174 of the corresponding ensemble means), the distribution of  $\phi\psi$  dihedrals, and the hydrogen bond networks (data not shown). On the other hand, the order in the N-terminal part, particularly from residue Val124 to Ala130, is significantly increased due to the additional  $\text{Ca}^{2+}$  distance constraints. However, this observation is only based on the structure modeling and could not be related to any increase in the number of NOE in the corresponding spectra (see above).

## DISCUSSION

The family of EGF modules is extremely widespread, and probably more than 600 different versions of the EGF module have so far been identified from sequence analysis (14). As sequence comparisons of various EGF modules reveal that the number of amino acids located in between the six conserved cysteines varies consistently (14,46), EGF-like modules may be best characterized by the presence of two double-stranded antiparallel  $\beta$ -sheets, connected by loops of varying lengths.

Figure 7 shows a sequence alignment for selected EGF-like modules with known three-dimensional structure. Besides the six cysteines involved in the three disulfide bonds, there are only very few residues which are conserved among the different EGF-like modules. Indeed, the glycine at position 162 in C1r-EGF seems to be the only strictly conserved residue, this being always followed by an aromatic amino acid. In addition, the sub-class of  $\text{Ca}^{2+}$ -binding EGF-like modules (cbEGF) contains the characteristic  $\text{Ca}^{2+}$ -binding consensus sequence highlighted in Figure 7. Table 2 resumes the structural comparison of different EGF-like modules, for which the trace atoms of the reduced core have been superimposed. The core of the EGF-like modules is defined as the region between the third cysteine and the conserved aromatic amino acid (residues Cys148 to Tyr163 in C1r-EGF) and includes the major  $\beta$ -sheet and the adjacent turn (30). However, in several EGF-like modules residue insertions at the top of the major  $\beta$ -sheet lead to the replacement of the tight turn by an  $\beta$ -loop. Therefore, we introduce a reduced core which excludes the

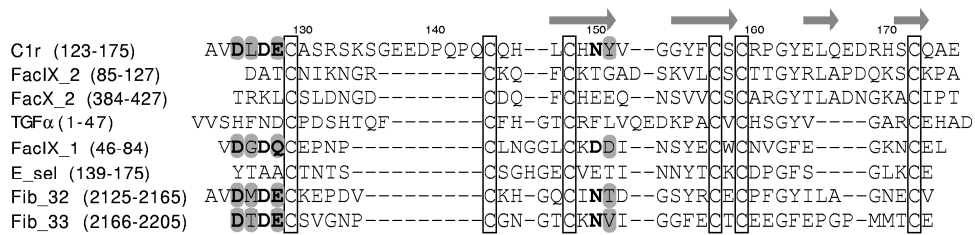


Figure 7: Sequence alignment of selected EGF-like modules. The six conserved cysteines are boxed. Residues belonging to the Ca<sup>2+</sup>-consensus sequence are in bold letters and Ca<sup>2+</sup>-binding residues are shaded. Arrows illustrate the  $\beta$ -sheet regions observed in C1r and the numbering corresponds to the C1r sequence. All sequences are from human species except for the Fac\_IX sequences which are from porcine factor IX.

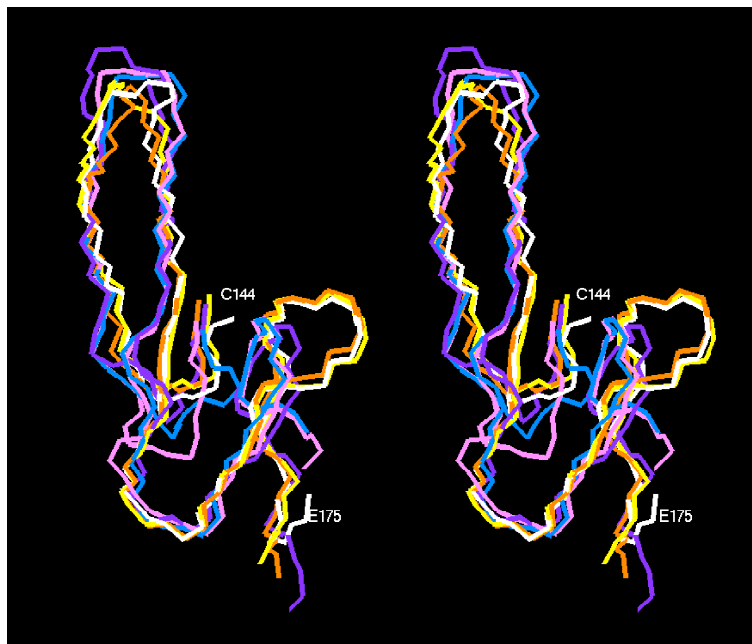


Figure 8: Superposition of the backbone atoms of several EGF-like modules in stereo representation. C1r-EGF is shown in white, the second EGF-like modules of factors IX and X in yellow and orange, respectively, the first EGF-like module of factor IX in pink, E-selectin in blue and TGF- $\beta$  in violet. This representation results from the superposition of the reduced core on the C1r-EGF coordinates, as defined in the text.

central residues of the turn or  $\beta$ -loop. It becomes obvious that the structure of C1r-EGF is very similar to FacIX\_2 and FacX\_2 (r.m.s.d. < 0.9 Å), whereas the r.m.s.d. with the other EGF-like modules varies between 1.2 and 1.9 Å. The three-dimensional structures of some representative EGF modules superposed to that of C1r-EGF are shown in Figure 8, the N-terminal part of the molecules being discarded for clarity. It can be seen that the backbone conformations of C1r-EGF, FaxIX\_2 and FacX\_2 are nearly identical, whereas the folds of TGF- $\beta$ , E-selectin and FacIX\_1 differ from that of C1r-EGF, mainly in the shape and orientation of the minor  $\beta$ -sheet.

Indeed, sequence comparison shows that in these particular cases the two C-terminal cysteines (Cys(5) and Cys(6)) are separated either by 12 residues, as in the case of C1r, or by 8 residues, as for example in the human EGF peptide hormone (see Figure 7). In what follows we will discuss conserved structural motifs for each of these two groups, referred to as C1r and hEGF homologs for differentiation. Other examples of C1r homologs are the second EGF-like modules of factors IX (FacIX\_2) and X (FacX\_2), that do not possess the Ca<sup>2+</sup> binding consensus sequence, whereas human transforming growth factor- $\beta$  (TGF- $\beta$ ), heregulin- $\beta$ , the first

EGF-like module of factors IX (FacIX\_1) and X (FacX\_1), and the EGF-like modules from E- and P-selectin as well as from human tissue-type plasminogen activator (t-PA) are hEGF homologs.

Figure 9 shows a schematic representation of the minor  $\beta$ -sheets of 3 C1r and 3 hEGF homologs. By comparing the individual structures within each group it can be noticed that long-range hydrogen bond networks are well conserved. In both groups, a  $\beta$ -bulge immediately precedes the tight turn. The nomenclature of Richardson et al. (47) has been adopted, in which residues on the bulged side are labeled 1,2,3... whereas the residue on the opposite strand is called X. The bulge in the C1r homologs belongs to the S3 class (44) in which the amide of residue X is linked to the carbonyl of residue 1, and the amide of residue 3 to the residue X carbonyl. In the three C1r homologs, residue 1 is characterized by a positive  $\beta$ -angle (33,34). In most EGF-like modules containing the Cys(5)-Xxx<sub>12</sub>-Cys(6) pattern there are two residues which are highly conserved. A residue with a long and bulky side chain is found at the position before residue X, which most often corresponds to a leucine (Leu165 in C1r-EGF), but can be replaced by methionine, valine, glutamine or isoleucine. As in our

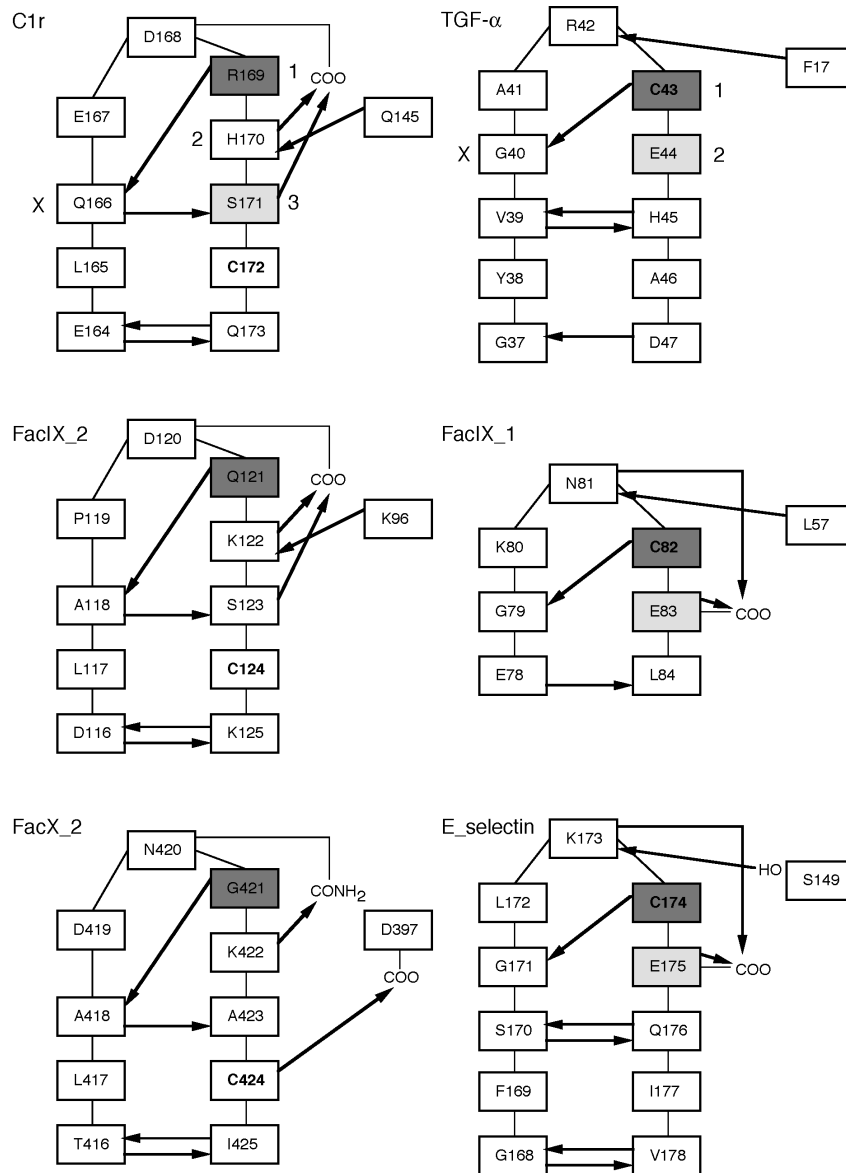


Figure 9: Schematic representation of the minor  $\beta$ -sheet of three C1r (left) and three hEGF homologs (right). Hydrogen bonds are indicated by arrows from the donor to the acceptor. The functional groups of side chains involved in such hydrogen bonds are shown. Residues with a positive  $\beta$  angle are shaded in dark grey and those whose amide proton is characterized by a strong downfield shift are marked in light grey. The nomenclature used in the text is explicated for the two structures at the top.

structures this residue is found to be in van der Waals contact with Cys157, its possible function might be the stabilization of the minor  $\beta$ -sheet. The second conserved residue corresponds to C1r-Asp168, situated in the tight turn. A residue with a polar side chain is conserved in most cases. It may be that the corresponding side chain is involved in hydrogen bonds, as it was found in the case of C1r-EGF, FacIX\_2 and FacX\_2 (see Figure 9). Note that most EGF-like modules that do not possess a bulky residue (L,M,I,V,Q) at position X-1 do not have the second conserved residue in the tight turn. This strongly suggests that these two residues are characteristic of the structural motif found in the C1r homologs.

On the other hand, the bulges of the hEGF homologs do not fall into one of the classes described by Chan et al. (44). In this group, residue X always corresponds to a glycine. It receives a hydrogen bond from residue 1, which corresponds to Cys(6), and has a positive  $\beta$ -angle, as in the C1r homologs. The same hydrogen bond pattern also occurs in

heregulin- $\beta$  and in t-PA. As in the C1r family, two residues are highly conserved in the  $\beta$ -bulge structure of the hEGF homologs. The first one is the above mentioned glycine, the second one being a glutamic acid, located at position 2. As in the C1r homologs, this residue can be replaced by residues with polar side chains (Q,D,N,T,S), able to form hydrogen bonds. Again, a correlation exists between the conservation of the glycine and this polar side chain, indicating their importance for the corresponding structural motif.

An interesting consequence of the occurrence of these two different structural motifs is the position of Cys(6) relative to the  $\beta$ -bulge: in both cases this lies on the bulged strand, but corresponds to residue 1 in the hEGF homologs, whereas it is located immediately after the bulge structure in the C1r homologs. It is clear therefore that there is sequential but no structural homology for this cysteine between these two types of EGF-like modules.



Two other striking features are the conservation of a hydrogen bond between the minor  $\beta$ -sheet and the residue following the second cysteine (see Figure 9), as well as the downfield shift observed for an amide proton in the  $\beta$ -bulge structure. In the hEGF homologs it corresponds to the conserved glutamic acid following Cys(6), (E44 in TGF- $\beta$ ). The chemical shift of its amide proton has been found above 9.6 ppm in several hEGF homologs (36,39-43,48). There is no chemical shift information available on the C1r homologs except C1r, in which the Ser171 amide proton has been detected at 10.4 ppm. A possible reason might be a hydrogen bond with a carboxyl group, as evoked by Freedman et al. (39) in the case of P-selectin. Such hydrogen bonds involving the conserved glutamic acid have been found in at least two hEGF homologs, FacIX\_1 (33) and E-selectin (31), whereas in the solution structures of hEGF, TGF- $\beta$ , heregulin- $\beta$  and u-PA the glutamic acid side chain seems to be unconstrained. In the three C1r homologs with known 3-D structure, analogous hydrogen bonds exist which are formed with the side chain of the conserved polar residue preceding the  $\beta$ -bulge (Asp168 in C1r-EGF, see above).

C1r-EGF contains an unusual loop connecting the first and second cysteines (Cys129 and Cys144). With its 14 amino acids, it is much larger than the corresponding segment of most EGF-like modules identified so far [maximal 7 residues (14)], with known exceptions being the 29th EGF-like module of the notch gene product from *Drosophila melanogaster* (49) and the recently identified tunicate MASP protein from *Halocynthia roretzi* (50). From the structural studies presented here it is clear that this loop does not possess a unique structure, at least in the isolated EGF-like module. The high number of charged residues (two basic and three acidic amino acids) and the absence of hydrophobic residues strongly suggest that this loop is surface-exposed, in agreement with the observation that the arginyl bond at position 134 is cleaved upon limited proteolysis of intact C1r by trypsin (6). A likely hypothesis is that this loop participates *via* electrostatic contacts in domain-domain interactions within C1r, or in protein-protein interactions within the C1 complex. Indeed, various data provide support for the occurrence of a  $\text{Ca}^{2+}$ -dependent interaction between the  $\beta$  region of C1r and its C-terminal catalytic region (7,51). Another possibility is that the extended loop of C1r-EGF participates in the interaction between the C1s-C1r-C1r-C1s tetramer and C1q (7). Whatever the precise interaction role of this loop, its large size and apparent flexibility make it particularly well adapted for such a function.

EGF modules with the consensus sequence for  $\text{Ca}^{2+}$  binding occur in a number of extracellular proteins with diverse functions. In order to understand the structural and functional consequences of  $\text{Ca}^{2+}$  binding, numerous structural investigations have been performed on single modules or module pairs in the presence and absence of  $\text{Ca}^{2+}$  (29,30,33,40,41,45,52-55). In this study, we have shown that the binding of  $\text{Ca}^{2+}$  to the EGF-like module from C1r does not induce major structural modification of its solution structure. This observation is in accordance with previous results obtained on the single cbEGF modules from blood

coagulation factors IX (29,52) and X (41) as well as from human fibrillin (55). From the NMR investigation of  $\text{Ca}^{2+}$ -bound factor X EGF there was some indication that  $\text{Ca}^{2+}$  might stabilize the peptide N-terminal end but the authors evoked the possibility that the lower resolution of the *apo* form might be due to the lack of experimental constraints (41). In our case, there is no experimental evidence for  $\text{Ca}^{2+}$ -induced stabilization of the N-terminal end, but modeling studies with additional  $\text{Ca}^{2+}$ -binding constraints suggest an increased order in this region. Therefore, the present study does not allow to draw a firm conclusion on this question. As C1r-EGF possesses the unusually large loop in its N-terminal part, one other aim of our study was to determine whether  $\text{Ca}^{2+}$  binding would have some influence on the stability of this loop. This hypothesis can be ruled out on the basis of both the experimental data and modeling studies.

Nevertheless, chemical shift changes are observed upon  $\text{Ca}^{2+}$  addition and their profile (see Figure 5) suggests that C1r-EGF binds  $\text{Ca}^{2+}$  in the same way as blood coagulation factors IX and X or fibrillin EGF modules (41,45,52). Interestingly, all studies performed on isolated  $\text{Ca}^{2+}$ -binding EGF modules show  $\text{Ca}^{2+}$  binding constants that are much weaker than those determined on the entire molecule or functional parts of it (13,56-59). It was initially assumed that the reason for this may be the incomplete ligation of the  $\text{Ca}^{2+}$  ion by the EGF-like module on its own, as this provides only six of the ideally seven  $\text{Ca}^{2+}$  ligands (29,41), and that the additional ligand could be provided by the preceding module. Experimental support for this hypothesis came from the crystal structure of the first factor IX cbEGF module in which an asparagine belonging to a second molecule in the crystal lattice participates in  $\text{Ca}^{2+}$  ligation (29). More recently, NMR structures of two module pairs,  $\beta$ -carboxyglutamate (GLA)-cbEGF from factor X(45) and cbEGF-cbEGF (Fib\_32,Fib\_33) from human fibrillin-1 (30) became available. In these studies the donation of an additional  $\text{Ca}^{2+}$  ligand from the N-terminal module could not be confirmed. Indeed, the difference in  $\text{Ca}^{2+}$  binding affinity, also observed on the two  $\text{Ca}^{2+}$  sites of the Fib\_32,Fib\_33 double module (59), seems to reflect the stability of the  $\text{Ca}^{2+}$  binding site (30). The module-module interface would thus contribute to the maintenance of the correct conformation for high-affinity  $\text{Ca}^{2+}$  binding (30). In addition, it may be assumed from these two studies that  $\text{Ca}^{2+}$  intervenes at the module-module interface by inducing a relative reorientation of the module pair as seen in factor X (45) and/or by the stabilization of the module-module contacts. The observation that  $\text{Ca}^{2+}$  binding stabilizes Fibrillin-1 against proteolytic degradation, underlines this hypothesis (60). With respect to C1r, recent experiments on the recombinant N-terminal CUB-cbEGF module pair (Thielens, Enri , Lacroix, Esser and Arlaud, unpublished data) show that this assembly exhibits a high-affinity  $\text{Ca}^{2+}$  binding site and is able to interact with C1s in a  $\text{Ca}^{2+}$ -dependent fashion. It is clear therefore that the N-terminal CUB module contributes in some way to  $\text{Ca}^{2+}$  binding. Additional structural studies on this particular module pair are required to elucidate the mechanism of  $\text{Ca}^{2+}$  binding and its functional role in human C1r. These would also provide

a further model for the study of Ca<sup>2+</sup>-binding by an EGF module within a module pair.

## ACKNOWLEDGMENT

We are very grateful to Dr. M.J. Blackledge for interesting discussions and his help with the structure calculations. We thank F. Hatt for her participation in NOE cross-peak assignment and initial structure calculations.

This work has been supported by the Centre National de la Recherche Scientifique, the Commissariat à l'Energie Atomique (CEA), the European Union Biotechnology Programme (contract N° BIO4-CT96-0662) and Molecular Simulations, Inc. This is publication no. 460 from the Institut de Biologie Structurale - Jean-Pierre Ebel. A preliminary report of this study was presented at the XVIth International Complement Workshop in Boston, MA, June 1996.

## REFERENCES

- Cooper, N.R. (1985) *Adv. Immunol.* 37, 151-216.
- Schumaker, V.N., Zavodszky, P., and Poon, P.H. (1987) *Annu. Rev. Immunol.* 5, 21-42.
- Volanakis, J.E. & Arlaud, G.J. (1997) in *Human Complement System in Health and Disease* (Frank, M. & Volanakis, J.E., Eds.), Marcel-Dekker, New York (in press).
- Arlaud, G.J., Colomb, M.G., and Gagnon, J. (1987) *Immunol. Today* 8, 106-111.
- Villiers, C.L., Arlaud, G.J., and Colomb, M.G. (1985) *Proc. Natl. Acad. Sci. USA* 82, 4477-4481.
- Thielens, N.M., Aude, C.A., Lacroix, M.B., Gagnon, J., and Arlaud, G.J. (1990) *J. Biol. Chem.* 265, 14469-14475.
- Thielens, N.M., Illy, C., Bally, I.M., and Arlaud, G.J. (1994) *Biochem. J.* 301, 509-516.
- Busby, T.F. and Ingham, K.C. (1990) *Biochemistry* 29, 4613-4618.
- Handford, P.A., Mayhew, M., Baron, M., Winship, P.R., Campbell, I.D., and Brownlee, G.G. (1991) *Nature* 351, 164-167.
- Arlaud, G.J., van Dorsselaer, A., Bell, A., Mancini, M., Aude, C., and Gagnon, J. (1987) *FEBS Lett.* 222, 129-134.
- Thielens, N.M., van Dorsselaer, A., Gagnon, J., and Arlaud, G.J. (1990) *Biochemistry* 29, 3570-3578.
- Cseh, S., Gal, P., Sarvari, M., Dobo, J., Lorincz, Z., Schumaker, V.N., and Zavodszky, P. (1996) *Mol. Immunol.* 33, 351-359.
- Hernandez, J.-F., Bersch, B., Pétillet, Y., Gagnon, J., and Arlaud, G.J. (1997) *J. Pept. Res.* 49, 221-231.
- Campbell, I.D. and Bork, P. (1993) *Curr. Opin. Struct. Biol.* 3, 385-392.
- Leytus, S.P., Kurachi, K., Sakariassen, K.S., and Davie, E.W. (1986) *Biochemistry* 25, 4855-4863.
- Journet, A. and Tosi, M. (1986) *Biochem. J.* 240, 783-787.
- Arlaud, G.J. and Gagnon, J. (1987) *Biochem. J.* 241, 711-720.
- Luo, C., Thielens, N.M., Gagnon, J., Gal, P., Sarvari, M., Tseng, Y., Tosi, M., Zavodszky, P., Arlaud, G.J., and Schumaker, V.N. (1992) *Biochemistry* 31, 4254-4262.
- Selander-Sunnerhagen, M., Persson, E., Dahlqvist, I., Drakenberg, T., Stenflo, J., Mayhew, M., Robin, M., Handford, P., Tilley, J.W., Campbell, I.D., and Brownlee, G.G. (1993) *J. Biol. Chem.* 268, 23339-23344.
- Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* 117, 479-485.
- Davis, D.G. and Bax, A. (1985) *J. Am. Chem. Soc.* 107, 2820-2821.
- Griesinger, C., Otting, G., Wüthrich, K., and Ernst, R.R. (1988) *J. Am. Chem. Soc.* 110, 7870-7872.
- Macura, S., Huang, Y., Suter, D., and Ernst, R.R. (1981) *J. Magn. Res.* 43, 259-281.
- Medvedeva, S., Simorre, J.-P., Brutscher, B., Guerlesquin, F., and Marion, D. (1993) *FEBS Lett.* 333, 251-256.
- Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) *J. Magn. Res.* 85, 393-399.
- Piotto, M., Saudek, V., and Sklenár, V. (1992) *J. Biomol. NMR* 2, 661-665.
- Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, John Wiley, New York.
- Blackledge, M.J., Medvedeva, S., Poncin, M., Guerlesquin, F., Bruschi, M., and Marion, D. (1995) *J. Am. Chem. Soc.* 117, 661-681.
- Rao, Z., Handford, P.A., Mayhew, M., Knott, V., Brownlee, G.G., and Stuart, D. (1995) *Cell* 82, 131-141.
- Downing, A.K., Knott, V., Werner, J.M., Cardy, C.M., Campbell, I.D., and Handford, P.A. (1996) *Cell* 85, 597-605.
- Graves, B.J., Crowther, R.L., Chandran, C., Rumberger, J.M., Li, S., Huang, K.-S., Presky, D.H., Familletti, P., Wolitzky, B.A., and Burns, D.K. (1994) *Nature* 367, 532-538.
- Smith, B.O., Downing, A.K., Driscoll, P.C., Dudgeon, T.J., and Campbell, I.D. (1995) *Structure* 3, 823-833.
- Brandstetter, H., Bauer, M., Huber, R., and Lollar, P. (1995) *Proc. Natl. Acad. Sci. USA* 92, 9796-9800.
- Padmanabhan, K., Padmanabhan, K.P., Tullinsky, A., Park, C.H., Bode, W., Huber, R., Blankenship, D.T., Cardin, A.D., and Kisiel, W. (1993) *J. Mol. Biol.* 232, 947-966.
- Jacobsen, N.E., Abadi, N., Siwkowski, M.X., Reilly, D., Skelton, N.J., and Fairbrother, W.J. (1996) *Biochemistry* 35, 3402-3417.
- Hommel, U., Harvey, T.S., Driscoll, P.C., and Campbell, I.D. (1992) *J. Mol. Biol.* 227, 271-282.
- Harvey, T.S., Wilkinson, A.J., Tappin, M.J., Cooke, R.M., and Campbell, I.D. (1991) *Eur. J. Biochem.* 198, 555-562.
- Bairoch, A., Bucher, P., and Hofman, K. (1995) *Nucleic Acids Res.* 24, 189-196.

39. Freedman, S.J., Sanford, D.G., Bachovchin, W.W., Furie, B., and Baleja, J.D. (1996) *Biochemistry* 35, 13733-13744.
40. Baron, M., Norman, D.G., Harvey, T.S., Handford, P.A., Mayhew, M., Tse, A.G.D., Brownlee, G.G., and Campbell, I.D. (1992) *Protein Sci.* 1, 81-90.
41. Selander-Sunnerhagen, M., Ullner, M., Persson, E., Teleman, O., Stenflo, J., and Drakenberg, T. (1992) *J. Biol. Chem.* 267, 19642-19649.
42. Tappin, M.J., Cooke, R.M., Fitton, J.E., and Campbell, I.D. (1989) *Eur. J. Biochem.* 179, 629-637.
43. Hansen, A.P., Petros, A.M., Meadows, R.P., Nettesheim, D.G., Mazar, A.P., Olejniczak, E.T., Xu, R.X., Pederson, T.M., Henkin, J., and Fesik, S.W. (1994) *Biochemistry* 33, 4847-4864.
44. Chan, A.W.E., Hutchinson, E.G., Harris, D., and Thornton, J.M. (1993) *Protein Sci.* 2, 1574-1590.
45. Sunnerhagen, M., Olah, G.A., Stenflo, J., Forsén, S., Drakenberg, T., and Trehwella, J. (1996) *Biochemistry* 35, 11547-11559.
46. Bork, P., Downing, A.K., Kieffer, B., and Campbell, I.D. (1996) *Quart. Rev. Biophys.* 29, 119-167.
47. Richardson, J.S., Getzoff, E.D., and Richardson, D.C. (1978) *Proc. Natl. Acad. Sci. USA* 75, 2574-2578.
48. Smith, B.O., Downing, A.K., Dudgeon, T.J., Cunningham, M., Driscoll, P.C., and Campbell, I.D. (1994) *Biochemistry* 33, 2422-2429.
49. Wharton, K.A., Johansen, K.M., Xu, T., and Artavanis-Tsakonas, S. (1985) *Cell* 43, 567-581.
50. Ji, X., Azumi, K., Sasaki, M., and Nonaka, M. (1997) *Proc. Natl. Acad. Sci. USA* 94, 6340-6345.
51. Pelloux, S., Thielens, N.M., Hudry-Clergeon, G., Pétillet, Y., Filhol, O., and Arlaud, G.J. (1996) *FEBS Lett.* 386, 15-20.
52. Huang, L.H., Cheng, H., Pardi, A., Tam, J.P., and Sweeney, W.V. (1991) *Biochemistry* 30, 7402-7409.
53. Ullner, M., Selander, M., Persson, E., Stenflo, J., and Drakenberg, T. (1992) *Biochemistry* 31, 5974-5983.
54. Sunnerhagen, M., Forsén, S., Hoffrén, A.-M., Drakenberg, T., Teleman, O., and Stenflo, J. (1995) *Nat. Struct. Biol.* 2, 504-509.
55. Wu, Y.-S., Bevilacqua, V.L.H., and Berg, J.M. (1995) *Curr. Biol.* 2, 91-97.
56. Handford, P., Downing, A.K., Rao, Z., Hewett, D.R., Sykes, B.C., and Kielty, C.M. (1995) *J. Biol. Chem.* 270, 6751-6756.
57. Astermak, J., Björk, I., Öhlin, A.K., and Stenflo, J. (1991) *J. Biol. Chem.* 266, 2430-2437.
58. Valcarce, C., Selander-Sunnerhagen, M., Tämlitz, A.M., Drakenberg, T., Björk, I., and Stenflo, J. (1993) *J. Biol. Chem.* 268, 26673-26678.
59. Knott, V., Downing, A.K., Cardy, C.M., and Handford, P. (1996) *J. Mol. Biol.* 255, 22-27.
60. Reinhardt, D.P., Ono, R.N., and Sakai, L.Y. (1997) *J. Biol. Chem.* 272, 1231-1236.

## 2.3 Approche pour des protéines doublement marquées de petite taille □ Introduction

Sans marquage isotopique, l'approche homonucléaire est limitée pratiquement aux protéines de moins que 100 acides aminés. De plus, des études telles que celles présentées précédemment sont très coûteuses en temps □ par exemple, l'étude complète du cytochrome *c551* (attribution, détermination des NOE, calculs de structure et interprétation des données) a nécessité plus d'un an de travail, et ceci malgré une qualité exceptionnelle des spectres.

Depuis les années 1990, la biologie structurale a vécu un changement significatif. Avec le séquençage de génomes complets est né la motivation de déterminer et de caractériser toutes les familles de repliements des protéines présentes dans la nature. Plusieurs initiatives, regroupant différentes équipes de recherches, se sont formées à travers le monde afin d'aborder ce problème à grande envergure d'une manière systématique.

En général, la RMN est très bien adaptée à la détermination de la structure tridimensionnelle des protéines de petite taille, d'autant plus quand elles contiennent des parties flexibles, ce qui empêche souvent leur cristallisation. Cependant, pour être attractif aussi bien dans le cadre d'un projet de génomique structural que dans le souhait de pouvoir fournir rapidement des informations structurales en complément d'une caractérisation biochimique d'une protéine, le processus entier de l'attribution au calcul de structure doit être optimisé et rendu plus efficace.

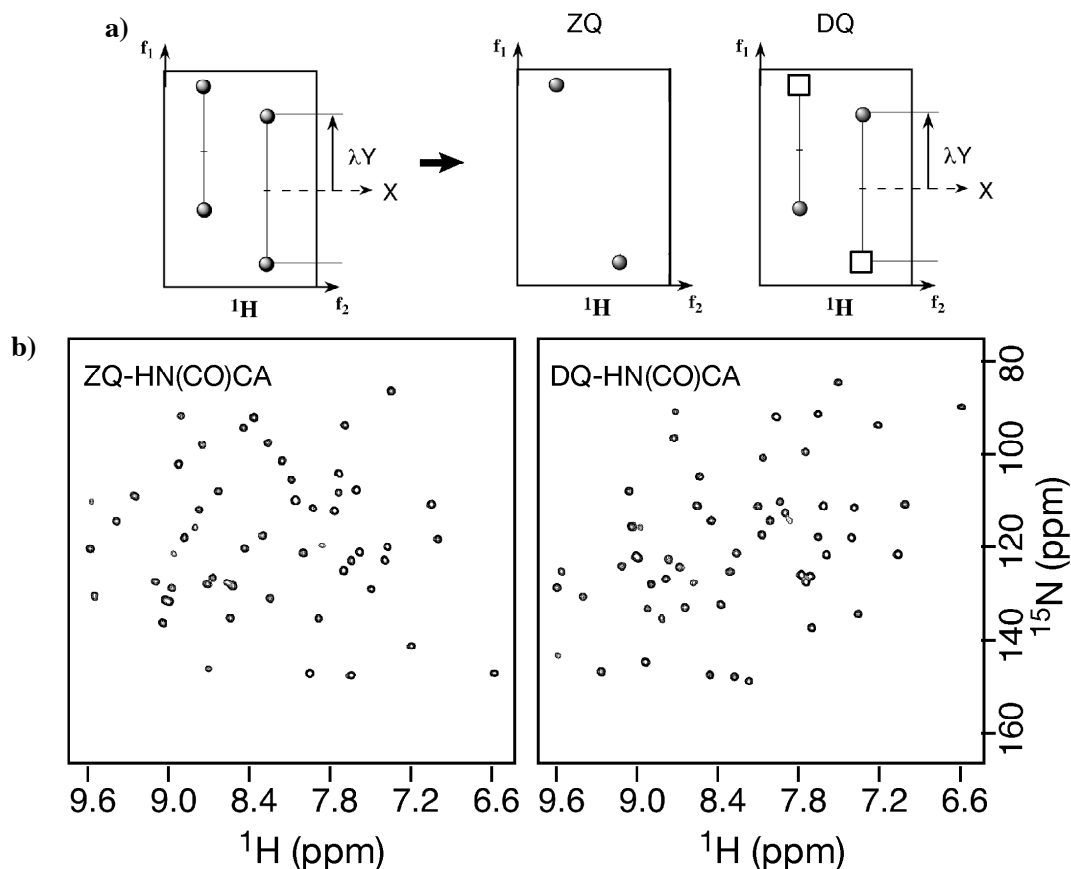
Dans la partie qui suit, je présente une approche de détermination de structure rapide, développée en collaboration avec Bernhard Brutscher (LRMN) et son application à la détermination du repliement de la chaîne principale de MerAa, le domaine N-terminal de 68 résidus de la réductase du mercure de *Ralstonia metallidurans* (voir 2.3.3 pour une introduction de la thématique biologique). Cette approche implique l'attribution séquentielle à l'aide des expériences triple-résonances à dimensionalité réduite, la proposition d'un premier modèle structural en exploitant les constantes de couplage dipolaire résiduel mesurées pour des noyaux de la chaîne principale (voir 2.3.1), et la détermination du repliement de la chaîne principale à l'aide des NOE méthyle-méthyle non-ambigus et des couplages dipolaires résiduels (voir 2.3.2). Dans le cas de cette dernière étude, nous nous sommes limités à la détermination du repliement de la chaîne peptidique, afin de confirmer son homologie avec d'autres protéines déjà bien caractérisées. Cependant, une fois le repliement global connu, la structure peut aisément être affinée en attribuant des NOE supplémentaires en se basant sur les coordonnées moléculaires [en utilisant par exemple le programme ARIA (41)].

### 2.3.1 Approche optimisée d'attribution et d'obtention des informations structurales dans le cas d'une petite protéine doublement marquée

Résumé de l'article (42) □ B. Bersch, E. Rossy, J. Covès, B. Brutscher. Optimized set of two-dimensional experiments for fast sequential assignment, secondary structure determination, and backbone fold validation of  $^{13}\text{C}/^{15}\text{N}$ -labelled proteins. *J. Biomol. NMR* 27, 57-67 (2003).

Dans le cas d'une protéine doublement marquée, l'attribution séquentielle est basée sur les transferts efficaces entre hétéronoyaux *via* les couplages  $J$  élevés. Chaque paire  $^1\text{H}$ - $^{15}\text{N}$  de la chaîne peptidique est ainsi corrélée avec des fréquences carbone ( $\text{C}\alpha$ ,  $\text{C}\beta$ ,  $\text{C}'$ ) ou proton ( $\text{H}\alpha$ ) du même résidu ( $i$ ) ou de celles du résidu précédent ( $i-1$ ). Pour l'attribution séquentielle, on cherche à superposer les triplets ou quadruplets de fréquences en  $i$  et  $i-1$ . Ces informations sont généralement obtenues à partir des expériences tridimensionnelles standards. Cependant, afin d'obtenir une résolution spectrale suffisante, le temps expérimental requis pour l'enregistrement de ces expériences

est élevé (deux à trois jours par expérience avec une sonde standard). Très récemment, plusieurs approches ont été élaborées par différentes équipes afin de réduire le temps expérimental, considérant que la sensibilité de l'expérience RMN n'est pas le facteur limitant [voir (43) pour une revue récente]. Une manière efficace de diminuer le temps expérimental est la réduction du nombre de dimensions, en éditant deux ou plusieurs fréquences différentes dans une même dimension. Cette approche de dimensionalité réduite avait été initialement développée dans les années 90 d'une manière indépendante par T. Szyperski (44), et par B. Brutscher et J.-P. Simorre (45, 46). Elle suscite un regain d'intérêt par l'introduction de nouvelles avancées (47, 48) et est maintenant plus connue sous le nom de GFT-spectroscopy [G-matrix and Fourier transform NMR spectroscopy, (49)]. Dans le cadre du projet MerAa, nous avons proposé un ensemble optimisé d'expériences à dimensionalité réduite de type HSQC [(42) et Figure 3]. Deux modifications ont été introduites par rapport aux expériences utilisées antérieurement au laboratoire (1) l'insertion d'un élément de séquence d'impulsion transférant sélectivement la cohérence du HN vers le C $\alpha$ (i) (50), ce qui supprime les pics inter-résiduels dans les spectres détectant les corrélations intra-résiduelles et, (2) la séparation des composantes double- et zéro quantum dans deux spectres distincts par l'incorporation d'un filtre de type DIPSAP (51). Cela permet, en minimisant le nombre de pics dans les spectres, une exploitation plus facile voire une automatisation de l'attribution séquentielle.



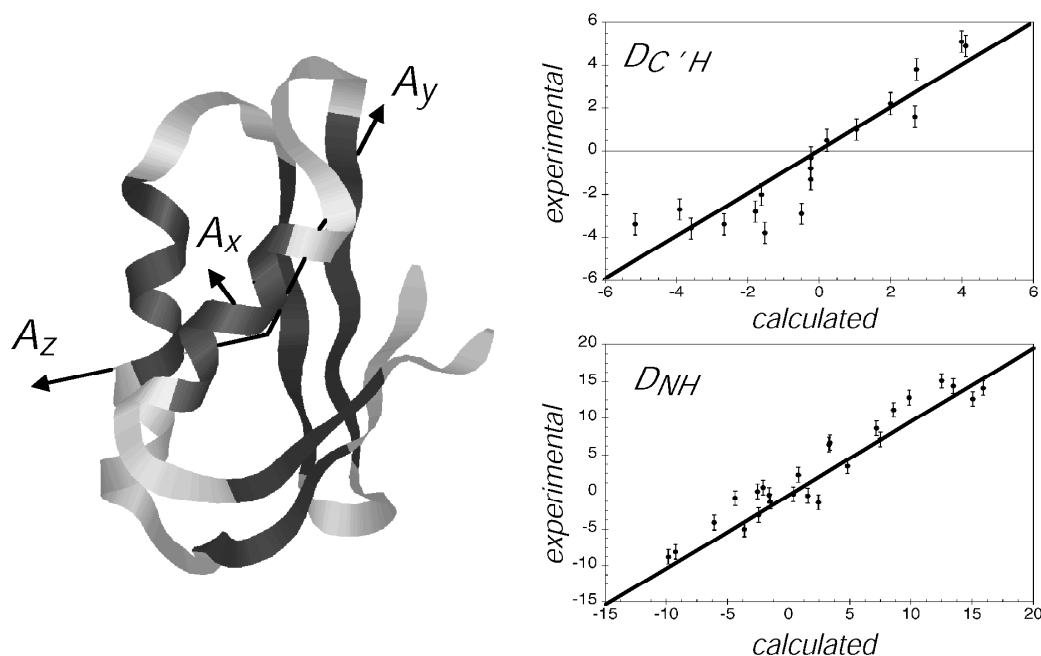
**Figure 3.** A. Principe de l'approche à dimensionalité réduite utilisée pour l'attribution séquentielle de MerAa. Les fréquences de trois noyaux  $^1\text{H}$ ,  $X(=^{15}\text{N})$  et  $Y$  sont obtenues à partir des composantes zéro- et double quantum séparées en deux expériences de type HSQC.  $Y$  peut correspondre à  $\text{CO}(i-1)$ ,  $\text{C}\alpha(i)$ ,  $\text{C}\alpha(i-1)$ ,  $\text{C}\beta(i)$ ,  $\text{C}\beta(i-1)$ ,  $\text{H}\alpha(i)$  ou  $\text{H}\alpha(i-1)$ . B. Exemples des spectres obtenus sur un échantillon de MerAa de 1.5 mM à 600 MHz dans un temps expérimental de 30min

L'attribution de la protéine a ainsi été effectuée dans un temps expérimental de moins de 15h grâce aux expériences suivantes (42)

| expérience     | noyaux détectés  | temps expérimental |
|----------------|--|--------------------|
| IntraHNCA      | $^1\text{H}$ , $^{15}\text{N}$ , $^{13}\text{C}$ (i)   | 1h                 |
| HN(CO)CA       | $^1\text{H}$ , $^{15}\text{N}$ , $^{13}\text{C}$ (I-1) | 30min              |
| IntraHN(CA)HA  | $^1\text{H}$ , $^{15}\text{N}$ , $^1\text{H}$ (i)      | 3h                 |
| HN(COCA)HA     | $^1\text{H}$ , $^{15}\text{N}$ , $^1\text{H}$ (I-1)    | 1h30               |
| Intra-HN(CA)CB | $^1\text{H}$ , $^{15}\text{N}$ , $^{13}\text{C}$ (i)   | 5h20               |
| HN(COCA)CB     | $^1\text{H}$ , $^{15}\text{N}$ , $^{13}\text{C}$ (I-1) | 2h40               |
| HNCO           | $^1\text{H}$ , $^{15}\text{N}$ , $^{13}\text{C}$ '     | 25min              |

Une fois l'attribution connue, l'analyse des déplacements chimiques des C $\alpha$  et C $\beta$  permet de déterminer et de localiser les éléments de structure secondaire.

La mesure des constantes de couplage dipolaire résiduel permet d'obtenir des informations structurales dès l'attribution des résonances de la chaîne principale (8) avant même d'avoir attribué les chaînes latérales ou des NOE. Une version modifiée de la séquence d'impulsion HNCO a été développée pour mesurer des couplages scalaires entre proton amide et nitrogène ( $^1J_{HN}$ ) et entre proton amide et carbonyle ( $^2J_{CH}$ ) dans une même expérience bidimensionnelle (42). Ainsi, deux constantes de couplage dipolaire résiduel peuvent être déterminées par plan peptidique après avoir dissous la protéine dans un milieu anisotrope. L'analyse de ces données utilisant le programme MODULE [développé par M. Blackledge au LRMN, (52)] a permis de comparer la topologie de MerAa avec une molécule homologue, le domaine N-terminal de l'ATPase Ccc2a de la levure [(53),Figure 4].



**Figure 4.** A. Modèle structural de MerAa basé sur la structure en solution du premier domaine de l'ATPase Ccc2a de la levure. Seulement les rdc des résidus montrés en noir sur la représentation en ruban ont été utilisées pour la validation du modèle structural. Les flèches indiquent l'orientation du tenseur avec le meilleur accord par rapport aux données expérimentales. B. Plots de corrélation entre les rdc expérimentales et recalculées à partir du modèle. Figure modifiée à partir de (42).

La comparaison des données expérimentales avec celles recalculées à partir de la structure de Ccc2a montre que les deux domaines possèdent le même repliement de la chaîne principale. Par la suite, ces couplages résiduels dipolaires ont été utilisés pour l'affinement de la structure de la chaîne peptidique (2.3.2). En conclusion, il nous a été possible de proposer un modèle structural de MerAa une semaine seulement après les premiers enregistrements de données.

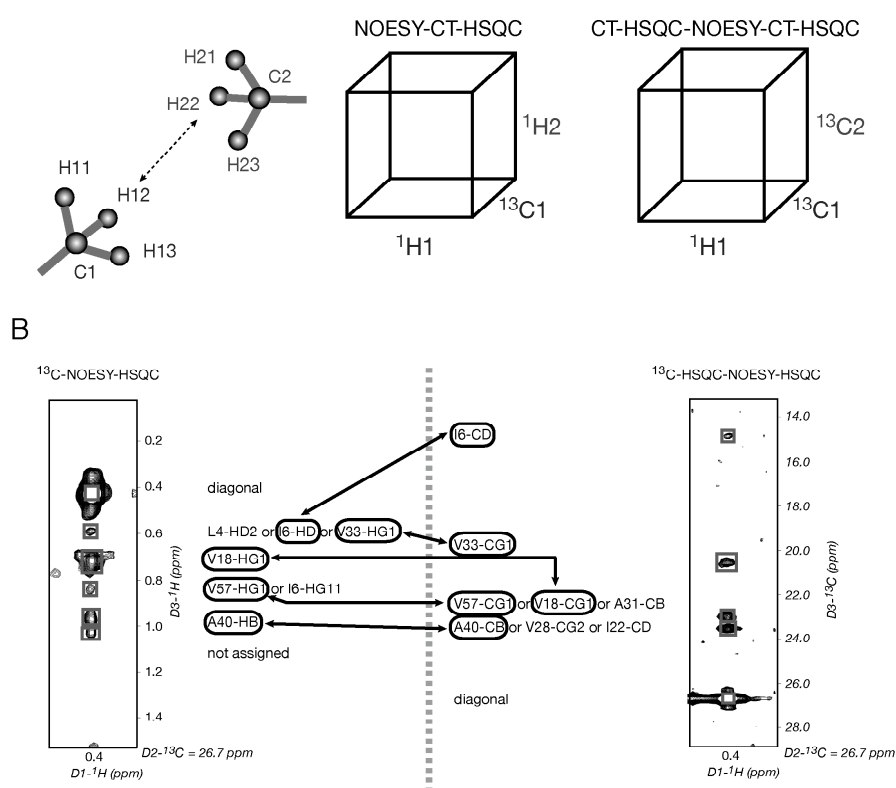
### 2.3.2 Détermination optimisée de structure à partir des données RMN non-ambiguës.

Résumé de la partie RMN d'un article soumis à Journal of Biological Inorganic Chemistry (Article 4). Cet article décrit la caractérisation biochimique et structurale de MerAa, le module N-terminal de la réductase du mercure (voir aussi 2.3.3).

L'introduction du marquage isotopique permet également d'optimiser le processus de calcul de structure *ab initio* à partir des données RMN. Je vais brièvement décrire la stratégie que nous avons utilisée pour le calcul de la structure de MerAa (les résultats sont décrits plus en détail dans l'Article 4).

L'attribution des résonances de la chaîne principale permet d'identifier et de délimiter les éléments de structure secondaire. Cette information est utilisée sous formes de contraintes d'angles dièdres canoniques. Les mesures des constantes de couplage dipolaire résiduel fournissent de l'information orientationnelle qui en elle-même est dégénérée (8), mais qui est attribuée d'une manière non-ambiguë

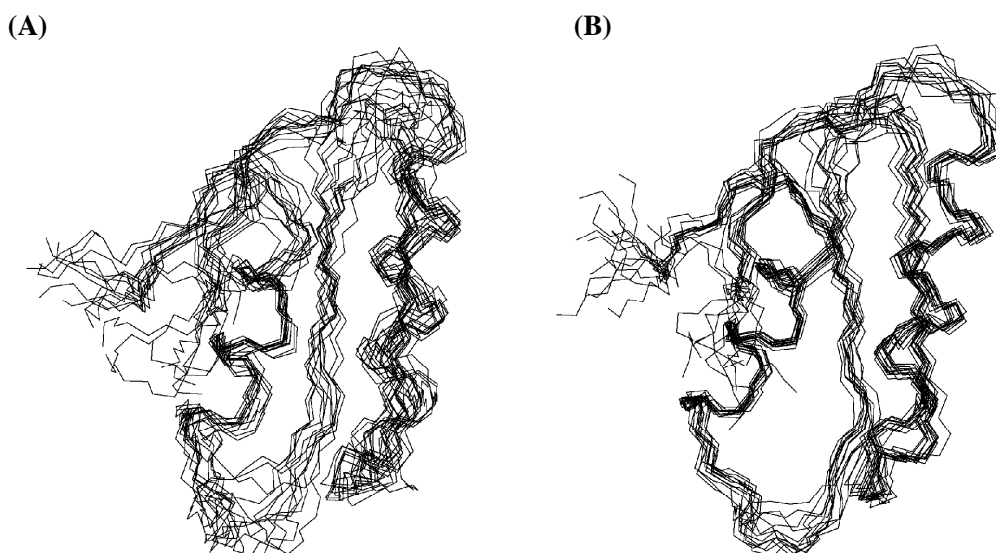
**Figure 5.** Attribution non-ambiguë des NOEs méthyle-méthyle. A. Principe des deux



expériences complémentaires. B. Exemple de l'attribution des NOE du groupement méthyle de la leucine-22 de MerAa.

à des positions dans la chaîne peptidique. La détermination de deux couplages par plan peptidique permet déjà un bon échantillonnage de l'espace orientationnel. Au cas où les couplages peuvent être

mesurées dans deux milieux orientés différents, la dégénérescence de l'information orientationnelle peut être levée (58). Les rdc permettent d'orienter les éléments de structure secondaire et d'affiner la conformation des boucles. Dans le protocole utilisé pour la détermination de la structure de la chaîne principale de MerAa, cette information a été utilisée dans l'étape d'affinement de la structure. Cependant, les rdc ne contiennent pas d'information translationnelle. Pour déterminer l'arrangement relatif des éléments de structure secondaire, nous avons déterminé quelques NOE clés à longue distance d'une manière non-ambiguë. Il est bien connu que les NOE méthyle-méthyle fournissent des informations précieuses sur le coeur hydrophobe des protéines (9, 28, 29). De plus, ces résonances sont bien résolues dans des spectres  $^1\text{H}$ - $^{13}\text{C}$ -HSQC, intenses à cause de la présence de trois spins équivalents, faciles à attribuer (expériences de type HCONH-TOCSY) et elles possèdent des propriétés de relaxation favorables dues à la rotation rapide des groupements  $\text{CH}_3$ . Pour l'attribution non-ambiguë des NOE méthyle-méthyle, nous avons utilisé deux expériences NOESY complémentaires, dans lesquelles soit une dimension  $^1\text{H}$ , soit une dimension  $^{13}\text{C}$  est ajoutée au spectre de corrélation  $^1\text{H}$ - $^{13}\text{C}$ . Ainsi, chaque paire de méthyles est définie par 4 fréquences ( $^1\text{H}$  et  $^{13}\text{C}$ ). Cette approche est illustrée dans la Figure 5. Dans le cas de la protéine MerAa, 56 distances méthyle-méthyle ont ainsi pu être définies. Le calcul de la structure a ensuite été effectué en deux étapes. Le programme Discover (Accelrys Inc.) a été utilisé pour calculer le repliement de la chaîne principale à partir des 56 contraintes de distances méthyle-méthyle, de 62 contraintes dièdres pour les régions de structure secondaire identifiées à partir des déplacements chimiques  $^{13}\text{C}$ , et 47 contraintes pour les liaisons hydrogènes, basées sur l'observation de NOE  $\text{H}^{\text{N}}-\text{H}^{\text{N}}$ . Le rmsd calculé sur les N, C $^{\alpha}$  et C' des résidus 4 à 66 était de 1.16 Å (Figure 6.A). Une étape d'affinement de la conformation de la chaîne principale a été ensuite effectuée en utilisant le programme SCULPTOR (54) et un jeu de contraintes comprenant 99 rdc mesurés sur la chaîne principale. L'ensemble de structures ainsi obtenu était caractérisé par un rmsd de 0.91 Å (Figure 6.B).



**Figure 6.** Repliement de MerAa. A. Ensemble de 15 structures calculé avec les contraintes suivantes: 56 méthyle-méthyle NOE, 62 angles dièdres (éléments de structure secondaire) et 47 liaisons hydrogènes (éléments de structure secondaire et appariement des brins anti-parallèles). B. Ensemble après affinement utilisant le programme Sculptor avec 99 contraintes orientationnelles (rdc) supplémentaires.



### 2.3.3 Le domaine N-terminal de la réductase du mercure de *Ralstonia metallidurans* □ repliement de la chaîne principale et liaison au mercure.

Collaboration avec J. Covès du Laboratoire de Chimie et Biochimie des Centres Rédox Biologiques au DRFMC, CEA-Grenoble (affectation actuelle □ Laboratoire de Protéines Membranaires, IBS).

La toxicité du mercure est due à la capacité du cation mercurique d'interagir avec des thiols, comme, par exemple les cystéines dans les protéines. Cependant, beaucoup de bactéries sont capables de résister à des concentrations fortes de ce métal toxique. Le mécanisme principal de cette résistance implique le transport actif du Hg(II) à travers la membrane cytoplasmique et sa réduction à l'intérieur de la cellule sous forme Hg(0) beaucoup plus inerte et volatile. Plusieurs protéines participent à ce processus dont les gènes correspondants sont regroupés dans l'opéron mer [voir (55) pour revue récente].

La réductase du mercure (MerA) peut posséder un ou deux domaines N-terminaux, qui montrent une forte homologie de séquence avec MerP. La structure tridimensionnelle de MerA a été déterminée par cristallographie aux rayons X, mais il a été impossible d'obtenir des informations structurales sur l'extension N-terminale, certainement dû à sa flexibilité par rapport au reste de la protéine. La fonction de ce domaine n'est pas connue. Cependant, la présence d'un motif de fixation du métal, GMTCCXC, ainsi que la topologie des éléments de structure secondaire [(42) et 2.3.1], suggèrent qu'il fait partie d'une famille de protéines regroupant des transporteurs de métaux tels que CopZ, MerP, Atx1, pour ne citer que quelques-uns (56). En utilisant les informations de la RMN et d'autres techniques spectroscopiques (spectrométrie de masse, dichroïsme circulaire), nous avons pu confirmer cette homologie structurale et mettre en évidence la liaison de l'ion mercurique par les deux cystéines conservées. Ces résultats nous permettent de proposer un rôle du domaine N-terminal de la réductase du mercure dans le transfert de l'ion mercurique vers le site actif de MerA.

## Biophysical Characterization of the MerP-like Amino-Terminal Extension of the Mercuric Reductase from *Ralstonia metallidurans* CH34

Emmanuel ROSSY, Ludovic CHAMPIER, Beate BERSCH, Bernhard BRUTSCHER, Martin BLACKLEDGE and Jacques COVES

E. Rossy – L. Champier – J. Covès (✉) : Laboratoire de Chimie et Biochimie des Centres Redox Biologiques, CEA-Grenoble, DRDC/CB, UMR 5047 CNRS - CEA - Université Joseph Fourier, 17, Avenue des Martyrs, 38054 Grenoble Cedex 9, France

J.C. Present address : IBS/LPM, 41, Rue Jules Horowitz, 38027 Grenoble Cedex, France. E-mail : [jacques.coves@ibs.fr](mailto:jacques.coves@ibs.fr). Telephone : 33-(0)4-38-78-24-03 FAX : 33-(0)4-38-78-54-94.

B. Bersch (✉) – B. Brutscher – M. Blackledge ✉ Laboratoire de Résonance Magnétique Nucléaire, Institut de Biologie Structurale – Jean-Pierre Ebel, UMR 5075 CNRS – CEA – Université Joseph Fourier, 41, Rue Jules Horowitz, 38027 Grenoble Cedex, France.

The purified native mercuric reductase (MerA) from *Ralstonia metallidurans* CH34 contains a N-terminal sequence of 68-amino acids predicted to be homologous to MerP, the periplasmic mercury-binding protein. This MerP-like protein has now been expressed independently. The protein was named MerAa by homology with Ccc2a, the first soluble domain of the copper-transporting ATPase from yeast. MerAa has been characterized by using a set of biophysical techniques. The binding of mercury was followed using circular dichroism spectroscopy and electrospray mass spectrometry. The two cysteine residues contained in the consensus sequence GMTCCXC are involved in the binding of one mercury atom with an apparent affinity comparable to that of MerP for the same metal. The metal binding site is confirmed by NMR chemical shift changes observed between *apo*- and metal-bound MerAa in solution. NMR shift and NOE data also indicate that only minor structural changes occur upon metal binding. Further NMR investigation of the fold of MerAa using long-range methyl-methyl NOE and backbone residual dipolar coupling data confirms the expected close structural homology with MerP. <sup>15</sup>N relaxation data show that MerAa is a globally rigid molecule. An increased backbone mobility was observed for the loop region connecting the first  $\beta$ -strand and the first  $\beta$ -helix and comprising the metal-binding domain. Although significantly reduced, this loop region keeps some conformational flexibility upon metal-binding. Altogether, our data suggest a role of MerAa in mercury trafficking.

The toxicity of mercury comes from its ability to strongly react with thiols such as the cysteine residues in proteins. However, several bacteria are able to resist high concentrations of mercury by expressing a set of specific proteins encoded by genes found in the *mer* operons. The basic bacterial mechanism of resistance includes an uptake of the toxic before the reductive enzymatic detoxification of Hg(II) to Hg(0) [1-4]. This is possible because of the rapid induction upon mercury stress of the expression of the structural genes *merTPA* under the positive control of product of the *merR* gene [5, 6]. The result of this expression is to set up a specific transport system composed of the periplasmic protein MerP, the integral membrane protein MerT, and the cytosolic mercuric reductase MerA. The protein MerP was suggested to act as a sink for Hg(II) in the periplasm [7] and then to give the mercuric ion to MerT which controls its transport through the membrane

[8]. Hg(II) must then reach the cytosol where it will be reduced by the mercuric reductase at the expense of NADPH [9]. One hypothesis suggests that the transport protein can transfer the mercuric ion directly to the reductase through an interaction with the N-terminal extension frequently found in MerA which shows high sequence homology with MerP. This hypothesis is still controversial since the N-terminal extension was found neither to be involved in catalysis *in vitro* [9, 10] nor

---

**Abbreviations** : CCA,  $\beta$ -cyano-4-hydroxy-trans-cinnamic acid; CSI, chemical shift index; HSQC, <sup>1</sup>H-detected heteronuclear single quantum coherence; MerAa, the 68-amino acids N-terminal extension of the mercuric reductase; NOE, nuclear Overhauser effect; RDC's, residual dipolar couplings; TCEP-HCl, Tris-[2-Carboxyethyl]-Phosphine Hydrochloride.

**Key-words** : Mercuric reductase, metallochaperones, detoxification, *Ralstonia metallidurans* CH34.

essential for mercury detoxification *in vivo* [10, 11]. However, even if this extension can be absent in certain mercuric reductases [12], it is generally found in one or several copies [4, 13], showing thus a high degree of evolutionary conservation. The crystallographic structure of MerA from *Bacillus* sp. strain RC607 has also been solved [13]. This protein consists of three modules : the N-terminal region (residues 1-166), the core (167-616), a structural homologue of the glutathione reductase and other members of FAD-containing disulfide oxidoreductase family, and the C-terminal extension (617-631). Due to considerable mobility with respect to the rest of the molecule, the N-terminal module did not appear in the electron density maps and had not thus been structurally defined. In the case of MerA from *Bacillus* sp. strain RC607, this module corresponds to the repetition of two domains of 80-85 amino acids which are highly homologous to MerP. Interestingly, the solution structures of oxidized or reduced MerP and MerP with bound Hg(II) have been solved by two different groups [14, 15].

The recurring structural motif GMTCXXC, where X represents any amino acid residue, is a general metal-binding motif occurring in many proteins involved in heavy-metal binding or transport proteins in bacteria or mammals, including MerP and the N-terminal extension of MerA [4, 16, 17]. This is the case for instance for the copper chaperones CopZ from *Enterococcus hirae* or *Bacillus subtilis*, for Atx1 that shuttles Cu(I) in yeast, or for parts of transport ATPases such as the Menkes and Wilson ATPase or the Ccc2 ATPase. The structures of all these proteins or protein fragments have been solved in solution or by crystallography, with or without a bound metal [for reviews, see 16, 17 and references therein]. They share with MerP a common structural fold,  $\alpha\alpha\alpha\alpha\alpha\alpha$ , defined as the characteristic “ferredoxin-like” fold with the metal-binding site situated in a loop between the first  $\alpha$ -helix and the first  $\beta$ -sheet.

In this study, we have focused on the N-terminal extension of the mercuric reductase from *Ralstonia metallidurans* CH34. *R. metallidurans* CH34 is a facultative autotrophic bacterium harboring two large plasmids pMOL28 (180 kb) and pMOL30 (240 kb) carrying respectively resistance to Ni<sup>2+</sup>, Co<sup>2+</sup> and Hg<sup>2+</sup> ions and resistance to Cd<sup>2+</sup>, Co<sup>2+</sup>, Hg<sup>2+</sup> and Zn<sup>2+</sup> [18]. The insertion in pMOL28 conferring a narrow-spectrum resistance to mercury is named Tn4378 [18] and contains a *mer* operon very homologous to the operon found in Tn501 (originating from the *Pseudomonas aeruginosa* plasmid pVS1) in terms of organization and sequence. As a consequence, the *R. metallidurans* CH34 mercuric reductase is highly homologous to the protein from *P. aeruginosa* that contains only one MerP-like N-terminal extension. This module is called MerAa in this paper by analogy with the first cytosolic domain of the yeast Ccc2 ATPase called Ccc2a [19]. We report now on spectroscopic evidence that MerAa is able to bind one mercury atom per monomer. We have also built a three dimensional structural model of metal-bound MerAa using long-range methyl-methyl NOE constraints combined with dihedral and hydrogen bond restraints for the previously identified secondary structural

elements [20]. As expected, the backbone structure of MerAa shows very close homologies with MerP, Atx1 or Ccc2a [16, 17]. A plausible role of this domain in mercury trafficking will be discussed.

## MATERIALS AND METHODS

**Bacterial Strains and plasmids:** The transposon Tn4378 from *R. metallidurans* CH34 was previously fully transferred on plasmid R1CM (Chloramphenicol<sup>R</sup>). The resulting plasmid R1CM::Tn4378 was then used to transform a derivative of the *E. coli* strain CSH23 [*recA1 spcR supE thisrl::Tn10*  $\square$  *pro lac*]. The resulting strain *E. coli* CSH23 (R1CM::Tn4378) is thus Hg<sup>R</sup>, Cm<sup>R</sup>, and Tc<sup>R</sup> (unpublished results) and was a kind gift of Prof. Max Mergeay (CEN/SCK, Mol, Belgium).

**Construction of the MerAa expression plasmid:** R1CM::Tn4378 was extracted and purified from its host strain by using the Flexiprep kit (Amersham-Pharmacia Biotech) and was then used as a template for PCR amplification of a synthetic DNA fragment expected to encode for a 68-residue protein defined as merAa. The sense primer 5'-GGG GAA TTC *ATA TGA CCC ATC TAA AAA TCA CCG*-3' was designed to incorporate a *Tru*II restriction site (*italic*) including the start codon (*underlined*) of the MerA gene. The antisense primer 5'-GGC *GGA TCC TCA ATC GGC TAG CGT TGC CTT GTA*-3' was designed to incorporate a *Bam*HI restriction site (*italic*) downstream of the stop codon (*underlined*) in the coding sequence of MerA to construct a protein truncated between aspartate-68 and alanine-69 (this numbering includes the initial methionine). The nucleotides in boldface are complementary of the corresponding coding sequence in MerA. The amplification of the target sequence was carried out upon 100 ng of R1CM::Tn4378 with 2.5 units of PWO DNA polymerase (Roche) and 100 pmol of each primer in the recommended buffer. The amplification was performed according to the following scheme with a Robocycler Gradient 40 (Stratagene): 60 s at 94 °C, 60 s at 55 °C and 60 s at 68 °C for 20 cycles. The authenticity of the DNA insert generated by PCR was confirmed by sequencing (GENOME EXPRESS, Meylan, France). It was then directly ligated into the plasmid pCR-4Blunt TOPO with the Zero Blunt TOPO Cloning kit for Sequencing (Invitrogen). After transformation in *E. coli* One Shot Competent Cell (Invitrogen), a positive clone was selected and the recombinant plasmid pCR-TOPO-MerAa was extracted using the Flexiprep kit. The *Tru*II-*Bam*HI fragment obtained by digestion of plasmid pCR-TOPO-MerAa was cloned into plasmid pT7.7 cut with the *Nde*I and *Bam*HI restriction enzymes to yield the expression plasmid pT7.7-MerAa.

**Expression of recombinant MerAa:** The expression plasmid pT7.7-MerAa was introduced into the T7 RNA polymerase-containing host *E. coli* B121(DE3)pLysS. Cells were grown overnight in 100 mL LB medium supplemented with ampicilline (100  $\square$ g.mL<sup>-1</sup>) and chloramphenicol (34  $\square$ g.mL<sup>-1</sup>). Then, 5 L of the same medium was inoculated at 1% with the overnight preculture. MerAa expression was induced at an absorbance of 0.4 at 600 nm by addition of 0.4 mM isopropyl thio- $\square$ -galactopyranoside (IPTG). Cells were harvested by centrifugation 4 h after induction and

resuspended in 40 mL of buffer A (50 mM Tris-HCl pH 7.5) supplemented with 10 mM DTT and lysed by sonic oscillation. The total protein extract was recovered by centrifugation at 45000 rpm during 90 min in a 50.2 TI rotor (Beckman). Ammonium sulfate was added to 30% saturation and the precipitated material was removed by centrifugation and discarded. The supernatant was brought to 80% ammonium sulfate saturation. The pellet so obtained was dissolved in a minimum volume of buffer A plus 1 mM DTT and loaded on a Superdex 75 column (Amersham-Pharmacia Biotech) previously equilibrated with the same medium. Elution was run at 0.8 mL.min<sup>-1</sup> and 1 mL fractions were collected. Fractions were assayed for protein (absorbance at 280 nm) and for the presence of MerAa (SDS/PAGE analysis). The MerAa-containing fractions were pooled and the concentration of DTT was brought to 10 mM before concentration using a Diaflo cell equipped with a YM-3 membrane (Amicon Corp.). The MerAa solution obtained (3 mL, 68 mg) was aliquoted and stored at - 80 °C for further use. At this stage, MerAa was electrophoretically pure. Alternatively, MerAa was purified in the absence of DTT and the pure protein was then fully reduced by incubation in the presence of a 5 to 10 molar excess of DTT or TCEP-HCl (Tris[2-Carboxyethyl]-Phosphine Hydrochloride, Pierce, Rockford, IL) during 1 hour at room temperature

For the NMR experiments, the freshly transformed bacteria were grown at 37°C in 1 L of M9 minimal mineral medium with glucose (2 g/L) as carbon source and supplemented with MnCl<sub>2</sub> (0.1 mM), ZnSO<sub>4</sub> (0.05 mM), FeCl<sub>3</sub> (0.05 mM) and a vitamin solution. Isotopically labeled protein was prepared by growing the cells with <sup>15</sup>NH<sub>4</sub>Cl (1 g/L) and <sup>13</sup>C<sub>6</sub>-glucose (2 g/L) as sole nitrogen and carbon sources. Purification of the labeled protein was performed as described above. Typical yield were about 70 mg protein per liter of culture medium.

Protein concentration was determined using bovine serum albumin as a standard and the Micro BCA protein assay (Pierce, Rockford, IL). The denatured molecular mass of the proteins was estimated by 0.1% SDS-15% polyacrylamide gel electrophoresis [21]. N-terminal sequence analyses were performed by using an Applied Biosystems gas-phase sequencer model 477A with on-line analysis of the phenylthiohydantoin derivatives.

**Alkylation of MerAa:** Alkylation of the thiol groups was performed in buffer A at room temperature. Excess of TCEP-HCl or DTT was removed by filtration on G25 Sephadex column and the reduced MerAa protein was incubated during 3 h in the dark in the presence of a 10 molar excess of a freshly prepared solution of iodoacetamide in the same buffer. The reaction was stopped by freezing the solution in liquid nitrogen.

To demonstrate the involvement of the cysteines in the mercury binding of mercury, alkylated MerAa or reduced MerAa were digested by about 1% w/w endoproteinase Lys-C at 37°C for 4 h before MALDI mass spectrometry analysis of the resulting peptides. Recombinant MerAa was run in parallel as a control.

**Addition of Hg<sup>2+</sup> to MerAa:** After elimination of the reductant as described above, the protein was diluted to the

desired concentration depending on the used technique. The reduced protein was then titrated by addition of small amounts of HgCl<sub>2</sub> for CD or mass spectrometry experiments. Spectra were recorded 1 min after each addition of mercury. In the following, the protein without mercury or with mercury will be called *apo*- or metal-bound MerAa respectively. For the NMR studies, metal-bound protein was prepared by a single addition of HgCl<sub>2</sub> corresponding to 1.2 mercury equivalents.

**Mass spectrometry Analysis of MerAa:** Electrospray mass spectrometry (ES-MS) was performed using Q-TOF Micro mass spectrometer (Micromass, Manchester, U.K.) equipped with an electrospray ion source, operating with a needle voltage of 3 kV, sample cone and extraction cone voltages respectively of 60 V and 3.5 V. Sample concentration was 600 nM in water/acetonitrile (1/1-v/v) with 0.2% formic acid and continuously infused at a flow rate of 5 μL/min. The mass spectra were recorded in the 750-2200 range of mass-to-charge (*m/z*). Data were acquired in the positive mode and calibration was performed using the multiply charged states produced by a separate injection of heart horse myoglobin dissolved in water/acetonitrile (1/1-v/v) with 0.2% formic acid. Spectra were acquired and data were processed with MassLynx 3.5 or 4.0 (Micromass).

MALDI mass spectrometry was performed on a Perseptive Voyager XL (Perseptive Biosystems) time-of-flight spectrometer. A solution of CCA in 50/50 water/acetonitrile containing 0.3 % TFA was used as a matrix. A volume of 1 μL of crude peptide mixture was mixed on the target with an equal volume of matrix, and the spot was allowed to air-dry. Ions were accelerated with an extraction voltage of 20 kV. Spectra were obtained by accumulation of 200 laser shots (Nd:YAG laser, λ = 337 nm).

**Spectroscopic methods:** The near UV region CD spectra were recorded at 20°C between 310 and 250 nm at a scan speed of 1 nm.sec<sup>-1</sup>, a band width of 4 nm and a response time of 1 s using a Jasco J-810 spectropolarimeter interfaced with the requested software for spectra analysis. Spectra were recorded 1 min after each addition of HgCl<sub>2</sub>. Each spectrum has been accumulated thirty times. The experiment was conducted at protein concentrations of 40 μM or 25 μM in a volume of 1.5 mL (path-length 1 cm) or at 3.5 μM in a volume of 20 mL (path-length 10 cm). The concentration of the mercury solution was calculated so that the added volume at the end of the experiment was negligible (10 μL and 100 μL respectively) and data were not corrected for dilution at the end of the experiment. Stability of reduced MerAa during the time of CD experiments was checked by alkylation of the protein at different times and mass spectrometry. As a control of the structural stability of the protein during mercury binding, far UV region CD spectra were recorded between 250 and 190 nm.

UV-visible spectra were obtained with a Cary 1-Bio spectrophotometer (Varian).

**NMR spectroscopy:** For the NMR experiments, the <sup>13</sup>C/<sup>15</sup>N-labelled protein was dissolved to a final concentration of 1.5 mM in 90% H<sub>2</sub>O, 10% D<sub>2</sub>O containing 50 mM Tris buffer pH 7 and 1 mM DTT in case of the *apo*-

form of MerAa or 1.2 molar equivalents of  $\text{HgCl}_2$  for the metal-bound form. All NMR experiments were acquired on a Varian INOVA 600 spectrometer equipped with a triple-resonance probe ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ) and shielded pulsed field gradients along the z-axis. The sample temperature was set to 25°C for side-chain resonance assignment, RDC measurements, and NOESY experiments, and to 20°C for backbone assignment and the  $^{15}\text{N}$  relaxation study. Data processing and peak picking were performed using Felix program version 2000 (Accelrys Inc.). Mirror image linear prediction [22] was applied to the  $^{13}\text{C}$  and  $^{15}\text{N}$  constant time (CT) dimensions to further increase the apparent spectral resolution. Squared cosine apodization was used in all dimensions prior to zero-filling and Fourier transformation.

Backbone resonance assignments for the *apo*- and the metal-bound form were performed using an original approach based on a series of two-dimensional (2D) H-N-X correlation spectra (with X corresponding to intra-residue or sequential  $\text{C}^\alpha$ ,  $\text{C}^\beta$  or  $\text{H}^\alpha$  nuclei), as described in detail previously [20]. The complete data set required for sequential backbone resonance assignment was acquired in an overall experimental time of only 14 hours per sample.

For the metal-bound form, additional  $^1\text{H}$  and  $^{13}\text{C}$  assignment of aliphatic side chains was obtained from 3D H(CCO)NH-TOCSY and 3D (H)C(CO)NH-TOCSY [23], and 2D (H)C(C)H-TOCSY experiments. The 3D data sets were acquired with 80( $\text{H}^\text{C}$ , C)  $\times$  55(N)  $\times$  512( $\text{H}^\text{N}$ ) complex points for spectral widths of 6000Hz ( $\text{H}^\text{C}$ ), 10000Hz (C), 2000Hz (N), and 7018Hz ( $\text{H}^\text{N}$ ) in an experimental time of 24 hours per experiment. Unambiguous  $\text{H}^i\text{-H}^k$  distance restraints were derived from the combined interpretation of two 3D spectra correlating the three nuclei ( $\text{H}^k$ ,  $\text{C}^i$ ,  $\text{H}^i$ ) and ( $\text{C}^k$ ,  $\text{C}^i$ ,  $\text{H}^i$ ).  $^{13}\text{C}$  frequency editing was performed in a constant time (CT) manner [24] for optimized spectral resolution. The 3D NOESY-CT-HSQC and 3D CT-HSQC-NOESY-CT-HSQC spectra were acquired with 130( $\text{H}^k$ )  $\times$  105( $\text{C}^i$ )  $\times$  512( $\text{H}^i$ ) and 105( $\text{C}^k$ )  $\times$  105( $\text{C}^i$ )  $\times$  512( $\text{H}^i$ ) complex points for spectral widths of 3000Hz ( $\text{H}^k$ ), 4000Hz ( $\text{C}^i$ ), and 7018Hz ( $\text{H}^i$ ) in an experimental time of 48 hours and 60 hours, respectively. An additional  $^{15}\text{N}$ -edited 3D NOESY-HSQC spectrum was recorded for the identification of short inter-strand amide  $\text{H}^\text{N}\text{-H}^\text{N}$  distances in the anti-parallel  $\beta$ -sheet. A total of 128(H)  $\times$  100(N)  $\times$  512( $\text{H}^\text{N}$ ) complex points for spectral widths of 6000Hz ( $\text{H}^k$ ), 2000Hz (N), and 7018Hz ( $\text{H}^i$ ) were acquired in an experimental time of 30 hours. The NOE mixing time was set to 100ms for all experiments.

Two residual dipolar coupling (RDC) constants,  $D_{\text{NH}}$  and  $D_{\text{CH}}$  were measured per peptide plane from 2D J-HNCO(H) spectra recorded on samples of metal-bound MerAa dissolved in isotropic and anisotropic solution in an experimental time of 45 min per experiment [20]. For the anisotropic sample the protein was dissolved in a mixture of 5%  $\text{C}_{12}\text{H}_5$ /hexanol with  $r=0.96$  [25].

$^{15}\text{N}$  relaxation experiments  $R_1$ ,  $R_{1\rho}$  and  $\{^1\text{H}\}\text{-}^{15}\text{N}$  NOE were performed using standard 2D pulse sequences [26]. Additional  $\text{C}^\alpha$  and CO decoupling was used during  $^{15}\text{N}$  frequency editing to suppress spin evolution due to  $J_{\text{NC}}$  scalar couplings. For  $R_{1\rho}$  relaxation measurements a spin lock field of  $|\omega_{\text{B}}\mathbf{B}_1|/2J = 2.5\text{kHz}$  was applied. The

relaxation decay was sampled at the following time points: 20, 100, 200, 300, 400, 500, 600, 700, and 1000ms for  $R_1$ , and 8, 40, 72, 104, 134, 168, and 200ms for  $R_{1\rho}$  in an experimental time of 10 hours and 15 hours, respectively. For the  $\{^1\text{H}\}\text{-}^{15}\text{N}$  NOE measurements, two spectra with on- and off-resonance  $^1\text{H}$  saturation were recorded in an interleaved manner. The saturation time was set to 3.5 s and the recycle delay to 7 s. The total experimental time was 28 hours.

**Relaxation data analysis:** For the calculation of relaxation rates and  $\{^1\text{H}\}\text{-}^{15}\text{N}$ -NOEs, peak intensities were measured rather than volumes using the program Felix (Accelrys Inc.). The relaxation rates were extracted using a non-linear least-squares fit of the intensities to a two-parameter single exponential function  $I(t) = I(0) \exp(Rt)$ , with  $I(t)$  the peak intensity after a relaxation delay of time  $t$  and  $I(0)$  the initial peak intensity. The transverse relaxation rate constants  $R_2$  were calculated from  $R_{1\rho}$  rate constants using the relation:  $R_{1\rho} = \cos^2(\Delta)R_2 + \sin^2(\Delta)R_1$  with  $\Delta = \tan^{-1}(2\pi\Delta\nu/k_B)$  and  $\Delta\nu$  the frequency difference between the  $^{15}\text{N}$  carrier and the frequency of the observed nitrogen. The steady state heteronuclear NOE were determined from the ratio of the peak intensities measured in experiments acquired with and without proton saturation. Data were analysed in terms of molecular rotational diffusion and internal mobility using the Tensor2 program as previously described [27].

**Structure calculation:** Structure calculations were performed using a two-step restrained molecular dynamics procedure. The first step, using the program Discover 2.98, utilizes a classical simulated annealing calculation from randomized initial Cartesian coordinates, and has been described in detail elsewhere [28]. Structural restraints for this step comprised 31  $\phi$  and 31  $\psi$  canonical dihedral angle restraints, applied in regions of the primary sequence where the presence of secondary structural elements could be confidently identified, and 26 hydrogen bond restraints in the helical regions, derived from the same observations. An additional 20 hydrogen bonding restraints defining the  $\beta$ -sheet pairing were identified from short range  $^1\text{H}^\text{N}\text{-}^1\text{H}^\text{N}$  distances. 56 long-range methyl-methyl distances were introduced as unambiguous distance restraints, and calibrated from known intra-residual methyl-methyl distances from Leu and Val side chains. In all cases the lowest volume cross-peak of the symmetric pair was used for the distance restraint. The 15 best structures, as defined from the total experimental target function, were selected for refinement against RDC restraints.

The second step refines the lowest energy structures calculated from the first step, using all of the restraints mentioned above, combined with 49  $^1\text{D}_{\text{HN}}$  and 50  $^2\text{D}_{\text{HC}}$  RDC restraints. This calculation was performed using the program SCULPTOR [29] and comprises three steps. The molecular coordinates are fixed during the initial step, while the alignment tensor components  $A_a$  and  $A_r$  evolve under the influence of the RDC's measured from sites present in secondary structural elements. This step comprises a sampling period of 4 ps (8000 0.25fs time-steps) at 300K, followed by 3 ps cooling to 200K and conjugate gradient minimization, and defines the most appropriate starting

point for the tensor corresponding to each structure. The molecule is then released, and the tensor and molecule evolve under the influence of the same RDC restraints and the complete restraint set from step 1, during 9.3 ps at a temperature of 1000 K. During this period  $k_{\text{RDC}}$  is raised from the initial (0.01 kcal.mol<sup>-1</sup>.Hz<sup>2</sup>) to the final values ( $k_{\text{rdc},ij}$ ). The energetic weighting of the different types of RDC was estimated on the basis of the precision of the different experimental measurements ( $\sigma_{ij}=1.0$  Hz for <sup>N</sup>H-N 0.5Hz for <sup>N</sup>H-C') such that final force constants were  $k_{\text{RDC,HN}}=0.50$  kcal.mol<sup>-1</sup>.Hz<sup>2</sup> and  $k_{\text{RDC,HC}}=1.0$  kcal.mol<sup>-1</sup>.Hz<sup>2</sup>. This is followed by a 2.7 ps cooling period to 100K and energy minimization. This step refines both local structure and the tensor simultaneously [30]. At this point the remaining RDC restraints are introduced and both tensor and conformation evolve freely during 24 ps at 1000 K, during which time  $k_{\text{NOE,h-bond}}$ ,  $k_{\text{dihc}}$  and  $k_{\text{RDC}}$  were raised from their initial (0.5, 2.0, 0.01) to their final values (50.0 kcal.mol<sup>-1</sup>.Å<sup>-2</sup>, 200.00 kcal.mol<sup>-1</sup>.deg<sup>-2</sup>,  $k_{\text{RDC,HN}}=0.50$  kcal.mol<sup>-1</sup>.Hz<sup>2</sup> and  $k_{\text{RDC,HC}}=1.0$  kcal.mol<sup>-1</sup>.Hz<sup>2</sup>). Following the sampling period, the system was cooled to 100 K over 2.7 ps and the system again minimized using a conjugate gradient algorithm. This calculation was repeated 3 times for each initial structure using different initial atomic velocity distributions, and the structure with the lowest total target function comprising NOE, backbone torsion, hydrogen bonding and RDC restraints from each initial structure was selected as a member of the final ensemble.

## RESULTS

### Cloning, expression and characterization of MerAa:

The purified native mercuric reductase from *R. metallidurans* CH34 was shown to be sensitive to proteolysis and identification of the C-terminal border of MerAa was guided by the identification of the cleavage site (unpublished results) and by the sequence homology with MerP (Fig. 1). The two proteins share 42% similarity including 32% identity. The DNA sequence coding for the 68 N-terminal residues of MerA was cloned. *E. coli* BL21(DE3)pLysS was transformed with the expression plasmid pT7.7-MerAa. Maximal expression was observed 4 h after IPTG induction. MerAa was recovered in the soluble extracts as judged from SDS-PAGE analysis that shows a predominant band at an apparent molecular mass of 7 kDa corresponding to the expected mass of MerAa. The overproduced protein was purified by a one-step procedure. The final ammonium sulfate pellet was resuspended in a minimal volume of Tris-HCl buffer and the solution was fractionated on a Superdex-75 filtration column. After SDS-PAGE analysis, fractions containing MerAa were pooled. They correspond to proteins with a native molecular mass inferior to 10 kDa, thus showing that MerAa is a monomer

under native conditions. A yield of about 30-40 mg of pure protein per liter of culture was routinely obtained. Isotopically labeled protein was produced for NMR experiments. The extent of isotopic labeling was evaluated to more than 98% for <sup>15</sup>N and <sup>13</sup>C by electrospray ionization mass spectroscopy analyses by comparison of the molecular mass of pure unlabeled MerAa with that of the labeled protein.

**Mass spectrometry characterization of MerAa and of its mercury-binding capacity:** Figure 2A shows the reconstructed mass spectrum of MerAa purified under reductive conditions. Two species with a mass difference of 131 Da are detected. The highest mass at 6840 Da corresponds to the theoretical mass deduced from the amino acid sequence. This species will be called MerAa-met in the following. The second species corresponds to the N-terminal methionine-excised form. The existence of these two species has been confirmed by amino-terminal sequencing of the protein. Alkylation was used as a probe for checking the redox state and the accessibility of the two cysteines of MerAa. Incubation of the reduced protein in the presence of an excess of iodoacetamide, the alkylating agent, led to the conversion of the two previous species to new species with 114 Da mass increases (not shown). They correspond to the addition of two alkylating groups to both forms, MerAa and MerAa-met. It should be noted that when the same experiment was conducted with MerAa purified without DTT added in the buffers, only a part of the protein, the proportion of which depending on the purification and storage conditions, can react with iodoacetamide (not shown). Further analysis using MALDI mass spectrometry after proteolytic cleavage with endoproteinase Lys-C confirmed that the alkylation occurs on the two cysteines in position 11 and 14. Two different peptides, Ile6-Lys19 and Met1-Lys19 (Fig. 3), containing the two cysteines, both showed a mass increase corresponding to two acetamides. As the two peptides do not contain the same number of histidines (His3 and His16), this result also suggests that we can exclude an additional alkylation on histidine residues which could be observed particularly after long reaction times [31]. The peptide Val25-Lys37, which does not contain any cysteines, and therefore is not subjected to alkylation, was taken as a control (Fig. 3).

Fully reduced MerAa was used to follow the binding of mercury (Fig. 2). Increasing amounts of HgCl<sub>2</sub> were added to a protein solution and ES-MS spectra were recorded after incubation during 10 min at room temperature. We can see that upon the first addition of 0.25 mercury equivalent, two extra peaks appeared with a mass difference of 200 Da with respect to MerAa or MerAa-met native masses. This corresponds to the addition on each protein of one mercury atom. As further mercury was added, the proportion of

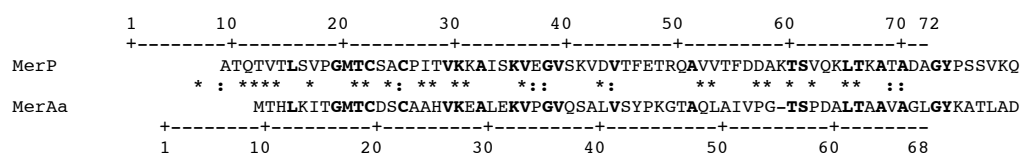
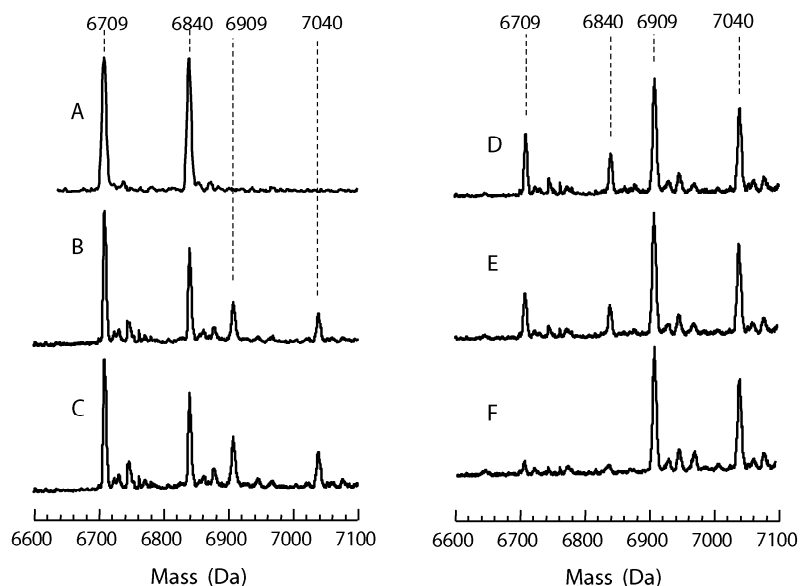
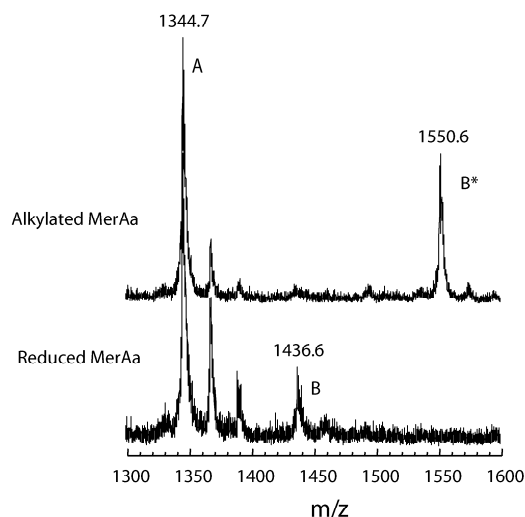


Fig. 1. Sequence comparison between MerP and MerAa. The identical amino acids are shown in bold (\*). Similarities are identified by (:).



**Fig. 2.** ES-MS titration of MerAa with  $\text{HgCl}_2$ . A, recombinant reduced MerAa; B, C, D, E, F, MerAa + respectively 0.25, 0.5, 0.75, 1 and 2 molar equivalent of  $\text{HgCl}_2$ . At the head of each panel, are indicated in increasing order the exact mass of the N-terminal excised form of MerAa, full-length MerAa and the metal-bound form of each protein. The conditions of analysis are described in the “Materials and Methods” section.

proteins with bound-mercury increased while the proportion of reduced MerAa decreased. Finally, after addition of one mercury equivalent, almost all the MerAa population has been converted to Hg-bound MerAa in a perfect agreement with the results of CD spectroscopy (see below). A full conversion is shown for two mercury equivalents. The binding of the mercury is strong enough to resist the ES-MS conditions. We have also verified that fully alkylated protein did not react with  $\text{HgCl}_2$  and that the Hg-bound protein did not react with iodoacetamide (not shown). This strongly argues in favor of a mercury-binding site involving the two cysteines as expected.

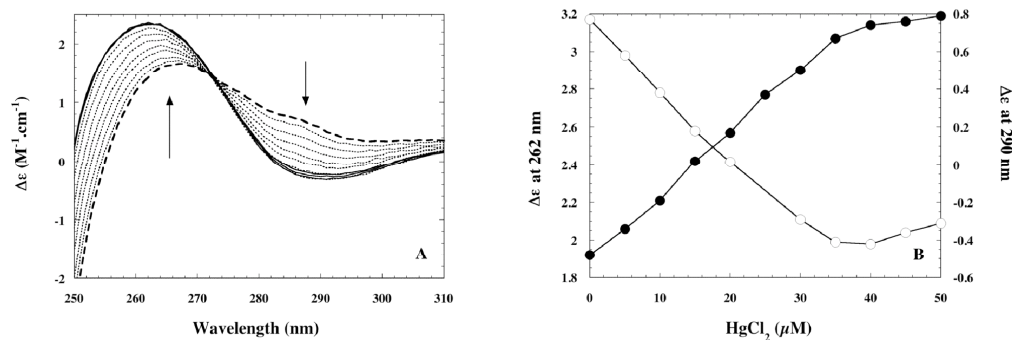


**Fig. 3.** MALDI mass spectra of the peptides obtained after enzymatic cleavage with endoproteinase Lys-C of MerAa under the alkylated (upper spectrum) or the reduced form (lower spectrum). Alkylation with iodoacetamide was performed as described in the “Materials and Methods” section. The different peaks correspond to singly charged ions from several peptides of the proteins: (A) V25-K37, control peptide; (B) I6-K19, peptide containing the cysteines 11 and 14; (B\*) I6-K19 plus two alkylating groups.

#### UV-visible and CD spectroscopic characterizations of mercury-binding to MerAa:

Because MerAa only contains two tyrosines and no tryptophane (Fig. 1), the absorbance of the protein is very low in the near UV region of an electronic absorption spectrum (the theoretical extinction coefficient is about  $2600 \text{ M}^{-1}\cdot\text{cm}^{-1}$  at 280 nm). However, upon addition of mercury, the absorbance at 267 nm increases as a function of the amounts of added metal and the maximum signal was reached upon addition of one mercury equivalent. A plot of  $\Delta A_{267 \text{ nm}}$  as a function of mercury added displays only one phase indicating probably the regular binding of the metal and the exposition of an aromatic residue such as a tyrosine. However, it is noteworthy that we were unable to detect any change in fluorescence upon mercury addition.

CD spectroscopy has been proved to be a suitable method to follow metal binding to metalloproteins [32, 33]. Moreover, this technique enables the determination of dissociation constants ( $K_d$ ) of the protein for the metal of interest. The spectral changes upon addition of  $\text{HgCl}_2$  to *apo*-MerAa were recorded between 310 nm and 250 nm. Changes in the near UV region allow the identification of tertiary structural changes. Figures 4A and B show the results obtained in a typical experiment conducted with 40  $\mu\text{M}$  MerAa. The spectrum of the *apo*-MerAa does not show major features above about 255 nm. Addition of mercury in sub-stoichiometric amounts causes a multi-phasic CD response characterized by three parameters: a large increase in the molar ellipticity around 262 nm, a slight decrease around 290 nm and the existence of an isodichroic point at 272 nm. The maximum signal was obtained upon addition of one mercury equivalent and the CD spectrum did not change as further mercury was added. The same kind of results was obtained at different protein concentrations (not shown). These features are compatible with the existence of only one Hg-binding site in MerAa and with the co-existence of only two protein states (the *apo*- and Hg-bound forms) during



**Fig. 4.** (A) CD titration of MerAa with  $\text{HgCl}_2$ . CD spectra were recorded after the addition of 0 to 1.25  $\text{Hg}^{2+}$  per MerAa ( $40 \mu\text{M}$ ). The spectrum in dashed line corresponds to the *apo*-protein. Addition of 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875  $\text{Hg}^{2+}$  equivalent is shown in dotted lines. Addition of 1, 1.125 and 1.25  $\text{Hg}^{2+}$  equivalent is shown in solid lines. The arrows point the direction of spectral changes. (B) Plot of the  $\Delta\epsilon$  at 262 nm and 290 nm as a function of the concentration of  $\text{HgCl}_2$  for MerAa at  $40 \mu\text{M}$  (the values are extracted from Fig. 5A).

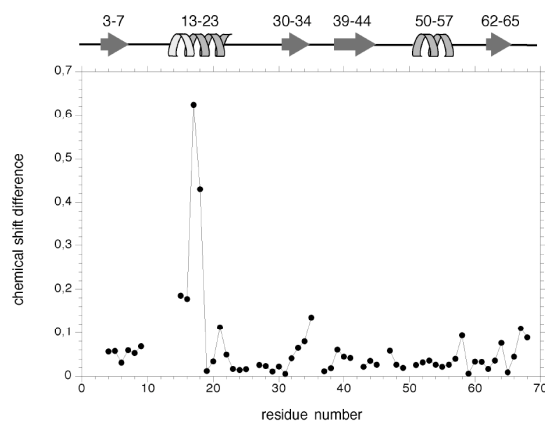
addition of the metal. The CD spectra recorded at different protein concentrations were analyzed with the use of the SPECFIT global fitting software (Spectrum Software Association). This allows the calculation of a  $K_d$  of about  $2.1 \pm 1.9 \mu\text{M}$ . It should be noted that the far UV CD signal recorded as a control during the addition of mercury did not indicate any significant change in the secondary structure of the protein.

#### NMR comparison of *apo*- and metal-bound MerAa:

Backbone assignment of the *apo*- and the metal-bound forms of MerAa has been reported previously [20]. The distribution of the secondary structure elements along the sequence (Fig. 5), as identified from  $^{13}\text{C}$  secondary chemical shifts, confirms that MerAa possesses the typical ferredoxin-like fold found in metallo-chaperones and metal-transporting ATPases [17]. This result is not surprising in view of the high sequence homology of MerAa with MerP (Fig. 1). The fold of *apo*- and metal-bound MerAa are very similar as can be deduced from the chemical shift changes observed upon metal binding. The weighted average chemical shift difference for the backbone amides ( $\langle \Delta_{\text{avg}}^{\text{HN}} = \{(\langle \Delta_{\text{H}} \rangle^2 + (\langle \Delta_{\text{H}} / \langle \Delta_{\text{N}} \rangle)^2)\}^{1/2}$ ) as a function of the protein sequence is shown in Fig. 5. It can be seen that major differences only occur at residues directly adjacent to the metal-ligating cysteines and at the end of  $\beta$ -sheet two (residues 34 and 35). In the MerP-like structures, this part is situated in proximity to the metal-binding site. Note, however, that in our case the largest chemical shift variations occur on histidine-17 and not on the cysteines. This might be due to the reorientation of the aromatic side-chain upon metal-binding. The most significant chemical shift changes are observed for the region comprising residues 10-18. No cross peak could be observed for residues Thr10 to Ala15 in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum indicating that this part of the protein is highly flexible and solvent exposed in the *apo*-form. In the metal-bound form most of the  $^1\text{H}$ - $^{15}\text{N}$  correlation peaks in this protein region become observable, and only Asp12 is still absent from the spectrum. The C-terminal part of helix 1 (residues 13-15) is only well structured in the presence of a bound metal ion. All these observations indicate that the two cysteine residues C11 and C14 are responsible for metal coordination, in agreement with the results from the mass spectrometry. Interestingly, the residues close to the metal ion (loop and C-terminal part of helix 1) keep some

conformational flexibility, as the observed  $^1\text{H}$ - $^{15}\text{N}$  correlation peaks in this region are exchange broadened, and strong water exchange peaks are detected for the amide protons of residues Thr10, Ser13, and His17. This may have some importance for the function of this protein fragment.

**Fold determination of metal-bound MerAa:** In order to obtain a structural model of MerAa, we calculated the fold of the metal-bound form based on long-range methyl-methyl distances derived from a limited number of unambiguously assigned  $^1\text{H}$ - $^1\text{H}$  NOEs, and orientational restraints from backbone RDC data. Complete assignment of methyl  $^1\text{H}$  and  $^{13}\text{C}$  resonances was obtained by correlating the  $^1\text{H}$ , and  $^{13}\text{C}$  side chain frequencies with the previously assigned amide  $^1\text{H}$  and  $^{15}\text{N}$  resonances in 3D (H)C(CO)NH- and H(CCO)NH-TOCSY spectra. Ambiguities were solved by analyzing 2D CT  $^1\text{H}$ - $^{13}\text{C}$  HSQC and 2D (H)C(C)H-TOCSY spectra. The additional 2D spectra also allowed identification of side chain resonances preceding proline residues. As most methyl groups are resolved in the CT  $^1\text{H}$ - $^{13}\text{C}$  HSQC correlation spectrum, a 4D double- $^{13}\text{C}$ -edited  $^1\text{H}$ - $^1\text{H}$  NOESY experiment would allow unambiguous assignment of all NOEs involving resolved methyl groups. This strategy, however, suffers from the intrinsic low



**Fig. 5.** Comparison between *apo*- and metal-bound MerAa. The weighted average chemical shift difference for the backbone amide groups is plotted in function of the protein sequence. Secondary structure elements are indicated on top of the figure. The helical segment shown in light gray has only been identified in metal-bound MerAa.

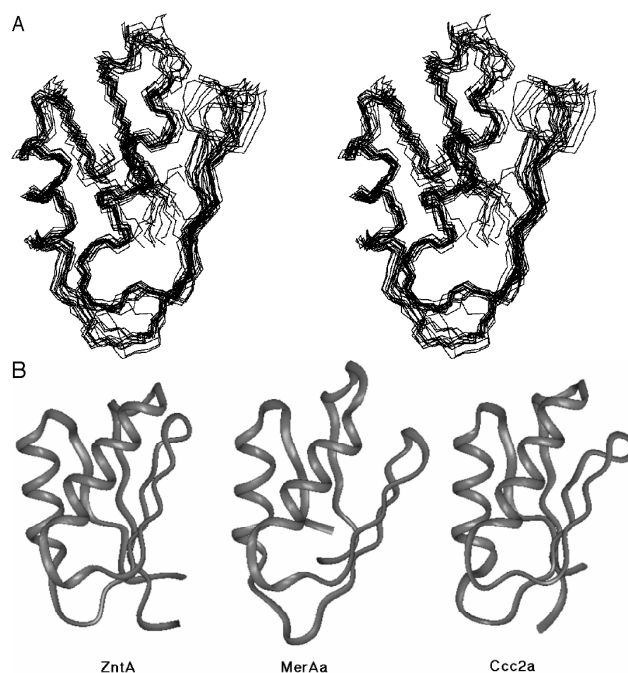


resolution of 4D NMR experiments recorded in a reasonable amount of time. In the case of the 68-residue protein MerAa, the combined use of two 3D NOESY data sets, where either a  $^1\text{H}$  or  $^{13}\text{C}$  NOE dimension is added to the  $^1\text{H}$ - $^{13}\text{C}$  correlation spectrum, proved to be sufficient to obtain a significant number (56) of unambiguously assigned long-range inter-methyl distance restraints. They were used as input to the program Discover (Accelrys Inc.) for the structure calculations together with dihedral constraints for the secondary structure elements identified from the  $^{13}\text{C}$  secondary chemical shifts and hydrogen bond constraints determining the  $\beta$ -sheet topology deduced from the analysis of  $\text{H}^{\text{N}}$ - $\text{H}^{\text{N}}$  NOEs. This initial model (results not shown) was then refined using the program Sculptor [29] by introducing  $D_{\text{NH}}$  and  $D_{\text{CH}}$  RDC data measured along the protein backbone as additional orientational constraints with respect to the molecular alignment frame.

15 structures constitute the final structural ensemble (experimental energies of  $16.2 \pm 1.4$  kcal/mol and  $2.07 \pm 0.44$  kcal/mol from RDC and other constraints, respectively). The backbone-fold of these structures is very well defined with the exception of three loops: the metal-binding loop and the loop connecting  $\beta$ -strand three and helix two are ill defined as they include several residues for which no experimental constraints could be determined. In addition, the loop connecting  $\beta$ -strands two and three shows some slightly different conformations across the ensemble. The average rmsd calculated over the  $\text{C}^{\alpha}$ ,  $\text{N}^{\text{H}}$  and  $\text{C}^{\gamma}$  atoms of residues 4 to 64 compared to the coordinates of the mean structure is  $1.16 \pm 0.21$  Å for the structure determined using NOESY restraints, and  $0.91 \pm 0.13$  Å for the RDC-refined structure. This result reiterates the particular importance of methyl-methyl distances for the definition of the topology of proteins in NOESY-based structure calculations [32-35] and provides a clear demonstration of their use for efficient fold determination.

Figure 6a shows the superposition of the backbone coordinates of the final ensemble. The spatial arrangement of the secondary structural elements confirms the structural homology with MerP. The superposition of the backbone coordinates of our lowest-experimental energy structure with those of the metal-bound first domain of the Ccc2 ATPase [19] and the *E. coli* zinc-transporting protein ZntA [36] is shown in Figure 6b. The rmsd calculated over the backbone atoms of structurally homologous residues are 1.57 and 1.33 respectively.

**Backbone mobility from  $^{15}\text{N}$  relaxation data:** The absence of  $^1\text{H}$ - $^{15}\text{N}$  HSQC peaks for several residues in the *apo*- and metal-bound protein prompted us to study the backbone mobility of the metal-bound form in more detail.  $R_1$ ,  $R_2$ , and  $\{^1\text{H}\}^{15}\text{N}$ -NOE were measured at 600 MHz  $^1\text{H}$  frequency. Relaxation data could be obtained for 53 out of 59 assigned  $^1\text{H}$ - $^{15}\text{N}$  HSQC peaks. Along the protein sequence, the relaxation rates are quite homogenous, with average values of  $2.5 \text{ s}^{-1}$  for  $R_1$ ,  $9 \text{ s}^{-1}$  for  $R_2$  and 0.75 for the heteronuclear NOE. The relaxation data were interpreted according to the model-free approach [37, 38] using the Tensor2 program [39]. The overall rotational motion was characterized using relaxation rates  $R_1$  and  $R_2$  measured for

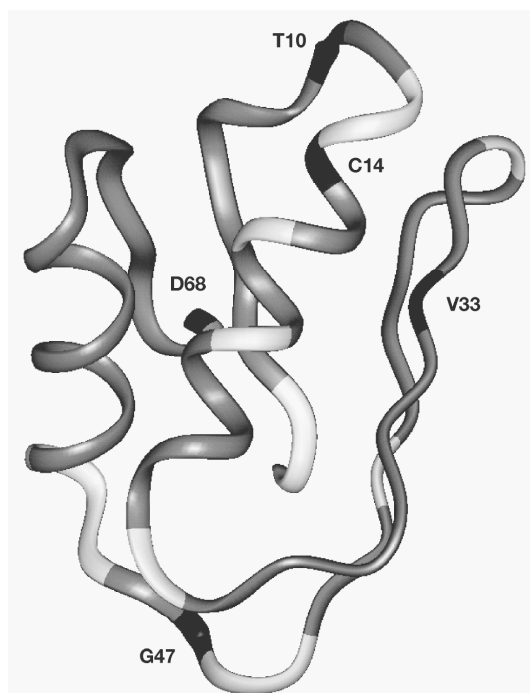


**Fig.6.** Backbone fold of metal-bound MerAa and comparison with homologous structures. (A) Superposition of the backbone atoms of the 15 selected structures. The structural ensemble was obtained from the superposition of the  $\text{N}^{\text{H}}$ ,  $\text{C}^{\alpha}$ , and  $\text{C}^{\gamma}$  atoms of residues 4 to 66 on the structure with the lowest experimental energy. (B) Structural comparison of MerAa (center) with the first soluble metal-binding module of the yeast Ccc2-ATPase (right) and the ZntA protein from *E. coli* (left). The structural alignment has been effectuated using the program Deep View (48).

residues located in secondary structure elements. The data could well be fitted assuming isotropic rotational diffusion with a correlation time of  $\tau_c = 6.2$  ns at  $20^\circ\text{C}$ . No significant improvement of the fit was obtained using an axially symmetric or fully anisotropic diffusion tensor. The results of the model free analysis, generalized order parameters  $S^2$ , internal correlation time constants  $\tau_i$  and conformational exchange rates  $R_{\text{ex}}$ , are shown in Figure S1 of the Supporting Information. All residues with order parameters smaller than a cutoff of  $S^2 = 0.85$  are highlighted in Fig. 7 on the ribbon representation of the MerAa backbone. It can be seen that MerAa globally is a rigid molecule. However, for residues around the metal-binding site, reduced order parameters are observed for residues Thr10 and Cys14 with internal correlation times in the ns range. The relaxation rates of His17 could not be fitted to any of the different models, indicating multiple time-scale flexibility for this N-H bond vector. Cys14 also has a significant conformational exchange contribution  $R_{\text{ex}}$ , indicating additional mobility of this protein region at the  $\mu\text{s}$  to ms time scale. These observations confirm that the metal-binding site keeps some conformational flexibility upon metal binding.

## DISCUSSION

We have expressed a protein corresponding to the first 68 amino acids of the mercuric reductase of *R. metallidurans* CH34 to define both its ability to bind mercury and its structural characteristics. Actually, the C-



**Fig. 7.** Backbone mobility of metal-bound MerAa assuming isotropic rotational diffusion. Order parameter  $S^2$  projected on the ribbon representation of the metal-bound MerAa backbone. Residues with order parameters between 0.85 and 1.0 are shown in gray, with order parameters below 0.85 in black. Residues for which no relaxation information could be obtained are given in white.

terminal border was guided by the identification of a cleavage site giving rise to the partial proteolysis of MerA and from sequence comparison with MerP (Fig. 2). The resulting domain appears to be stable, well folded and compact, as could be deduced from its good expression in *E. coli* and its NMR spectral characteristics. The structural model we constructed using NMR-derived experimental constraints confirmed that MerAa belongs to the family of metallo-chaperones and metal-transporting ATPases as expected. This structural homology and the presence of the metal-binding motif (GMTCXXC) suggest that MerAa could play a role in mercury trafficking.

In order to go further in this hypothesis, we have determined its metal-binding properties using ES-MS and CD spectroscopy. The full conversion of reduced *apo*-MerAa to mercury-bound MerA can be detected directly by using ES-MS, the binding of mercury being strong enough to resist ES-MS conditions. This technique was also used to demonstrate that the two cysteines were involved in metal-binding as alkylation was found to interfere with metal-binding whereas metal-bound MerAa did not react with iodoacetamide. We also determined the apparent dissociation constant of the metal for the protein using circular dichroism spectroscopy in the near UV region. This technique offers the advantage that strict equilibrium conditions can be kept all along the experimental time and that there is no need for radioactive mercury. CD has already been proved to be a suitable method to follow metal binding to metalloproteins [40, 41] such as metallothioneins and was also used recently to characterize the copper binding to CopZ, a copper chaperone from *Bacillus subtilis* [42]. From our CD experiments, we can conclude that only

one site of Hg-binding exists in MerAa and that the transition between the *apo*-form and the metal-bound form occurs without any intermediates as judged from the very tight isodichroic point during titration of MerAa with mercury. In earlier mercury binding studies conducted with MerP, it was observed that the protein exhibits a strong binding site leading to the formation of a 1:1 complex only in the presence of external thiol compounds [7]. This led the authors to suggest the existence of a ternary complex between the protein-mercuric ion complex and a thiol compound. However, in our study it is noteworthy that under both conditions, ES-MS and CD, the reduced protein solution did not contain any external thiols and only one mercury-binding site was titrated. This result makes the hypothesis of a ternary complex unlikely, at least for MerAa. In the structure of mercury-bound MerP, the metal is found in an approximately linear S-Hg-S bound [14]. This linear bicoordination has also been probed by  $^{199}\text{Hg}$  NMR spectroscopy [14]. Due to the considerable similarities between MerP and MerAa, it is very likely that the metal binding-site adopts a similar geometry in this latter case. Analysis of the CD spectra recorded at different protein concentrations allowed the calculation of an apparent  $K_d$  in the  $\mu\text{M}$  range. This value is in perfect agreement with that of the same parameter evaluated for MerP. A  $K_d$  of  $3.7 \pm 1.3 \mu\text{M}$  was determined in pseudo-equilibrium conditions using microconcentrators and radioactive mercury ( $^{203}\text{Hg}$ ) [7]. A  $K_d$  of  $2.8 \mu\text{M}$  determined from the reduction of the CD absorbance at 222 nm upon mercury binding was also reported [14]. In terms of affinity, the value of about  $2.1 \pm 1.9 \mu\text{M}$  determined for MerAa is low enough to be biologically relevant as it also would allow the exchange with a partner such as the catalytic core of the mercuric reductase.

Analysis of protein dynamics showed that metal-bound MerAa globally is a rigid molecule except for the metal binding motif which displays reduced order parameters even in the presence of bound mercury. This behavior could be relevant of a physiological role for this part of the protein in the transfer of the mercury between MerT and the mercuric reductase. It is proposed that the metal exchange involves a series of two- and three-coordinate intermediates between paired thiols in this family of proteins [8, 43]. In this case, the transfer should be considered between the pair of cysteines located in the cytoplasmic loop of the transporter situated between the second and the third transmembrane helices and the cysteine pair of MerAa. Interestingly Cys-14 shows a significant conformational exchange contribution to its  $^{15}\text{N}$  relaxation and could play a major role in such a transfer mechanism. We have also demonstrated that additional residues situated close to the metal binding motif retain some conformational flexibility in the metal-bound form. These residues located in the first loop and in the beginning of helix 1 may be implicated in the global mercury transfer mechanism. A relation between the flexibility of this region and the acquisition of different metals by peptides mimicking the mercury-binding domain of MerP was already underlined in a recent NMR study [44].

The exact role of MerAa remains now to be elucidated. Very recently, a model has been presented showing the

hypothetical interaction between the catalytic core of *Bacillus* MerA and MerP which has been chosen to represent the N-terminal MerA domain [4]. This model was built based on the complementary shapes of MerP and of the cleft where the accessible C-terminal cysteine pair of MerA is situated. Our results clearly confirm the structural homology between MerP and MerAa, and therefore suggest that MerAa could be able to interact with the catalytic core and to directly deliver the mercury ion to the catalytic site of the mercuric reductase.

**Acknowledgments:** We are very grateful to David Lemaire and David Lascoux for their help in mass spectroscopy experiments, to Jean-Pierre Andrieu for N-terminal sequencing, to Elizabeth Mintz and Vincent Forge for the access to the CD apparatus and the use of the *SPECFIT* software and finally to Tatiana Valleys and Max Mergeay for fruitful discussions and constant encouragements. ER and LC were supported by PhD grants from the Région Rhône-Alpes and the Commissariat à l’Energie Atomique, respectively.

## REFERENCES

- Silver S, Phung LT (1996) *Ann Rev Microbiol* 50:753-789
- Hobman JL, Brown NL (1997) *Metal Ions Biol Syst* 34:527-568
- Liebert CA, Hall RM, Summers AO (1999) *Microbiol Mol Biol Rev* 63:507-522
- Barkay T, Miller SM, Summers AO (2003) *FEMS Microbiol Rev* 27:355-384
- O’Halloran TV, Walsh CT (1987) *Science* 235:211-214
- Brown NL, Stoyanov JV, Kidd SP, Hobman JL (2003) *FEMS Microbiol Rev* 27:145-163
- Sahlman L, Jonsson B-H (1992) *Eur J Biochem* 205:375-381
- Brown NL, Camakaris J, Lee BT, Williams T, Morby AP, Parkhill J, Rouch DA (1991) *J Cell Biochem* 46:106-114
- Fox B, Walsh CT (1982) *J Biol Chem* 257:2498-2503
- Moore MJ, Walsh CT (1989) *Biochemistry* 28:1183-1194
- Brown NL, Shih Y-C, Leang C, Glendinning KJ, Hobman JL, Wilson JR (2002) *Biochem Soc Trans* 30:715-718
- Sedlmeier R, Altenbuchner J (1992) *Mol Gen Genet* 236:76-85
- Shiering N, Kabsch W, Moore MJ, Distefano MD, Walsh CT, Pai EF (1991) *Nature* 352:168-172
- Steele RA, Opella SJ (1997) *Biochemistry* 36:6885-6895
- Qian H, Sahlman L, Eriksson P-O, Hambræus C, Edlund U, Sethson I (1998) *Biochemistry* 37:9316-9322
- Opella SJ, Desilva TM, Veglia G (2002) *Curr Opin Chem Biol* 6:217-223
- Arnesano F, Banci L, Bertini I, Ciofi-Baffoni S, Molteni E, Huffman DL, O’Halloran TV (2002) *Genome Res* 12:255-271
- Mergeay M, Monchy S, Vallaëys T, Auquier V, Benotmane A, Bertin P, Taghavi S, Dunn J, van der Lelie D, Wattiez R (2003) *FEMS Microbiol Rev* 27:385-410
- Banci L, Bertini I, Ciofi-Baffoni S, Huffman DL, O’Halloran TV (2001) *J Biol Chem* 276:8415-8426
- Bersch B, Rossy E, Covès J, Brutscher B (2003) *J BioMol NMR* 27:57-67
- Laemmli UK (1970) *Nature* 227:680-685
- Zhu G, Bax A (1990) *J Magn Reson* 101:114-119
- Grzesiek S, Anglister J, Bax A (1993) *J Magn Reson B* 101:114-119
- Bax A, Mehlkopf AF, Smidt J (1979) *J Magn Reson* 35:167-169
- Rückert M, Otting G (2000) *J Am Chem Soc* 122:7793-7797
- Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE (1994) *Biochemistry* 34:868-878
- Tsan P, Hus J-C, Caffrey M, Marion D, Blackledge M (2000) *J Am Chem Soc* 121:2311-2312
- Blackledge M, Medvedeva S, Poncin M, Guerlesquin F, Bruschi M, Marion D (1995) *J BioMol NMR* 245:661-681
- Hus J-C, Marion D, Blackledge M (2000) *J Mol Biol* 298:927-936
- Sibille N, Pardi A, Simorre J-P, Blackledge M (2001) *J Am Chem Soc* 123:12135-12146
- Gurd F (1972) *Methods Enzymol* 25:424-438
- Metzler WJ, Wittekind M, Goldfarb V, Mueller L, Farmer BT (1996) *J Am Chem Soc* 119:6800-6801
- Smith B, Ito Y, Raine A, Teichmann S, Ben-Tovim L, Nietlispach D, Broadhurst RW, Terada T, Kelly M, Oschkinat H, Shibata T, Yokoyama S, Laue E (1996) *J BioMol NMR* 8:360-369
- Gardner K, Rosen MK, Kay LE (1997) *Biochemistry* 36:1389-1401
- Gardner K, Kay LE (1998) *Ann Rev Biophys Biomol Struc* 27:357-406
- Banci L, Bertini I, Ciofi-Baffoni S, Finney LA, Outten CE, O’Halloran TV (2002) *J Mol Biol* 323:883-897
- Lipari G, Szabo A (1982a) *J Am Chem Soc* 104:4546-4559
- Lipari G, Szabo A (1982a) *J Am Chem Soc* 104:4559-4570
- Dosset P, Hus J-C, Blackledge M, Marion D (2000) *J BioMol NMR* 16:23-28
- Lu W, Zelazowski AJ, Stillman MJ (1993) *Inorg Chem* 32:919-926
- Lu W, Stillman MJ (1993) *J Am Chem Soc* 115:3291-3299
- Kihlken MA, Leech AP, Le Brun NE (2002) *Biochem J* 368:729-739
- Morby AP, Hobman JL, Brown NL (1995) *Mol Microbiol* 17:25-35
- DeSilva T, Veglia G, Porcelli F, Pranter AM, Opella SJ (2002) *Biopolymers* 64:189-197
- Guex N, Peitsch MC (1997) *Electrophoresis* 18:2714-2723

## 2.4 Conclusion

Le double-marquage des protéines permet d'accélérer significativement le processus d'attribution et d'obtention de contraintes structurales. En absence de marquage au carbone-13, l'attribution complète des systèmes de spin est nécessaire avant d'accéder à l'information structurale. Dans le cas d'une protéine doublement marquée, l'attribution des résonances de la chaîne principale devient l'étape clé. L'utilisation des expériences à dimensionalité réduite de type HSQC permet de réduire le temps expérimental significativement par rapport à l'approche standard basée sur les expériences à trois dimensions. Une fois l'attribution de la chaîne principale connue, les éléments de structure secondaire peuvent être localisés [chemical shift index, CSI (57)] et leur l'orientation relative peut être comparé avec d'autres structures à l'aide des constantes de couplage dipolaire résiduel. En parallèle, on peut obtenir des informations sur la dynamique de la protéine par des mesures de relaxation des azotes de la chaîne principale. L'exemple de l'étude de MerAa montre que le marquage des protéines aux  $^{15}\text{N}$  et  $^{13}\text{C}$ , combiné avec des approches originales pour l'acquisition des données et les calculs de structure, permet d'obtenir des informations sur le repliement dans des délais très courts, de l'ordre de la semaine.

En principe, il est possible de calculer le repliement de la chaîne principale uniquement à partir des couplages dipolaires résiduels, si ceux-ci peuvent être mesurés dans deux milieux différents (58,59). Cependant, l'utilisation des distances proton-proton déterminées à partir de NOE apporte toujours de l'information précieuse sur le repliement global de la protéine. Dans le cas d'une petite protéine doublement marquée, la détermination de NOE est relativement facile□ la relaxation est peu importante et les résonances sont en général bien résolues, surtout si on utilise une édition azote-15 ou carbone-13. Nous avons montré que des NOE à longue distance entre des groupements méthyle peuvent être attribués d'une manière non-ambiguë en exploitant deux expériences NOESY édités carbone-13 complémentaires. Cette information en combinaison avec la localisation des éléments de structure secondaire et la topologie de feuillets-□ permet de calculer rapidement le repliement de la chaîne principale avec une bonne résolution.

En conclusion, l'utilisation d'un échantillon doublement marqué est toujours préférable, même pour les protéines avec des masses inférieures à 10kDa. Cependant, la biologie moléculaire a également ses limites. Les étapes de clonage et de la mise au point des conditions de surexpression peuvent nécessiter beaucoup de temps et il n'est pas toujours possible d'obtenir des quantités suffisantes en protéine à un coût acceptable. Dans ce cas, l'approche homonucléaire reste une alternative comme le montre les deux premiers exemples, qui sont représentatifs pour un grand nombre d'études similaires effectuées depuis une vingtaine d'années dans tous les laboratoires de RMN biomoléculaire.

### 3 Etudes des protéines > 15kDa par RMN

Avec l'augmentation de la taille des protéines étudiées par RMN, l'obtention des données structurales par l'approche habituelle décrite ci-dessus devient beaucoup plus difficile et une autre stratégie doit être adaptée.

Premièrement, la molécule doit être deutériée, au moins partiellement, afin de diminuer la relaxation des carbones et d'obtenir des spectres avec une meilleure sensibilité (9, 10). Dans ces conditions, l'attribution de la chaîne principale est relativement aisée, même pour des protéines de plus de 200 acides aminés. Cependant, pour des protéines de cette taille, l'attribution des chaînes latérales est beaucoup plus difficile car elle est en général basée sur la présence des hydrogènes attachés aux carbones. Dans le cas de la protéine Sir-FP18 de 18kDa, l'attribution complète des chaînes latérales a pu être effectué en utilisant un échantillon protoné (voir 3.2), mais ceci n'a pas été possible pour la protéine Fur, un dimère de 2x17kDa (voir 3.1) et pour la protéine MsrA de 25kDa (voir 3.3) en utilisant les méthodes standard et des marquages uniformes.

Deuxièmement, l'information structurale ne pourra plus venir essentiellement des NOE. Comme nous l'avons vu, l'attribution correcte des NOE nécessite une attribution complète de tous les protons de la molécule (ou de tous les groupements méthyles si on se limite à ce type d'interaction). Or, pour des grosses protéines, ceci est difficile à réaliser. Le nombre croissant de protons entraîne une dégénérescence plus importante des fréquences et le nombre de NOE non-ambigus diminue, ce qui peut empêcher la détermination du repliement correct de la protéine. Enfin, la diffusion de spin, qui augmente également avec le poids moléculaire, peut introduire des larges erreurs dans les contraintes de distance dues au transfert relayé entre protons. Dans le cas d'une deutération partielle ou totale de la protéine, la relaxation et la diffusion de spin sont diminuées significativement. Ceci permet d'utiliser des temps de mélange plus longs et d'observer des NOE entre protons à des distances de plus que 6Å (60). Cependant, le nombre des NOE observables diminue et l'information disponible ne peut pas permettre le calcul de la structure avec une bonne résolution.

Une solution possible à ce dernier problème est la réintroduction de protons en certains sites dans une molécule par ailleurs perdeutériée. Différentes approches ont été décrites dans la littérature. L'utilisation de  $^{13}\text{C}$ -pyruvate comme seule source de carbone en combinaison avec du  $\text{D}_2\text{O}$  permet d'obtenir une protéine dans laquelle les groupements méthyles des leucines, alanines, valines et les méthyles  $\beta$  des isoleucines sont protonés (61). Cependant, différents isotopomères des méthyles sont produits, diminuant la résolution et la sensibilité de l'expérience RMN. De plus, le rendement de surexpression est deux fois plus faible qu'avec du glucose. Une autre façon est l'addition d' $\alpha$ -ketobutyrate et d' $\alpha$ -ketoisovalérate (62), précurseurs spécifiques des leucines, valines et isoleucines, en complément du  $^2\text{H}$ -,  $^{13}\text{C}$ -glucose au milieu du culture. Ceci résulte en une incorporation de leur groupement méthyle dans  $\geq 90\%$  des alanines, valines, leucines et isoleucines (seulement en  $\beta$ ). Dans un premier temps, ces précurseurs ont été marqués uniformément  $^{13}\text{C}$  avec des  $^2\text{H}$  en position 3 (62), résultant en une protéine marquée uniformément  $^{13}\text{C}$  et  $^2\text{H}$ , mais portant des protons sur les méthyles des acides aminés cités. Plus tard, le marquage sélectif au  $^{13}\text{C}$  des méthyles de ces précurseurs dans une molécule protonée a été proposée (63, 64). L'utilisation de ceux-ci en complément du  $^2\text{H}$ -,  $^{12}\text{C}$ -glucose dans du  $\text{D}_2\text{O}$ , permet de produire une protéine dans laquelle seuls les groupements méthyles sont marqués  $^{13}\text{C}$  et  $^1\text{H}$ . Cette dernière méthode a comme avantage d'augmenter encore la résolution et la sensibilité car les couplages  $J$   $^{13}\text{C}$ - $^{13}\text{C}$  sont absents. D'une manière équivalente, la phénylalanine marquée sélectivement au  $^{13}\text{C}$  en position  $\alpha$  a été utilisée pour l'attribution et l'obtention de contraintes

NOE (65) et très récemment, une nouvelle approche a été proposée utilisant l'acide shikimique protoné comme précurseur pour une protonation complète des cycles aromatiques (66).

Même s'il a été montré que le repliement d'une protéine peut être obtenu seulement à partir des NOE entre amides, méthyles et aromatiques combinés avec des angles dièdres à partir du CSI (28, 29), la détermination des structures à haute résolution reste très improbable par cette méthode. Dans ce contexte, l'utilisation d'un nouveau type d'information structurale basée sur les mesures des constantes de couplage dipolaire résiduel a révolutionné la RMN (8, 67). L'amplitude de ces couplages dipolaires est reliée à l'orientation d'un vecteur liant deux spins par rapport à un système référentiel commun à la molécule. Il en résulte que ces contraintes ont un caractère orientant et ordonnant, car elles sont toutes reliées au même système de référence, indépendamment de leur localisation dans la molécule. Cette propriété les rend complémentaires des NOE et couplages J qui ne contiennent que de l'information locale. Différents couplages peuvent aisément être mesurés pour les noyaux de la chaîne principale ( $H^N$ , N,  $C\alpha$ ,  $C'$ ,  $C\beta$ ) fournissant de l'information structurale précieuse, même dans le cas d'une protéine perdeutériée. Par exemple, nous avons récemment mesuré des constantes de couplage dipolaire résiduel pour les groupements méthyles d'une protéine deutériée à 75% dans deux milieux différents. Ces données permettent de raffiner l'orientation des chaînes latérales et, dans le cas de valines, de déterminer l'angle  $\chi_1$  ainsi que l'attribution stéréospécifique des deux méthyles (68).

Dans les trois parties qui vont suivre, je présenterai trois de mes sujets principaux depuis 2000 : l'étude par RMN des protéines bactériennes Fur (ferric uptake regulator), MsrA (méthionine sulfoxide réductase) et SiR-FP18, le domaine de liaison du FMN de la sulfite réductase. Ces travaux ont été effectués par des étudiants en thèse que j'ai encadrés en co-direction avec des chercheurs de notre laboratoire ou de laboratoires extérieurs.

Pour deux des protéines étudiées (Fur et MsrA), des structures cristallographiques ont été déterminées avant que nos études soient suffisamment avancées pour calculer une structure. Cependant, l'étude de ces protéines par RMN a pu fournir des informations précieuses complémentaires : dans le cas de la protéine Fur, nous avons pu élucider le mécanisme de dimérisation et pour la protéine MsrA, les mesures des couplages dipolaires résiduels ont permis d'obtenir des informations détaillées sur le site actif de l'enzyme. Dans le cas de la protéine SiR-FP18, des couplages dipolaires résiduels ont été utilisés pour l'affinement d'un modèle structural et également pour le calcul d'une structure à haute résolution en combinaison avec l'information «classique» (NOE, couplages J et liaisons hydrogènes).

### 3.1 Etude par RMN de la protéine Fur d'*Escherichia coli* du monomère au dimère

Collaboration avec Isabelle Michaud-Soret du Laboratoire de Physicochimie des Métaux en Biologie (UMR 5155 CEA-CNRS-UJF), CEA-Grenoble. Une partie du travail sur la protéine Fur a été effectuée par Ludovic Pecqueur, étudiant d'abord en DEA et maintenant en thèse que j'encadre avec I. Michaud-Soret.

Bien qu'indispensable à la vie de presque toutes les cellules vivantes, le fer est presque insoluble à l'état ferrique et toxique à l'état ferreux (induction de la formation de radicaux libres très mutagènes). C'est pourquoi la régulation de la concentration intracellulaire en fer est essentielle. Chez les bactéries Gram négatives (telle que *Escherichia coli*), Fur (ferric uptake regulator) est la protéine clé de cette



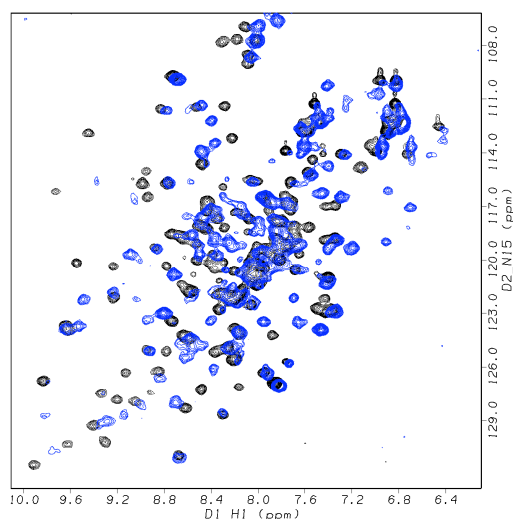


Figure 8. Activation de la protéine Fur\_Ecoli par le zinc. Spectres  $^{15}\text{N}$ -HSQC de la protéine dimérique avant (noir) et après (bleu) addition de 2 équivalents de zinc.

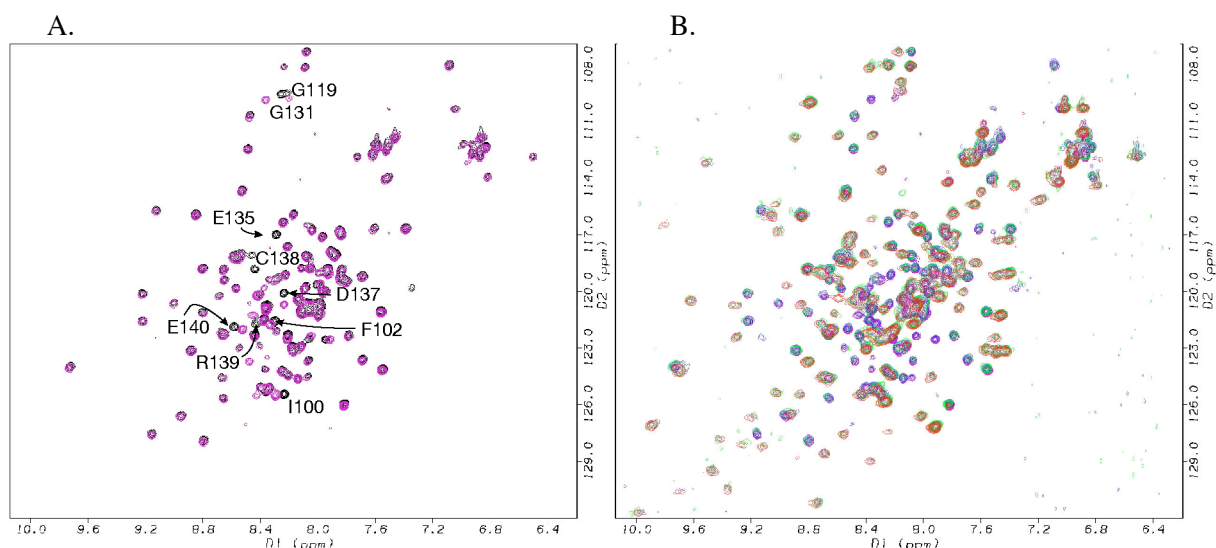
90% des résonances de la chaîne principale et  $\approx 40\%$  des résonances  $^1\text{H}$  et  $^{13}\text{C}$  des chaînes latérales de la forme *apo* dimérique ont été attribuées à l'aide des expériences triple résonance. Pour l'attribution de la chaîne principale, un échantillon deutérié à 50% a été utilisé. L'attribution des chaînes latérales a été très difficile dû la relaxation importante de la protéine. La connaissance des déplacements chimiques  $^{13}\text{C}$  a permis de déterminer la structure secondaire de la protéine. Ainsi, nous avons pu confirmer que Fur fait partie des protéines contenant un motif hélice-coude-hélice ailé (78) dans leur partie N-terminale, responsable de leur interaction avec l'ADN (Figure 7). Une première analyse des NOE  $\text{H}^{\text{N}}\text{-H}^{\text{N}}$  a également montré la topologie des feuillets  $\beta$  les petits brins sont organisés en deux paires antiparallèles et le grand brin C-terminal forme un feuillet antiparallèle à l'interface entre les deux unités monomériques. D'une manière générale, ces données sont en bon accord avec la structure cristallographique. Cependant, les déplacements chimiques des carbones ainsi que les données de relaxation mesurés sur Fur\_Ecoli dimérique et non-activée ne sont pas en accord avec la présence d'une hélice  $\alpha$  en N-terminal du motif hélice coude-hélice ailé. Les NOE hétéronucléaires montrent même une mobilité très importante des résidues 5 à 8 (voir Figure 10). Pourtant, l'analyse de la séquence de Fur\_Ecoli avec l'algorithme Phd (B. Rost, Columbia University) confirme qu'il y a une forte probabilité pour que les résidus Asp3 à Lys10 adoptent une conformation en hélice  $\alpha$  (voir ci-dessous).

L'activation de la protéine par le zinc a été suivie par RMN (Figure 8). Nous avons utilisé le zinc plutôt que le fer pour ses propriétés diamagnétiques. La comparaison des spectres HSQC montre que la protéine subit un changement de conformation léger, mais ceci dans sa globalité. Nous envisageons d'attribuer la forme métallée afin de mieux pouvoir analyser les variations structurales induites par l'activation.

Lors de la purification de la protéine Fur, une fraction de la protéine est obtenue sous forme monomérique. De plus, nous avons observé une transformation très lente de la protéine dimérique vers la forme monomérique dans nos échantillons RMN. Ceci nous a conduit à étudier les facteurs déterminants pour la dimérisation.

Des études de mutagenèse dirigée ont montré l'importance des deux cystéines 93 et 96 pour l'activité de la protéine Fur (79). Par la suite il a été démontré que ces deux cystéines lient un atome de

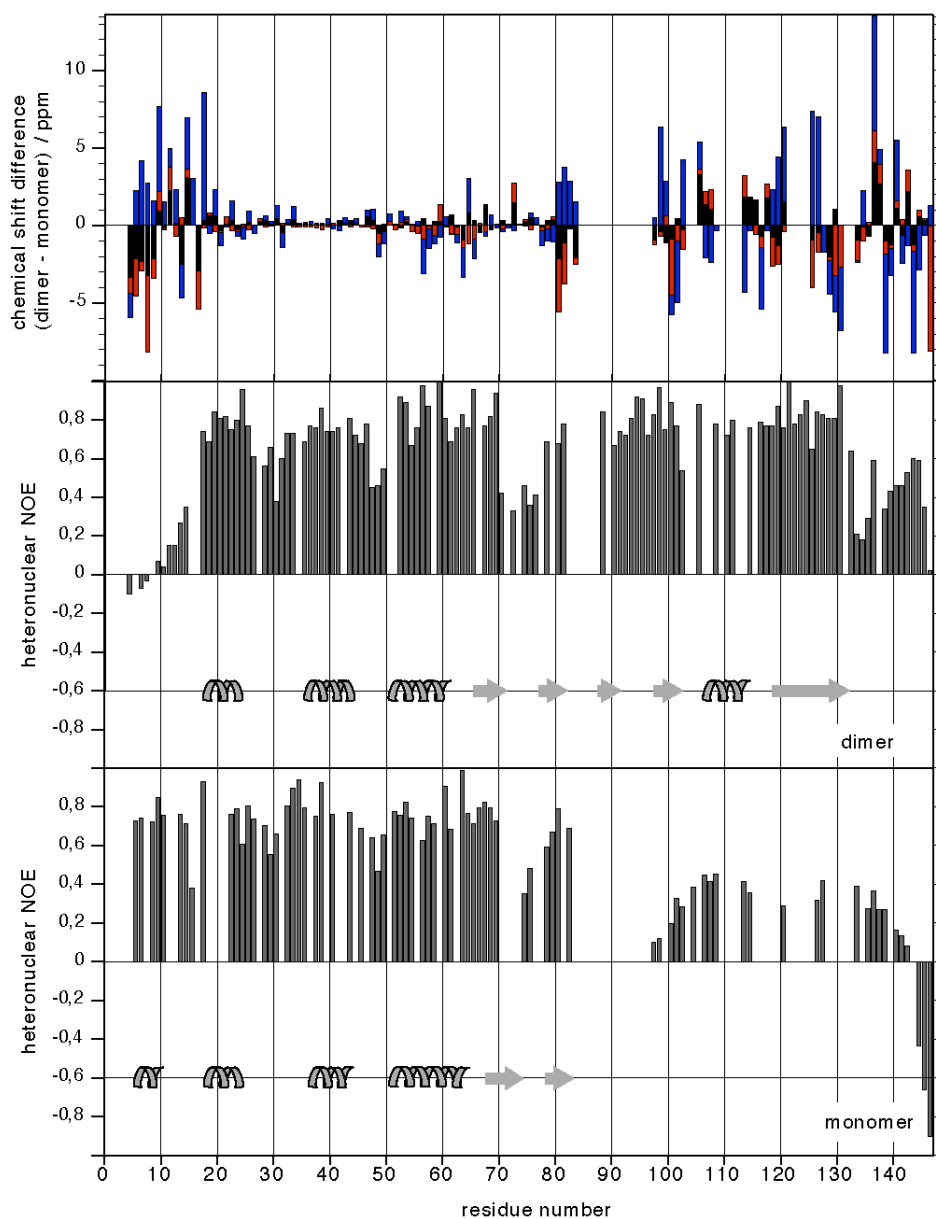




**Figure 9.** Effets structuraux dus à l'addition de DTT (gauche) et du zinc (droite) à *Fur\_Ecoli* monomérique. Noir  $\square$  *Fur\_Ecoli* monomère oxydé  $\square$  mauve  $\square$  *Fur\_Ecoli* monomère réduite  $\square$  bleu  $\square$  *Fur\_Ecoli* monomère réduite + 0.5 équivalents molaires de zinc, vert  $\square$  + 1.5 équivalents molaires de zinc  $\square$  rouge  $\square$  *Fur\_Ecoli* dimère, forme apo.

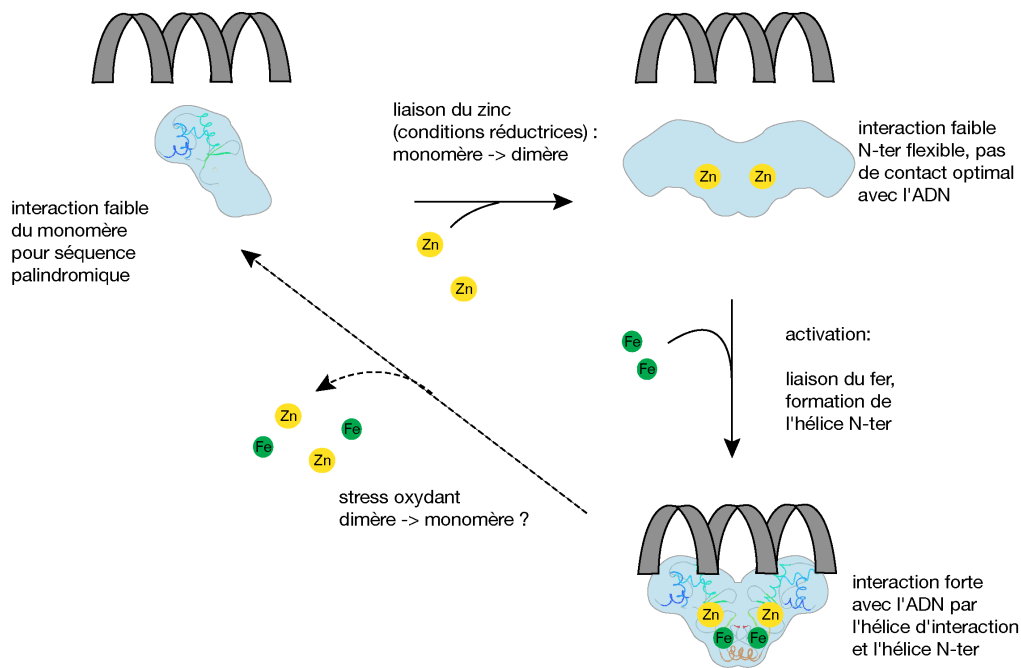
zinc dont le rôle est la stabilisation conformationnelle de la protéine. Les cystéines étant sensibles à l'oxydation, nous avons supposé que la perte du zinc, suite à une oxydation des deux cystéines, pourrait induire le passage dimère-monomère que nous observions. En effet, une transformation monomère-dimère a pu être observée par gel filtration après l'addition d'un réducteur (DTT) à un échantillon de la protéine monomérique. Nous avons donc étudié les variations structurales de *Fur* monomère en fonction de l'état d'oxydation, en présence et en absence de zinc. Tandis que l'ajout du DTT n'entraîne que quelques changements localisés dans la séquence primaire autour des cystéines, l'ajout du zinc à la protéine réduite induit la dimérisation de la protéine (Figure 9). En effet, l'augmentation de la concentration en zinc conduit à l'apparition du spectre HSQC de la forme dimérique *apo* (sans métal dans le site d'activation, Figure 9B). Ceci indique également que les formes mono- et dimérique sont en échange lent par rapport à l'échelle de la RMN.

Nous avons attribué la majeure partie de la chaîne principale de la forme monomérique et mesuré les NOE hétéronucléaires. L'ensemble de ces résultats est résumé dans la Figure 10. La comparaison des déplacements chimiques indique que le motif hélice-coude-hélice ailé est certainement maintenu dans une conformation très similaire entre les deux formes (mono- et dimère non-activés). La partie C-terminale de la protéine subit des changements structuraux plus importants. En absence du zinc, cette partie paraît beaucoup plus flexible  $\square$  l'analyse du CSI (chemical shift index) ne permet pas d'établir la présence des éléments de structure secondaire et les valeurs plus faibles des NOE hétéronucléaires indiquent une mobilité accrue de la chaîne principale. Cependant, le résultat le plus surprenant était la confirmation de la présence de l'hélice  $\square$  en N-terminal du motif hélice-coude-hélice ailé dans la forme monomérique. Ceci a clairement pu être établi par l'analyse des déplacements chimiques des carbones ainsi que par les NOE hétéronucléaires, qui indiquent que les résidus 6 à 10 font partie du noyau rigide de la protéine monomérique.



**Figure 10.** Du haut en bas □ variations des déplacements chimiques entre les deux formes. Bleu □  $^{15}\text{N}$  □ rouge □  $^{13}\text{C}$  □, □ noir □  $^{13}\text{C}'$ . NOE hétéronucléaire mesuré pour la forme dimérique non-activée et la forme monomérique. Ne sont données que les valeurs pour lesquelles l'intensité du pic a pu être mesurée avec fiabilité. Les éléments de structure secondaire déterminés à partir de l'analyse des déplacements chimiques sont montrés pour les deux formes.

En conclusion, les résultats que nous avons obtenus par RMN ont permis une meilleure caractérisation de la protéine Fur dans ses différents états. Nous avons pu démontrer l'importance du zinc structural pour la stabilisation de la forme dimérique et une restructuration lors de l'activation de la protéine avec une perte de l'hélice  $\alpha$  N-terminale. Ce dernier point reste à être confirmé par l'étude de la protéine dimérique et active en solution. Basé sur nos données, nous pouvons proposer le schéma illustré dans la Figure 11.



**Figure 11.** Représentation schématique de la régulation de l'activité de la protéine Fur. Fur ayant lié le fer est active et se lie spécifiquement à l'ADN (bas). La perte du fer induit un changement de conformation et une déstructuration de la partie N-terminale ce qui provoque une baisse d'affinité et une dissociation de l'ADN (en haut à droite). Dans des conditions de stress oxydant, la perte du zinc est à l'origine de la formation de la protéine monomérique n'étant plus capable d'interagir avec l'ADN avec une forte affinité (gauche). Ce processus est réversible car la réduction des deux cystéines permet de lier à nouveau le zinc et à la protéine de dimériser (haut). L'implication de la transformation monomère-dimère dans le processus de régulation de l'activité est pour l'instant hypothétique.

Une interaction forte avec l'ADN nécessite la présence d'un dimère en combinaison avec le positionnement correct des éléments qui interagissent avec l'ADN. Dans ce contexte, deux facteurs déterminant l'activité de la protéine peuvent être envisagées (i) la liaison du fer modifie la conformation ou le positionnement relatif des deux motifs hélice-coude-hélice et (ii) la liaison du zinc permet la dimérisation de la protéine.

Une différence majeure entre la forme dimérique inactivée que nous avons étudiée par RMN et la forme dimérique active étudiée par cristallographie est l'hélice  $\alpha$  N-terminale. Mon modèle de travail, schématisé dans la Figure 11, prédit une interaction de cette hélice avec l'ADN. En effet, cette hélice contient à son C-terminal un ou deux résidus basiques (Arg8 et Lys9 dans Fur\_Psae et Lys9, Lys10 dans Fur\_Ecoli) qui sont hautement conservés. Une interaction de ces résidus avec l'ADN chargé négativement est donc fortement probable. Une telle interaction entre la partie en N-terminal du motif hélice-coude-hélice avec l'ADN a déjà été observée pour les protéines FadR (80), E2F4 (81) et a récemment pu être démontrée au laboratoire pour la protéine BlaI, un régulateur de transcription de la  $\beta$ -lactamase. Dans ce dernier cas, trois zones ont été délimitées qui interagissent directement avec l'ADN les résidues en N-terminal, l'hélice 3 et le feuillet  $\beta$  (82). Dans leur modèle, les auteurs proposent que les résidus N-terminaux interagissent d'une manière spécifique avec le petit sillon de l'ADN (82).

Concernant le rôle du zinc dans la dimérisation de la protéine Fur\_Ecoli, il est tentant de suggérer que des conditions de stress oxydant puissent induire l'inactivation de la protéine par sa

transformation à l'état monomérique. La protéine Fur est impliquée dans la réponse au stress oxydant, mais actuellement il n'y a pas d'évidence pour l'existence du monomère *in vivo*. Cependant, dans le cas de la protéine FurS de *Streptomyces reticuli*, une protéine Fur-like, il a été démontré que la capacité de se lier à l'ADN dépend de l'état redox de deux cystéines (83), qui correspondent à Cys93 et Cys96 dans la séquence de Fur\_Ecoli. Les auteurs ont également démontré qu'une de ces deux cystéines est impliquée dans la liaison d'un atome de zinc et proposent, qu'un changement conformationnel de la protéine se produit suite à l'oxydation et la perte du zinc, empêchant ainsi la liaison de la protéine à l'ADN (84). Pour l'instant, il n'a pas encore été vérifié si la protéine FurS est monomérique en absence du zinc, ce qui pourrait expliquer sa très faible affinité pour l'ADN. Dans la littérature, des exemples ont été décrits où l'affinité des protéines contenant un motif hélice-coude-hélice ailé pour l'ADN est régulée par une transformation dimère-monomère. Ceci est le cas par exemple pour deux inhibiteurs de transcription, MecI (85) et BlaI (86) de *Staphylococcus aureus* où un clivage protéolytique induit la dissociation du dimère actif. Au vu de ces données, on ne peut donc pas exclure que cette voie soit également utilisée dans le cas de la protéine Fur. L'équipe d'I. Michaud-Soret étudie actuellement la réponse de la protéine Fur\_Ecoli au stress oxydant *in vivo*.

En ce qui concerne les différents états de la protéine Fur, la RMN nous a aidé à avancer en employant des méthodes plutôt classiques (détermination des déplacements chimiques et leur comparaison, analyse des NOEs et de la relaxation). Au stade actuel nous envisageons de compléter les résultats par une analyse de la forme métallée afin de vérifier la présence de l'hélice  $\alpha$  N-terminal en solution. La possibilité de pouvoir obtenir les différentes formes (activée ou non) permet de nous intéresser plus précisément au mécanisme d'activation. Nous supposons qu'il implique une réorganisation des éléments de structure secondaire, et notamment du motif hélice-coude-hélice, au sein du dimère. L'analyse de couplages dipolaires résiduels permettrait de déterminer l'orientation relative de ces éléments de structure secondaire. Ces informations seront complétées par des NOE méthyls-méthyls (compare 2.3.2) pour tenter d'obtenir une structure tridimensionnelle. Enfin, je souhaite caractériser l'interaction de la protéine Fur avec l'ADN. Pour cela j'envisage d'isoler le domaine N-terminal et de l'étudier en interaction avec la séquence du promoteur connue. Malgré la faible affinité de ce fragment monomérique, cette approche a permis de caractériser l'interaction spécifique entre la protéine BlaI et son site de reconnaissance sur l'ADN à des concentrations employées pour la RMN. J'espère que dans l'année à venir nous serons capables de rassembler les morceaux de ce puzzle que nous tenons actuellement en main.

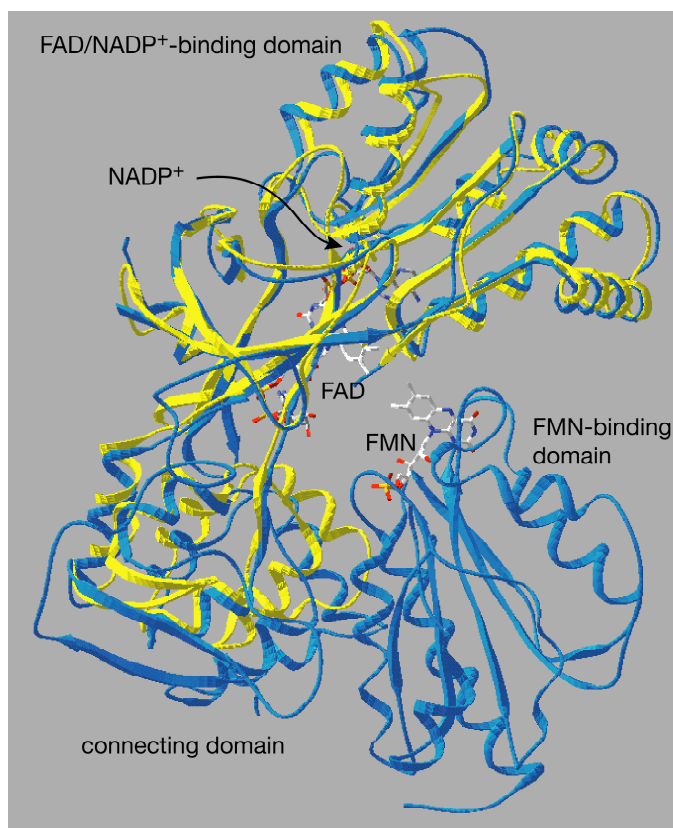
### 3.2 Le domaine flavodoxine de la sulfite réductase d'*Escherichia coli*

L'étude de la sulfite réductase fait partie du travail de thèse de Nathalie Sibille, que j'ai co-encadrée avec Bernhard Brutscher dans le cadre d'une collaboration avec J. Covès (Laboratoire de Chimie et Biochimie des Centres Redox, DRFMC, CEA-Grenoble, maintenant au Laboratoire des Protéines Membranaires à l'IBS) qui a travaillé sur cet enzyme depuis plusieurs années (87, 88).

La sulfite réductase d'*Escherichia coli* est essentielle pour l'assimilation du sulfate dans ces bactéries et catalyse la réduction à six électrons du sulfite ( $\text{SO}_3^-$ ) en sulfure ( $\text{S}^2-$ ), seule forme du soufre utilisable dans la biosynthèse de la cystéine, la méthionine et leurs dérivés soufrés. Cet enzyme est une hémoflavoprotéine multimérique, composé de 8 sous-unités  $\alpha$  et 8 sous-unités  $\beta$  (88). Chaque sous-unité  $\alpha$  contient un centre  $\text{Fe}_4\text{-S}_4$  et un sirohème. La sous-unité  $\beta$  contient un FAD et un FMN comme groupements prosthétiques ainsi qu'un site de liaison pour le NADP. Au sein de la sous-unité  $\beta$ , les électrons sont transférés du NADPH vers le FAD et le FMN qui les donne aux centres métalliques de

la sous-unité  $\beta$  où a lieu la réduction du sulfite. La séquence et les cofacteurs suggèrent que la sous-unité  $\beta$  de la sulfite réductase appartient à la même famille de protéines que la cytochrome-P450 réductase, la NO synthase et le flavocytochrome-P450 BM3 bactérien. Dernièrement, la structure tridimensionnelle d'une forme monomérique et active de la sous-unité  $\beta$  a été résolue au Laboratoire de Cristallographie et Cristallogénèse des Protéines, LCCP, à l'IBS (89). Cependant, dans les quatre formes cristallines obtenues, le domaine N-terminal qui lie le FMN était absent des cartes de densité électronique (voir Figure 12), bien que différentes analyses effectuées sur des cristaux dissous indiquaient sa présence dans la protéine. Ceci a été interprété par une mobilité du domaine N-terminal par rapport au reste de la protéine. Vu la taille de ce domaine (167 résidues), j'ai entrepris une collaboration avec J. Covès afin de compléter la structure par l'étude RMN du domaine manquant. J'ai donc cloné la partie de l'ADN codant pour les résidus A52 à A218, correspondant au domaine flavodoxine, nommé SiR-FP18 d'après son poids moléculaire apparent sur gel SDS en conditions dénaturantes, et j'ai mis au point l'expression de la protéine en milieu minimum. La purification ainsi que la caractérisation biochimique ont été effectuées par L. Champier au DRDC (90, Article 5).

Le travail de thèse de Nathalie Sibille est divisé en deux parties : une partie consacrée à l'étude structurale de SiR-FP18, une deuxième partie plus méthodologique (mesures des couplages dipolaires résiduels pour les méthyles des chaînes latérales, 68). Concernant la structure, nous avons tout d'abord déterminé le repliement global à l'aide des couplages dipolaires résiduels mesurés sur la chaîne principale (90, voir Article 5). Actuellement, Nathalie Sibille est en train de finir l'affinement de la structure dont je ne parlerai pas ici.



**Figure 12.** Superposition des structures cristallographiques de la cytochrome-P450 réductase du rat (bleu) et de SiR-FP60 (jaune). Le domaine FMN de SiR-FP60 est absent de la structure.

## Reactivity, Secondary Structure and Molecular Topology of the *Escherichia coli* Sulfite Reductase Flavodoxin-like Domain

Ludovic CHAMPIER<sup>‡,||</sup>, Nathalie SIBILLE<sup>§,||</sup>, Beate BERSCH<sup>§\*</sup>, Bernhard BRUTSCHER<sup>§</sup>, Martin BLACKLEDGE<sup>§</sup> and Jacques COVES<sup>‡\*</sup>

<sup>‡</sup>Laboratoire de Chimie et Biochimie des Centres Rédox Biologiques, CEA-Grenoble, DBMS/CB, UMR 5047 CNRS-CEA-UJF, 17 Avenue des Martyrs, 38054 Grenoble Cedex 9, France

<sup>§</sup>Laboratoire de Résonance Magnétique Nucléaire, Institut de Biologie Structurale - Jean-Pierre Ebel, UMR 5075 CNRS-CEA-UJF, 41, rue Jules Horowitz, 38027 Grenoble Cedex, France

The flavodoxin-like domain, missing in the three-dimensional structure of the monomeric, simplified model of the *Escherichia coli* sulfite reductase flavoprotein component (SiR-FP) has now been expressed independently. This 168-amino acids protein was named SiR-FP18 with respect to its native molecular weight and represents the FMN-binding domain of SiR-FP. This simplified biological object has kept the main characteristics of its counterpart in the native protein. It could incorporate FMN exclusively and stabilize a neutral air-stable semiquinone radical. Both the radical and the fully reduced forms of SiR-FP18 were able to transfer their electrons to DCPIP or cytochrome *c* quantitatively. SiR-FP18 was able to form a highly stable complex with SiR-HP, the hemoprotein component of the sulfite reductase containing an iron-sulfur cluster coupled to a siroheme. In agreement with the postulated catalytic cycle of SiR-FP, only the fully reduced form of SiR-FP18 could transfer one electron to SiR-HP, the transferred electron being localized exclusively on the heme. As isolated SiR-FP18 has kept the main characteristics of the FMN-binding domain of the native protein, a structural analysis by NMR was performed in order to complete the partial structure obtained previously. Structural modeling was performed using sequence homologues, cytochrome P450 reductase (CPR; 29% identity) and bacterial cytochrome P450 (P450-BM3; 26% identity) as conformational templates. These sequences were anchored using common secondary structural elements identified from heteronuclear NMR data measured on the protein backbone. The resulting structural model was validated, and subsequently refined using residual ( $C^{\alpha}-C^{\beta}$ ,  $N-H^N$  and  $C^{\beta}-H^N$ ) dipolar couplings measured in an anisotropic medium. The overall fold of SiR-FP18 is very similar to that of bacterial flavodoxins and of the flavodoxin-like domain in CPR or P450-BM3.

Sulfite reductase (SiR) from *Escherichia coli* is a multimeric and soluble hemoflavoprotein that mediates the transfer of electrons from NADPH to sulfite to produce sulfide. This reaction is a key step of the sulfate assimilation pathway leading to the biosynthesis of organic sulfur compounds such as L-cysteine (1). SiR was originally

described as an  $\alpha_8\beta_4$  polymer (2). The homo-octameric  $\alpha_8\beta_4$  complex is the flavoprotein component (SiR-FP) of the enzyme as each  $\beta$ -chain binds one FAD and one FMN and, in addition, contains an NADPH-binding domain (3, 4). The module constituted of the FAD- and NADPH-binding domains is homologous to the ferredoxin-NADP<sup>+</sup> reductases whereas the FMN-binding domain is homologous to bacterial flavodoxins. This kind of organization is characteristic of the family of proteins that includes NADPH-cytochrome P450 reductase (CPR), nitric oxide synthase and *Bacillus megaterium* cytochrome P450 (P450-BM3) (5-7). The role of SiR-FP is to transfer electrons from NADPH via the FAD and the FMN prosthetic groups to the  $\alpha$ -subunit (8). SiR-FP is also able to transfer electrons either from the FMN site to artificial acceptors such as ferricyanide, dichlorophenol indophenol or cytochrome *c*, while FAD can reduce directly AcPyADP<sup>+</sup> or free exogenous flavins (3, 8, 9). Each  $\beta$ -protein contains a Fe<sub>4</sub>S<sub>4</sub> cluster coupled to a siroheme through a cysteine bridge and corresponds to the hemoprotein component (SiR-HP) of the

\* To whom correspondence should be addressed

† This work was supported by the Centre National de la Recherche Scientifique (CNRS), and the Commissariat à l'Energie Atomique (CEA). The authors are grateful to Accelrys Inc. for continuing collaboration.

|| LC and NS have equally contributed to this work

Abbreviations:

CPR, cytochrome P450 reductase; CSI, chemical shift index; DCPIP, dichlorophenol indophenol; FAD, flavin adenine dinucleotide; FMN, flavin mononucleotide; HSQC, <sup>1</sup>H-detected heteronuclear single quantum coherence; NOE, nuclear Overhauser effect; P450-BM3, *Bacillus megaterium* cytochrome P450; RDC, residual dipolar couplings; SiR, sulfite reductase; SiR-FP, sulfite reductase flavoprotein component; SiR-HP, sulfite reductase hemoprotein component.

protein. The siroheme finally transfers the electrons to sulfite for its reduction to sulfide (8, 10).

The crystal structure of SiR-HP has already been determined (10). However, native or recombinant SiR holoenzyme has so far resisted crystallization. Therefore, we have developed a molecular dissection strategy of the SiR molecule in order to create new simplified biological objects corresponding to the different functional modules of the protein (For a review, see (11)). Among them, a monomeric simplified model of SiR-FP was constructed by deletion of the first 51 N-terminal amino acids of the  $\alpha$ -polypeptide, thus removing the determinants required for octamerization (12). The truncated protein, previously called SiR-FP60, was demonstrated to be a monomer with FAD and FMN as cofactors and displaying all the catalytic diaphorase activities of SiR-FP. Moreover, we have demonstrated that incubation of SiR-FP60 with SiR-HP led to a stable and functional  $\alpha_1\beta_1$  complex able to oxidize NADPH at the expense of sulfite (13).

The crystal structure of SiR-FP60 has been determined at 1.94 Å resolution (14). Unfortunately, in the four different crystal forms obtained, the FMN-binding domain was absent from the electron density although its presence was demonstrated within the crystal. This was interpreted as a flexibility of the flavodoxin-like domain required for optimal electron transfer. In order to complete the crystal structure, we report the expression of a monomeric version of the flavodoxin-like domain of SiR-FP. This new protein was named SiR-FP18 with respect to its native molecular weight. We describe here the specific characterization of SiR-FP18 and its comparison with SiR-FP23, the polymeric FMN-binding domain of SiR-FP (15). The overall chemical reactivity of this protein makes it a very good model of its polymeric counterpart. These results are complemented by structural information obtained on SiR-FP18. Using multidimensional NMR, the secondary structure and the topology of SiR-FP18 have been determined. A structural model was built using sequence similarity and then validated and refined using heteronuclear residual dipolar couplings (RDC) measured in an anisotropic medium. It is demonstrated that, using RDC, an experimentally verified low resolution structural model can be obtained rapidly after assignment of the backbone resonances. As expected, SiR-FP18 shows very close structural homology with the corresponding modules in CPR and P450-BM3 whose three-dimensional structures have been determined by X-ray crystallography (16-18).

## EXPERIMENTAL PROCEDURES

### Construction of the expression plasmid, expression and purification of the recombinant SiR-FP18 protein

The DNA fragment expected to encode for the monomeric form of the flavodoxin-like domain of SiR-FP was amplified by PCR with suitable restriction sites for subcloning in the expression vector pET30. The sense primer 5'-ACGGAATTCATATGGCAACGCCAGCGCCAGCTGCAGAAATGC contains the nucleotides (in bold) coding for a N-terminal amino acid sequence starting with alanine-52 (ATPAPAAEM). The overhanging sequence includes an NdeI site (underlined) carrying the initiator

codon ATG. The antisense primer 5'-GGCGGATCC TCAAGCGACGGATTGCGAAGGTGC was designed to incorporate a BamHI site (underlined) and a stop codon in the coding sequence of SiR-FP to construct a protein truncated between alanine 219 and threonine 220. The nucleotides (in bold) are complementary of the corresponding coding sequence of SiR-FP. The authenticity of the DNA insert generated by PCR was confirmed by sequencing. The resulting expression plasmid pET-SiR-FP18 was used to transform the T7 RNA polymerase-containing host *E. coli* BL21(DE3). The freshly transformed bacteria were grown at 37°C in M9 minimal mineral medium with glucose (2 g/L) as carbon source and supplemented with MnCl<sub>2</sub> (0.1 mM), ZnSO<sub>4</sub> (0.05 mM), FeCl<sub>3</sub> (0.05 mM) and a vitamin solution according to Jansson et al (19). Isotopically labeled protein was prepared by growing the cells with <sup>15</sup>NH<sub>4</sub>Cl (1 g/L) and <sup>13</sup>C<sub>6</sub>-glucose (2 g/L) as sole nitrogen and carbon sources. A triply labeled sample (<sup>2</sup>H/<sup>15</sup>N/<sup>13</sup>C) was obtained in the same medium prepared with 90% <sup>2</sup>H<sub>2</sub>O. Protein expression was induced by the addition of 0.5 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside when cells reached an A<sub>600</sub> between 0.5 and 0.6. Cells were harvested by centrifugation after 4 h of incubation at 37°C. In the specific case of the triply labeled sample, incubation after induction was for 8 h at 37°C.

The pellet from 1 L of culture was suspended in 100 mM Tris/HCl (pH 7.5) and lysed by sonic oscillation. The total protein extract was recovered by centrifugation during 90 min at 45 000 rpm in a 50.2 TI rotor (Beckman) and used for further purification. The total protein extract (100 mg protein) was treated as previously described for the purification of SiR-FP (3). The pellet obtained after ammonium sulfate precipitation (80% final saturation) was dissolved in a minimal volume of 100 mM potassium phosphate buffer (pH 7.0). Potassium ferricyanide was added at a final concentration of 1 mM before the protein solution was subjected to filtration on a Superdex-75 column (Amersham-Pharmacia Biotech.) previously equilibrated with the same buffer. Elution was run at 0.8 mL.min<sup>-1</sup> and 1 mL fractions were collected. Fractions were assayed for protein (absorbance at 280 nm) and for flavin content (absorbance at 450 nm). Flavin-containing fractions were pooled and concentrated by ultrafiltration using a Diaflo cell equipped with a YM-10 membrane (Amicon Co.) and Centricon-10 microconcentrators. At this stage, SiR-FP18 was electrophoretically pure. The yield for a typical purification procedure is 30-35 mg of pure SiR-FP18 from 1 L of minimal medium.

Protein concentration was determined using bovine serum albumin as a standard (20) and the commercial Bio-Rad protein assay solution. Denatured molecular mass of SiR-FP18 was estimated by 0.1 % SDS-15 % polyacrylamide gel electrophoresis (21) and its native molecular mass was ascertained by elution of a calibrated Superdex-75 filtration column.

### Cofactor analysis and reconstitution of SiR-FP18

A sample of pure SiR-FP18 was boiled for 10 min in the dark, chilled on ice, and microcentrifuged for 10 min to pellet the denatured protein. An aliquot of the supernatant

was analyzed by thin-layer-chromatography on silica-gel 60 F254 (Merck) and developed with butan-1-ol/acetic acid/water (12:3:5, by vol.). Pure FMN and FAD were run separately or as a mixture under the same conditions as a control.

For reconstitution experiments, SiR-FP18 was incubated with a two-fold excess of either FMN, FAD or riboflavin during 60 min at ice temperature in the dark. Excess flavin was removed in a first step by filtration on Sephadex G-25. Non-specifically bound flavin was then removed by two cycles of dilution in Tris/HCl 50 mM pH 7.5 containing 250 mM NaCl, followed by concentration on centricon-10. Optical spectra of SiR-FP18 were recorded before and after reconstitution. In each case, concentration of protein-bound flavin was calculated after heat-denaturation of the protein from the absorbance of the free cofactor. An extinction coefficient of  $12.2 \text{ mM}^{-1}\cdot\text{cm}^{-1}$  at 450 nm was used for FMN and riboflavin, and of  $11.3 \text{ mM}^{-1}\cdot\text{cm}^{-1}$  for FAD at the same wavelength (22).

For NMR experiments, SiR-FP18 in potassium phosphate buffer was reconstituted by dilution with a slight excess of pure commercial FMN (Sigma). The protein sample was then concentrated on Centricon-10, washed once with the same buffer to remove the excess of FMN and transferred to the NMR tube for further use.

#### **Reduction of SiR-FP18 and reactivity of the reduced protein**

For the titration experiments, the protein sample was made oxygen-free by overnight incubation in a glove box under overpressure of nitrogen. Sodium dithionite solutions were prepared in anaerobic buffers. Dithionite concentration was calculated using an extinction coefficient of  $8 \text{ mM}^{-1}\cdot\text{cm}^{-1}$  at 315 nm. At the end of the experiments, the stability of the titrant was checked using the same method. After each addition of titrant, the solution was left for at least 10 min at room temperature to reach the thermodynamic equilibrium before recording a spectrum. Protein conditions are detailed in the Results section or in the legend of the corresponding figures.

Photochemical reductions in the presence of deazaflavin-EDTA were carried out as previously described (15). The one-electron-reduced form of SiR-FP18 ( $\text{FMNH}^\bullet$ ) was generated by oxidation of the fully reduced form upon air admission.

The chemical reactivity of the two-electron-reduced form ( $\text{FMNH}_2$ ) or of the one-electron-reduced form ( $\text{FMNH}^\bullet$ ) of SiR-FP18 was assayed as outlined previously for SiR-FP23 with cytochrome *c* as an electron acceptor (15). DCPIP was also used in place of cytochrome *c*. In this case, the absorbance change was followed at 600 nm using an absorption coefficient of  $21 \text{ mM}^{-1}\cdot\text{cm}^{-1}$ .

In the same way, SiR-HP, the hemoprotein component of the sulfite reductase, was used as an electron acceptor from the reduced forms of SiR-FP18. The one-electron reduced form obtained by air oxidation of the fully reduced form was deaerated by prolonged incubation in the glove box before its reactivity was checked under anaerobiosis. The absorbance changes upon electron transfer were followed spectrophotometrically by recording the absorption

spectra of the mixture of a slight molar excess of SiR-FP18 with respect to SiR-HP. Once the reaction finished the mixture was transferred to an EPR tube. The tubes were frozen under anaerobiosis, capped with a rubber septum and analyzed by EPR spectroscopy as described below and in the legend of the corresponding figures. Then, the content of two tubes was thawed under anaerobiosis, and each solution was mixed with an excess of sulfite ( $500 \mu\text{M}$ ) before being analyzed again by EPR spectroscopy.

Alternatively, visual observation of the oxidation of the reduced forms of SiR-FP18 was performed with potassium ferricyanide as an oxidant.

#### **Affinity chromatography**

The capacity of SiR-FP18 and SiR-HP to form a stable complex was checked by affinity chromatography. SiR-FP18 was coupled to CNBr-activated Sepharose 4B (Amersham-Pharmacia Biotech) as described by the manufacturer. The gel was then incubated overnight under constant agitation at room temperature with an excess of SiR-HP. A column of 200  $\mu\text{L}$  was packed and unbound SiR-HP was removed by washing the gel with 3 volumes of Tris/HCl 50 mM pH 7.5 (buffer A). The column was then washed with 3 volumes of buffer A containing 0.5 M NaCl and, then with 3 volumes of buffer A containing 1 M NaCl. The time of contact of the gel with these high ionic strength solutions was at least 15 min. Finally the gel was incubated overnight in the presence of 5 M urea before being washed again with buffer A. The beads were then removed from the column and the suspension was boiled in electrophoresis denaturing buffer during 10 min.

#### **Spectroscopic methods**

UV-visible spectra of aerobic samples were recorded at 25 °C in a quartz cell (10 mm light path) using a Cary 1 Bio (Varian) or a Uvikon 930 (Kontron) spectrophotometer. For recording the absorption spectra of anaerobic samples, a cuvette holder inside a glove box under overpressure of nitrogen was connected with optical fibers to a Hewlett-Packard 8452A diode array spectrophotometer.

EPR spectra were recorded with a Bruker EMX spectrometer operating at 9.45 GHz and equipped with an ESR 900 helium flow cryostat (Oxford instruments). Neutral semiquinone radical on flavin was quantified by comparison with a sample of SiR-FP60 containing a known amount of FMN radical (12).

All NMR experiments were performed on Varian INOVA 600 and INOVA 800 spectrometers, both equipped with a triple-resonance ( $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$ ) probe and shielded z-gradients. The FMN-reconstituted oxidized protein was dissolved to a final concentration of about 1.2 mM in 90%  $\text{H}_2\text{O}$ , 10%  $\text{D}_2\text{O}$  containing 100 mM sodium phosphate, pH 7.0 and 0.02% sodium azide. The sample temperature was set to 30°C. Quadrature detection in the indirect dimensions of the multidimensional experiments was achieved by the echo/antiecho detection scheme for  $^{15}\text{N}$ , and by the TPPI-States method for  $^1\text{H}$ . A three-dimensional  $^{15}\text{N}$  edited NOESY spectrum was obtained at 800 MHz with the NOESY mixing-time set to 100 ms. Eight scans were recorded per ( $t_1$ ,  $t_2$ ) increment. For the deuterium exchange,



the protein was lyophilized in a SpeedVac (Savant) and suspended in D<sub>2</sub>O buffer. <sup>1</sup>H, <sup>15</sup>N HSQC spectra were acquired after 15 min, 12 h and 1 month to identify the slowly exchanging amide protons.

Hydrogen bonds were established according to Cordier & Grzesiek (23) by the detection of cross-hydrogen bond <sup>15</sup>N-<sup>13</sup>C' scalar couplings (<sup>3</sup>hJ<sub>NC'</sub>) in a long-range HNCO experiment at 800 MHz. Residual dipolar couplings (RDC) were collected on a 1.6 mM uniformly <sup>13</sup>C, <sup>15</sup>N and 70% <sup>2</sup>H labeled sample suspended in a liquid crystalline medium consisting of 14 mg/ml of the filamentous phage Pf1 (Asla Ltd, Riga, Latvia) in 100 mM potassium phosphate buffer pH 7.0 containing 2 mM MgCl<sub>2</sub>. Three different types of dipolar couplings were measured at 600 MHz: <sup>1</sup>D<sub>NH</sub>, <sup>1</sup>D<sub>C $\square$ C'</sub>, and <sup>2</sup>D<sub>C $\square$ H $\square$</sub>  couplings were obtained using 3D HNCO-type experiments(24, 25).

Data processing and peak picking were performed using FELIX program version 2000 (MSI Technologies).

### Homology modeling

A structural model for SiR-FP18 was built by homology modeling using SWISS-MODEL and the Swiss-PdbViewer (26). The crystal structures of the flavodoxin-like domains of human CPR (pdb code 1B1C, (17)) and P450-BM3 (pdb code 1BVY, (18)) were used as templates. For the initial project file, a structural alignment was built using the Swiss-PdbViewer considering protein sequence similarity and identification of secondary structure elements. Secondary structure information was derived from the analysis of the deviation of the experimental <sup>13</sup>C chemical shifts from random-coil values (chemical shift index, CSI, (27)), the intensity distribution of short- and medium range NOEs involving H<sup>N</sup> protons and the presence of hydrogen bonds. This alignment was then submitted to SWISS-MODEL for final modeling.

### Model Validation using Module

The homology-based model, corresponding to residues 60 to 210 of SiR-FP, was further validated by using long range structural information obtained from RDC data. The recently developed program Module (28) was used to evaluate the accordance between the structural model calculated with SWISS-MODEL and the experimental RDC data. Module determines alignment tensor parameters and provides a graphical interface for the visualization of the tensor relative to the three-dimensional atomic coordinates as well as correlation plots of the measured and calculated couplings. In the initial stage, the secondary structure elements were treated as separate structural entities. The alignment tensor is characterized by five parameters: the axial and rhombic components, A<sub>a</sub> and A<sub>r</sub>, measure the extent of residual alignment due to the restricted orientational sampling in the anisotropic medium, whereas the Euler angles, { $\alpha$ ,  $\beta$ ,  $\gamma$ } define the non-averaged orientation of the molecule in the three-dimensional space. For each entity, these five parameters were determined using non-linear least-squares minimization of the target function over all couplings associated with a given domain:

$$\chi^2 = \sum_n \left\{ D_{ij}^{\text{exp}} - D_{ij}^{\text{calc}} \right\}^2 / \sigma_{ij}^2 \quad (1)$$

where D<sub>ij</sub> are the residual dipolar couplings between spins *i* and *j* and  $\sigma_{ij}$  is the uncertainty in the experimentally measured coupling.  $\sigma_{ij}$  was set to 1 Hz for all couplings.

The quality of the model structure can be evaluated by comparing measured dipolar couplings with dipolar couplings back-calculated from the best-fit alignment tensor. Additional information can be obtained from the comparison of the amplitude and orientation of the individual alignment tensors for the different structural elements. In the case of a rigid molecule, A<sub>a</sub> and A<sub>r</sub> should be identical for each structural entity. Differences would indicate differential flexibility. On the other hand, the orientation of the individual alignment tensors should be coaxial, if the relative orientation of the secondary structure elements in the model structure is correct.

### Refinement using Sculptor.

The initial structural model was further refined with the program Sculptor (29, 30), using a restrained molecular dynamics protocol, incorporating residual dipolar couplings as experimental restraints.

The homology-based model was initially fixed, and the alignment tensor optimized using an energy minimization protocol under the influence of residual dipolar couplings from the secondary structural elements. This alignment tensor is then identical to the tensor determined using the program Module. Following this, the molecule was released and allowed to evolve under the influence of RDC's, backbone dihedral angle restraints in the secondary structural elements, measured NOE's from the  $\beta$ -sheet motif, and hydrogen bonds identified from the direct measurement via scalar coupling. As RDC's contain only orientational information and no translational restriction, a tethering potential was applied on the C $\square$  atoms of the peptide chain in order to avoid simple expansion of the molecular model. This energy term is applied as a sum over all C $\square$  atoms, global unfolding is thus avoided while allowing local reorientation. A potential of the form

$$E_{\text{teth}} = k_{\text{teth}} \sum_i \sqrt{(x_i - x_i^0)^2} / N \quad (2)$$

was incorporated into the potential energy function. *x<sub>i</sub>* are the Cartesian coordinates of the atoms to be tethered and *x<sub>i</sub><sup>0</sup>* the target coordinates. The force constant *k<sub>teth</sub>* (20 kcal.mol<sup>-1</sup>.Å<sup>-2</sup>) is used to restrain the N=156 atoms to the coordinates of the initial model structure.

The system was then equilibrated at 300 K for 0.5 ps of molecular dynamics, and gradually increased to 500 K over a period of 2 ps, during which time *k<sub>NOE,h-bond</sub>*, *k<sub>dih</sub>* and *k<sub>RDC</sub>* were raised from their initial (0.1, 0.1, 0.001) to their final values (50.0 kcal.mol<sup>-1</sup>.Å<sup>-2</sup>, 100.00 kcal.mol<sup>-1</sup>.deg<sup>-2</sup>, *k<sub>rdc,i,j</sub>*). The energetic weighting of the different types of RDC was initially estimated on the basis of the precision of the different experimental measurements ( $\sigma_{ij}$ =1.0 Hz for <sup>N</sup>H-N, C $\square$ -C' and <sup>N</sup>H-C') and then empirically adjusted such that the final structures had a total reduced  $\chi^2$  ( $\chi^2/N$  where N is the number of couplings) close to 1. Final force constants of *k<sub>RDC,N-H</sub>*=0.63 kcal.mol<sup>-1</sup>.Hz<sup>-2</sup>, *k<sub>RDC,C $\square$ -C'</sub>*=1.11 kcal.mol<sup>-1</sup>.Hz<sup>-2</sup> and *k<sub>RDC,NH-C</sub>*=1.11 kcal.mol<sup>-1</sup>.Hz<sup>-2</sup> were used.

During the sampling period, the molecule and tensors evolved freely for 2 ps before the system was cooled to 100

K over 7 ps and the system minimized using a conjugate gradient algorithm. 30 calculations were performed using different initial atomic velocity distributions, the 10 structures with the lowest total target function comprising NOE, backbone torsion, hydrogen bonding and RDC restraints were selected for final analysis. This set of calculations was repeated using no RDC data, to test for force-field, or protocol dependent effects, and serves as a control for the calculations that included the RDC.

## RESULTS

### Overexpression, purification and characterization of SiR-FP18

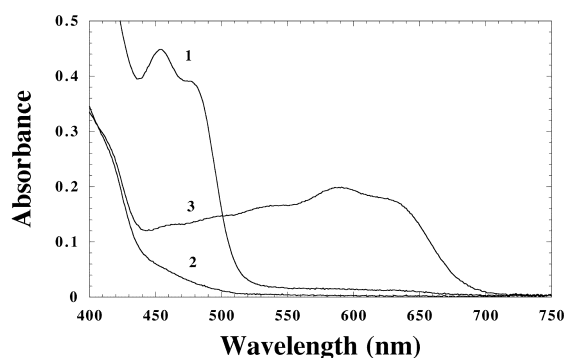
*E. coli* BL21(DE3) was transformed with the expression plasmid pET-SiR-FP18 encoding SiR-FP18, a protein corresponding to the FMN-binding domain of SiR-FP but with alanine-52 as the first residue and alanine-219 as the last one. Maximal expression was observed between 4 and 8 hours after addition of IPTG to the cell cultures depending on the characteristics of the growth medium (4 hours in H<sub>2</sub>O and 8 hours in <sup>2</sup>H<sub>2</sub>O). SiR-FP18 was recovered in the soluble extracts which were blue-grey. Their UV-visible spectrum was dominated by a long wavelength absorption band at 589 nm and a shoulder at 630 nm characteristic of the presence of FMNH<sup>•</sup>. This indicated the large accumulation of a flavoprotein under the blue neutral semiquinone state as previously observed for SiR-FP23 (15) and SiR-FP60 (12). SDS-PAGE analysis of the crude extracts showed a predominant band at an apparent molecular weight of 18 kDa corresponding to the expected mass of SiR-FP18 (not shown). The overproduced protein was then purified to homogeneity by a one-step procedure. The ammonium sulfate pellet was resuspended in a minimal volume of potassium phosphate buffer and 1 mM ferricyanide was added in order to oxidize the protein. The solution turned instantaneously yellow under these conditions and was fractionated on a Superdex-75 filtration column. Flavin containing fractions corresponding to a native molecular weight of about 20 kDa were pooled. SDS-PAGE analysis of these fractions revealed the presence of an 18 kDa protein corresponding to the overproduced protein contained in the extracts. This showed that SiR-FP18 is a monomer as expected. A yield of about 30 mg of pure protein per liter of culture was routinely obtained. Isotopically labeled protein was produced for NMR experiments. The extent of isotopic labeling was checked by electrospray ionization mass spectroscopy analyses by comparison of the molecular mass of pure unlabeled SiR-FP18 with that of the labeled proteins. The extent of labeling was > 97% for <sup>15</sup>N and <sup>13</sup>C and 70% for <sup>2</sup>H. Analysis of the flavin cofactor in pure SiR-FP18 was done by thin-layer chromatography of the supernatant of a heat-denatured protein sample. This confirmed that, as expected, SiR-FP18 contained FMN exclusively. Quantification of the FMN content was performed by spectrophotometric analysis of the same supernatant using an extinction coefficient of 12.2 mM<sup>-1</sup>.cm<sup>-1</sup> at 450 nm. We calculated a ratio of 0.25 to 0.6 mol of FMN per mole of SiR-FP18 depending on the protein preparation. This shows that, as previously observed for SiR-FP23 (15), SiR-FP18 can be largely depleted of its

flavin cofactor, likely due to the large overexpression of the protein. An extinction coefficient of 11.3 mM<sup>-1</sup>.cm<sup>-1</sup> at 456 nm (the absorption maximum) was estimated in the fully oxidized state. As for SiR-FP23, the visible spectrum of SiR-FP18 also presents a characteristic shoulder at 480-490 nm (Fig. 1, spectrum 1).

Incubation of SiR-FP18 with an excess of FMN led to a reconstituted protein containing 1 mol of FMN per mole of protein. As previously observed with SiR-FP23, reconstitution of SiR-FP18 with riboflavin or FAD was shown to be ineffective. This indicates that SiR-FP18 presents the correct fold allowing for the exclusive binding of FMN.

### Reduction of SiR-FP18 and reactivity of the reduced protein

Photoreduction of fully oxidized SiR-FP18 (85 μM total flavin concentration) was achieved by illumination at pH 7.5 and under anaerobic conditions in the presence of 20 μM deazaflavin and 10 mM EDTA. The reduction proceeds in two distinct steps (not shown) exactly as observed with SiR-FP23 (15). During the first step, the absorbance at 456 nm decreased while a long-wavelength absorption band at 589 nm with a shoulder at 630 nm, characteristic of the neutral blue semiquinone previously observed in the total extracts, increased in parallel. These spectral changes occurred with the isosbestic point at 504 nm which was demonstrated to be characteristic of the transition from the oxidized form of the protein to its one-electron reduced form (12, 15). The second step is characterized by the concomitant decrease of both bands at 456 and 589 nm until the protein was reduced to the hydroquinone state (Fig. 1, spectrum 3). Upon air admission, the fully reduced form was converted almost instantaneously to the air-stable radical form (Fig. 1, spectrum 2). This stability allows an accurate quantification of the spin concentration by EPR spectroscopy. The X-band EPR spectrum of the flavin semiquinone radical was recorded at 20.5 K and then compared to that of a solution of SiR-FP60 containing a known amount of FMNH<sup>•</sup> (12).



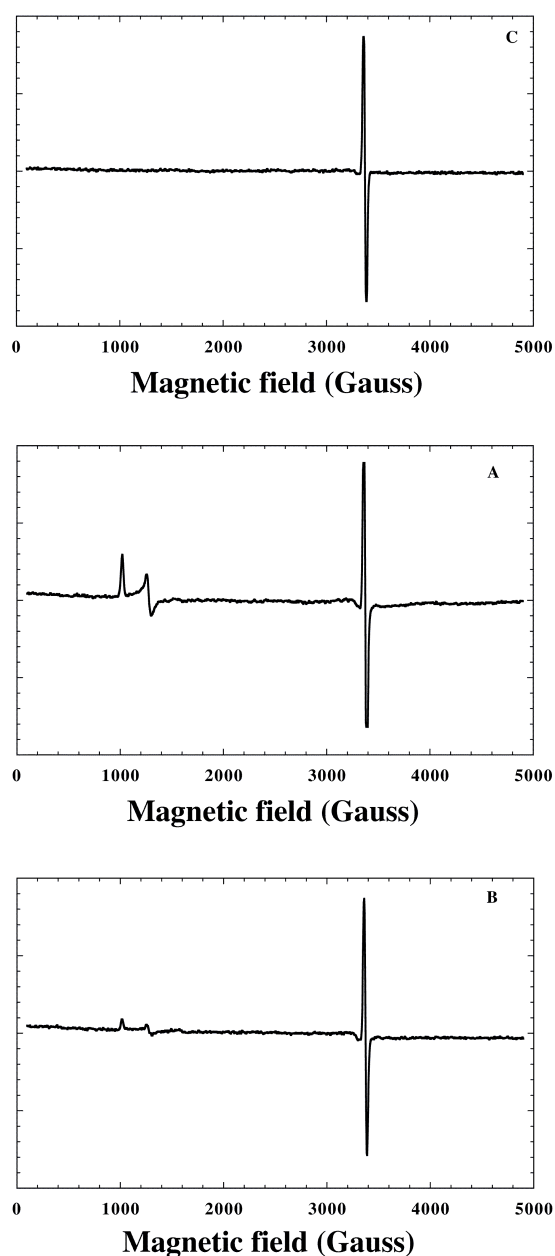
**Figure 1.** Absorption spectra of the three redox states of SiR-FP18 : *spectrum 1*, oxidized SiR-FP18 (40 μM in FMN); *spectrum 2*, fully reduced state obtained by anaerobic photoreduction in the presence of deazaflavin-EDTA as described in Experimental Procedures, and *spectrum 3*, air-stable semiquinone state obtained from a short incubation of fully reduced SiR-FP18 with air. Note that the contribution of the oxidized deazaflavin-EDTA system is visible below 430 nm in spectrum 1.

The experiment was run with SiR-FP18 obtained after purification and thus partially depleted in FMN as well as with a fully reconstituted protein. Starting with a preparation of SiR-FP18 containing 300  $\mu\text{M}$  FMN in the first case and 100  $\mu\text{M}$  the second case, 290  $\mu\text{M}$  and 92  $\mu\text{M}$  radical were obtained respectively for the partially depleted and the reconstituted protein. We can thus conclude that respectively 96% and 92% of the FMNH<sub>2</sub> was converted to the semiquinone form. An average value of the extinction coefficient at 589 nm of 4.5  $\text{mM}^{-1}\cdot\text{cm}^{-1}$  was estimated from these results. This value is situated between the value of 4.88  $\text{mM}^{-1}\cdot\text{cm}^{-1}$  obtained with SiR-FP23 and of 4  $\text{mM}^{-1}\cdot\text{cm}^{-1}$  obtained with SiR-FP (9).

The spectral changes associated with dithionite titration of SiR-FP18 were also recorded spectrophotometrically under anaerobic conditions (not shown). As expected, addition of one electron-equivalent (0.5 mol of dithionite per mol of protein-bound FMN) resulted in spectral changes characteristic of the formation of FMNH<sup>•</sup>. We could calculate that about 100% FMNH<sup>•</sup> was generated under these conditions. After a second electron-equivalent was added, a spectrum characteristic of the hydroquinone form of SiR-FP18 was obtained.

The ability of the reduced forms of SiR-FP18 to transfer electrons to dichlorophenol indophenol has been assayed. Fully photoreduced or semi-reduced SiR-FP18, containing 10  $\mu\text{M}$  FMNH<sub>2</sub> or 10  $\mu\text{M}$  FMNH<sup>•</sup> respectively were mixed with an excess of oxidized DCPIP (30  $\mu\text{M}$ ). The spectral changes due to DCPIP reduction were recorded spectrophotometrically. Using an extinction coefficient of 21  $\text{mM}^{-1}\cdot\text{cm}^{-1}$  at 600 nm for DCPIP, it was easy to calculate that respectively 21  $\mu\text{M}$  or 10  $\mu\text{M}$  DCPIP were reduced almost instantaneously. This shows that the two reduced forms of SiR-FP18 can transfer quantitatively their electrons to such an acceptor. It should be noted that, as neither the protein nor the dye has a significant absorbance at 600 nm when oxidized, no spectrum correction is required as is the case when the same experiment is run with cytochrome *c*. Indeed, the same quantitative and immediate electron transfer was also observed with cytochrome *c* as an oxidant. Ferricyanide was also used qualitatively to oxidize the protein present under the semiquinone state in the soluble extracts before purification.

Finally, fully or one-electron reduced SiR-FP18 (100  $\mu\text{M}$ ) were also mixed under anaerobiosis with oxidized SiR-HP (80  $\mu\text{M}$ ), the physiological partner of SiR-FP in native sulfite reductase. The reaction between the two proteins was followed spectrophotometrically. In both cases, the resulting experimental spectra resembled the calculated spectrum resulting from the addition of SiR-HP plus one-electron reduced SiR-FP18. In addition, subtraction of the SiR-HP spectrum from the experimental ones resulted in spectra characteristic of the semiquinone species (not shown). This suggests that only the fully reduced SiR-FP18 had transferred one electron to SiR-HP and that the one-electron reduced form of SiR-FP18 was not reactive with oxidized SiR-HP. In order to validate this hypothesis, the EPR spectra of the two different mixtures were recorded at helium temperature (Fig. 2). In both cases (Fig 2A and B), the



**Figure 2.** EPR characterization of the reaction between reduced SiR-FP18 and SiR-HP. The concentrations of SiR-FP18 and SiR-HP were respectively 100  $\mu\text{M}$  in FMN and 80  $\mu\text{M}$  in the three experiments. A, X-band EPR spectrum obtained after anaerobic incubation of semi-reduced SiR-FP18 with SiR-HP. B, X-band EPR spectrum obtained after anaerobic incubation of fully reduced SiR-FP18 with SiR-HP. C, 500  $\mu\text{M}$  sodium sulfite was added to the previous solution before recording the X-band EPR spectrum. EPR conditions were as follows : temperature, 9 K; modulation amplitude, 1 milliTesla; microwavpower, 1 milliwatt; and microwave frequency, 9.449 or 9.448 or 9.443 for spectrum A to C.

presence of the organic radical characteristic of the flavin semiquinone was evident at  $g = 2$ . Quantification of these signals at 30K gave a value of about 92  $\mu\text{M}$  FMN under the semireduced state. The EPR spectrum of the solution resulting from the mixture SiR-HP plus one-electron reduced SiR-FP18 contained in addition the signal characteristic of a high spin  $S = 5/2$  ferrisiroheme with the

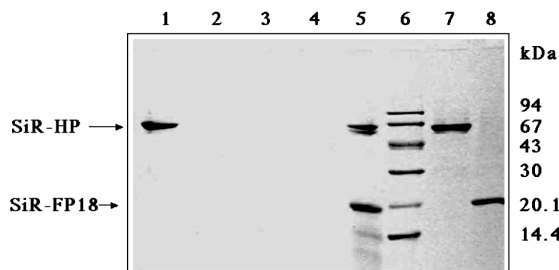


Figure 3. SiR-FP18 and SiR-HP form a tight complex. SiR-HP has been incubated with SiR-FP18 previously coupled to CNBr-activated Sepharose-4B as described in Experimental Procedures. The fractions issued from the elution procedure were analyzed by SDS-PAGE. Lane 1, unbound SiR-HP eluted with Tris/HCl 50 mM pH 7.5 (buffer A); lane 2 to 4, fractions eluted with buffer A containing in addition 0.5 M NaCl or 1 M NaCl or 5 M urea respectively; lane 5, proteins obtained after incubation of the gel beads in the electrophoresis denaturing buffer; lane 6, low molecular weight markers; lane 7 and 8, pure SiR-HP and SiR-FP18 respectively run as a control.

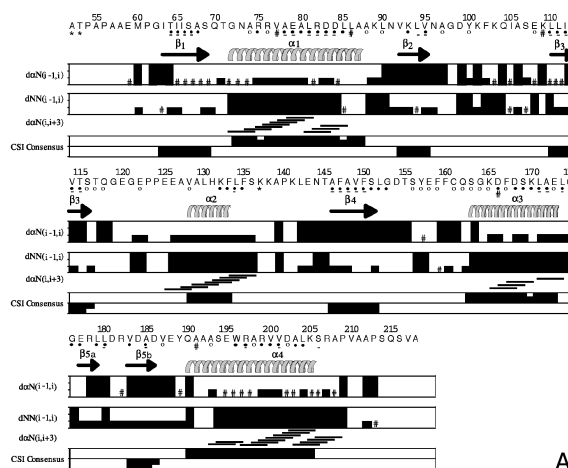
features at  $g = 6.65$  and  $5.23$  clearly visible (Fig. 2A). This signal was almost absent after reaction of SiR-HP with fully reduced SiR-FP18 (Fig. 2B) demonstrating that the majority of the ferrisiroheme was converted to ferrosiroheme (probably to the low spin state  $S = 0$ ). The spectrum of a  $S = 1/2$  ( $\text{Fe}_4\text{S}_4$ )<sup>+</sup> cluster was absent, showing that SiR-HP was only reduced by one electron with this electron localized on the heme. This electron originates from the hydroquinone form of SiR-FP18 which was converted in turn to the semiquinone form. Moreover, after addition of sulfite, the remaining signal of the ferrisiroheme disappeared completely (Fig. 2C). This experiment demonstrates clearly that (i) SiR-HP and SiR-FP18 do interact, (ii) electron transfer is possible between these two proteins, (iii) only the hydroquinone form can give one electron to SiR-HP.

**SiR-FP18 and SiR-HP form a stable complex**

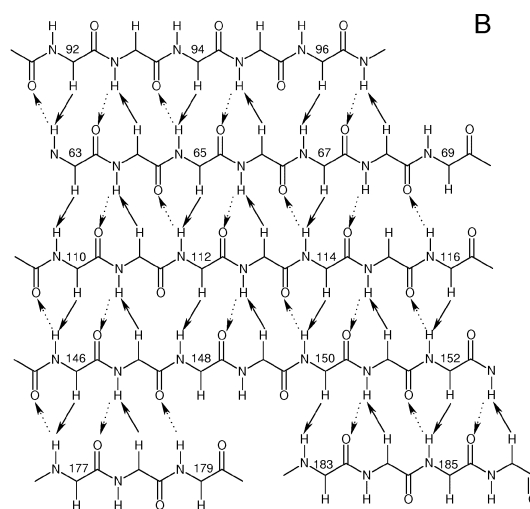
SiR-HP was incubated with SiR-FP18 previously coupled to CNBr-activated Sepharose 4B and packed to a small column. The column was first washed with a standard buffer in order to remove unbound SiR-HP (Fig. 3, lane 1). Then, the column was extensively washed with the same buffer containing 0.5 M NaCl (Fig 3, lane 2) and 1 M NaCl (Fig 3, lane 3). The use of a chaotropic agent (5 M urea) was also inefficient in eluting bound SiR-HP (Fig 3, lane 4). SiR-HP could be recovered only after incubation of the gel beads with denaturing electrophoresis buffer (Fig 3, lane 5). This demonstrates unambiguously the very tight association of the two proteins under these conditions. The same experiment was repeated with BSA in place of SiR-HP as a control, and no binding of BSA on SiR-FP18 was observed (not shown).

**Secondary structure and global fold**

The complete NMR backbone and side-chain assignment of SiR-FP18 has been previously reported (31). Secondary structure information was derived from the analysis of different NMR parameters. These results are summarized in Figure 4a. SiR-FP18 is composed of five  $\beta$  strands (residues 63-69, 92-95, 110-116, 146-151 and 183-186) and four  $\beta$ -helices (residues 72-87, 128-133, 163-173 and 190-205). A small, additional  $\beta$  strand named 5a (residues 177-179) can also be proposed based on the observation of hydrogen bonds, sequential NOEs and in analogy to the flavodoxin-like domain of CPR (17).



A



B

Figure 4. Secondary structure information obtained from NMR. A, Amide protons protected from solvent exchange for 15 min, 12 h, and one month, are represented respectively by  $\circ$ ,  $\bullet$ , and  $\square$ ; sequential and intraresidue NOEs are indicated by black bars, where the height corresponds to the intensity (# NOEs not observed because of resonance overlap, \* missing assignments); positive and negative values of the consensus chemical shift index are indicative for  $\beta$ -helices and  $\beta$ -strands, respectively. Schematic representation of the NOEs and hydrogen bonds used to deduce the topology of the  $\beta$ -sheet.  $\text{H}^N\text{-H}^N$  NOEs are shown as arrows, dashed arrows represent hydrogen bonds as identified by scalar  $^3J_{\text{NC}}$  couplings.

bond donors. In order to complete this information, cross hydrogen-bond scalar couplings were measured to directly correlate hydrogen bond donors to the corresponding acceptors. The complete hydrogen bond networks stabilizing the  $\beta$  helices as well as those connecting the individual  $\beta$  sheets have thus been detected. In combination with long-range  $H^N-H^D$  NOE these data allowed the determination of the topology of the central  $\beta$ -sheet (Fig 4b). The strands form a parallel  $\beta$ -sheet with strands ordered as follows - (2, 1, 3, 4, 5). The distribution of the slowly exchanging amide protons indicates that  $\beta$ -strands 2 and 5 are partially solvent exposed (see Figure 4a and b). These observations show that the topology of the central  $\beta$ -sheet is close to the bacterial flavodoxins and the flavodoxin-like domains of CPR and P450-BM3. The NMR spectra obtained with a largely FMN-depleted sample of SiR-FP18 were very similar to those obtained with a fully reconstituted protein (not shown). This indicates that folding of SiR-FP18 is not dependent on the presence of the FMN cofactor.

In order to obtain a more detailed structural model of SiR-FP18, we have undertaken homology modeling followed by a validation and refinement process based on experimental long range structural information obtained from RDC data. The model structure (residues 60 to 210) obtained from SWISS-MODEL is very close to the two template structures (heavy atom rmsd calculated over the secondary structure elements of 0.52 Å and 0.71 Å for the corresponding domains in human CPR and P450-BM3 respectively).

This model structure was then compared to the experimental RDC data. Three sets of residual dipolar couplings defining the backbone conformation of the peptide chain were used as input data in the program Module (28). Individual analyses were performed using RDCs from the 5 different secondary structural elements identified above. They correspond to the central  $\beta$ -sheet, which is treated as a single structural entity, and the surrounding four  $\beta$ -helices. RDCs from residues 92, 177-179, and 190 were omitted from the analysis as the corresponding residues were not found to be in a  $\beta$  conformation in the model structure. The magnitude of the alignment tensors for the individual domains are similar, and the orientation of the tensor axes are found to be almost co-axial relative to the common molecular frame (Figure 5a). This provides strong evidence that the relative orientation of the  $\beta$ -helices and  $\beta$ -sheet in the homology based model of SiR-FP18 is similar to that found in solution, an observation confirmed by the analysis of these elements as one structural domain, which also shows close agreement between calculated and experimental couplings (Figure 5b-d).

Having verified that the overall fold of the initial model is in agreement with independently measured RDCs, and therefore probably representative of the solution structure, we have refined the conformation using all of the measured experimental NMR data, i.e. secondary structure restraints and RDC. In order to avoid the translational shift of secondary structural elements during the calculation the global topology is weakly tethered to the initial model topology. The backbone rmsd of the lowest energy

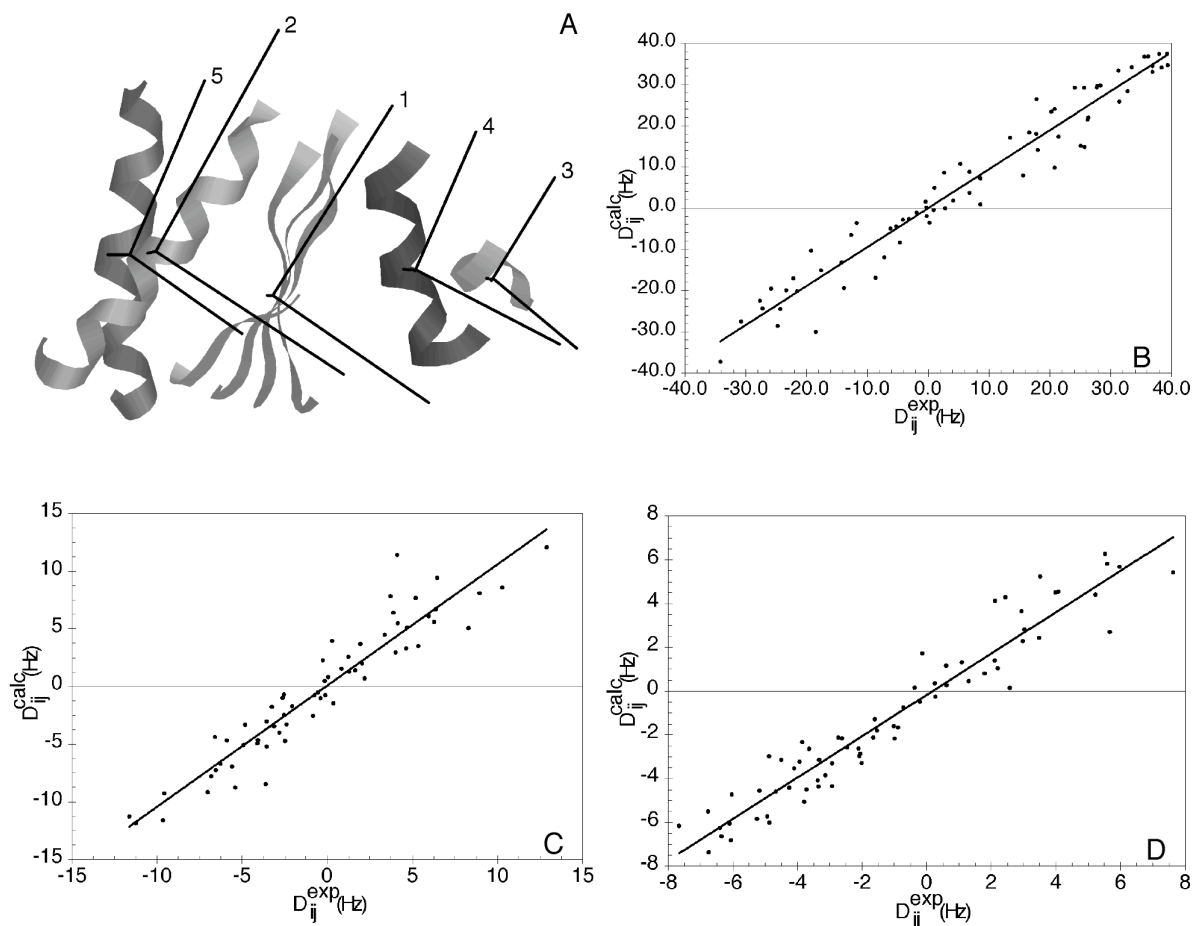
structures from the repeated calculations compared to the mean coordinates was  $(1.3 \pm 0.2)$  Å. The ensemble derived from the control calculation, identical except for the absence of RDC data, refined to an rmsd over the backbone of  $(3.2 \pm 0.5)$  Å over the ensemble. The lowest energy conformer from the RDC-refinement is shown in Figure 6, this has a backbone rmsd of 1.9 Å compared to the initial structural model. It can be seen that the typical flavodoxin-like fold is well conserved. The rmsd compared to human CPR and P450-BM3, taken again over the backbone atoms of the secondary structure elements is 1.98 Å and 1.95 Å respectively.

## DISCUSSION

The crystal structure of SiR-FP60, the monomeric model of the flavoprotein component of the *E. coli* sulfite reductase has been previously determined (14). However, the first 172 N-terminal amino acid residues (residues 52 to 224), corresponding to the flavodoxin-like domain, were not defined in any of the electron density maps obtained. In order to complete the crystal structure of the flavoprotein component, we have now expressed a protein corresponding to the residues 52 to 219, called SiR-FP18 with respect to its theoretical molecular mass of 17,870 Da. The choice of the domain borders was guided by previous results from limited proteolysis of SiR-FP60, during which an analogous fragment had been obtained (12).

SiR-FP18 was purified as a monomeric, partially depleted FMN-containing protein. Reconstitution could only be achieved with FMN, demonstrating its selectivity towards the naturally bound flavin. Its photoreduction in the presence of deazaflavin-EDTA proceeded in two distinct steps with accumulation of the one-electron reduced form  $FMNH^\bullet$  before full reduction to the hydroquinone form ( $FMNH_2$ ). The  $FMNH^\bullet$  form of the protein can be defined as the air stable semiquinone form of the protein since it is generated immediately after air admission on the anaerobically fully reduced SiR-FP18. These results suggest a strong conservation in SiR-FP18 of the difference between the midpoint potential of the  $FMNH^\bullet/FMN$  couple ( $-130 \pm 10$  mV) and the  $FMNH_2/FMNH^\bullet$  couple ( $-335 \pm 10$  mV) previously determined in SiR-FP23 (15). The 200 mV difference between these two couples explains the high degree of thermodynamic stabilization of the  $FMNH^\bullet$  radical SiR-FP18, either under the semiquinone or the hydroquinone form, could be fully oxidized by DCPIP, ferricytochrome *c* or ferricyanide. The reaction, quantified here using DCPIP, was fast and quantitative. It should be noticed that  $FMNH_2$  gives its two electrons to DCPIP without the detectable transient appearance of  $FMNH^\bullet$ . The overall reactivity of SiR-FP18 was thus very similar to that of SiR-FP23, showing that this monomeric fragment is a good model of its polymeric counterpart.

We had previously shown that SiR-FP and SiR-FP60 interact strongly with cytochrome P450c17 *in vitro* but only SiR-FP could transfer electrons to this artificial partner (32). This was interpreted as a requirement of the N-terminal part of the protein in the electron transfer process. With the hemoprotein component of SiR as the electron acceptor,

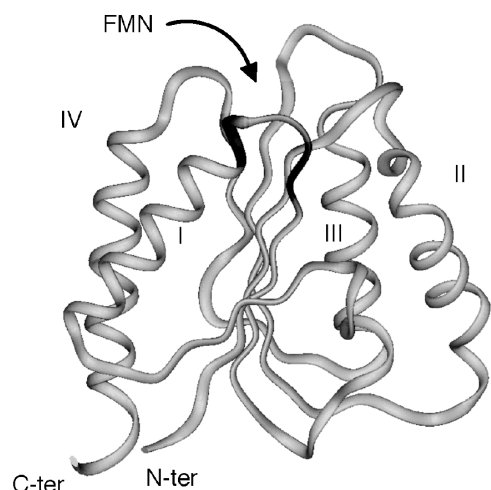


**Figure 5.** Model refinement using the program MODULE. A, Relative orientation of the five different secondary structure elements in the SWISS MODEL structure and their associated alignment tensors. Secondary structure elements are defined in the text. Numbers from 1 to 5 correspond to the central  $\beta$ -sheet, and the helices I, II, III, and IV, respectively. The amplitude of the axes shown corresponds to the eigenvalues of the alignment tensors. The axial and rhombic components as well as the total reduced  $\sigma^2$  of the five secondary structure elements are: 1,  $A_a=16.07e^{-4}$ ,  $A_r=10.53e^{-4}$ ,  $\sigma^2=9.7$ ; 2,  $A_a=17.62e^{-4}$ ,  $A_r=10.50e^{-4}$ ,  $\sigma^2=7.5$ ; 3,  $A_a=16.89e^{-4}$ ,  $A_r=10.49e^{-4}$ ,  $\sigma^2=1.1$ ; 4,  $A_a=18.54e^{-4}$ ,  $A_r=9.09e^{-4}$ ,  $\sigma^2=6.6$ ; 5,  $A_a=16.82e^{-4}$ ,  $A_r=6.99e^{-4}$ ,  $\sigma^2=9.0$ . B-D, Correlation plots between  $D_{ij}^{exp}$  and  $D_{ij}^{back\ cal}$  for  $^1D_{NH}$  (b),  $^1D_{CHN}$  (c), and  $^1D_{CCl}$  (d). The five different secondary structure elements were fitted as a single entity. The axial and rhombic components and the total reduced  $\sigma^2$  are  $A_a=17.47e^{-4}$ ,  $A_r=11.37e^{-4}$ , and  $\sigma^2=9.1$ .

both a very tight interaction and electron transfer were possible from SiR-FP60 (13). We have demonstrated now that it is also the case with SiR-FP18, although these latter proteins missed the first 51 amino acids. Interestingly, only the fully reduced form of SiR-FP18 was able to transfer one electron to SiR-HP, generating the one-electron reduced form of the flavoprotein (FMNH $^{\bullet}$ ). This resulted also in the reduction of the ferrisiroheme as judged from the almost complete disappearance of its EPR signal in the equilibrium state so obtained, a small part of the siroheme signal, estimated at about one-third of the original one, persisted along with the signal of the FMN semiquinone. The addition of sulfite, the natural substrate of the protein, only led to the disappearance of the remaining heme signal, the FMN semiquinone signal being stable. It has been shown that the complex between reduced SiR-HP and  $SO_3^{2-}$  was stable under anaerobic conditions (33) and that sulfite could be reduced only if additional reducing equivalents became available. In other words, under these conditions, sulfite

could not oxidize reduced SiR-HP in the absence of an external reductant at a lower potential than FMNH $^{\bullet}$ . It has also been shown that oxidized SiR-HP could slowly form an EPR-silent complex with sulfite (33) whose nature is unclear. However, in the present study, the remaining heme signal disappears almost instantaneously upon addition of  $SO_3^{2-}$ , probably because the thermodynamic equilibrium resulting from the distribution of the electrons after reaction of fully reduced SiR-FP18 with SiR-HP is shifted by the presence of sulfite, thus allowing the complete reduction of the siroheme.

From these experiments it became clear that SiR-HP can only accept one electron from fully reduced SiR-FP18 although one more electron could theoretically be transferred to the iron-sulfur cluster. However, the EPR signal of a  $S = 1/2$   $(Fe_4S_4)^{+1}$  cluster was never observed during these experiments. This is consistent with the hypothesis that in SiR-FP, the FMN cofactor cycles



**Figure 6.** Three-dimensional structure of SiR-FP18 after refinement using Sculptor. The backbone of SiR-FP18 is shown in ribbon representation, roman numerals designate the four  $\alpha$ -helices. Residues situated close to the presumed FMN-binding site (arrow) and whose amide protons are shifted above 9.5 ppm are darkened.

between the hydroquinone and the semiquinone forms (34), as already demonstrated with the homologous protein CPR (35). This can also be justified by the  $E'_m$  values of the different moieties present in the electron transfer complex (FMNH couples, see above;  $(\text{Fe}_4\text{S}_4)^{+1}/(\text{Fe}_4\text{S}_4)^{+2} = -405$  mV; ferrosiroheme/ferrisiroheme =  $-340$  mV (36, 37)). A distribution of the electrons between the heme and the FMN is thus plausible with the iron-sulfur cluster serving either as a relay between  $\text{FMNH}_2$  and ferriheme or as a barrier between  $\text{FMNH}^\bullet$  and the active site. The massive addition of sulfite triggered the full reduction of the heme but not of the iron-sulfur cluster as the quantity of FMN semiquinone was not significantly changed. To be catalytically active in SiR-FP,  $\text{FMNH}^\bullet$  must be reduced to  $\text{FMNH}_2$  very likely by  $\text{FADH}_2$  that, in turn, must oxidize to  $\text{FADH}^\bullet$ . The presence of the one-electron reduced form of the FAD moiety has never been detected during steady-state or stopped-flow analyses of SiR-FP (34) very likely because of its extreme reactivity and instability. However, using the isolated FAD-binding domain (12, 38), we have previously demonstrated that the FAD semiquinone state could exist and was stable enough to be detected during air oxidation of anaerobically fully reduced  $\text{FADH}_2$ .

We then were interested in the reasons for the absence of the electron density of the flavodoxin-like domain of SiR-FP60 (14). In the NMR experiments, SiR-FP18 appears as a properly folded, stable protein. The  $^1\text{H}, ^{15}\text{N}$  correlation spectrum of SiR-FP18 was found to be very well dispersed (31) and no indication could be obtained from the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shift data concerning significant chemical exchange or the presence of unstructured regions, even in the absence of FMN. Interstrand distances and hydrogen bonds were detected for the entire central  $\beta$  sheet indicating a high structural stability. These observations suggest that the absence of the electron density in the previous crystal study of the entire flavoprotein could be explained by the motion of the flavodoxin-like domain with respect to the rest

of SiR-FP60. An analogous behavior has also been suggested in the case of the eukaryotic cytochrome P450 reductase. Using site-directed mutagenesis, residues involved in electron transfer to cytochrome P450 have been identified (17). However, the same structural region was covered by the FAD of the FAD/NADPH-binding domain in the crystal structure of the entire rat CPR (16). Consequently, it was suggested that the binding of cytochrome P450 requires a movement of the FMN-binding domain facilitated by the presence of the flexible linker between this domain and the rest of the molecule (17). Identical results have very recently been reported on the homologous bacterial electron transfer system formed of ferredoxin-NADP<sup>+</sup> reductase, flavodoxin and cobalamin-dependent methionine synthase (39). NMR chemical shift mapping was used to show that both physiological partners bind to overlapping sites on flavodoxin, indicating again the need for relative domain-domain reorientation during the electron transfer reaction.

The final aim of the NMR study of SiR-FP18 is the completion of the high resolution structure of SiR-FP as well as the investigation of the electron transfer complex with the hemoprotein ( $\beta$ -subunit). However, at the present state of the work, we can use the information provided by NMR to propose a low-resolution model of SiR-FP18 based on sequence similarity considerations and validated by experimental NMR data. It has been proposed that *E. coli* sulfite reductase, CPR, nitric oxide synthase and P450-BM3 share a common fold (16). The initial structure prediction process was therefore based on a combination of experimentally identified secondary structural elements and the high resolution crystallographic structures of homologous molecules, the flavodoxin-like domains of human CPR and P450-BM3. These two domains show 29% and 26% sequence identity with SiR-FP18 respectively and could be used as templates. The resulting model was further validated by independently obtained experimental RDC data. Using the program MODULE, the model was shown to be in very good agreement with measured RDC data, both with respect to the relative orientation of the previously identified secondary structure elements in the model, and the effective alignment experienced by each individual element (Figure 5). The final structural model, comprising residues 60 to 210, was refined in the presence of all of the measured experimental NMR data (Figure 6). As expected the relative orientation of the secondary structural elements does not change significantly during the calculation. Helix 2 is tilted slightly under the influence of the RDC data, while the C-terminal helix is less-curved. Interestingly the region 96-100, an extended loop in the model structure, is also found to fold into a single helical turn due to the RDC restraints.

Some features in the C-terminal part of the molecule merit further discussion. In CPR but not in P450-BM3, there is a small supplementary  $\beta$  strand, named 5a, which is nearly perpendicular to the other strands but is connected by three hydrogen bonds to  $\beta$  strand 4. These hydrogen bonds are absent in the P450-BM3 structure. In the case of SiR-FP18, the analogous hydrogen bonds have been identified in the long-range, cross hydrogen-bond HNC0 experiment and are also present in the refined structure of SiR-FP18. The

dihedral angles of residues 177-179 are very close to those found in  $\beta$ -strand 5a in CPR. In addition, the analysis of the SWISS-MODEL structure revealed some ambiguities at the level of the loop connecting the fifth  $\beta$  strand and the C-terminal helix. In the SiR-FP18 sequence there is a two-residue deletion presumably in the loop region. Depending on the position of these two deletions either the  $\beta$  strand or the  $\beta$  helix of the SWISS-MODEL structure agree less well with the RDC data than the rest of the molecule. This indicates that the local structure is slightly different from the model structure. This is confirmed in the refined structure, where a change of conformation occurs in the short  $\beta$  loop between the final strand of the  $\beta$  sheet and the C-terminal helix, resulting in a slight shift of the position of the helix. It should be noted that this treatment of the available experimental data only allows us to judge the correctness of the backbone conformation of the SiR-FP18 model, and to refine this model. More detailed studies using NOE data or RDC measured on side-chain atoms are necessary in order to obtain a high resolution structure of SiR-FP18.

As expected, SiR-FP18 shows the same fold as bacterial flavodoxins which also possess an FMN as cofactor. At the present state of the work, no attempts were made to identify the exact localization of the prosthetic group in the protein. However, in bacterial flavodoxins as well as in human CPR, the amide proton chemical shifts in the phosphate binding loop are shifted to low-field (40, 41). A very analogous behavior is observed in the present study where the amide protons of S69, Q70, N73 and A74 are found between 9.5 and 11.5 ppm. These results confirm the structural conservation of the FMN-binding already suggested by the selective binding of FMN by SiR-FP18. A further key feature of FMN binding is the conservation of two aromatic residues which flank the isoalloxazine ring (Y80 and Y118 in human CPR and Y536 and W574 in P450-BM3, crystal structure numbering). However, in the sequence of SiR-FP18, the first of these residues is not conserved and replaced by a glutamine. A more detailed structural analysis is needed to elucidate the binding of FMN in case of SiR-FP18.

**ACKNOWLEDGEMENT.** We thank Dr. Stéphane Ménage for his help during EPR experiments and David Lascoux for the determination of the exact mass of the protein.

## REFERENCES

- Kredich, N. M. (1996) in *In Escherichia coli and Salmonella thyphimurium: Cellular and Molecular Biology*. (Neidhard et al., Eds.) pp 514-527, A.S.M. Press, Washington, D.C.
- Siegel, L. M., and Davis, P. S. (1974) *J. Biol. Chem.* 249, 1587-1598.
- Eschenbrenner, M., Covès, J., and Fontecave, M. (1995) *J. Biol. Chem.* 270, 20550-20555.
- Eschenbrenner, M., Covès, J., and Fontecave, M. (1995) *FEBS Lett.* 374, 82-84.
- Karplus, P. A., and Bruns, C. M. (1994) *J. Bioenerg. Biomembr.* 26, 89-99.
- Porter, T. D., and Kasper, C. B. (1986) *Biochemistry* 25, 1682-1687.
- Bredt, D. S., Hwang, P. M., Glatt, C. E., Lowenstein, C., Reed, R. R., and Snyder, S. H. (1991) *Nature* 351, 714-718.
- Siegel, L. M., Davis, P. S., and Kamin, H. (1974) *J. Biol. Chem.* 249, 1572-1586.
- Ostrowski, J., Barber, M. J., Rueger, D. C., Miller, B. E., Siegel, L. M., and Kredich, N. M. (1989) *J. Biol. Chem.* 264, 15796-15808.
- Crane, B. R., Siegel, L. M., and Getzoff, E. D. (1995) *Science* 270, 59-67.
- Covès, J., Zeghouf, M., and Fontecave, M. (2000) *Recent Res. Devel. Biochem.* 2, 97-107.
- Zeghouf, M., Fontecave, M., Macherel, D., and Covès, J. (1998) *Biochemistry* 37, 6114-6123.
- Zeghouf, M., Fontecave, M., and Covès, J. (2000) *J. Biol. Chem.* 275, 37651-37656.
- Gruez, A., Pignol, D., Zeghouf, M., Covès, J., Fontecave, M., Ferrer, J. L., and Fontecilla-Camps, J. C. (2000) *J. Mol. Biol.* 299, 199-212.
- Covès, J., Zeghouf, M., Macherel, D., Guigliarelli, B., Asso, M., and Fontecave, M. (1997) *Biochemistry* 36, 5921-5928.
- Wang, M., Roberts, D. L., Paschke, R., Shea, T. M., Masters, B. S., and Kim, J. J. (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 8411-8416.
- Zhao, Q., Modi, S., Smith, G., Paine, M., McDonagh, P. D., Wolf, C. R., Tew, D., Lian, L. Y., Roberts, G. C., and Driessen, H. P. (1999) *Protein Sci.* 8, 298-306.
- Sevrioukova, I. F., Li, H., Zhang, H., Peterson, J. A., and Poulos, T. L. (1999) *Proc. Natl. Acad. Sci. U. S. A.* 96, 1863-1868.
- Jansson, M., Li, Y. C., Jendeberg, L., Anderson, S., Montelione, B. T., and Nilsson, B. (1996) *J. Biomol. NMR* 7, 131-141.
- Bradford, M. M. (1976) *Anal. Biochem.* 72, 248-254.
- Laemmli, U. K. (1970) *Nature* 227, 680-685.
- Whitby, L. G. (1953) *Biochem. J.* 53, 437-442.
- Cordier, F., and Grzesiek, S. (1999) *J. Am. Chem. Soc.* 121, 1601-1602.
- Yang, D., Venters, R. A., Mueller, G. A., Choy, W. Y., and Kay, L. E. (1999) *J. Biomol. NMR* 14, 333-343.
- Brutscher, B. (2001) *J. Magn. Reson.* 151, 332-8.
- Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* 18, 2714-2723.
- Wishart, D. S., and Sykes, B. D. (1994) *J. Biomol. NMR* 4, 171-180.
- Dosset, P., Hus, J. C., Marion, D., and Blackledge, M. (2001) *J. Biomol. NMR* 20, 223-231.
- Hus, J. C., Marion, D., and Blackledge, M. (2000) *J. Mol. Biol.* 298, 927-936.
- Sibille, N., Pardi, A., Simorre, J. P., and Blackledge, M. (2001) *J. Am. Chem. Soc.* 123, 12135-12146.
- Sibille, N., Covès, J., Marion, D., Brutscher, B., and Bersch, B. (2001) *J. Biomol. NMR* 21, 71-72.
- Zeghouf, M., Defaye, G., Fontecave, M., and Covès, J. (1998) *Biochem. Biophys. Res. Commun.* 246, 602-605.
- Janick, P. A., and Siegel, L. M. (1983) *Biochemistry* 22, 504-515.

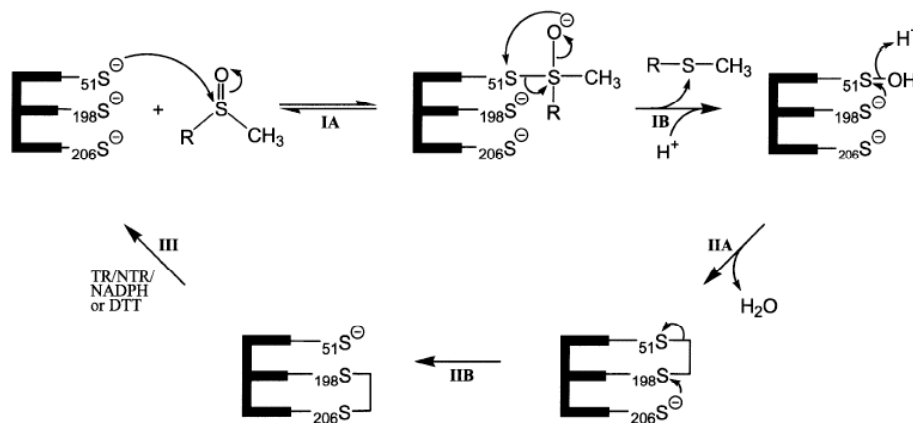


34. Siegel, L. M., Faeder, E. J., and Kamin, H. (1972) *Z. Naturforsch. [B]* 27, 1087-1089.
35. Vermilion, J. L., Ballou, D. P., Massey, V., and Coon, M. J. (1981) *J. Biol. Chem.* 256, 266-277.
36. Janick, P. A., and Siegel, L. M. (1982) *Biochemistry* 21, 3538-3547.
37. Siegel, L. M., Rueger, D. C., Barber, M. J., Krueger, R. J., Orme-Johnson, N. R., and Orme-Johnson, W. H. (1982) *J. Biol. Chem.* 257, 6343-6350.
38. Covès, J., Lebrun, C., Gervasi, G., Dalbon, P., and Fontecave, M. (1999) *Biochem. J.* 342, 465-472.
39. Hall, D. A., Vander Kooi, C. W., Stasik, C. N., Stevens, S. Y., Zuiderweg, E. R., and Matthews, R. G. (2001) *Proc. Natl. Acad. Sci. U. S. A.* 98, 9521-9526.
40. Peelen, S., Wijmenga, S., Erbel, P. J., Robson, R. L., Eady, R. R., and Vervoort, J. (1996) *J. Biomol. NMR* 7, 315-30.
41. Barsukov, I., Modi, S., Lian, L. Y., Sze, K. H., Paine, M. J., Wolf, C. R., and Roberts, G. C. (1997) *J. Biomol. NMR* 10, 63-75.

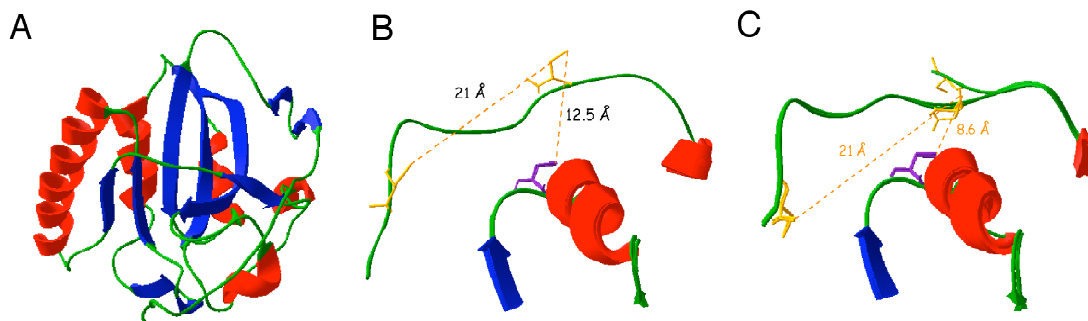
### 3.3 La protéine MsrA d'*Erwinia chrysanthemi* : les bases structurales du mécanisme réactionnel revisités.

Collaboration avec Jean-Pierre Chambost et Fred Barras du Laboratoire de Chimie Bactérienne de l'IBSM à Marseille. Le travail sur la MsrA réduite était le sujet de thèse de Sabine Béraud que j'ai co-dirigée avec Dominique Marion.

Les dommages oxydatifs créés sur les protéines par les radicaux libres ou autres dérivés oxygénés non radicalaires peuvent avoir des conséquences dramatiques pour la cellule. La méthionine, qu'elle soit libre ou liée à une protéine, s'oxyde très facilement en méthionine sulfoxyde. Cette modification, qui peut causer la perte de l'activité de la protéine, peut toutefois être inversée physiologiquement grâce à une enzyme ubiquitaire: la méthionine sulfoxyde réductase (MsrA). En plus de ses fonctions réparatrices et anti-oxydantes, cette enzyme pourrait faire partie d'un système de régulation d'activité cellulaire via des réactions d'oxydoréduction. MsrA joue aussi un rôle primordial pour la virulence de plusieurs bactéries telle que la bactérie pathogène de plante *E. chrysanthemi* (91). Le mécanisme réactionnel de MsrA fait intervenir trois cystéines et postule la formation de ponts disulfures entre elles (92,93 et Figure 13). Deux structures de MsrA réduite [*E. coli* (94) et bovin (95)] ont été résolues par cristallographie pendant le travail de thèse de S. Béraud.



**Figure 13.** Un des deux mécanismes catalytiques de MsrA proposés, faisant intervenir l'acide sulfénique. Les trois cystéines correspondent à C53, C200 et C208 dans la séquence d'*E. chrysanthemi*). Figure tirée de (92).



**Figure 14.** Structures cristallographiques de MsrA. A. Chaîne principale de MsrA d'*E. coli*. B. et C. Détail structural montrant l'orientation relative des trois cystéines impliquées dans la réaction catalytique dans la MsrA d'*E. coli* (B) et de bovin (C).

Les deux structures sont très similaires entre-elles. Elles sont caractérisées par un tonneau  $\beta/\beta$ , comportant 41% des résidues et deux larges boucles N- et C-terminales. La cystéine du site actif (C53 dans *E. chrysanthemi*) est accessible à la surface. Les deux cystéines C-terminales ne sont résolues que dans la moitié des différentes molécules (MsrA d'*E. coli* en forme trimérique, MsrA de bovin, formes mono- et dimérique). Cependant, la distance entre les cystéines est incompatible avec le mécanisme d'action sans une large réorientation de la boucle C terminale (Figure 14). Ceci nous a motivé d'examiner de plus près la position relative des cystéines ainsi que la dynamique du bras C-terminal en solution.

Nous avons entrepris l'étude structurale et dynamique de MsrA de *E. chrysanthemi* (221 acides aminés) en solution. Dans un premier temps, nous avons produit un échantillon de MsrA triplement marqué ( $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ ) puis effectué l'attribution de sa chaîne principale et de ses C $\alpha$  (96). Quatre types de paramètres RMN ont permis la localisation des éléments de structure secondaire dont les positions coïncident avec celles identifiées sur les deux structures cristallographiques.

Par la suite, la structure tridimensionnelle de MsrA en solution a été caractérisée par une stratégie originale faisant intervenir des constantes de couplage dipolaire résiduel. Deux jeux de données contenant les couplages  $^1\text{D}_{\text{HN}}$ ,  $^1\text{D}_{\text{COCA}}$ ,  $^2\text{D}_{\text{COH}}$  et  $^1\text{D}_{\text{CACB}}$  ont été mesurés dans deux milieux anisotropes différents. La comparaison de ces données avec celles calculées théoriquement à partir des structures cristallographiques montre une bonne corrélation pour l'ensemble des résidus sauf pour une région située dans la queue C terminale et comprenant une cystéine importante pour la catalyse. L'utilisation des seules contraintes de couplages dipolaires résiduelles a permis de déterminer *de novo* la conformation de cette région en solution [(97) et Article 6]. Celle-ci, différente de celle de MsrA d'*E. coli*, est plus favorable à la formation du premier pont disulfure. La structure obtenue a servi de modèle pour l'interprétation de la mobilité de la molécule. L'analyse des données de relaxation obtenues à deux champs différents (600 et 800 MHz) indique un ordre élevé pour toute la protéine. Ce résultat montre que les modèles catalytiques proposés nécessitent certainement une déstabilisation de la protéine.

# Direct structure determination using residual dipolar couplings: reaction-site conformation of methionine sulfoxide reductase in solution.

J|A|C|S

A R T I C L E S

Published on Web 10/25/2002

## Direct Structure Determination Using Residual Dipolar Couplings: Reaction-Site Conformation of Methionine Sulfoxide Reductase in Solution

Sabine Béraud,<sup>†</sup> Beate Bersch,<sup>†</sup> Bernhard Brutscher,<sup>†</sup> Pierre Gans,<sup>†</sup> Frédéric Barras,<sup>‡</sup> and Martin Blackledge<sup>\*†</sup>

Contribution from the Institut de Biologie Structurale, Jean-Pierre Ebel, Centre National de la Recherche Scientifique–Commissariat à l’Energie Atomique UJF, 41 rue Jules Horowitz, 38027 Grenoble Cedex, France, and the Laboratoire de Chimie Bactérienne, Centre National de la Recherche Scientifique, Institut Biologie Structurale et Microbiologie, 31 Chemin Joseph Aiguier, 13402, Marseille, France

Received May 13, 2002

**Abstract:** Residual dipolar couplings (RDC) from partially aligned molecules provide long-range structural data and are thus particularly well adapted to rapid structure validation or protein fold recognition. Extensive measurements in two alignment media can also provide precise *de novo* structure from RDC alone. We have applied a novel combination of these approaches to the study of methionine sulfoxide reductase (MsrA) from *Erwinia chrysanthemi*, a 27 kDa enzyme essential for repairing oxidative stress damage. The tertiary fold was initially validated by comparing backbone RDC to expected values from the crystal structure of the homologous MsrA from *Escherichia coli*. Good agreement was found throughout the chain, verifying the overall topology of the molecule, with the exception of the catalytically important peptide P196–L202, where strong and systematic RDC violation was observed. No evidence for local differential mobility in this region was detected, implying that the structure of the strand differs in the two molecules. We have therefore applied the *de novo* approach *meccano* to determine the conformation of this peptide using only RDC. A single conformation is found that is in agreement with all measured data. The aligned peptide can be docked onto the expected covalence of the rest of the template molecule while respecting its strictly defined relative orientation. In contrast to the structure of MsrA from *E. coli*, the reactive side chain of Cys200 is oriented toward the interior of the molecule and therefore closer to the catalytic Cys53, obviating the need for previously proposed conformational reorganization prior to formation of this disulfide intermediate. This analysis requires only backbone assignment and uses unambiguously assigned and readily measurable structural data, thereby greatly economizing investigation time compared to established nuclear Overhauser effect- (nOe-) based structure calculation methods.

Classical NMR structure determination requires extensive assignment of backbone and side-chain resonances followed by unambiguous identification of nuclear Overhauser effect (nOe) correlations between these assigned frequencies, before even low-resolution structural models can be derived.<sup>1</sup> Complete structural analysis can thus be particularly time-consuming for large proteins. In contrast, residual dipolar couplings (RDC) in macromolecules aligned in dilute liquid crystalline media<sup>2–4</sup> can be measured routinely from pairs of nuclei distributed throughout the molecule immediately following assignment of the backbone resonances. RDC provide coherent, long-range struc-

tural data, allowing entirely new approaches to structural biology in solution.<sup>5–8</sup> This coherence makes RDC particularly appropriate for comparison with available structural homologues, enabling rapid validation of initial molecular models or identification of folds from database structures<sup>9–11</sup> and subsequent refinement of these initial models.<sup>12</sup> Although this approach requires the presence of a structural homologue in accessible databases, the identification of even a low-resolution molecular fold from primary data is a major breakthrough for solution-state NMR.

\* Corresponding author: telephone (33) 4 38 78 95 54; fax (33) 4 38 78 54 94; e-mail martin@rnm.ibs.fr.

<sup>†</sup> Institut de Biologie Structurale, Jean-Pierre Ebel, CNRS-CEA.

<sup>‡</sup> Laboratoire de Chimie Bactérienne, CNRS, Institut Biologie Structurale et Microbiologie.

(1) Withrich, K. *NMR of proteins and nucleic acids*; Wiley: New York, 1986.

(2) Gayathri, C.; Bothner-by, A. A.; van Zyl, P. C. M.; Maclean, C. *Chem. Phys. Lett.* **1982**, *87*, 192–196.

(3) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9279–9283.

(4) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111–1114.

(5) Prestegard, J. H. *Nat. Struct. Biol.* **1998**, *5*, 517–522.

(6) Clore, G. M.; Gronenborn, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5891–5898.

(7) Tjandra, N. *Structure* **1999**, *7*, R205–R211.

(8) Bax, A.; Kontaxis, G.; Tjandra, N. *Methods Enzymol.* **2001**, *339*, 127–174.

(9) Annala, A.; Aitio, H.; Thulin, E.; Drakenberg, T. *J. Biomol. NMR* **1999**, *14*, 223–230.

(10) Meiler, J.; Peti, W.; Griesinger, C. *J. Biomol. NMR* **2000**, *17*, 283–294.

(11) Andreo, M.; Du, P.; Levy, R. M. *J. Am. Chem. Soc.* **2001**, *123*, 1222–1229.

(12) (a) Chou, J. J.; Li, S.; Bax, A. *J. Biomol. NMR* **2000**, *18*, 217–227. (b) Champier, L.; Sibille, N.; Bersch, B.; Brutscher, B.; Blackledge, M.; Covese, J. *Biochemistry* **2002**, *41*, 3770–3780.

In cases where sufficient measurements can be made from throughout the peptide chain in the presence of two alignment media, protein structure can be determined from RDC alone, either in comparison with short fragments from structural databases<sup>13</sup> or by sequentially building the peptide chain from unambiguously oriented peptide units.<sup>14</sup> The requirement of continuous data from peptide planes and/or tetrahedral junctions has so far limited the number of applications of these *de novo* methods to particularly well-behaved systems. In the absence of continuous RDC data, locally well-defined substructures may be isolated in the primary sequence and cannot easily be placed with respect to the rest of the molecule unless additional structural information, for example long-range NOEs, is also available.<sup>15</sup>

Methionine sulfoxide reductase (MsrA) catalyzes the reduction of free and protein-bound oxidized methionine residues and is consequently required for the repair of important enzymes.<sup>16–18</sup> The enzyme, found in nearly all living organisms, is present at high concentration in tissues and cells susceptible to oxidative stress damage<sup>19</sup> and has thus been linked to aging processes, lifespan determination,<sup>20</sup> and diverse degenerative pathologies such as Alzheimer's disease.<sup>21</sup> Two crystal structures of MsrA have recently been elucidated: MsrA from *Escherichia coli*<sup>22</sup> (MsrA<sup>Ecob</sup>) and *Bos taurus*<sup>23</sup> (MsrA<sup>bov</sup>). Although very similar, these structures differ significantly in the C-terminal strand involved in freeing the catalytic site for further function. To further understand the mechanism by which MsrA provides protection against oxidative stress, we have studied the solution structure of reduced MsrA from *Erwinia chrysanthemi* (MsrA<sup>Echmi</sup>, 221 amino acids), a plant pathogen that requires the enzyme for full virulence,<sup>24</sup> using RDC measured in partially aligned liquid crystal media.

To characterize the catalytically important C-terminal loop domain of MsrA<sup>Echmi</sup>, we have developed a novel combination of the two RDC-based approaches presented above: The tertiary fold is initially validated by RDC, providing low-resolution structural information and positioning the molecule in the common alignment frames. We then focus on the region of interest, which is determined from RDC alone by the *meccano* approach. This region can be unambiguously positioned relative to the remaining molecular scaffold by use of the graphic modeling tool Module, a program specifically developed for RDC analysis to facilitate the manipulation of oriented molecular domains relative to a common alignment frame.<sup>25</sup>

This is a highly efficient form of conformational analysis via NMR: unambiguously assigned structural data, requiring only backbone resonance assignment and derived from comparatively simple experimental techniques, provide precise structural detail in regions of specific interest in the molecule. In the example shown here, the characterization of the reaction site of MsrA provides potentially important insights into the molecular mechanism of the catalytic cycle of the enzyme.

## Methods

Samples of <sup>15</sup>N-labeled and <sup>13</sup>C,<sup>15</sup>N,<sup>2</sup>D-labeled MsrA<sup>Ecob</sup> were prepared as previously described.<sup>26</sup> Both samples were studied under reducing conditions with 10 mM dithiothreitol.

**Residual Dipolar Coupling Measurement.** All NMR experiments were performed on a Varian Inova 600 spectrometer, equipped with a triple-resonance (<sup>1</sup>H, <sup>15</sup>N, <sup>13</sup>C) probe and shielded z-gradients. Sample temperature was set to 298 K for all studies. RDC were collected on a 1.0 mM uniformly <sup>13</sup>C-, <sup>15</sup>N-, and 82% <sup>2</sup>H-labeled sample suspended in a liquid crystalline medium consisting of 20 mg/mL of the filamentous phage Pfl (Asla Ltd., Riga, Latvia) in 80 mM potassium phosphate buffer, pH 7.5. RDC were also collected on an identical sample suspended in a liquid crystalline medium consisting of 5% C12E6/hexanol<sup>27</sup> in 25 mM potassium phosphate buffer, pH 7.5. Four different types of dipolar couplings were measured as previously described<sup>28,29</sup> at 600 MHz, in both media: <sup>1</sup>D<sub>NH</sub>, <sup>1</sup>D<sub>CCα</sub> and <sup>2</sup>D<sub>CHβ</sub> couplings were obtained from 3D HNCQ-type experiments, and <sup>1</sup>D<sub>CCβ</sub> were measured in a 3D HNCoCA-type experiment. Data processing and peak-picking were performed with FELIX version 2000 (Accelrys Inc.).

**Relaxation Experiments and Data Analysis.** The <sup>15</sup>N R<sub>1</sub> and R<sub>1ρ</sub> relaxation and <sup>1</sup>H–<sup>15</sup>N NOE measurements were performed on the <sup>15</sup>N-labeled sample (1.4 mM concentration in 25 mM potassium phosphate buffer) at <sup>1</sup>H frequency of 600 MHz and temperature of 298 K with the classical <sup>1</sup>H-detected pulse sequence based on established methods<sup>30</sup> and described in detail elsewhere.<sup>31</sup> Data were analyzed by the model-free Lipari–Szabo approach with the program TENSOR2.<sup>32</sup>

**Fold Validation.** The crystal structure of MsrA<sup>Ecob</sup> (molecule A in PDB file 1FF3) was used to analyze and validate the solution structure of MsrA<sup>Ecob</sup>. Secondary structure was identified from experimental <sup>13</sup>C chemical shifts from random-coil values (chemical shift index, CSI<sup>33</sup>), and the intensity distribution of short- and medium range NOEs involving <sup>1</sup>H<sup>α</sup> protons.

The program Module<sup>25</sup> was used to evaluate the accordance between the MsrA<sup>Ecob</sup> structure and the experimental RDC data. Residual dipolar couplings can be expressed in terms of the orientation { $\theta$ ,  $\phi$ } of the internuclear vector relative to a common alignment tensor for the molecule:

$$D_{ij} = -\frac{\gamma_i \gamma_j \mu_0^2 \hbar^4}{16\pi^2 r_{ij}^3} \left( A_a (3 \cos^2 \theta - 1) + \frac{3}{2} A_r \sin^2 \theta \cos 2\phi \right) \quad (1)$$

where  $A_a$  and  $A_r$  are the axial and rhombic components of the alignment tensor,  $r_{ij}$  is the internuclear distance, and  $S$  is the order parameter.

- (13) Delaglio, F.; Kontaxis, G.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 2142–2143.
- (14) Hsu, J. C.; Marion, D.; Blackledge, M. J. *Am. Chem. Soc.* **2001**, *123*, 1541–1542.
- (15) Fowler, C. A.; Tian, F.; Al-Hashimi, H. A.; Prestegard, J. H. *J. Mol. Biol.* **2000**, *304*, 447–460.
- (16) Davis, D. A.; Newcomb, F. M.; Moskovitz, J.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J.; Fales, H. M.; Levine, R. L.; Yarchoan, R. *Biochem. J.* **2000**, *346*, 305–311.
- (17) Sigalov, A. B.; Stern, L. *FEBS Lett.* **1998**, *433*, 196–200.
- (18) Sun, H.; Gao, J.; Ferrington, D. A.; Biesiada, H.; Williams, T. D.; Squier, T. C. *Biochemistry* **1999**, *38*, 105–112.
- (19) Brot, N.; Weissbach, H. *Biopolymers* **2000**, *55*, 288–296.
- (20) Moskovitz, J.; Bar-Noy, S.; Williams, W. M.; Reguena, J.; Berlett, B. S.; Stadman, E. R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 12920–12925.
- (21) Gabbita, S. P.; Aksenov, M. Y.; Lovell, M. A.; Markesbery, W. R. *J. Neurochem.* **1999**, *73*, 1660–1666.
- (22) Tete-Favier, F.; Cobessi, D.; Boschi-Muller, S.; Azza, S.; Branlant, G.; Aubry, A. *Structure* **2000**, *8*, 1167–1178.
- (23) Lowther, W. T.; Brot, N.; Weissbach, H.; Matthews, B. W. *Biochemistry* **2000**, *39*, 13307–13312.
- (24) Hassouni, M. E.; Chambost, J. P.; Expert, D.; Van Gijsegem, F.; Barras, F. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 887–892.

- (25) Dosset, P.; Hsu, J. C.; Marion, D.; Blackledge, M. J. *Biomol. NMR* **2001**, *20*, 223–231.
- (26) Béraud, S.; Chambost, J. P.; Bersch, B.; Gans, P.; Barras, F.; Marion, D. *J. Biomol. NMR* **2001**, *20*, 97–98.
- (27) Rüdicke, M.; Otting, G. *J. Am. Chem. Soc.* **2000**, *122*, 7793–7797.
- (28) Yang, D.; Venters, R. A.; Mueller, G. A.; Choy, W. Y.; Kay, L. E. *J. Biomol. NMR* **1999**, *14*, 333–343.
- (29) Brutscher, B. *J. Magn. Reson.* **2001**, *151*, 332–338.
- (30) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1994**, *33*, 5984–6003.
- (31) Tsan, P.; Hsu, J.-C.; Caffrey, M.; Marion, D.; Blackledge, M. J. *Am. Chem. Soc.* **2000**, *122*, 5603–5612.
- (32) Dosset, P.; Hsu, J.-C.; Blackledge, M.; Marion, D. *J. Biomol. NMR* **2000**, *16*, 23–28.
- (33) Wishart, D. S.; Sykes, B. D. *Methods Enzymol.* **2001**, *239*, 363–392.

The alignment tensor is characterized by five parameters: the axial and rhombic components,  $A_{\parallel}$  and  $A_{\perp}$ , measure the extent of residual alignment due to the restricted orientational sampling in the anisotropic medium, whereas the Euler angles,  $\{\alpha, \beta, \gamma\}$ , define the nonaveraged orientation of the molecule relative to an external reference frame. For each entity, these five parameters were determined by nonlinear least-squares minimization of the target function over all couplings associated with a given domain:

$$\chi^2 = \sum_j \{D_{ij}^{\text{exp}} - D_{ij}^{\text{calc}}\}^2 / \sigma_{ij}^2 \quad (2)$$

where  $D_{ij}$  are the residual dipolar coupling between spins  $i$  and  $j$  and  $\sigma_{ij}$  is the uncertainty in the experimentally measured coupling. The average estimated  $\sigma_{ij}$  is on the order of 2 Hz. Alignment tensor parameters were determined and visualized relative to the three-dimensional atomic coordinates.

The quality of the fit to experimental RDC was inspected qualitatively with correlation plots of the measured and calculated couplings from the best-fit alignment tensor and quantitatively with the total  $\chi^2$  target function. Amplitude and orientation of the individual alignment tensors for the different structural elements were compared for evidence of differential flexibility. In a rigid molecule,  $A_{\parallel}$  and  $A_{\perp}$  should be identical for each structural entity, and the individual alignment tensors should be coaxial if the relative orientation of the secondary structure elements in the model structure is correct.

**De Novo Structure Calculation.** The recently published *meccano* approach<sup>14</sup> was slightly modified. The molecular dynamics (MD) program Sculptor<sup>34</sup> was incorporated into the sequential positioning algorithm designed to place the peptide units. This provides randomized sampling of conformational space, which facilitates inspection of the conformational precision in the final ensemble. The use of a molecular dynamics force field also allows the straightforward introduction of repulsive nonbonding interactions. In this case a specific force field was written to include only experimental RDC, reinforced backbone covalent terms, and simple repulsive nonbonding terms. Only (C, C $\alpha$ , C $\beta$ , H $\alpha$ , H $\beta$ , N) atoms were used in the calculation.

The experimental RDC for each peptide unit [from C $\alpha$  junction ( $i$ ) to C $\alpha$  junction ( $i + 1$ )] were successively applied to place the unit in the calculation frame.  $A_{\parallel}^i$ ,  $A_{\perp}^i$ ,  $A_{\parallel}^{i+1}$ , and  $A_{\perp}^{i+1}$ , and their relative orientations  $\{\alpha, \beta, \gamma\}$  are known from the analysis of the core region of the molecule and are fixed throughout the calculation. A short RDC-restrained molecular dynamics calculation consisting of 1000 0.1-fs heating steps (to a nominal 500 K) and 1000 0.1-fs cooling steps was applied for each peptide unit. One structure calculation takes 15 s on a Linux 1 GHz PC. Two thousand conformers were calculated by sequential positioning of peptide planes and C $\alpha$  junctions. The final ensemble was chosen on the basis of the total residual in the RDC target function  $\chi^2$  (eq 2). Covalent strain in the peptide planes or tetrahedral junctions in the final peptide structures was measured from the residual terms compared to a standard molecular force field (AMBER4).<sup>35</sup>

**Docking the Meccano Peptide to the Protein Scaffold.** The MsrA<sup>Ecol</sup> P196-L202 peptide was positioned relative to the crystal coordinates of MsrA<sup>Ecol</sup> with the molecular modeling tools available in the program Module. This program allows positioning of different oriented modules using only the equivalent orientations and three-dimensional translational freedom available in a common alignment frame. The crystal structure and peptide were aligned relative to their common frame and the peptide was positioned laterally such that the expected covalence was best satisfied. Module works in a single-alignment tensor subspace, so the procedure was repeated for both datasets to verify the common solution.

(34) Huis, J. C.; Marion, D.; Blackledge, M. J. *Mol. Biol.* **2000**, *298*, 927–936.  
(35) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 4.0*, University of California, San Francisco, CA, 1991.

All structure comparisons of the aligned peptide were performed with the rms difference between the coordinates in the common alignment frame rather than superimposing the coordinates to a best fit, as is the standard procedure, to ensure that the orientational information is preserved. The crystal structure of MsrA<sup>Bov</sup> (molecule B; PDB code 1FVB) was compared to the *meccano*-constructed MsrA<sup>Ecol</sup> P196-L202 peptide.

The chimeric structure, containing the *meccano* peptide with all atoms present and the backbone conformation of the MsrA<sup>Ecol</sup> crystal structure, was refined by low-temperature (300 K) RDC-restrained MD with Sculptor to regularize covalence. During this calculation, carried out with the force field AMBER4, the backbone atoms of the MsrA<sup>Ecol</sup> were tethered to their crystal structure conformation by use of the additional force-field potential:

$$E_{\text{teth}} = k_{\text{teth}} \sum_i \sqrt{(x_i - x_i^0)^2} / N \quad (3)$$

with a  $k_{\text{teth}}$  value of 20 kcal·mol<sup>-1</sup>·Å<sup>-2</sup>. All measured RDCs were used as standard restraints in the presence of the two independent tensors.

## Results and Discussion

We have assigned the backbone resonances<sup>26</sup> and measured extensive dipolar couplings (N–H $\beta$ , C'–H $\beta$ , C $\alpha$ –C', and C $\alpha$ –C $\beta$ ) from the reduced form of <sup>13</sup>C-, <sup>15</sup>N-, and <sup>2</sup>H-labeled MsrA<sup>Ecol</sup> in two liquid crystalline solutions. These couplings were initially compared to expected values from the recently published crystal structure of the 75% identical primary sequence homologue MsrA<sup>Ecol</sup> by use of the alignment tensor optimization and molecular modeling program Module. The reduced form of MsrA<sup>Ecol</sup> consists of an  $\alpha/\beta$  roll core structure, comprising 80/123 residues in secondary structural elements, and predominantly coil N- and C-terminal regions wrapped around this core. The 45 amino acid C-terminal region is of particular interest as two of the three cysteines, Cys200 and Cys208, present in the proposed catalytic cycle<sup>36,37</sup> are found in this strand (MsrA<sup>Ecol</sup> numbering is used unless otherwise stated). This cycle involves initial nucleophilic attack of Cys53 on Met-SO, followed by a disulfide cascade implicating Cys53–Cys200 and then Cys200–Cys208 disulfide bridges, the latter step freeing the catalytic sulfur (Cys53) for further function.

**Validation of the Backbone Fold in Solution.** Initial comparison, concentrating on the secondary structural elements of the molecule, reveals that the central core is folded very similarly in MsrA<sup>Ecol</sup> in solution and in the MsrA<sup>Ecol</sup> crystal structure (Figure 1 and Table 1). Alignment tensor axial and rhombic components from nonmobile sites in the molecule are reproduced similarly in the different structural motifs, and the relative orientations of the alignment tensor axes in the combined helical and combined  $\beta$ -sheet regions are in agreement with the MsrA<sup>Ecol</sup> crystal structure. This analysis was performed for RDC measured in both alignment media, further validating the expected fold of the core of the molecule. The best-fitting structural element is the combined helical region, which has a total  $\chi^2$  of 279 for 294 couplings in the two media ( $\chi^2/N = 0.95$ ). Alignment parameters were then determined for the C-terminal loop alone and in combination with the core structure. This again shows satisfactory agreement with experi-

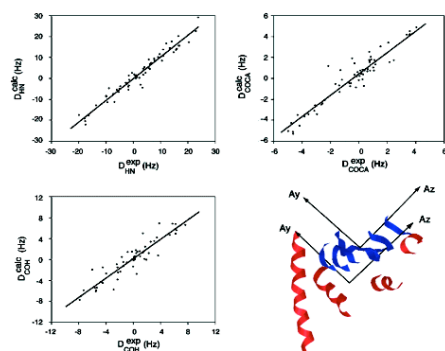
(36) Boschi-Muller, S.; Azza, S.; Sanglier-Cianferani, S.; Talfournier, F.; Van Dorsselaar, A.; Branlant, G. *J. Biol. Chem.* **2000**, *275*, 35908–35913.

(37) Lowther, W. T.; Brot, N.; Weissbach, H.; Honek, J. F.; Matthews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 6463–6468.

**Table 1.** Alignment Tensor Characteristics in the Two Liquid Crystal Media<sup>a</sup>

| medium                     |                        | N <sup>b</sup> | A <sub>z</sub> (10 <sup>-4</sup> ) | A <sub>x</sub> (10 <sup>-4</sup> ) | α (deg) | β (deg) | γ (deg) | χ <sup>2</sup> | χ <sup>2</sup> /N |
|----------------------------|------------------------|----------------|------------------------------------|------------------------------------|---------|---------|---------|----------------|-------------------|
| phages <sup>c</sup>        | α-helices <sup>d</sup> | 144            | 13.8 ± 0.3                         | 6.0 ± 0.3                          | 110 ± 5 | 157 ± 2 | 126 ± 5 | 139            | 0.96              |
|                            | β-sheets               | 110            | 14.8 ± 0.4                         | 6.4 ± 0.5                          | 108 ± 6 | 156 ± 3 | 128 ± 4 | 228            | 2.07              |
|                            | C-loop <sup>e</sup>    | 204            | 13.2 ± 0.2                         | 6.2 ± 0.3                          | 109 ± 4 | 157 ± 2 | 130 ± 3 | 291            | 1.43              |
| C12E6/hexanol <sup>f</sup> | α-helices              | 150            | 11.5 ± 0.3                         | 7.3 ± 0.3                          | 140 ± 2 | 87 ± 2  | 161 ± 2 | 140            | 0.93              |
|                            | β-sheets               | 106            | 11.8 ± 0.3                         | 6.9 ± 0.3                          | 139 ± 5 | 88 ± 2  | 162 ± 3 | 197            | 0.96              |
|                            | C-loop                 | 210            | 10.9 ± 0.2                         | 6.3 ± 0.2                          | 140 ± 2 | 87 ± 2  | 162 ± 2 | 355            | 1.69              |

<sup>a</sup>  $^1D_{NH}$ ,  $^1D_{C\alpha C\alpha}$ ,  $^2D_{CHN}$  and  $^1D_{C\alpha C\beta}$  RDC from MsrA<sup>Echmi</sup> were fitted to the conformation of MsrA<sup>Ecoli</sup> by use of the program Module. <sup>b</sup> Number of couplings present in the structural motif. <sup>c</sup> All secondary structural elements were also fitted together: in this case  $A_z = (14.0 \pm 0.3) \times 10^{-4}$ ,  $A_x = (6.4 \pm 0.3) \times 10^{-4}$ , and  $\chi^2 = 380$  (254 couplings). <sup>d</sup> α-helices (56–62, 85–89, 110–119, and 145–164) and β-sheets (45–50, 68–75, 95–104, 140–142, 175–177, and 183–185) using the numbering of the MsrA<sup>Echmi</sup> sequence. <sup>e</sup> C-loop region contains the C-terminal helix and the coil region until C208, but without the P196-L202 RDC data. <sup>f</sup> All secondary structural elements were also fitted together: in this case  $A_z = 11.62 \times 10^{-4}$ ,  $A_x = 6.89 \times 10^{-4}$ , and  $\chi^2 = 356$  (256 couplings considered).

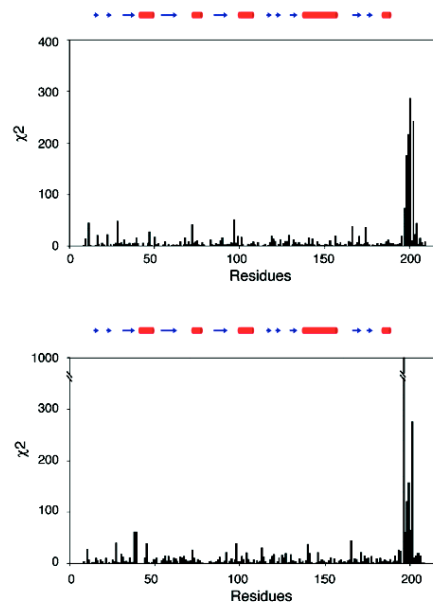


**Figure 1.** Comparison of measured RDC from MsrA<sup>Echmi</sup> with expected values for the crystal structure coordinates from MsrA<sup>Ecoli</sup> (a, bottom right) Relative orientation of the alignment tensors for the different secondary structure elements in MsrA<sup>Ecoli</sup> for the RDC from the C12E6/hexanol aligned sample. In this figure the alignment tensors of the central β-sheet (46–50, 68–75, 95–103, 139–141, 174–176, and 182–184), and α-helices (55–62, 84–88, 109–118, and 145–164) are compared relative to the crystal structure. The axial and rhombic components and χ<sup>2</sup> of the five secondary structure elements are given in Table 1. The different secondary structure elements were also fitted as a single domain using all couplings simultaneously as shown in panels b–d. (b, top left) Correlation between  $D_{NH}^{exp}$  and  $D_{NH}^{calc}$  in C12E6/hexanol for all core secondary structural elements shown in panel a. (c, bottom left) Correlation between  $^2D_{CHN}^{exp}$  and  $^2D_{CHN}^{calc}$  in C12E6/hexanol for all core secondary structural elements. (d, top right) Correlation between  $^1D_{C\alpha C\alpha}^{exp}$  and  $^1D_{C\alpha C\alpha}^{calc}$  in C12E6/hexanol for all core secondary structural elements.

mental RDC throughout the primary sequence, excepting the segment from Pro196 to Leu202, where strong disagreement is found (Figure 2). This local inconsistency is systematically found for each individual coupling type, in both alignment media. Removal of these data from the fit reproduces very similar alignment tensor values for the coil region alone (Table 1) as were found for the core region and verifies the relative orientation of these regions in the solution state.

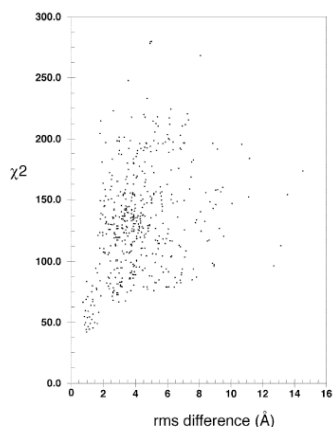
The observed systematic disagreement could be explained by differential dynamics in this region, producing time-averaged RDC values that are in disagreement with a single conformational model, or by a different local conformation in MsrA<sup>Echmi</sup> compared to the MsrA<sup>Ecoli</sup> crystal structure. <sup>15</sup>N relaxation data measured at 600 MHz <sup>1</sup>H frequency present no evidence for large amplitude motion on the rapid (pico- to nanosecond) or intermediate (micro- to millisecond) dynamic time scale in this region. By use of the Lipari–Szabo approach, the lowest order parameter in this region is 0.75 and the average is 0.84 (Figure S1, Supporting Information). None of the residues in the strand

13712 J. AM. CHEM. SOC. ■ VOL. 124, NO. 46, 2002



**Figure 2.** Local quality factor of the fit of RDC data from MsrA<sup>Echmi</sup> to the crystal structure of MsrA<sup>Ecoli</sup>. (a, top) Total χ<sup>2</sup> for the four different couplings measured in phage alignment medium with respect to the primary sequence. Residues showing significant rapid backbone dynamics were excluded from the plot (<sup>1</sup>H–<sup>15</sup>N heteronuclear nOe < 0.7). (b, bottom) Total χ<sup>2</sup> for the four different couplings measured in 5% C12E6/hexanol alignment medium with respect to the primary sequence. Residues showing significant rapid backbone dynamics were excluded from the plot (<sup>1</sup>H–<sup>15</sup>N heteronuclear nOe < 0.7). Secondary structure elements are represented along the sequence at the top of each graphic.

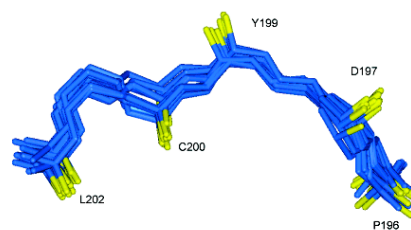
requires an exchange contribution to  $R_2$ . It should be noted that this analysis does not exclude the possibility of differential dynamics occurring in this region on a time scale that is not easily detectable by <sup>15</sup>N relaxation measurements. RDC quenching due to broad sampling of orientational space appears to be unlikely in view of the presence of large measured RDCs throughout the strand relative to the available range from the known tensors (Table S2, Supporting Information). In the absence of direct evidence for mobility, we have therefore attempted to calculate the local conformation of this peptide segment using only RDC, to determine whether a single conformation can be found that satisfies the available data and expected covalent geometry.



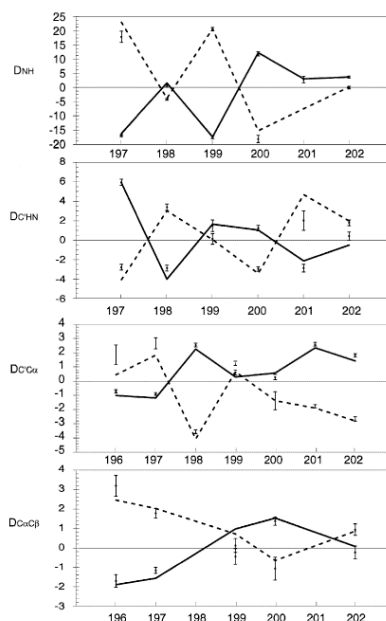
**Figure 3.** Conformational sampling of the structure calculation algorithm. The root-mean-square difference of the structures in the ensemble compared to the lowest  $\chi^2$  structure ( $\chi^2 = 40$ ) is plotted with respect to each individual  $\chi^2$ . Only conformers that do not violate any expected covalent angle by more than  $10^\circ$  are shown in the figure. Structures whose  $\chi^2$  falls below 62 form the ensemble shown in Figure 4.

**De Novo Structure Determination with *meccano*.** This analysis was performed via the *meccano* approach, previously demonstrated for the *de novo* determination of the backbone conformation of the protein ubiquitin.<sup>14</sup> In the case of MsrA<sup>Echm</sup>, the remainder of the protein structure is assumed to be identical to MsrA<sup>Ecolh</sup>, as suggested from the RDC analysis in the core region. Only the structure of the P196–L202 peptide is determined *de novo*. The *meccano* method requires that the eigenvalues and relative orientation of the alignment tensors are known, or can be determined, and that sufficient RDC are available from two differently orienting media to sequentially build the backbone conformation in the molecular calculation frame. In the case of MsrA<sup>Echm</sup> the two tensors are determined from the analysis of the secondary structural core of the molecule, as described above, in comparison with the coordinates of MsrA<sup>Ecolh</sup>. Despite the incompleteness of the RDC dataset over the whole sequence (only 90% of all possible RDC could be measured), nearly all potential RDC (47 from 48 potential couplings) were available from the peptide region, comprising six peptide planes and seven C $\alpha$  junctions (Table S2, Supporting Information).

Conformational sampling of the algorithm is illustrated in Figure 3, where the root-mean-square difference of the structural ensemble compared to the lowest  $\chi^2$  structure is plotted with respect to each  $\chi^2$  (for simplicity only those structures with valid covalent angles are included in the figure). The 20 lowest target function ( $\chi^2$ ) conformers, selected on the basis of the total  $\chi^2$  only, are shown in Figure 4 (average pairwise rms difference in this alignment frame is  $(1.2 \pm 0.3)$  Å over all atoms). The highest target function conformer in the final ensemble has  $\chi^2 = 62$ , and the best-fitting structure has a  $\chi^2$  of 40, close both to the expected experimental error and to the average from the core region ( $\chi^2/N = 0.85$  compared to  $\chi^2/N < 0.95$  for the helical regions of the crystal structure). While it is difficult to quantify the degree of confidence in the proposed model, the data appear to define a unique conformation as all structures with  $\chi^2 < 70$



**Figure 4.** Ensemble of 20 lowest target function structures representing the backbone conformation of peptide P196–L202, determined by use of only RDC with the algorithm *meccano*. The conformers were placed relative to the common alignment frame, rather than superimposing their coordinates, because their relative orientation is known a priori. The average pairwise rmsd in this alignment frame is  $(1.2 \pm 0.3)$  Å over the (C $\alpha$ , C $\beta$ , C $\gamma$ , H $\alpha$ , H $\beta$ , H $\gamma$ , N) atoms. For clarity only the (C $\alpha$ , C $\beta$ , C $\gamma$ , N) atoms are shown, and the C $\beta$  atoms are colored in yellow.



**Figure 5.** RDC values for each residue in the lowest energy *meccano*-derived conformation. Lines represent the values calculated from the model structure; the error bars are centered on the experimental values. Dashed lines correspond to the sample aligned in the alcohol medium, and solid lines, to the phage-aligned sample. These values are given in the Supporting Information.

have the same fold as the minimum  $\chi^2$  conformer (all-atom rms difference of the aligned conformer is less than 1.9 Å). Nevertheless over the whole 2000-conformer ensemble,  $\chi^2$  ranges to a maximum of 640, and conformational space is sampled very broadly (pairwise backbone rmsd of aligned conformers  $10.5 \pm 4.2$  Å), illustrating that algorithm efficiency is low.

The experimental and calculated values from the lowest target function model are shown in Figure 5. The covalent distortion present in the molecule due to possible overrestrained vector orientations appears to be minimal in the final ensemble as illustrated from the distances and dihedral and covalent angles



**Table 2.** Interatomic Distances and Covalent and Planar Dihedral Angles in the Lowest Energy *Meccano* Conformer

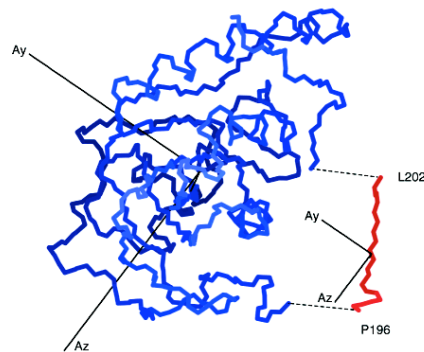
| atoms                               | distance (Å)              |
|-------------------------------------|---------------------------|
| N—H <sup>N</sup>                    | 1.02 ± 0.01               |
| C <sub>α</sub> —C <sub>β</sub>      | 1.53 ± 0.01               |
| C'—C <sub>α</sub>                   | 1.52 ± 0.02               |
| C <sub>α</sub> —N                   | 1.450 ± 0.003             |
| C'—N                                | 1.335 ± 0.002             |
| atoms                               | angles <sup>a</sup> (deg) |
| C'—C <sub>α</sub> —C <sub>β</sub>   | 109.2 ± 0.4               |
| N—C <sub>α</sub> —C <sub>β</sub>    | 109.3 ± 0.3               |
| C'—C <sub>α</sub> —N                | 109.6 ± 0.1               |
| C'—N—C <sub>α</sub>                 | 121.2 ± 0.2               |
| C <sub>α</sub> —C'—N                | 118.7 ± 0.2               |
| C'—N—H <sup>N</sup>                 | 118.4 ± 0.2               |
| C <sub>α</sub> —N—H <sup>N</sup>    | 121.2 ± 0.2               |
| C <sub>α</sub> —C'—N—C <sub>α</sub> | 179.4 ± 0.3               |

<sup>a</sup> No deviations of angles greater than 2.6° from the force field values were found.

shown in Table 2. Total RDC violation ( $\chi^2$ ) and molecular strain (as measured by the residual energy in the covalent terms of the molecular force field) are correlated over all structures ( $r = 0.89$ , data not shown), illustrating that the experimental data are coherent with conformers of expected peptide chain geometry and in disagreement with conformers of incorrect geometry. All dihedral angles are in most-favored or additionally-allowed regions of the Ramachandran plot.

**Insertion of the *meccano* Peptide into the MsrA<sup>EcoRI</sup> Scaffold.** Importantly, the length between the fragment termini P196-C<sub>α</sub> and L202-C<sub>α</sub> over the final ensemble ( $14.7 \pm 0.4$  Å) closely reproduces the equivalent length in MsrA<sup>EcoRI</sup> ( $14.9$  Å). To dock the *meccano* peptide, the equivalent peptide in MsrA<sup>EcoRI</sup> was removed from the X-ray model. The procedure for positioning the fragment must then respect the stringent orientational degrees of freedom available for the two structural domains relative to the common alignment tensor frames. Under these conditions a single relative alignment of the two domains exists,<sup>38</sup> imposing an unambiguous orientation of the peptide fragment relative to the crystal structure. The translational docking must then be performed so that this relative alignment is respected, while best satisfying the expected covalence. This procedure is illustrated with respect to one of the alignment tensors by use of the program Module (Figure 6). We find that the fragment can be directly accommodated into the crystal conformation with no serious violation of known covalence. The final ensemble falls into two  $\chi^2$ -equivalent subfamilies that differ at the Pro  $\psi$  backbone dihedral angle, and these subfamilies can be docked with covalent distances of  $\{d^{(N^{195}C, P^{196}N)} = 1.4 \pm 0.1$  Å and  $d^{(L^{202}C, G^{203}N)} = 1.35 \pm 0.07$  Å} and  $\{d^{(N^{195}C, P^{196}N)} = 2.0 \pm 0.1$  Å and  $d^{(L^{202}C, G^{203}N)} = 2.1 \pm 0.1$  Å)}. The expected peptide bond distance between C<sup>j</sup> and N<sup>j+1</sup> is approximately 1.33 Å. The fact that the minimum  $\chi^2$  conformation of this peptide reproduces all measured RDC and is compatible with the MsrA<sup>EcoRI</sup> structural scaffold provide a qualitative measure of our confidence in this approach to local structure determination.

**Focused *de Novo* Structure Determination Compared to Restrained MD Refinement.** Structure refinement of imprecise structural models with RDC has already been described, by use of restrained molecular dynamics in the presence of dipolar



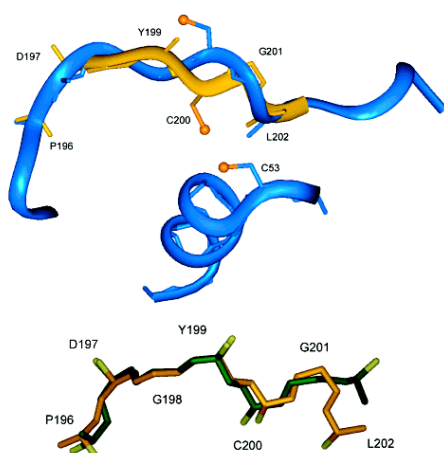
**Figure 6.** Representation of the positioning of the P196–L202 *meccano*-peptide relative to the crystal structure of MsrA<sup>EcoRI</sup> by use of the program Module to place the fragment with respect to the alignment tensor in phase. Only transverse degrees of freedom are available in the common coordinate system. The peptide can be easily accommodated at the C'-K195 and N-G203 positions without significantly violating known covalence.

couplings measured in a single alignment medium.<sup>12</sup> It may be instructive to underline the differences between our analysis and this kind of study: In contrast to restrained structure refinement, which is subject to stringent constraints, for example on the terminal positions of the peptide of interest, our approach is *de novo*, so that the conformation of the peptide is determined only by RDC and local covalence. Once the local structure has been determined, the ability to replace the oriented peptide in the molecular scaffold then provides an independent measure of the probability that the conformation is realistic. This *de novo* approach is of course only made possible because the available data is sufficient to unambiguously define the molecular conformation (eight RDC per peptide unit).<sup>14</sup> This is not true if RDC from a single alignment medium are available, where multiple conformations can exist for the same measured data. In such cases molecular dynamics-based refinement of an available structure becomes the most appropriate method. The different conformational sampling characteristics of the two methods are evident.

**Comparison with the Available Crystal Structures.** The differences between the *meccano* conformation of MsrA<sup>EcoRI</sup> and the X-ray crystal MsrA<sup>EcoRI</sup> structure in this region provide potentially important information concerning the proposed functional cycle of MsrA. While the Leu202 C<sub>α</sub>—C<sub>β</sub> vector direction is the same in both conformers, implying that the hydrophobic interactions for this side chain are conserved, the peptide chain around Cys200 has a very different conformation. Most importantly, the C<sub>α</sub>—C<sub>β</sub> vector is pointing away from the core in the X-ray structure and points into the core in the solution model (Figure 7, top). The Cys200 and Cys53 C<sub>β</sub> atoms are therefore 3.2 Å closer together in the chimeric model ( $d_{C\beta-C\beta} = 7.9$  Å) of MsrA<sup>EcoRI</sup>, and the disulfide-forming sulfurs are concomitantly closer, apparently obviating the need for previously suggested large-scale conformational rearrangement before the formation of this bond.<sup>22</sup> Note that no information concerning the exact position of the S<sup>y</sup> is available, as the  $\chi^1$  angle is undefined in our study.

It may be relevant that the crystal lattice contains three monomers, of which only one monomer (A) contains coordi-

(38) Al-Hashimi, H.; Valafar, H.; Terrell, M.; Zartler, M.; Eidsness, M.; Prestegard, J. H. *J. Magn. Reson.* 2000, 143, 402–406.



**Figure 7.** *De novo* RDC derived and crystal structures of the P196–L202 peptide. (a, top) Comparison of the native  $MsrA^{EcoRI}$  conformation (blue) and the lowest target function  $MsrA^{EcoRI}$  model (yellow) after positioning of the *meccano* peptide onto the  $MsrA^{EcoRI}$  structure. The backbone dihedral angles are  $\{(\phi, \psi) = (-129^\circ, 1^\circ)_{D197}, (-43^\circ, 163^\circ)_{G198}, (-72^\circ, 170^\circ)_{Y199}, (-150^\circ, 98^\circ)_{C200}, (-86^\circ, 18^\circ)_{G201}, (-34^\circ, \text{undefined})_{L202}\}$  for the  $MsrA^{EcoRI}$  model, compared to  $\{(\phi, \psi) = (-78^\circ, 166^\circ)_{Y199}, (-71^\circ, 133^\circ)_{G198}, (-137^\circ, 120^\circ)_{Y199}, (-131^\circ, 46^\circ)_{C200}, (-116^\circ, 17^\circ)_{G201}, (-120^\circ)_{L202}\}$  for the  $MsrA^{EcoRI}$  crystal structures. Note that the  $MsrA^{EcoRI}$  numbering is used here for both molecules. The principal implication of the different backbone conformations involves the direction of the Cys200  $C^\alpha$ – $C^\beta$  vector, and the consequent inter- $S'$  distances between Cys53 (shown in the  $MsrA^{EcoRI}$  crystal configuration) and Cys200. The  $S'$  position is unknown from the RDC-defined structure and is shown here placed on the available cone with the assumption of a  $\chi^1 = -60^\circ$  conformation. (b, bottom) Comparison of the Pro196–L202 peptide backbone conformation in  $MsrA^{EcoRI}$  determined from the direct structure calculation with RDC (yellow) and the equivalent conformation in  $MsrA^{bov}$  (green). For clarity only the ( $C^\alpha$ ,  $C^\beta$ ,  $C^\gamma$ , N) atoms are shown.

notes for the C-terminal region beyond Lys194. Superposition of the common main chain of monomer A on monomer B reveals that the Cys200  $S'$  atoms from molecules A and B would be 5.5 Å apart if the two C-terminal conformations were identical. This suggests that the A-monomer structure may indeed be influenced by the close proximity of the Cys-containing strands, as is already the case for the intermonomer disulfide bridge observed between  $^A\text{Cys208}$  and  $^B\text{Cys88}$ .<sup>22</sup> Interestingly, the more distantly related  $MsrA^{bov}$  (51% identity with several insertions) differs most significantly from the  $MsrA^{EcoRI}$  conformation in precisely this region, where the Cys218, equivalent to Cys200 in  $MsrA^{EcoRI}$ , points toward the core of the molecule, as in our model of  $MsrA^{EcoRI}$ . Indeed, the backbone of the RDC-defined peptide closely reproduces the  $MsrA^{bov}$  conformation between residues Pro196 and Cys200 (0.73 Å backbone rmsd over this strand; Figure 7, bottom). The inter- $C^\beta$  distance between Cys200 and Cys53 ( $d_{C^\beta-C^\beta} = 7.3$  Å and 8.0 Å) from the two crystalline conformations in  $MsrA^{bov}$  is also close to that predicted from the  $MsrA^{EcoRI}$  model. Note that the Leu202 conformation differs significantly in  $MsrA^{bov}$ , resulting in a distinct L202  $C^\alpha$ – $C^\beta$  bond orientation compared to  $MsrA^{EcoRI}$  and  $MsrA^{EcoRI}$ .

The  $MsrA^{EcoRI}$  crystal contains high levels of disorder in the P196–L202 region with a maximal amplitude at Cys200. It has therefore been suggested that the C-terminal strand is inherently

flexible and that this flexibility is related to the formation of disulfide intermediates involving Cys53, -200, and -208. Using NMR relaxation measurements we find no evidence of large-amplitude mobility on the ps-pico- to nanosecond time scale in this strand. Our data suggest that in  $MsrA^{EcoRI}$  the conformation of this peptide is well-defined and sufficiently close to Cys53 to allow the Cys53–Cys200 disulfide bridge to occur without conformational rearrangement. This hypothesis is supported by the observation that, in contrast to  $MsrA^{EcoRI}$ , the equivalent peptide strand in  $MsrA^{bov}$  shows no differential disorder in the crystal compared to the rest of the peptide chain. Note that the *meccano* model is most similar to the  $MsrA^{bov}$  conformation, although it contains elements common to both crystal structures. Of course the second step of the proposed catalytic cycle, involving the Cys200–Cys208 bridge, still requires backbone reorganization as these Cys  $C^\beta$  atoms are on the order of 20 Å apart. This conformational change may be induced by destabilization of the packing of the C-terminal loop region onto the core following formation of the initial disulfide bridge.

### Conclusions

NMR structure determination of large molecules by classical methods can be time-consuming due to the need for extensive assignment of backbone and side-chain resonances and the unambiguous identification of nOe correlations between these assigned frequencies. Structural genomics projects are currently providing an immense database, relating primary sequence to expected protein fold, an effort that still remains largely unexploited by the biomolecular NMR community. The routine measurement of RDC has recently been shown to provide a promising tool for low-resolution protein fold validation by comparatively simple experimental methodology. More precise structure determination is also possible with RDC but remains elusive for large molecules, where RDC data may be harder to measure extensively. In this study we propose a combination of these RDC-based techniques, initially for fold validation by use of a primary sequence homologue, followed by a focused structure determination of the site of interest, *de novo*, by only RDC.

This analysis required only backbone assignment and used only unambiguously assigned structural data, thereby greatly economizing investigation time and effort in comparison to established nOe-based structure calculation techniques, and demonstrates the enormous potential of using RDC for solution-state structural biology. Using the reaction site of the methionine sulfoxide reductase as an example, we have shown that RDC can provide precise local structure in functionally important regions of relatively large, highly deuterated molecules. This approach vastly simplifies the characterization of backbone structure in large molecules in solution, suggesting that this kind of analysis will have a significant impact on functional studies of biomolecules by NMR in the future.

**Acknowledgment.** This work was supported by the Commissariat à l'Énergie Atomique and the Centre National de la Recherche Scientifique.

**Supporting Information Available:** Graph of order parameters from the region of interest (Figure S1) and a table of RDC data used in the analysis (Table S2). This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA0268783

## 4 Conclusion

A travers mes divers sujets de recherche, j'ai essayé de montrer les différentes applications de la RMN à l'étude des propriétés structurales et dynamiques des protéines. La multitude des paramètres expérimentaux accessibles par la RMN font que son application n'est pas limitée à la pure détermination des structures tridimensionnelles. Cette technique a donc un impact important quand il s'agit de déterminer les propriétés moléculaires plus reliées à la fonction des biomolécules, comme, par exemple la dynamique moléculaire, l'étude de complexes intermoléculaires ou même les variations structurales liées au cycle catalytique. Des informations importantes peuvent souvent être obtenues facilement par des mesures simples et rapides des déplacements chimiques ou de paramètres de relaxation (exemple Fur). Le développement de l'approche utilisant l'information orientationnelle à partir des couplages dipolaires résiduels donne rapidement accès à l'information structurale globale ou locale. Ainsi, on peut déterminer sous certaines conditions le repliement de la protéine (exemples SiR-FP18 ou MerAa), ou la structure détaillée d'un fragment indépendamment du reste de la molécule (exemple MsrA), et ceci dès l'attribution des résonances de la chaîne principale. Quand il s'agit de la détermination de la structure à haute résolution, la RMN a toujours été une alternative intéressante pour les molécules de petite taille et joue évidemment un rôle très important pour les molécules ou fragments moléculaires difficiles à cristalliser. Dans les cas des petites protéines, elle peut être très compétitive avec la cristallographie aux rayons X sous condition d'un double marquage isotopique (exemple MerAa). Mais, d'une manière générale, avec les évolutions permanentes de cette technique, la limite de taille a été constamment repoussée et le processus entier de détermination de structure est constamment rendu plus efficace et plus rapide.

Au sein du groupe RMN à l'IBS, j'ai eu l'occasion de vivre une partie de l'histoire de la RMN relié à l'étude de biomolécules dont ce document veut témoigner en partie. Je tiens à remercier tous ceux qui ont participé à ce travail qui est le résultat des collaborations intenses avec des personnes aux compétences très complémentaires. J'aimerais également remercier nos deux chefs d'équipe successifs qui se sont investis pour que le groupe fonctionne, et ceci pas seulement en trouvant les moyens techniques et financiers. Enfin, un grand merci à tous de l'IBS et d'ailleurs qui ont contribué, chacun à sa façon à la réussite de mon travail.

## 5 Références

- 1 McIntosh, L.P., Dahlquist, F.W. Biosynthetic incorporation of  $^{15}\text{N}$  and  $^{13}\text{C}$  for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Quart. Rev. Biophys.* 23, 1-38, 1990.
- 2 Lian, L-Y., Middleton, D.A. Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Progr. NMR Spectr.* 39, 171-267 (2001).
- 3 Goto, N.K., Kay, L. New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr. Op. Struct. Biol.* 10, 585-592 (2000)
- 4 Clore, G.M., Gronenborn, A.M. Multidimensional heteronuclear magnetic resonance of proteins. *Meth. Enzymol.* 239, 349-362 (1994).
- 5 Sattler, M., Schleucher, J., Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Res. Spectr.* 34, 93-158 (1999).
- 6 Wand, A.J., Short, J.H. Nuclear magnetic resonance studies of protein—peptide complexes. *Meth. Enzymol.* 239, 700-717 (1994).
- 7 Kay, L.E., Torchia, D.A., Bax, A. Backbone dynamics of proteins as studied by  $^{15}\text{N}$  inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28, 8972-8979 (1989).
- 8 Bax, A. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci.*, 12, 1-16 (2003).
- 9 Kay, L.E., Gardner, K.H. Solution NMR spectroscopy beyond 25 kDa. *Curr. Opin. Struct. Biol.* 7, 722-731 (1997)
- 10 Gardner, K.H., Kay, L.E. The use of  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* 27, 357-406 (1998).
- 11 Pervushin, K., Riek, R., Wider, G., Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biomolecules in solution. *Proc. Natl. Acad. Sci. USA*, 94, 12366-12371 (1997).
- 12 Desplancq, D., Kieffer, B., Schmidt, K., Posten, C., Forster, A., Oudet, P., Strub, J.-M., Van Dorsselaer, A., Weiss, E. Cost-effective and uniform  $^{13}\text{C}$ - and  $^{15}\text{N}$ -labeling of the 24-kDa N-terminal domain of the *Escherichia coli* gyrase B by the overexpression in the photoautotrophic cyanobacterium *Anabaena* sp. PCC 7120. *Protein Expr. Purif.* 23, 207-217 (2001)
- 13 Lustbader, J.W., Birken, S., Pollak, S., Pound, A., Chait, B.T., Mirza, U.A., Ramnarain, S., Canfield, R.E., Brown, J.M. Expression of human chorionic gonadotropin uniformly labeled with NMR isotopes in Chinese hamster ovary cells: an advance toward rapid determination of glycoprotein structures. *J. Biomol. NMR* 7, 295-304 (1996).
- 14 Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., Yokoyama, S. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett.* 442, 15-9 (1999).
- 15 Sawasaki, T., Ogasawara, T., Morishita, R., Endo, Y. 2002. A cell-free protein synthesis system for high throughput proteomics. *Proc. Ntl. Acad. Sci.* 99, 14652-14657.
- 16 Blanchard, L., Marion, D., Pollock, B., Voordouw, G., Wall, J., Bruschi, M., Guerlesquin, F. Overexpression of *Desulfovibrio vulgaris* Hildenborough cytochrome  $c_{553}$  in *Desulfovibrio*

- desulfuricans* G200. Evidence of conformational heterogeneity in the oxidized protein. Eur. J. Biochem. 218, 293-301 (1993).
- 17 Caffrey, M.S., Davidson, E., Cusanovich, M.A., Daldal, F. Cytochrome  $c_2$  mutants of *Rhodobacter capsulatus*. Arch. Biochem. Biophys. 292, 419-426 (1992).
  - 18 Meyer, T.E. Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium *Ectothiorhodospira halophila*. Biochim. Biophys. Acta 806, 175-183 (1985).
  - 19 Hernandez, J.-F., Bersch, B., Pétillet, Y., Gagnon, J., Arlaud, G.J. Chemical synthesis and characterization of the epidermal growth factor-like module of human complement protease C1r. J. Peptide Res., 49, 221-231 (1997).
  - 20 Koenig, B.W., Rogowski, M., Louis, J.M. A rapid method to attain isotope labeled small soluble peptides for NMR studies. J. Biomol. NMR 26, 193-202 (2003).
  - 21 Vella, F., Thielens, N.M., Bersch, B., Arlaud, G.J., Frachet, P. A recombinant chimeric epidermal growth factor-like module with high binding affinity for integrins. J. Biol. Chem. 278, 19834-19843 (2003).
  - 22 Wüthrich, K. NMR of proteins and nucleic acids. John Wiley, New York (1986).
  - 23 Neuhaus, D., Williamson, M.P. The nuclear Overhauser Effect in Structural and Conformational Analysis, 2nd Edition. John Wiley, New York (2000).
  - 24 Case, D.A., Dyson, J.H., Wright, P.E. Use of chemical shifts and coupling constants in nuclear magnetic resonance structural studies on peptides and proteins. Meth. Enzymol. 239, 392-416 (1994).
  - 25 Karplus, M. Contact electron-spin coupling of nuclear magnetic moments. J. Chem. Phys., 30, 11-15 (1959).
  - 26 Englander, S.W., Kallenbach, N.R. Hydrogen exchange and structural dynamics of proteins and nucleic acids. Quart. Rev. Biophys. 16, 521-655 (1983).
  - 27 Aszodi, A., Gradwell, M.J., Taylor, W.R. Global fold determination from a small number of distance restraints. J. Mol. Biol. 251, 308-326 (1995).
  - 28 Gardner, K.H., Rosen, M.K., Kay, L.E. Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. Biochemistry 36, 1389-1401 (1997).
  - 29 Smith, B.O., Ito, Y., Raine, A., Teichmann, S., Ben-Tovim, L., Nietlispach, D., Broadhurst, R.W., Terada, T., Kelly, M., Oschkinat, H., Shibata, T., Yokoyama, S., Laue, E.D. An approach to global fold determination using limited NMR data from large proteins selectively protonated at specific residue types. J. Biomol. NMR 8, 360-368 (1996).
  - 30 Zhao, D., Jardetzky, O. An assessment of the precision and accuracy of protein structures determined by NMR. Dependence on distance errors. J. Mol. Biol. 239, 601-607 (1994).
  - 31 Clore, G.M., Robien, M.A., Gronenborn, A.M. Exploring the limits of precision and accuracy of protein structures determined by nuclear magnetic resonance spectroscopy. J. Mol. Biol. 231, 82-102 (1993).
  - 32 Güntert, P. Structure calculation of biological macromolecules from NMR data. Quart. Rev. Biophys. 31, 145-237 (1998).
  - 33 Blackledge, M.J., Medvedeva, S., Poncin, M., Guerlesquin, F., Bruschi, M., Marion. Structure and dynamics of ferrocycytochrome  $c_{553}$  from *Desulfovibrio vulgaris* studied by NMR spectroscopy and restrained molecular dynamics. J. Mol. Biol. 245, 661-681 (1995).

- 34 Arlaud, G.J., Gaboriaud, C., Thielens, N.M., Rossi, V., Bersch, B., Hernandez, J.-F., Fontecilla-Camps, J.C. Structural biology of C1: dissection of a complex molecular machinery. *Immunol. Rev.* 180, 136-145 (2001).
- 35 Arlaud, G.J., Gaboriaud, C., Thielens, N.M., Budayova-Spano, M., Rossi, V., Fontecilla-Camps, J.C. Structural biology of the C1 complex of complement unveils the mechanisms of its activation and proteolytic activity. *Mol Immunol.* 39,383-94 (2002).
- 36 Bersch, B., Hernandez, J.-F., Marion, D., Arlaud, G.J. Solution structure of the epidermal growth factor (EGF)-like module of human complement protease C1r, an atypical member of the EGF family. *Biochemistry* 37, 1204-1214 (1998).
- 37 Gaboriaud, C., Rossi, V., Bally, I., Arlaud, G.J., Fontecilla-Camps J.C. Crystal structure of the catalytic domain of human complement C1s: a serine protease with a handle. *EMBO J.* 19, 1755-65 (2000).
- 38 Budayova-Spano, M., Lacroix, M., Thielens, N.M., Arlaud, G.J., Fontecilla-Camps, J.C., Gaboriaud, C. The crystal structure of the zymogen catalytic domain of complement protease C1r reveals that a disruptive mechanical stress is required to trigger activation of the C1 complex. *EMBO J.* 21, 231-9 (2002).
- 39 Budayova-Spano M, Grabarse W, Thielens NM, Hillen H, Lacroix M, Schmidt M, Fontecilla-Camps JC, Arlaud GJ, Gaboriaud C. Related Articles, Links Abstract Monomeric structures of the zymogen and active catalytic domain of complement protease c1r: further insights into the c1 activation mechanism. *Structure.* 10, 1509-1519, (2002).
- 40 Gregory, L.A., Thielens, N.M., Arlaud, G.J., Fontecilla-Camps, J.C., Gaboriaud, C. X-ray structure of the Ca<sup>2+</sup>-binding interaction domain of C1s: Insights into the assembly of the C1 complex of complement. *J Biol Chem.* (in press).
- 41 Nilges, M., Macias, M.J., O'Donoghue, S.I., Oschkinat, H. Automated NOESY interpretation with ambiguous distance restraints: the refined solution structure of the pleckstrin homology domain from beta-spectrin. *J. Mol. Biol.* 269, 408-422 (1997).
- 42 Bersch, B., Rossy, E., Covès, J., Brutscher, B. Optimized set of two-dimensional experiments for fast sequential assignment, secondary structure determination, and backbone fold validation of <sup>13</sup>C/<sup>15</sup>N-labelled proteins. *J. Biomol. NMR* 27, 57-67 (2003).
- 43 Freeman, R., Kupče, E. New methods for fast multidimensional NMR. *J. Biomol. NMR* 27, 101-113 (2003).
- 44 Szyperski, T., Wider, G., Bushweller, J.H., Wüthrich, K. Reduced dimensionality in triple-resonance NMR experiments. *J. Am. Chem. Soc.* 115, 9307-9308 (1993).
- 45 Brutscher, B., Simorre, J.-P., Caffrey, M., Marion D. *J. Magn. Reson. B*, 105, 77-82 (1994).
- 46 Simorre, J.-P., Brutscher, B., Caffrey, M., Marion D. *J. Biomol. NMR*, 4, 325-333 (1994).
- 47 Ding, K., Gronenborn, A.M. Novel 2D triple-resonance experiments for sequential resonance assignments of proteins. *J. Magn. Res.* 156, 262-268 (2002).
- 48 Szyperski, T., Yeh, D.C., Sukumaran, D.K., Moseley, H.N.B., Montelione, G.T. Reduced-dimensionality NMR spectroscopy for high-throughput protein resonance assignment. *Proceed. Natl. Acad. Sci. USA* 99, 8009-8014 (2002).
- 49 Kim, S., Szyperski, T. GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J. Am. Chem. Soc.* 125, 1385-1393 (2003).
- 50 Brutscher, B. Intraresidue HNCA and COHNCA experiments for protein backbone resonance assignment. *J. Magn. Res.* 156, 155-159 (2002).

- 51 Brutscher, B. Accurate measurement of small spin-spin couplings in partially aligned molecules using a novel *J*-mismatch compensated spin-state-selection filter. *J. Magn. Reson.* 151, 332-338 (2001).
- 52 Dosset, P., Hus, J.-C., Blackledge, M. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR* 20, 927-936 (2001).
- 53 Banci, L., Bertini, I., Ciofi-Baffoni, S., Moltoni, E., Huffman, D.L., O'Halloran, T.V. Solution structure of the yeast copper transporter domain Ccc2a in the apo and Cu(I)-loaded states. *J. Biol. Chem.* 276, 8415-8426 (2001).
- 54 Hus, J.-C., Marion, D., Blackledge, M. *De novo* determination of protein structure by NMR using orientational and long-range order restraints. *J. Mol. Biol.* 298, 927-936 (2000).
- 55 Barkay, T., Miller, S.M., Summers, A.O. Bacterial mercury resistance from atoms to ecosystems. *FEMS Microbiol. Rev.* 27, 355-384 (2003).
- 56 Arnesano, F., Banci, L., Bertini, I., Ciofi-Baffoni, S., Molteni, E., Huffman, D.L., and O'Halloran, T.V. Metallochaperones and metal-transporting ATPases: a comparative analysis of sequences and structures. *Genome Res.* 12, 255-271 (2002).
- 57 Wishart, D.S., Sykes, B.D. The  $^{13}\text{C}$  chemical-shift index: a simple method for the identification of protein secondary structure using  $^{13}\text{C}$  chemical-shift data. *J. Biomol. NMR* 4, 171-180.
- 58 Hus, J.-C., Marion, D., Blackledge, M. Determination of protein backbone structure using only residual dipolar couplings. *J. Am. Chem. Soc.* 123, 1541-1542 (2001).
- 59 Delaglio, F., Kontaxis, G., Bax, A. Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *J. Am. Chem. Soc.* 122, 2142-2143 (2000).
- 60 Venters, R.A., Metzler, W.J., Spicer, L.D., Mueller, L., Farmer, B.T. II Use of  $^1\text{H}_\text{N}$ - $^1\text{H}_\text{N}$  NOEs to determine protein global folds in perdeuterated proteins. *J. Am. Chem. Soc.* 117, 9592-9593 (1995).
- 61 Rosen, M.K., Gardner, K.H., Willis, R.C., Parris, W.E., Pawson, T., Kay, L.E. Selective methyl group protonation of perdeuterated proteins. *J. Mol. Biol.* 263, 627-636 (1996).
- 62 Goto, N., Gardner, K.H., Mueller, G.A., Willis, R.C., Kay, L.E. A robust and cost-effective method for production of Val, Leu, Ile( $\square$ 1) methyl-protonated  $^{15}\text{N}$ -,  $^{13}\text{C}$ ,  $^2\text{H}$ -labeled proteins. *J. Biomol. NMR* 13, 369-374 (1999).
- 63 Hajduk, P.J., Augeri, D.J., Mack, J., Mendoza, R., Yang, J., Betz, S.F., Fesik, S.W. NMR-based screening of proteins containing  $^{13}\text{C}$ -labeled methyl groups. *J. Am. Chem. Soc.* 122, 7898-7904 (2000).
- 64 Cross, J.D., Gelev, V.M., Wagner, G. A sensitive and robust method for obtaining intermolecular NOEs between side chains in large protein complexes. *J. Biomol. NMR* 25, 235-242 (2003).
- 65 Wang, H., Janowick, D.A., Schkeryantz, J.M., Liu, X., Fesik, S.W. A method for assigning phenylalanines in proteins. *J. Am. Chem. Soc.* 121, 1611-1612 (1999).
- 66 Rajesh, S., Nietlispach, D., Nakayama, H., Takio, K., Laue, E.D., Shibata, T., Ito, Y. A novel method for the biosynthesis of deuterated proteins with selective protonation at the aromatic rings of Phe, Tyr, and Trp. *J. Biomol. NMR* 27, 81-86 (2003).
- 67 Prestegard, J.H., Al-Hashimi, H.M., Tolman, J.R. NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Quart. Rev. Biophys.* 33, 371-424 (2000).
- 68 Sibille, N., Bersch, B., Covès, J., Blackledge, M., Brutscher, B. Side chain orientation from methyl  $^1\text{H}$ - $^1\text{H}$  residual dipolar couplings measured in highly deuterated proteins. *J. Am. Chem. Soc.* 124, 14616-14625 (2002).

- 69 Escolar, L., Pérez-Martin, J., De Lorenzo, V. Opening the iron box: transcriptional metalloregulation by the fur protein. *J. Bacteriol.* 181, 6223-6229 (1999).
- 70 Coy, M., Neilands, J.B. Structural dynamics and functional domains of the fur protein. *Biochemistry* 30, 8201-8210 (1991).
- 71 Stojiljkovic, I., Hantke, K. Functional domains of the *Escherichia coli* ferric uptake regulator protein (Fur). *Mol. Gen. Genet.* 199-205 (1995).
- 72 De Lorenzo, V., Wee, S., Herrero, M., Neilands, J.B. Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (fur) repressor. *J. Bacteriol.* 169, 2624-2630 (1987).
- 73 Adrait, A., Jacquamet, L., Le Pape, L., Gonzalez de Peredo, A., Aberdam, D., Hazemann, J.-L., Latour, J.-M., Michaud-Soret, I. Spectroscopic and saturation magnetization properties of the manganese- and cobalt-substituted Fur (ferric uptake regulation) protein from *Escherichia coli*. *Biochemistry*, 38,6248-6260 (1999).
- 74 Jacquamet, L., Aberdam, D., Adrait, A., Hazemann, J.-L., Latour, J.-M., Michaud-Soret, I. X-ray absorption spectroscopy of a new zinc site in the fur protein from *Escherichia coli*. *Biochemistry* 37, 2564-2571 (1998).
- 75 Althaus, E.W., Outten, C.E., Olson, K.E., Cao, H., O'Halloran, T.V. The ferric uptake regulation (Fur) repressor is a zinc metalloprotein. *Biochemistry* 38, 6559-6569 (1999).
- 76 Gonzalez de Peredo, A., Saint-Pierre, C., Adrait, A., Jacquamet, L., Latour, J.-M., Michaud-Soret, I., Forest, E. Identification of two zinc-bound cysteines in the ferric uptake regulation protein from *Escherichia coli*: Chemical modification and mass spectrometry analysis. *Biochemistry* 38, 8582-8589 (1999).
- 77 Pohl, E., Haller, J.C., Mijovilovich, A., Meyer-Klaucke, W., Garman, E., Vasil, M.L. Architecture of a protein central to iron homeostasis: crystal structure and spectroscopic analysis of the ferric uptake regulator. *Mol. Microbiol.* 47, 903-915 (2003).
- 78 Gajiwala, K.S., Burley, S.K. Winged helix proteins. *Curr. Opin. Struct. Biol.* 10, 110-116 (2000).
- 79 Coy, M., Doyle, C., Besser, J., Neilands, J.B. Site-directed mutagenesis of the ferric uptake regulation gene of *Escherichia coli*. *Biometals* 7, 292-298 (1994).
- 80 Van Aalten, D.M.F., DiRusso, C.C., Knudsen, J. The structural basis of acyl coenzyme A-dependent regulation of the transcription factor FadR. *Embo J.* 20, 2041-2050 (2001).
- 81 Zheng, N., Fraenkel, E., Pabo, C.O., Pavletich, N.P. Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev.* 13, 666-674 (1999).
- 82 Van Melckebeke, H., Vreuls, C., Gans, P., Filée, P., Llabris, G., Joris, B., Simorre, J.-P. Solution structural study of BlaI: implications for the repression of genes involved in  $\beta$ -lactam antibiotic resistance. *J. Mol. Biol.*, in press (2003).
- 83 Ortiz de Orué Lucana, D., Schrempf, H. The DNA-binding characteristics of the *Streptomyces reticuli* regulator FurS depend on the redox state of its cysteine residues. *Mol. Gen. Genet.* 264, 341-353 (2000).
- 84 Ortiz de Orué Lucana, D., Tröller, M., Schrempf, H. Amino acids involved in reversible thiol formation and zinc ion binding in *Streptomyces reticuli* redox regulator FurS. *Mol. Genet. Genomics* 268, 618-627 (2003).
- 85 Garcia-Castellanos, R., Marrero, A., Mallorci-Fernandez, G., Potempa, J., Coll, M., Gomis-Rüth, F.X. Three-dimensional structure of MecI : molecular basis for transcriptional regulation of staphylococcal methicillin resistance. *J. Biol. Chem.*, in press (2003).



- 86 Lewis, R.A., Curnock, S.P., Dyke, K.G. Proteolytic cleavage of the repressor BlaI of beta-lactamase synthesis in *Staphylococcus aureus*. *FEMS Microbiol. Lett.* 178, 271-275 (1999).
- 87 Covès, J., Zeghouf, M., Fontecave, M. Molecular dissection as a tool for simplification: the example of the *Escherichia coli* sulfite reductase. *Recent Res. Devel. Biochem.*, 2, 97-107 (2000).
- 88 Zeghouf, M., Fontecave, M., Covès, J. A simplified functional version of the *Escherichia coli* sulfite reductase. *J. Biol. Chem.* 275, 37651-37656 (2000).
- 89 Gruez, A., Pignol, D., Zeghouf, M., Covès, J., Fontecave, M., Ferrer, J.-L., Fontecilla-Camps, J.C. Four crystal structures of the 60kDa flavoprotein monomer of the sulfite reductase indicate a disordered flavodoxin-like module. *J. Mol. Biol.* 299, 199-212 (2000).
- 90 Champier, L., Sibille, N., Bersch, B., Brutscher, B., Blackledge, M., Coves, J. Reactivity, secondary structure, and molecular topology of the *Escherichia coli* sulfite reductase flavodoxin-like domain. *Biochemistry*, 41, 3770-80 (2002)
- 91 Hassouni, M.E., Chambost, J.-P., Expert, D., Van Gijsegem, F., Barras, F. The minimal gene member *msrA*, encoding peptide methionine sulfoxide reductase, is a virulence determinant of the plant pathogen *Erwinia chrysanthemi*. *Proc. Natl. Acad. Sci. USA* 96, 887-892 (1999).
- 92 Boschi-Muller, S., Azza, S., Sanglier-Cianferani, S., Talfournier, F., Van Dorsselaer, A., Branlant, G. A sulfenic acid enzyme intermediate is involved in the catalytic mechanism of peptide methionine sulfoxide reductase from *Escherichia coli*. *J. Biol. Chem.* 275, 35908-35913 (2000).
- 93 Lowther, W.T., Brot, N., Weissbach, H., Honek, J.F., Matthews, B.W. Thiol-disulfide exchange is involved in the catalytic mechanism of peptide methionine sulfoxide reductase. *Proc. Natl. Acad. Sci. USA* 97, 6463-6468 (2000).
- 94 Tête-Favier, F., Cobessi, D., Boschi-Muller, S., Azza, S., Branlant, G., Aubry, A. Crystal structure of the *Escherichia coli* peptide methionine sulphoxide reductase at 1.9 Å resolution. *Structure* 8, 1167-1178 (2000).
- 95 Lowther, W.T., Brot, N., Weissbach, H., Matthews, B. Structure and mechanism of peptide methionine sulfoxide reductase, an “anti-oxidation” enzyme. *Biochemistry* 39, 13307-13312 (2000).
- 96 Béraud, S., Chambost, J.-P., Bersch, B., Gans, P., Barras, F., Marion, D. Backbone H<sup>N</sup>, N, C<sup>α</sup>, C<sup>β</sup> and C<sup>γ</sup> assignment of the 25 kD peptide methionine sulfoxide reductase from *Erwinia chrysanthemi*. *J. Biomol. NMR* 20, 97-98 (2001).
- 97 Béraud, S., Bersch, B., Brutscher, B., Gans, P., Barras, F., Blackledge, M. Direct structure determination using residual dipolar couplings : Reaction site conformation of methionine sulfoxide reductase in solution. *J. Am. Chem. Soc.* 124, 13709-13715 (2002).



## Curriculum vitae

### Beate Bersch

Date de naissance: 14 juillet 1962 à Hambourg (RFA)

Nationalité: Allemande

Situation familiale: vie maritale, 2 enfants nés en 1994 et 1997

#### Adresse personnelle:

85 route du Vernon  
38410 St Martin d'Uriage

#### Adresse professionnelle:

Institut de Biologie Structurale - JP Ebel  
UMR 5075 CEA-CNRS-UJF  
41 rue Jules Horowitz  
38027 Grenoble Cedex 1  
e-mail ☐ beate.bersch@ibs.fr

**Fonction actuelle :** Chargée de recherche première classe au CNRS ; affectation à l'Institut de Biologie Structurale à Grenoble, UMR5075 (depuis octobre 1992) dans le Laboratoire de RMN de Protéines, dirigé par Jean-Pierre Simorre

#### Formation :

- juin 1981 : Abitur (équivalent au Baccalauréat), Hambourg, RFA
- 1981-1987 : Etudes de Biologie, Université de Constance, RFA.
  - juillet 1983: Vordiplom (équivalent à la Licence)
  - 1986: Diplomarbeit dans le laboratoire de Biophysique membranaire sous la direction du Prof. Dr. P. Läuger: "Etudes de potentiels membranaires dans des vésicules à l'aide d'un colorant fluorescent"
  - janvier 1987: Diplom der Biologie (équivalent au D.E.A.)
- février-juin 1987: auditrice libre à l'Université Paris VI, Module Biophysique moléculaire, niveau Maîtrise.
- octobre 1987-février 1988 : participation aux cours de D.E.A. de Biophysique (Prof. C. Gary-Bobo, Prof. P. Vigny) comme auditrice libre à l'Université Paris VI.
- 1988-1992 : Doctorat ès sciences sous la direction de Prof. G. Ourisson, Laboratoire de Chimie Organique des Substances Naturelles, URA 31, Université Louis Pasteur, Strasbourg: "Interaction peptides-membranes: études RMN de la conformation membranaire de neuropeptides par NOE transféré".
  - 1988 ☐ Boursière Maxwell de l'Académie des Sciences
  - 1989-1992 ☐ Bourse du groupe industriel Elf Aquitaine
- 23.10.1992 : Obtention du grade de Docteur de l'Université Louis Pasteur à Strasbourg.
- 1.10.92 : Recrutement comme CR2 au CNRS, affectation à l'UPR 9015 à Grenoble
- 1996 : Promotion au grade CR1
- depuis juin 1997 ☐ Travail à temps partiel (80%)

**Publications:**

- Rossy E., Champier L., **Bersch B.**, Blackledge M., Brutscher B. Covès J. The amino-terminal extension of the mercuric reductase from *Ralstonia metallidurans* CH34 is a MerP-like module. Submitted to Journal of Biological Inorganic Chemistry.
- **Bersch B.**, Rossy E., Covès J., Brutscher B. (2003) Optimized set of two-dimensional experiments for fast sequential assignment, secondary structure determination, and backbone fold validation of  $^{13}\text{C}/^{15}\text{N}$ -labelled proteins. J. Biomol. NMR 27, 57-67.
- Vella F., Thielens N.M., **Bersch B.**, Arlaud G.J., Frachet P. (2003) A recombinant chimeric epidermal growth factor-like module with high binding affinity for integrins. J. Biol. Chem. 278, 19834-19843.
- D'Autréaux B., Touati D., **Bersch B.**, Latour J.-M., Michaud-Soret I. (2002) A new transcriptional response to nitrosative stress : direct Fur inhibition by nitric oxide via nitrosylation of the iron site. Proc. Natl. Acad. Sci. 99, 16619-16624.
- Sibille N., **Bersch B.**, Covès J., Blackledge M., Brutscher B (2002) Side chain orientation from methyl  $^1\text{H}$ - $^1\text{H}$  residual dipolar couplings measured in highly deuterated proteins. J. Am. Chem. Soc. 124, 14616-14625.
- Béraud S., **Bersch B.**, Brutscher B., Gans P., Barras F., Blackledge M. (2002) Direct structure determination using residual dipolar couplings : Reaction site conformation of methionine sulfoxide reductase in solution. J. Am. Chem. Soc. 124, 13709-13715.
- Champier L., Sibille N., **Bersch B.**, Brutscher B., Blackledge M., Covès J. (2002) Reactivity, secondary structure, and molecular topology of the *Escherichia coli* sulfite reductase flavodoxin-like domain. Biochemistry, 41, 3770-80.
- Sibille, N., Covès J., Marion D., Brutscher B., **Bersch B.** (2001)  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  assignment of the flavodoxin-like domain of the *Escherichia coli* sulfite reductase. J. Biomol. NMR, 21, 71-72.
- Béraud S., Chambost J.-P., **Bersch B.**, Gans P., Barras F., Marion D. (2001) Backbone  $\text{H}^{\text{N}}$ ,  $\text{N}$ ,  $\text{C}^{\alpha}$ ,  $\text{C}'$  and  $\text{C}^{\beta}$  assignment of the 25 kD peptide methionine sulfoxide reductase from *Erwinia chrysanthemi*. J. Biomol. NMR, 20, 97-98.
- Arlaud G.J., Gaboriaud C., Thielens N.M., Rossi V., **Bersch B.**, Hernandez J.F., Fontecilla-Camps J.C. (2001) Structural biology of C1: dissection of a complex molecular machinery. Immunol Rev., 180, 136-45.
- Auge S, **Bersch B.**, Tropis M, Milon A (2000) Characterization of substance P-membrane interaction by transferred nuclear Overhauser effect. Biopolymers 54, 297-306.
- **Bersch B.**, Blackledge M, Brutscher B, Cordier F, Hus JC, Marion D (1999) Protein Structure and dynamics by NMR, in Proteome and Protein Analysis, Kyriakidis DA, Choli-Papadopoulou T, Kamp RM (eds), Springer Verlag Berlin Heidelberg New-York Tokio, pp 91-116.
- Thielens NM, **Bersch B.**, Hernandez JF, Arlaud GJ (1999) Structure and functions of the interaction domains of C1r and C1s: keystones of the architecture of the C1 complex. Immunopharmacology, 42, 3-13.

- Arlaud GJ, Rossi V, Thielens NM, Gaboriaud C, **Bersch B**, Hernandez JF (1998) Structural and functional studies on C1r and C1s: new insights into the mechanisms involved in C1 activity and assembly. *Immunobiology* 199(2):303-16.
- **Bersch B**, Hernandez JF, Marion D, Arlaud GJ (1998) Solution structure of the epidermal growth factor (EGF)-like module of human complement protease C1r, an atypical member of the EGF family. *Biochemistry* 37(5):1204-14.
- Sebban C, Blanchard L, **Bersch B**, Blackledge M, Marion D, Dolla A, Guerlesquin F (1998) <sup>1</sup>H-NMR study of the structural influence of Y64 substitution in *Desulfovibrio vulgaris* Hildenborough cytochrome *c*<sub>553</sub>. *Eur. J. Biochem.* 251, 787-794.
- Hernandez JF, **Bersch B**, Pétillet Y, Gagnon J, Arlaud G (1997) Chemical synthesis and characterization of the epidermal growth factor-like module of human complement protease C1r. *J. Pept. Res.* 49, 221-231.
- **Bersch B**, Blackledge M, Meyer T, Marion D (1996) *Ectothiorhodospira halophila* ferrocyclochrome c-551 : solution structure and comparison with bacterial cytochromes c. *J.Mol.Biol.* 264, 567-584.
- Tropis M, Bardou F, **Bersch B**, Daffé M, Milon A (1996) Composition and phase behaviour of polar lipids isolated from *Spirulina maxima* cells grown in a perdeuterated medium. *Biochim. Biophys. Acta* 1248, 196-202.
- Guy P, Jaquinod M, Réminy H, Andrieu JP, Gagnon J, **Bersch B**, Dolla A, Blanchard L, Guerlesquin F, Forest E (1996) New conformational properties induced by the replacement of Tyr-64 in *Desulfovibrio vulgaris* Hildenborough ferricytochrome *c*<sub>553</sub> using isotopic exchanges monitored by mass spectrometry. *FEBS Lett.* 395, 53-57.
- Guy P, Réminy H, Jaquinod M, **Bersch B**, Blanchard L, Dolla A, Forest E (1996). Study of the new stability properties induced by amino acid replacement of tyrosine64 in cytochrome *c*-553 from *Desulfovibrio vulgaris* Hildenborough using electrospray mass spectrometry. *Biochem. Biophys. Res. Commun.*, 218, 97-103.
- **Bersch B**, Brutscher B, Meyer TE, Marion D (1995) <sup>1</sup>H and <sup>13</sup>C NMR assignments and structural aspects of a ferrocyclochrome c-551 from the purple phototrophic bacterium *Ectothiorhodospira halophila*. *Eur. J. Biochem.*, 227, 249-260.
- Blanchard L, Dolla A, **Bersch B**, Forest E, Bianco P, Wall J, Marion D, Guerlesquin F. (1994) Effect of the Tyr64-substitution on the stability of cytochrome *c*<sub>553</sub>, a low oxidoreduction-potential cytochrome from *Desulfovibrio vulgaris*Hildenborough. *Eur. J. Biochem.*, 226, 423-432.
- **Bersch B**, Koehl P, Nakatani Y, Ourisson G, Milon A (1993) <sup>1</sup>H nuclear magnetic resonance determination of the membrane-bound conformation of senktide, a highly selective neurokinin B agonist. *J. Biomol. NMR*, 3, 443-461.
- **Bersch B**, Starck JP, Milon A, Nakatani Y, Ourisson G (1993) Total synthesis of perdeuterated phospholipids. *Bull. Soc. Chim. Fr.* 130, 575-583.

- Apell HJ, **Bersch B** (1988) Na,K-ATPase in artificial lipid vesicles: potential dependent transport rates investigated by a fluorescence method. Prog Clin Biol Res. 268A, 469-476.
- Apell HJ, **Bersch B** (1987) Oxonol VI as an optical indicator for membrane potentials in lipid vesicles. Biochim. Biophys. Acta 903, 480-494.

**Communications orales et conférences:**

(ne sont mentionnées que celles représentatives de mon travail)

- Rôle des cystéines et du zinc dans l'activation et la dimérisation de Fur d'*E. coli*. B. d'Autréaux, L. Pecqueur, B. Brutscher, B. Bersch, Isabelle Michaud-Soret., GdR 1879 sur la Dérégulation du métabolisme du fer □ chimie, biologie et thérapeutiques, Evian, 31 août-3 septembre 2002
- Etude structurale du domaine FMN de la sulfite reductase d'*E. coli* par mesures de couplages dipolaires résiduels. Sibille N., Bersch B, Covès J., Blackledge M., Brutscher B. Ve Rencontre Rhône-Alpes de RMN, Lyon, juin 2002
- The use of anisotropic orientational and reorientational molecular behaviour for the determination of macromolecular structure in solution. S. Béraud, B. Brutscher, B. Bersch, N. Sibille, J.-P. Simorre, D. Marion, P. Dosset, J.-C. Hus □ M. Blackledge. La RMN, un Outil pour la Biologie, Paris, février 2002.
- Fur, an iron metalloregulatory protein from bacteria □ conformational changes of Fur upon metal activation and DNA binding. I. Michaud-Soret □ B. d'Autréaux; A. Gonzalez de Peredo □ L. Jacquamet □ O. Horner □ C. Jeandey; J.-L. Oddou; L. Le Pape □ B. Bersch; E. Forest; J.M. Latour. 1<sup>st</sup> Symposium on Chemical Biology of Metal Sensors with Switching Functions, Kyoto (Japon), juin 2001.
- Etude structurale par RMN du domaine flavodoxine de la sulfite réductase d'*Escherichia coli*. N. Sibille, J. Covès, D. Marion, B. Brutscher, B. Bersch. XVII Congrès du GERM (Groupe d'Etudes de Résonance Magnétique), La Pommeraye, mai 2001.
- Fur, an iron metalloregulatory protein from bacteria ; structural aspects and reactivity with nitric oxide. I. Michaud-Soret; B. D'Autréaux □ A. Gonzalez de Peredo □ O. Horner □ C. Jeandey □ J.L. Oddou; B. Bersch; E. Forest; J.M. Latour. 5<sup>th</sup> Winter Research Conferences, Val Cenis, mars 2001
- Spectroscopic and structural characterization of the Fur protein, iron sensor in bacteria □ conformational changes of Fur upon metal activation and DNA binding. I. Michaud-Soret □ B. d'Autréaux; A. Gonzalez de Peredo □ L. Jacquamet □ O. Horner □ C. Jeandey; J.-L. Oddou; L. Le Pape □ B. Bersch; E. Forest; J.M. Latour. Gordon Conference on Metals in Biology □ Ventura (USA), janvier 2001
- Etude structurale par RMN d'une protéine de 25 kDa: une méthionine sulfoxyde reductase. S. Béraud, J.-P. Chambost, B. Bersch, P. Gans, F. Barras, D. Marion. IV<sup>ème</sup> Journée Rhône-Alpes de RMN, Lyon, novembre 2000.

- Protein Structure and Dynamics by NMR. B. Bersch, M. Blackledge, B. Brutscher, F. Cordier, H. Déméné, P. Gans, L. Guilhaudis, J.-C. Hus, D. Marion, J.-P. Simorre, P. Tsan. XIIth International Conference on Methods in Protein Structure Analysis, Halkidiki (Grèce), septembre 1998.
- Structural and functional studies on the N-terminal CUB-EGF module pair of human complement protease C1r. N. Thielens, B. Bersch, J.-F. Hernandez, D. Marion, G. Arlaud. EU Module Project, Midterm Meeting, Oxford (Grande-Bretagne), juillet 1998.
- Structural and functional characterization of the EGF-like module of human C1r. J.-F. Hernandez, B. Bersch, J. Gagnon, D. Marion, G.J. Arlaud. IVth International Workshop on C1 and the Collectins, Mainz (Allemagne), 1997.
- Chemical synthesis, structural and functional characterization of the epidermal growth factor-like modules of human complement proteases C1r and C1s. J.-F. Hernandez, F. Hatt, B. Bersch, Y. Petillot, D. Marion, J. Gagnon, G.J. Arlaud. XXIVth Symposium of the European Peptide Symposium, Edinburgh (Grande-Bretagne), 1996.
- *Ectothiorhodospira halophila* ferrocyclochrome c-551 : détermination de la structure par RMN et comparaison avec d'autres cyclochromes c d'origine bactérienne. B. Bersch, M. Blackledge, T. Meyer, D. Marion. Conférence au Laboratoire de Chimie Bactérienne, CNRS, Marseille, juillet 1996.
- Solution structure of *Ectothiorhodospira halophila* ferrocyclochrome c-551. B. Bersch, M. Blackledge, T. Meyer, D. Marion. Les 24 heures de Biologie Structurale, Villard de Lans, mai 1996.
- Interaction peptides-membranes: études RMN de la conformation membranaire de tachykinines par NOE transféré. B. Bersch, Y. Nakatani, G. Ourisson, A. Milon. Vème Réunion du GEIMM (Groupes d'Etudes des Interactions Molécules-Membranes) à Rouen, mars 1993.
- Interaction peptides-membranes: études RMN de la conformation membranaire de tachykinines par NOE transféré. B. Bersch, Y. Nakatani, G. Ourisson, A. Milon. Conférence donnée à Sanofi Recherche, Toulouse et au Laboratoire de Pharmacologie et Toxicologie Fondamentales, CNRS, Toulouse, juillet 1992.
- Attribution totale du spectre <sup>1</sup>H-RMN de la Substance P. B. Bersch. Journées strasbourgeoises de Communications Orales de la Société Française de Chimie, décembre 1990.

### Communications par affiche:

(ne sont mentionnées que celles représentatives de mon travail)

- Etude structurale du processus de dimérisation de la protéine Fur d '*Escherichia coli*. L. Pecqueur, B. d'Autréaux, B. Brutscher, B. Bersch et I. Michaud-Soret. Réunion du GdR 1879 sur la dérégulation du métabolisme du fer, Gif, septembre 2003
- Fur et le métal migrateur responsable de la dimérisation. B. d'Autréaux, L. Pecqueur, O. Horner, C. Jeandey, J.L. Oddou, J.M. Latour, B. Bersch, B. Brutscher, I. Michaud-Soret. Club métalloprotéines, Carry leRouet, septembre 2003
- Very fast NMR structure determination of  $^{13}\text{C}/^{15}\text{N}$  labeled small to medium-sized proteins. B. Bersch, E. Rossy, J. Covès, M. Blackledge, B. Brutscher. VIème Rencontre Rhône-Alpes de RMN, juin 2003.
- Etude structurale du processus de dimérisation de la protéine Fur d '*Escherichia coli*. L. Pecqueur, B. d'Autréaux, B. Brutscher, B. Bersch et I. Michaud-Soret. VIème Rencontre Rhône-Alpes de RMN, juin 2003.
- NMR study of the E.coli sulfite reductase flavodoxin-like domain using residual dipolar couplings. Sibille N., Bersch B, Covès J., Blackledge M., & Brutscher B. 44<sup>th</sup> Experimental Nuclear Magnetic Resonance Conference. Savannah, Georgia, NM (USA), mars/avril 2003.
- Structural studies of the ferric uptake regulation protein (Fur) using NMR. B. Bersch, B. Brutscher, L. Pecqueur, I. Michaud-Soret. EMBO-ILL Workshop on Deuterium Labelling for Bio-molecular NMR and Neutron Scattering, Grenoble, janvier 2003 and Fifth European Symposium of the Protein Society, Florence, Italie, mars 2003.
- NMR Study of a highly deuterated protein using residual dipolar couplings. N. Sibille, J. Covès, B. Bersch, M. Blackledge, B. Brutscher. EMBO-ILL Workshop on Deuterium Labelling for Bio-molecular NMR and Neutron Scattering, Grenoble, janvier 2003.
- L. Pecqueur, B. D'Autréaux, B. Brutscher, B. Bersch, I. Michaud-Soret. Rôle des cystéines et du zinc dans l'activité et la dimérisation de Fur. 2èmes journées scientifiques de l'IMBG, Autrans, décembre 2002
- L. Pecqueur, B. D'Autréaux, B. Brutscher, B. Bersch, I. Michaud-Soret. Rôle des cystéines et du zinc dans l'activité et la dimérisation de Fur. Réunion du Club métalloprotéines et modèles, Carry-le-Rouet, sept./oct. 2002
- NMR study of the flavodoxin-like domain of the *E. coli* sulfite reductase. N. Sibille, B. Bersch, J. Covès, M. Blackledge, B. Brutscher. Embo Pratical course □ Multidimensional NMR in Structural Biology, Il Ciocco, Italie, août 2002.
- Direct structure determination in deuterated proteins □ reaction site conformation of methionine sulfoxide reductase. S. Béraud, B. Brutscher, F. Barras, P. Gans, B. Bersch, M. Blackledge. Vème rencontre Rhône-Alpes de RMN, Lyon, octobre 2001.



- NMR study of a 25 kDa protein: a methionine sulfoxide reductase. S. Béraud, B. Bersch, M. Blackledge, P. Gans, D. Marion. NMR in Molecular Biology: from Structure to Function - Folding, Catalysis, Interaction and Signalling. Denmark, juin 2001.
- Etude structurale par RMN du domaine flavodoxine de la sulfite réductase d'*Escherichia coli*. N. Sibille, J. Covès, D. Marion, B. Brutscher, B. Bersch. Xe Journée Grenobloise de RMN, Le Fontanil, juin 2001.
- Rôle des cystéines et du zinc dans l'activité et la dimérisation de Fur. I. Michaud-Soret □ B. D'Autréaux □ B. Bersch □ J.M. Latour. Réunion du réseau Physique et Chimie des Métalloprotéines et de leurs Analogues Synthétiques, Autrans, 4-7 mars 2001.
- La protéine Fur : clonage, marquage isotopique et études par RMN 3D. B. Bersch; B. D'Autréaux, I. Michaud-Soret. Réunion du réseau Physique et Chimie des Métalloprotéines et de leurs Analogues Synthétiques, Giens, avril 2000.
- Domaines structuraux des protéases du complément humain: modélisation et méthode d'expression pour une étude RMN. S. Béraud, B. Bersch, N. Thielens, G. Arlaud et D. Marion. XVIème congrès du Groupement d'Etudes de Résonance Magnétique, Aussois, mai 1999.
- Solution structure of *Ectothiorhodospira halophila* ferrocyclochrome c-551 B. Bersch, M. Blackledge, D. Marion. XXIIth International Biophysics Congress, Amsterdam, août 1996 et XVIIth International Conference on Magnetic Resonance in Biological Systems, Keystone, Colorado, USA, août 1996.
- <sup>1</sup>H and <sup>13</sup>C nuclear magnetic resonance investigation and structural aspects of a ferrocyclochrome c-551 from the purple phototrophic bacterium *Ectothiorhodospira halophila*. B. Bersch, B. Brutscher, T. Meyer & D. Marion. Colloque IMABIO à Montpellier (septembre 1994).
- Conformational studies of cytochromes c by nuclear magnetic resonance and restrained molecular dynamics. B. Bersch, M. Blackledge, B. Brutscher, M. Caffrey, P. Gans, D. Marion, N. Morelle, M. Poncin & J.-P. Simorre. Réunion scientifique annuelle du GDR "Métalloprotéines et Analogues de Synthèse" à Villard de Lans (janvier 1994).
- Interaction peptides-membranes : études RMN de la conformation membranaire de tachykinines par NOE transféré. B. Bersch, Y. Nakatani, G. Ourisson & A. Milon. 5ème Réunion du Groupe Thématique Magnétisme Nucléaire et Biologie à Toulouse (octobre 1993).
- Relation structure / potentiel d'oxydo-reduction du cytochrome c553 de *Desulfovibrio*. L. Blanchard, A. Dolla, B. Bersch, S. Medvedeva, M. Blackledge, D. Marion & F. Guerlesquin. 5ème Réunion du Groupe Thématique Magnétisme Nucléaire et Biologie à Toulouse (octobre 1993).
- Synthesis of Perdeuterated Phospholipids with Varying Polar Head-Groups : Application in the <sup>1</sup>H-NMR Investigation of Peptide-Membrane Interaction. B. Bersch, J.-P. Starck, A. Milon, Y. Nakatani & G. Ourisson. 18<sup>e</sup> Symposium IUPAC sur la Chimie des Substances Naturelles, Strasbourg (30 août - 4 septembre 1992).

- Interaction peptides-membranes: synthèse de phospholipides et études RMN par NOE transféré. B. Bersch, A. Milon, Y. Nakatani et G. Ourisson. IVème Réunion du GEIMM (Groupes d'Etudes des Interactions Molécules-Membranes) à Strasbourg (novembre 1991).
- Interaction peptides-membranes: synthèse de phospholipides et études RMN par NOE transféré. B. Bersch, A. Milon, Y. Nakatani et G. Ourisson. Congrès de la Société Française de Chimie, Strasbourg 1991 (septembre 1991).

### Encadrements

- Avril 1994 à juin 1995 □ Encadrement d'un stagiaire en maîtrise physique (100%), «Modélisation moléculaire de la substance P en interaction avec des membranes □
- Octobre 1995 à juin 1996 □ Encadrement de Fabienne Hatt, stagiaire du DEA de RMN et Cristallographie Biologiques (100%), «Etude structurale du module EGF de la protéine C1r du complément humain C1 □
- Octobre 1997 à juin 1998 □ Encadrement de Sabine Béraud, stagiaire de DEA de RMN et Cristallographie Biologiques (100%), «Etapes d'une étude RMN □ de la surexpression d'un module de protéine à la caractérisation structurale d'un peptide □
- Octobre 1998 à mars 2002 □ Encadrement du travail de thèse de Sabine Béraud (soutenue le 20 mars 2002) □ «Caractérisations Structurale et Dynamique par Résonance Magnétique Nucléaire d'une Protéine de 25 kDa □ la Méthionine Sulfoxyde Réductase d'*Erwinia chrysanthemi* » □ (en co-direction avec Dominique Marion).
- à partir d'octobre 2000 □ Directeur du travail de thèse de Nathalie Sibille «Etude par RMN du domaine FMN de la sulfite réductase d'*Escherichia coli* □ (en co-direction avec Bernhard Brutscher).
- juillet 2001 □ Encadrement d'un stagiaire en licence de biochimie (100%) □ «Introduction à des techniques de bioinformatique et modélisation moléculaire basée sur l'homologie de séquences □.
- à partir d'octobre 2001 □ Encadrement de Ludovic Pecqueur pendant son stage de DEA de Biologie Structurale et Fonctionnelle et thèse (en co-direction avec Isabelle Michaud-Soret), «Structure de la protéine Fur d'*Escherichia coli* par RMN □.
- mars 2003 à octobre 2003 □ Encadrement du Dr Jianghua Feng, stagiaire post-doctoral financé par le Ministère de la Recherche dans le cadre de l'accueil de jeunes chercheurs étrangers, «Elucidation du cycle catalytique de la méthionine sulfoxyde réductase, enzyme impliqué dans la réparation des dommages oxydatifs de protéines □.