



HAL
open science

Utilisation de l'ingénierie des protéines pour l'étude des relations structure/fonction dans les protéines

Thierry Vernet

► **To cite this version:**

Thierry Vernet. Utilisation de l'ingénierie des protéines pour l'étude des relations structure/fonction dans les protéines. Biologie structurale [q-bio.BM]. Université Louis Pasteur de Strasbourg, 1994. tel-01302402

HAL Id: tel-01302402

<https://hal.univ-grenoble-alpes.fr/tel-01302402>

Submitted on 14 Apr 2016

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.

*Résumé des travaux de recherche, en vue de l'obtention du
diplôme d'habilitation à diriger des recherches*

(février 1994)

*Utilisation de l'ingénierie des protéines pour l'étude
des relations structure/fonction dans les
protéines*

Thierry Vernet

*Agent de Recherche Associé
au Conseil National de Recherches du Canada*

*Laboratoire de Génétique des Eucaryotes
(Dr. David Y. Thomas)*

**Institut de Recherche en Biotechnologie
Conseil National de Recherches du Canada
Montréal**

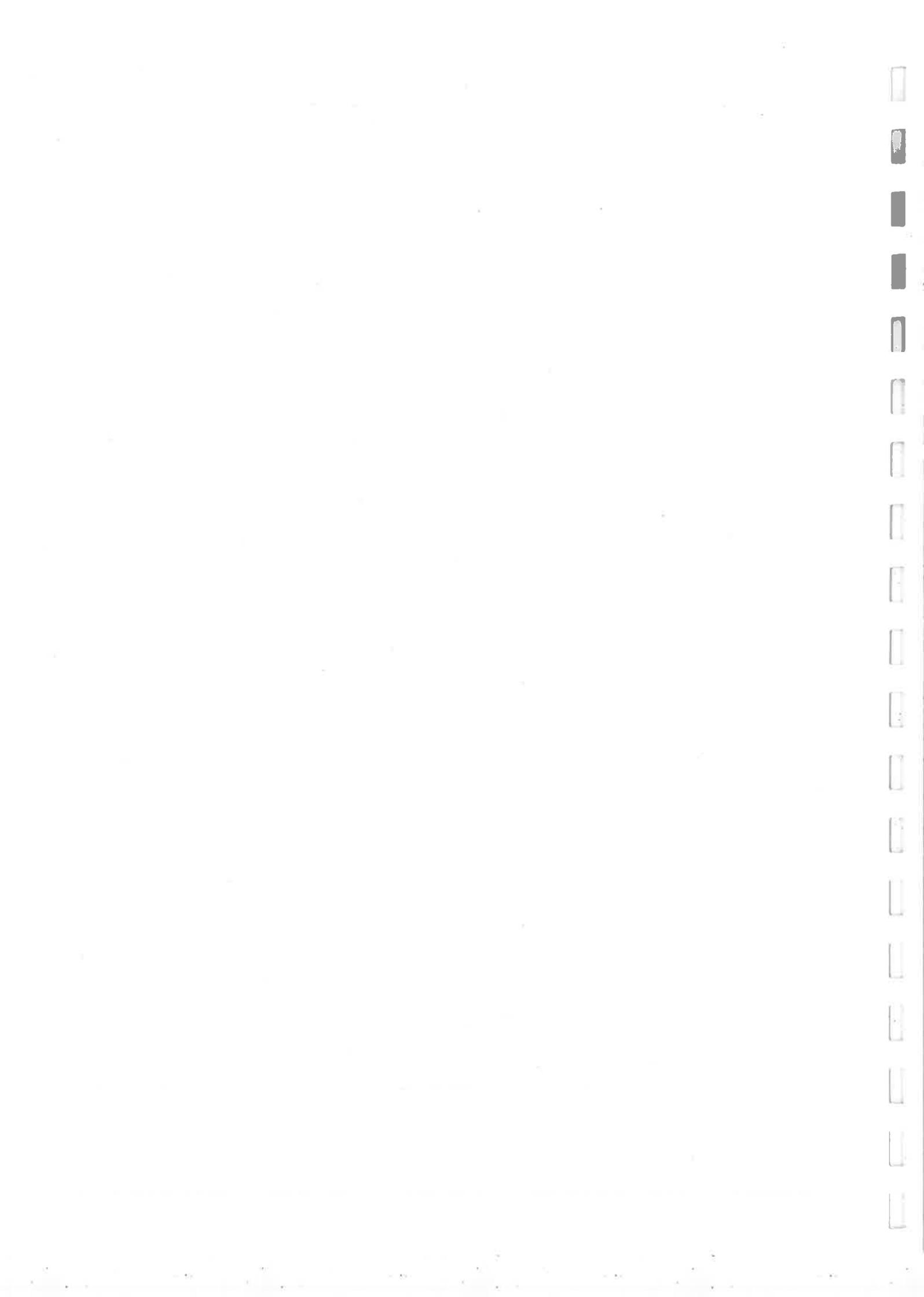
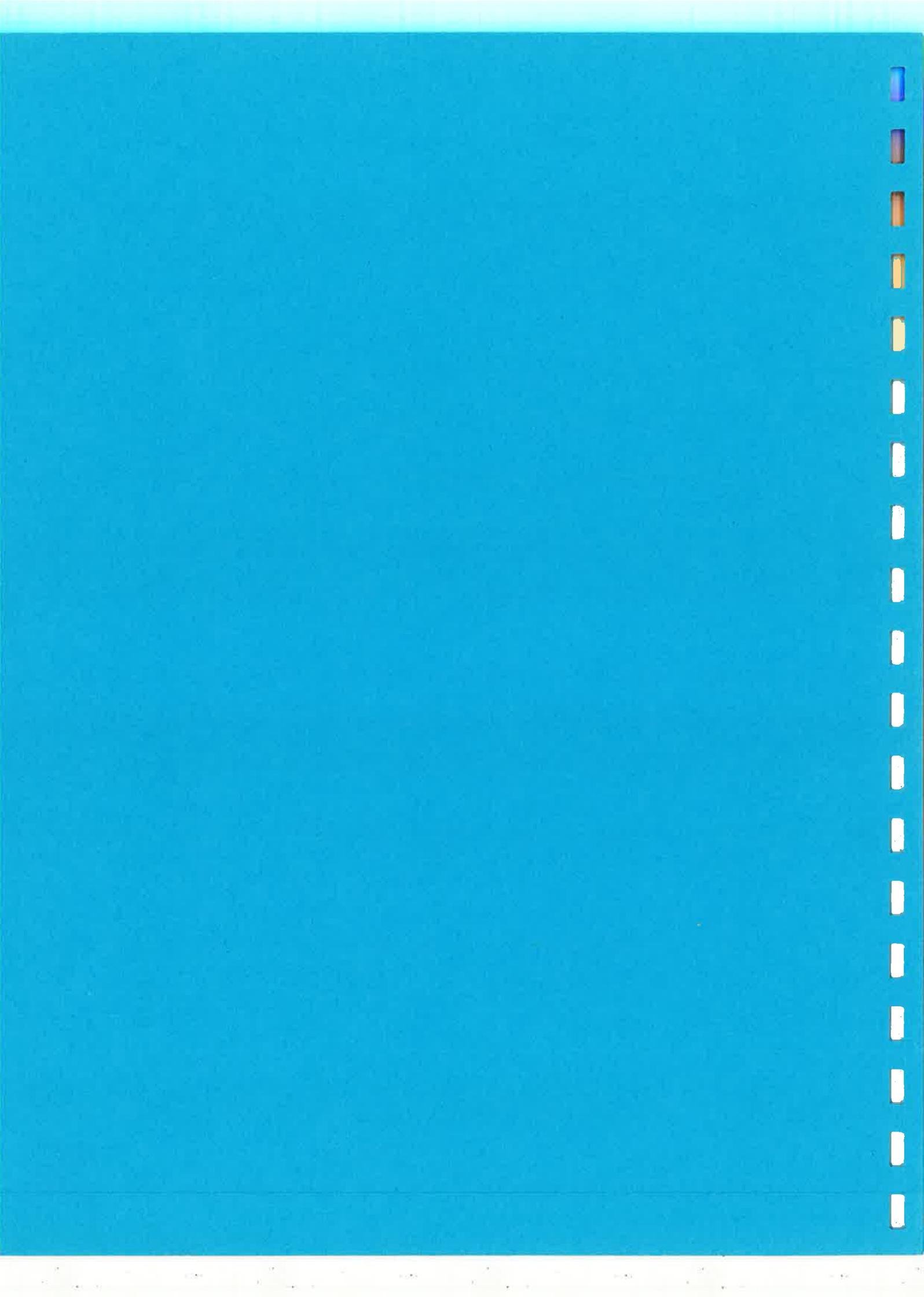


TABLE DES MATIERES

PREMIERE PARTIE:	
PRÉSENTATION DE L'ACTIVITÉ SCIENTIFIQUE	4
CURRICULUM VITAE.....	5
ACTIVITE SCIENTIFIQUE.....	7
LISTES DES PUBLICATIONS.....	18
SYNTHESE DE L'ACTIVITE SCIENTIFIQUE.....	24
1- <i>Préambule</i>	25
2- <i>Période 1978-1981</i>	
<i>Différenciation chez les plantes supérieures</i>	26
2.1- Contexte professionnel.....	26
2.2- Clonage d'une sonde génétique pour la petite sous-unité de la rubisco.....	26
2.3- Régulation de l'expression de la petite sous-unité.....	27
3- <i>Période 1981-1984</i>	
<i>Etude des plasmides colicinogènes et applications</i>	27
3.1- Contexte professionnel.....	27
3.2- Caractérisation des plasmides colicinogènes d' <i>E. coli</i>	29
3.3- Développement d'un vecteur d'expression régulé à sélection directe.....	29
3.4- Etude de la maintenance des plasmides en absence de pression de sélection.....	31
4- <i>Période 1984-1987</i>	
<i>Etude du système "killer" chez Saccharomyces cerevisiae</i>	33
4.1- Contexte professionnel.....	33
4.2- Construction d'une série de vecteurs d'expression pour la levure.....	35
4.3- Etude du système toxine "killer" de la levure.....	37
4.3.1- Introduction au système "killer".....	37
4.3.2- Elucidation du facteur d'immunité de la toxine K1.....	39
4.3.3- Application du système "killer".....	42
5- <i>Période 1987-1993</i>	
<i>Ingénierie des hydrolases, étude des relations structure/fonction des protéines</i>	42
5.1- Contexte professionnel.....	42
5.2- Ingénierie de la papaïne.....	43
5.2.1- Mise au point de systèmes d'expression.....	43
5.2.2- Caractérisation du précurseur de la papaïne.....	46
5.2.3- Définition des rôles de la région pro.....	49
5.2.4- Etude du site actif.....	55
5.2.5- Etude de la spécificité de la papaïne.....	59
5.3- Ingénierie des lipases.....	60
5.4- Mutations coordonnées des protéines monomériques.....	62

DEUXIEME PARTIE	
PRÉSENTATION DES TRAVAUX PUBLIÉS.....	67
1- Différenciation chez les plantes supérieures.....	69
2- Développement de vecteurs, techniques et méthodes.....	70
3- Etude des systèmes toxine/immunité chez <i>E. coli</i> et chez <i>Saccharomyces cerevisiae</i>	71
4- Relation structure/fonction des hydrolases.....	72
TROISIEME PARTIE	
PROGRAMME DE RECHERCHE.....	73
1- Préambule.....	74
2- Etude du précurseur de la papaïne.....	74
3- Relations structure/fonction des lipases GCL.....	75
4- Rôles des contacts domaine-domaine et entre sous-unités des protéines.....	76
5- Etude du repliement des protéines par analyse de répertoires de protéines à séquence aléatoire.....	77
6- Remerciements.....	79

Première partie: présentation de l'activité scientifique



CURRICULUM VITAE

ETAT CIVIL

Nom: Vernet
Prénom: Thierry
Date et lieu de naissance: 8 mai 1955, Coblenz (Allemagne)
Nationalités: française, canadienne
Situation de famille: Marié, deux enfants

Adresse personnelle: 5130 Grand Boulevard
Montréal, Qc, Canada, H3X 3S4
Tél: (514) 485-1743

Adresse professionnelle: Conseil National de Recherches du
Canada, Institut de Recherche en
Biotechnologie (IRB)
6100 avenue Royalmount
Montréal, Qc, Canada, H4P 2R2
Tél: (514) 496-6154
FAX: (514) 496-6213
e-mail: Vernet@BRI.NRC.CA

FORMATION UNIVERSITAIRE

1976: DEUG B, Université Louis Pasteur, Strasbourg.

1978: Maîtrise de biochimie, Université Louis Pasteur, Strasbourg.

1979: Diplôme d'études approfondies. *Option:* biochimie et biologie moléculaire, mention bien, Université Louis Pasteur, Strasbourg.

1981: Doctorat de Troisième Cycle. *Option:* biochimie et biologie moléculaire. Sous la direction du Dr. J. Fleck. "*Approche moléculaire de la différenciation chez les plantes supérieures*". Soutenu le 28 septembre 1981. Université Louis Pasteur, Strasbourg.

1994: Habilitation à diriger des recherches en préparation. Université Louis Pasteur, Strasbourg. Soutenance prévue en 1994.

BOURSES ET PRIX

Bourse de l'académie de Strasbourg DEA (1979)
Bourse de la DGRST (1980-1981)
Prix de la fondation Jean-Walter Zellindja (1981)
Bourse du président du CNRC (1987)

CARRIERE SCIENTIFIQUE

Associé de recherche, Division des Sciences Biologiques, Conseil National de Recherches du Canada, Ottawa (nov. 1981 - mai 1984)

Associé de recherche, Institut de Recherche en Biotechnologie, Conseil National de Recherches du Canada, Montréal (juin 1984 - juillet 1985)

Agent de recherche adjoint, Institut de Recherche en Biotechnologie, Conseil National de Recherches du Canada, Montréal (aout 1985 - juillet 1988)

Agent de recherche associé, Institut de Recherche en Biotechnologie, Conseil National de Recherches du Canada, Montréal (juillet 1988 - présent)

Chargé de cours, université de Montréal (1988-1992)

Professeur associé, Institut Armand Frappier, Université du Québec (janvier 1994- présent)

ACTIVITE SCIENTIFIQUE

1- APPARTENANCE A DES ORGANISATIONS PROFESSIONNELLES

Membre de l' American Association for the Advancement of Science
(depuis 1991)

2- ACTIVITE D'ENSEIGNEMENT

"*Molecular Genetics of Protein Secretion*" (3h), université d'Ottawa
(1985)

"*Protein secretion in microorganisms: applications to industrial
microbiology*" (2h), université Concordia, Montréal (1985)

"*Microbiologie et enzymologie industrielle*" (3h), école polytechnique,
Montréal (1986)

"*Biologie moléculaire et cellulaire*" (45 h/an), université de Montréal
(de 1988 à 1992)

"*Mutagenesis as a tool in molecular and cellular biology*" (24h/an),
cours de travaux pratiques et cours magistral, université McGill,
Montréal (de 1987 à 1993)

"*L'ingénierie des protéines*" (3h/an), université Laval, Québec (depuis
1990)

"*Répertoires moléculaires et biotechnologie*", (2h) Institut Armand
Frappier (1994)

3- INTERACTIONS AVEC L'INDUSTRIE

Construction d'une souche recombinante de levure industrielle
(1992-1993).

Octroi d'une licence non-exclusive à Stratagene® pour l'utilisation du
peptide signal de la mélitine d'abeille appliquée à la sécrétion de
protéines recombinantes exprimées dans les cellules d'insectes

(1993). Un vecteur dérivé du pVT-Bac est disponible dans leur catalogue depuis fin 1993.

Expression, sous contrat, d'une protéine recombinante dans la levure pour un partenaire industriel (1994)

4- ARBITRAGE ET JURY

Arbitre de publications soumises à *Biochemica et Biophysica Acta*, *Gene* et *Molecular and General Genetics*

Examineur externe des thèses de maîtrise (Msc) de M. Alain Moreau, Institut Armand Frappier (1988) et M^{me} Eugenia Zalce Institut Armand Frappier (1993)

Arbitre pour demandes d'octroi de recherche pour le Wellcome Trust (Angleterre) et le Conseil de Recherche Médicales du Canada

Agent de liaison scientifique pour le programme de coopération université-industrie du Conseil de Recherches en Sciences Naturelles et en Génie du Canada (1991-présent)

5- COLLABORATIONS NATIONALES ET INTERNATIONALES

Dr. S. A. Narang (Division des sciences biologiques, Ottawa). Expression de l'insuline humaine dans *E. coli* (1982-1983)

Dr. I. J. McDonald (Division des sciences biologiques, Ottawa). Stabilisation des plasmides lors des cultures d'*E. coli* en fermenteur (1982-1983)

Dr. H. Bussey (Université McGill, département de biologie). Le système killer de la levure (1984-1986)

Dr. R. Brousseau (IRB, Montréal). Développement d'un système de ligation d'oligonucléotides simple brin (1985)

Dr. C.Y. Kang (Université d'Ottawa, département de microbiologie). Expression de l'antigène de surface de l'hépatite B dans la levure (1986)

Dr. A. C. Storer (IRB, Montréal). Ingénierie des protéases à cystéine (depuis 1986)

Dr. C. Richardson (IRB, Montréal). Développement d'un système d'expression dans les cellules d'insectes (1987)

Dr. D. Altschuh (ULP, Strasbourg). Programme d'échange international entre le CNRC et le CNRS. Mutations coordonnées des protéines. Etude des interactions entre domaines et sous-unités des protéines (depuis 1988)

Collaborateur (1990-1993) et participant (1994) du réseau de centres d'excellence en génie protéique. Ingénierie des protéases impliquées dans des pathologies

Dr. J. Scharfstein (Université fédérale de Rio de Janéiro, Brésil) Caractérisation des protéases à cystéine de *Trypanosoma cruzi* (depuis 1990)

Dr. M. Cygler (IRB, Montréal). Etude par mutagenèse des lipases I et II de *Géotrichum candidum* (depuis 1991)

6- ORGANISATION D'ÉVÉNEMENTS SCIENTIFIQUES

Organisateur de l'atelier "*Expression of foreign proteins in E. coli*" Institut de recherche en biotechnologie (1^{er} décembre 1987)

Initiateur puis modérateur du comité d'organisation des "*journées de l'IRB*" (2 journées début mai, de 1988 à 1992)

Responsable de l'organisation de la réunion hebdomadaire "*les jeudis de génétique*" (1989-1991)

Membre du comité d'organisation des séminaires hebdomadaires IRB. Négociateur avec les commanditaires potentiels (depuis 1993)

7- SUPERVISION DE PERSONNEL SCIENTIFIQUE

Thomas Beardall: étudiant diplômé ("graduate student") du programme coopératif de l'université de Waterloo (Ontario). Expression de l'insuline humaine dans *E. coli* (janvier-août 1984) et développement de vecteurs navettes *E. coli* - *Saccharomyces cerevisiae* de type phagemide (mai-août 1985)

Glavan Marie-Anne: étudiante d'été, collège Colonel By, Ottawa (Ontario). Caractérisation de l'insuline humaine recombinante (juin-juillet 1984)

Kathrada Faizel: étudiant d'été, collège Colonel By, Ottawa (ontario). Etude du système colicine N (octobre 1984-janvier 1985)

Charles Boone: étudiant diplômé, université McGill (Montréal). Etude par mutagénèse dirigée du système killer de *Saccharomyces cerevisiae* (janvier-avril 1985)

Bénédicte Soldner: étudiante invitée, BTS Strasbourg. Expression de la papaïne (mai-novembre 1985)

Katherine Britt: étudiante diplômée du programme coopératif de l'université Concordia. Expression de la papaïne dans *E. coli* (septembre-décembre 1986 et mai-août 1987)

Dave Neustaedter: étudiant d'été diplômé, université de Calgary (Alberta). Expression de la papaïne dans le système de cellule d'insecte/baculovirus (mai-août 1989)

Lee Alexander Sciortino: étudiant d'été diplômé, université Concordia (Montréal). Caractérisation de mutants de la papaïne (mai-août 1990)

Céline Bonnefoi: étudiante ingénieure IAAL de Lille. Purification et caractérisation de la protéase de maturation Kex2 (avril-septembre 1990)

Ana Paula Lima: étudiante prédoctorale de l'université fédérale de Rio de Janéiro (Brésil). Clonage et étude par mutagénèse dirigée des protéases à cystéine de *Trypanosoma cruzi* (août 1990-octobre 1993)

Dr. Dieter Brömme: post-doc, supervision de la partie biologie moléculaire du projet, université de Halle (Allemagne). Ingénierie de la cathepsin S (janvier 1990-octobre 1993)

Jean Chatellier: étudiant DEA, université de Nancy. Caractérisation des mutants en position 175 de la papaïne (février 1991-septembre 1992). Etudiant predoctoral, boursier MRT (Strasbourg) contingent international, co-supervision avec D. Altschuh. Etude

du rôle des contacts inter sous-unités dans un Fab modèle (depuis 1993)

Francesco Lipari: étudiant du programme coopératif université Concordia (Montréal). Etude de mutants de la lipase II recombinante de *Géotrichum candidum* (septembre 1991-décembre 1991)

Dr. Maria Célia Bertolini: stage post-doctoral, université de Sao Paulo (Brésil). Clonage, expression et caractérisation de la lipase I de *Géotrichum candidum* (décembre 1991-décembre 1993)

Françoise Cailler: étudiante ingénieure INSA (Toulouse). Etude de la kinase Ste 20 de levure (septembre 1992-août 1993)

Pierre Ménard: étudiant invité, université de Montréal. Purification et caractérisation de deux mutants délétés de leurs boucles d'occlusion de la lipase II recombinante de *Géotrichum candidum* (mai 1993-août 1993)

8- SEJOURS DANS D'AUTRES LABORATOIRES

Laboratoire de Génétique des Eucaryotes, Strasbourg Prof. P. Chambon, juin-septembre 1980. Clonage de l'ADNc de la petite sous-unité de la rubisco.

Institut de biologie moléculaire et cellulaire du CNRS, Strasbourg Boursier programme d'échange international CNRS/NRC:

- Séjours dans le groupe de cristallographie du Dr. D. Moras Collaborateur Dr. Danièle Altschuh.

Du 16 avril au 13 mai 1987, du 1 au 18 avril 1988, du 4 mai au 12 juin 1989 et du 7 mai au 18 mai 1990. Les mutations concertées dans les protéines:

- Séjours dans le groupe d'immunochimie du Prof. Van Regenmortel Collaborateur Dr. Danièle Altschuh.

Du 18 décembre au 3 janvier 1991, du 19 au 30 novembre 1992 et du 13 septembre au 8 octobre 1993. Etude du rôle des contacts entre domaines et entre sous-unités dans les protéines.

Institut de Biotechnologie des plantes, Saskatoon, Canada. Dr. W. Crosby. Février 1994 (1 semaine). Manipulation de téra-bibliothèques de bactériophages exposant des anticorps:

9- SEMINAIRES DONNES DANS DES LABORATOIRES EXTERIEURS

Université Laval, Québec, Canada, Département de microbiologie (28 mai 1985)

Institut Pasteur, Paris. Groupe des Antigènes Bactériens du Dr. M. Mock, (septembre 1986)

Université Louis Pasteur, Strasbourg, institut de biologie moléculaire et Cellulaire, Strasbourg, (septembre 1986)

Hôtel Dieu de Québec, Québec, Canada (8 octobre 1987)

Institut Armand Frappier, Montréal, Canada (9 mars 1989)

Université de Bordeaux, département de génétique (12 mai 1989)

Transgène, Strasbourg, laboratoire du Dr. A. Mercenier (mai 1989)

Université Louis Pasteur, Strasbourg, institut de biologie moléculaire et cellulaire (30 mai 1989)

Université McGill, Montréal, Canada, département de Biologie (4 décembre 1989)

Université de Montréal, Canada, département de biochimie (17 avril 1990)

Université de Moncton, Nouveau Brunswick, Canada, département de chimie, biochimie (30 mars 1990)

Université de Sherbrooke, Canada, département de microbiologie (11 décembre 1991)

Université de Nancy, laboratoire d'enzymologie et de génie génétique (5 octobre 1993)

10- PARTICIPATION A DES CONGRES, SYMPOSIA ET ATELIERS

UCLA Symposia, *Protein Transport and Secretion*. Keystone, USA (avril 1983)

Annual Meeting of the Genetics Society of Canada, Ottawa, Canada (juin 1983)

Summer Symposia in Molecular Biology, Pennsylvania, USA (juillet 1983)

Poster: Expression of a synthetic human proinsulin gene in *E. coli* using a colicinogenic expression vector

Third Annual Queen's University Symposium, Kingston, Canada (août 1983)

Conférence: Expression of a synthetic human proinsulin gene in *E. coli* using a colicinogenic expression vector

Symposium on Plasmids in Bacteria, Illinois, USA (mai 1984).

Poster: Human proinsulin expression using a colicin E3 derivative vector for the direct selection of recombinant DNA

Publié dans *Plasmids in bacteria*, pp 964, 30. Edited by Helinski D.R. et al., Plenum Press et dans: *Genetics* 107, pS 111 (1984)

Réunion Annuelle de la Société de Génétique du Canada. Vancouver, Canada (août 1984)

Summer Symposium in Molecular Biology, Pennsylvania, USA (août 1984)

Poster: Tetracycline resistance gene and plasmid maintenance in chemostat cultures

Fifth Annual Congress for Recombinant DNA. San Francisco, USA (février 1985)

Poster: Induction of pColE3-CA38 hic-proinsulin gene fusion causes *E. coli* cell lysis

Toronto Yeast Meeting, Toronto, Canada (mai 1985)

Protein Engineering Workshop, Ottawa, Canada (mai 1985)

Yeast Meeting. Cold Spring Harbor, USA (août 1985)

Poster: Localization of the immunity domain of the yeast K1 preprotoxin

Forum de biologie moléculaire de la levure. Bordeaux-Bombannes (octobre 1985)

Conference: Le précurseur de la toxine "killer" de la levure: définition des domaines moléculaires

55th Annual Meeting of the Genetics Society of America. Urbana-Champaign, USA (juin 1986)

Poster: The structure and function of the yeast killer toxin precursor

Fifth International Symposium on Genetics of Industrial Microorganisms. Split, Yougoslavie (septembre 1986)

Conférencier invité: Site-directed mutants define functional domains of the yeast killer toxin

INSERM Conferences. *Migration and processing in secretory processes*. Blois (septembre 1986)

Poster: Localization of the immunity domain of the yeast killer toxin

Atelier du génie des protéines. Montréal, Canada (mai 1987)

Yeast Genetics and Molecular Biology Meeting, San Francisco, USA (juin 1987)

Poster: Expression and secretion from *S. cerevisiae* of a synthetic gene coding for the cysteine protease papain

Third Annual Biotechnology Workshop. *Stress in living systems*. London, Canada (octobre 1987)

First annual meeting of the Centre for Protein Structure and Design, Ottawa, Canada (janvier 1988)

Conférencier invité: Best host/vector system for the prepropapain gene

39ieme Mosbach colloquium. *Protein structure and protein engineering* Mosbach, Allemagne. (avril 1988)

UCLA symposia on molecular and cellular biology. *Protein and pharmaceutical engineering*. Park City, USA (janvier 1989)

Poster: Studies of the mechanism and substrate specificity of papain using site directed mutagenesis

Poster: Production of precursor of papain protein and its mutants in insect cells

Troisième colloque de biotechnologie de l'université de Montréal.
Montréal, Canada (avril 1989)

Conférencier invité: Ingénierie du précurseur d'une protéase à cystéine: la papaïne

Second annual meeting of the Centre for Protein Structure and Design, Cornwall, Canada (mars 1989)

Conférencier invité: Insect cells secrete a functional papain precursor

Third annual meeting of the Centre for Protein Structure and Design, Ottawa, Canada (mars 1990)

Conférencier invité: rational alteration of a highly constrained region of papain

Protein Engineering Network Centres of Excellence workshop. Montréal, Canada (juin 1990)

Conférencier invité: Alternative packing at the two domain interface of papain

Gordon research conference. *Glycoprotein and glycolipids*. Ventura, USA (février 1991)

La levure: modèle et outil. Paris (mars 1991)

Poster: manipulation de la structure de la papaïne exprimée dans la levure

Keystone symposia on molecular and cellular biology. *Proteolysis in regulation and disease. Protein folding, structure and function*. Keystone, USA (April 1991)

Poster: Contribution of the glutamine 19 side chain to transition-state stabilization in the oxyanion hole of papain

Poster: Intramolecular processing of papain precursor

Protein Engineering Network Centres of Excellence workshop. Ottawa, Canada (novembre 1991)

Conférencier invité: screening method for cysteine protease mutants and inhibitors

International industrial biotechnology conference. *Bio-recognition*. Montréal, Canada (juin 1992)

International Network of Protein Engineering Centres, scientific meeting. Montréal, Canada (juin 1992)

Poster: The catalytic triad of papain: is the Asn175 essential for activity?

Protein Engineering Network of Centres of Excellence. Annual General Meeting. Montréal, Canada (juin 1992)

Miami winter symposia. *Protein engineering and beyond*. Miami, USA (janvier 1993)

Poster: Dual role of Asn175 in the catalytic activity and the stability of the cysteine protease papain

Publié sous forme de communication brève dans: *Protein Engineering* 6 sup., 52

36^{ieme} congrès annuel. Fédération Canadienne des Sociétés de Biologie. Windsor, Canada (juin 1993)

Conférence: Early molecular events in papain precursor processing

Protein Engineering Network Centres of Excellence Annual General Meeting Edmonton, Canada (juin 1993)

Poster: Definition of the role of a conserved motif within the pro region of papain by *in vitro* random mutagenesis

EMBO workshop. *Molecular repertoires and methods of selection*. Gubbio, Italie (octobre 1993)

Poster: définition of the role of a conserved motif within the pro region of papain by random mutagenesis

11- PROMOTION DE LA SCIENCE DANS LE GRAND PUBLIC

Interview à CBC radio, "yeast in biotechnology" (janvier 1986)

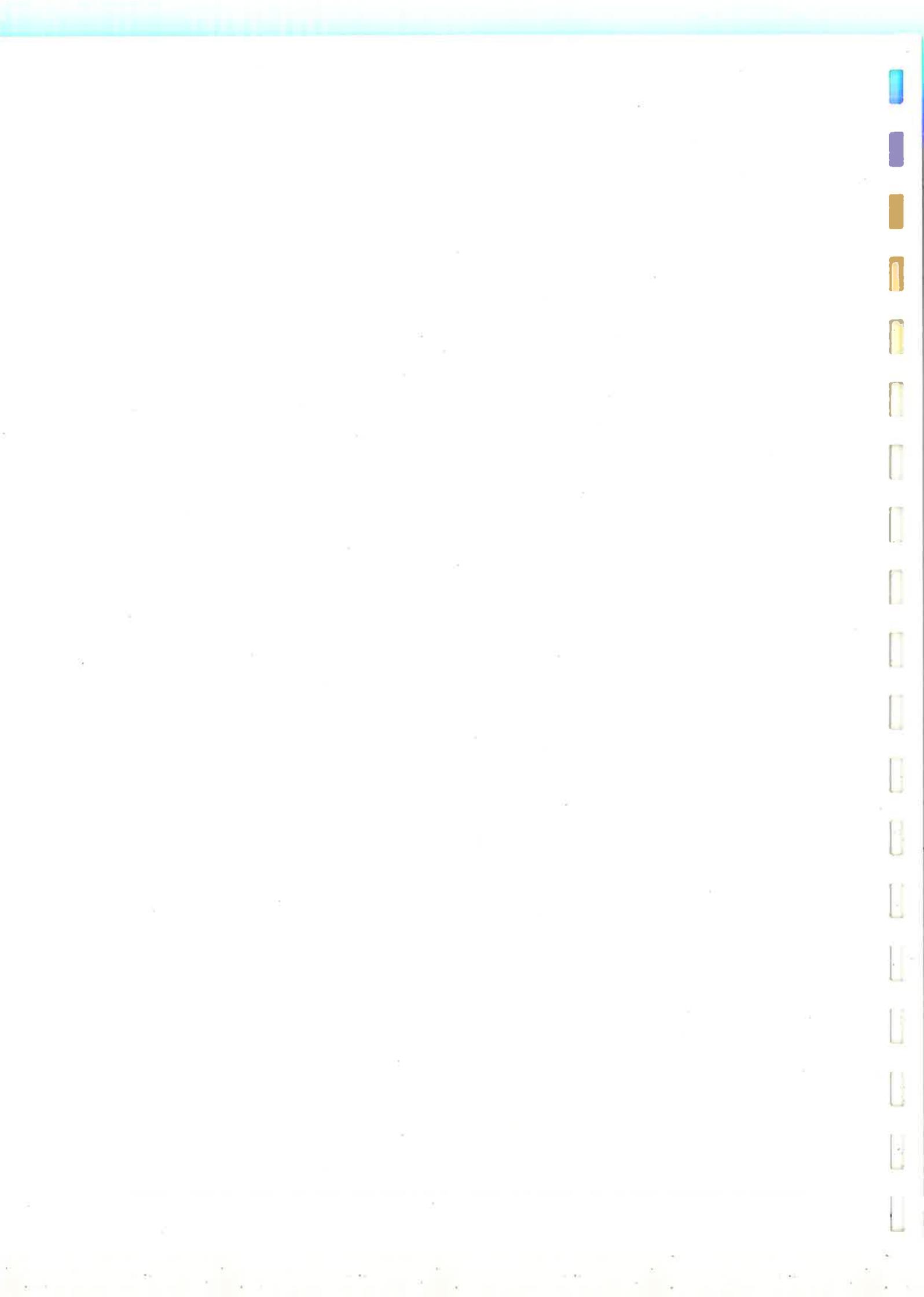
Interview à Radio Canada, "aujourd'hui la science" par Yanick Villedieu, "les levures tueuses" (29 avril 1990)

Interview du journal "the Gazette" (publiée dimanche 17 octobre 1993)

Organisation et supervision de deux projets primés et membre du jury dans le cadre des Expo Science (mars 1993)

Co-organisateur des journées portes ouvertes à l'IRB (mai 1991 et octobre 1993)

Fréquentes présentations des travaux de l'IRB pour des lycéens, collégiens et groupes de professeurs des écoles secondaires



LISTES DES PUBLICATIONS



PUBLICATIONS

- 1- Vernet, T., Fleck, J., Durr, A., Fritsch, C., Pinck, M. and Hirth, L.: Expression of the gene coding for the small subunit of ribulosebiphosphate carboxylase during differentiation of tobacco plant protoplasts. *Eur. J. biochem.* 126: 489-494 (1982).
- 2- Fleck, J., Durr, A., Vernet, T. and Hirth, L.: Osmotic-shock "stress proteins" in protoplasts of *Nicotiana sylvestris*. *Plant Science Letters* 26: 159-165 (1982).
- 3- Narang, S.A., Brousseau, R., Georges, F., Hanna, Z., Michniewicz, J., Préfontaine, G., Stawinski, J., Stepien, P., Sung, W., Thomas, D.Y., Tso, J., Vernet, T., Visentin, L.P. and Wu, R.: Total synthesis and expression of human preproinsulin gene. *Proceedings of the XV International Congress of Genetics*, 81-90 (1983).
- 4- Watson, R.J., Lau, P.C.K., Vernet, T. and Visentin, L.P.: Characterization and nucleotide sequence of a colicin-release gene in the *hic* region of plasmid Co1E3-CA38. *Gene* 29: 175-184 (1984).
- 5- Vernet, T., Lau, P.C.K., Narang, S.A. and Visentin, L.P.: A direct selection vector derived from pCo1E3-CA38 adapted for foreign gene expression. *Gene* 34: 87-93 (1985).
- 6- Watson, R.J., Vernet, T. and Visentin, L.P.: Relationships of the Col plasmids E2, E3, E4, E5, E6 and E7: restriction and colicin gene fusions. *Plasmid* 13: 205-210 (1985).
- 7- Vernet, T., McDonald, I.J., Cameron, D. and Visentin, L.P.: Stable maintenance in chemostat-grown *Escherichia coli* of pBR322 and pACYC1874 by disruption of the tetracycline resistance gene. *Bioscience reports* 5: 29-37 (1985).
- 8- Boone, C., Bussey, H., Greene, D., Thomas, D.Y. and Vernet, T.: Yeast killer toxin: site directed mutations implicate the precursor protein as the immunity component. *Cell* 46: 105-113 (1986).
- 9- Tessier, D.C., Brousseau, R. and Vernet, T.: Ligation of single-stranded oligodeoxyribonucleotides by T4 RNA Ligase. *Analytical Biochemistry* 158: 171-178 (1986).

- 10- Vernet, T., Boone, C., Greene, D., Thomas, D.Y. and Bussey, H.: Yeast killer toxin: site-directed mutations affecting toxin and immunity domains of the precursor gene. in, *Natural antimicrobial systems*. R.G. Board ed., Bath University Press, England p. 222-231 (1986).
- 11- Bussey, H., Boone, C., Dmochowska, A., Greene, D., Zhu, H., Lolle, S.J., Vernet, T., Dignard and D. Thomas, D.Y.: Secretion and Action of Yeast K1 Killer toxin: In, *Virus of Fungi and Simple eukaryotes*. eds. M.J. Liebowitz and Y. Koltin. Marcel Dekker, New York. p. 161-178. (1986).
- 12- Vernet, T., Dignard, D. and Thomas, D.Y.: A family of yeast expression vector containing the phage f1 intergenic region. *Gene* 52: 225-233 (1987).
- 13- Vernet, T., Thomas, D.Y. Boone, C. and Bussey, H.: The yeast killer-immunity system and its applications to biotechnology. *Proceedings of the Fifth international Symposium on the Genetics of industrial Microorganisms*, 275-483 (1987).
- 14- Thomas, D.Y., Vernet, T., Boone, C., Greene, D., Lolle, S. and Bussey, H.: The yeast killer toxin system: functional analysis and biotechnological applications. In, *CRC Reviews in Biotechnology*, R. Klein et G. Stewart ed., Boston, USA, Vol. I, p. 155-163 (1987).
- 15- Kang, C.Y., Vernet, T. and Thomas, D.Y.: Expression of hepatitis B virus surface antigen in mammalian, yeast, and insect cells. *Applied Virology Research*, Vol. 1, p. 31-42 (1988).
- 16- Bussey, H., Vernet, T. and Sdicu, A-M.: Mutual antagonism among killer yeasts: competition between K1 and K2 killers and a novel cDNA-based K1/K2 killer strain of *Saccharomyces cerevisiae* *Can. J. Microbiol.*, V. 34, p. 38-44 (1988).
- 17- Altschuh, D., Vernet, T., Berti, P., Moras, D. and Nagai, K.: Coordinated amino acid changes in homologous protein families. *Protein Engineering*, 2: 193-199 (1988).
- 18- Vernet, T., Tessier, D.C., Laliberté, F., Dignard, D. and Thomas, D.Y.: The expression in *Escherichia coli* of a synthetic gene coding for the precursor of papain is prevented by its own putative signal sequence. *Gene* 77: 229-236 (1989).

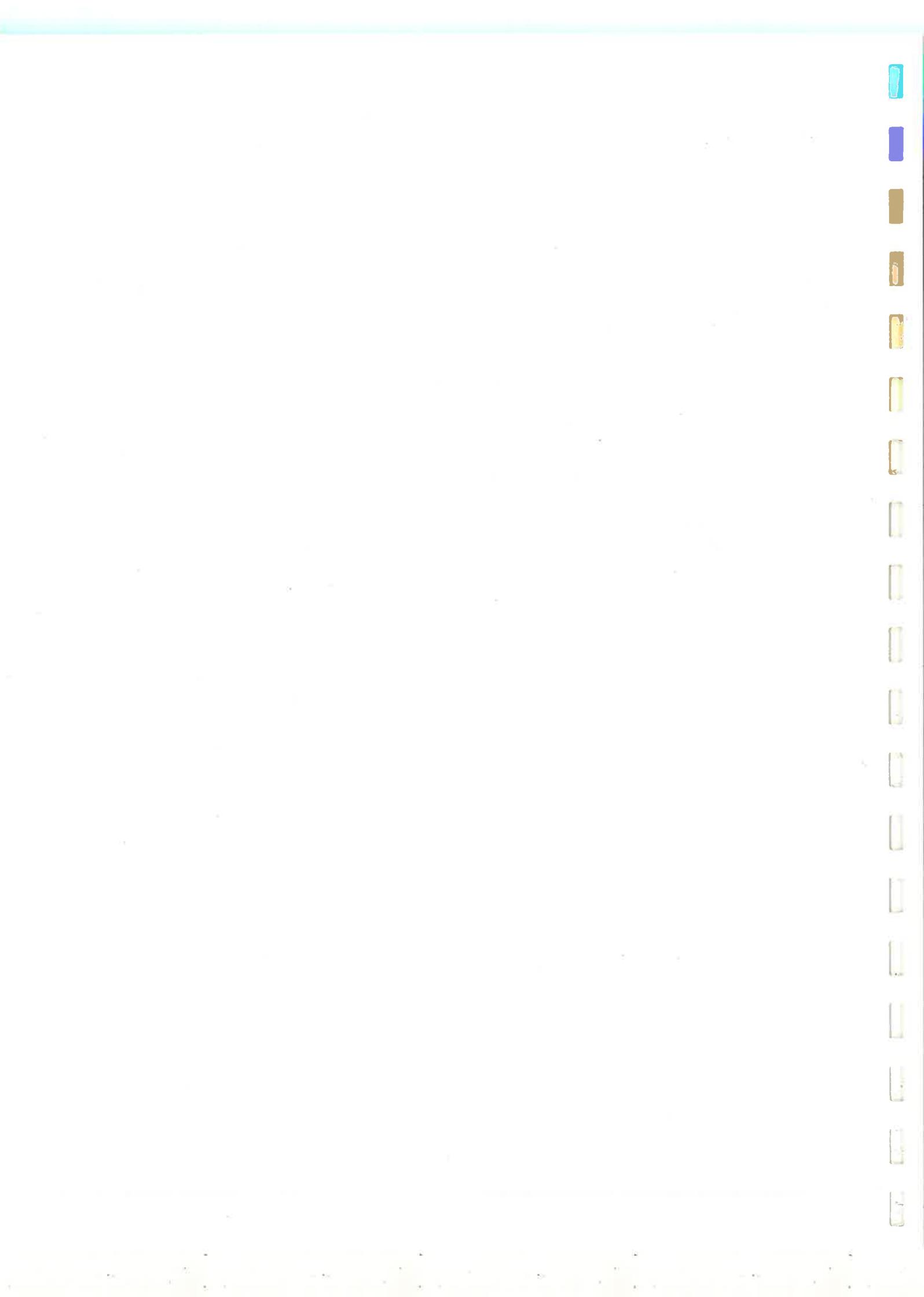
- 19- Vialard, J., Lalumière, M., Vernet, T., Briedis, D., Alkhalib, G., Henning, D., Levin, D. and Richardson, C.: Synthesis of the membrane fusion and hemagglutinin proteins of measles virus using a novel baculovirus vector containing the beta-galactosidase gene. *J. Virol.* 69: 37-50 (1990).
- 20- Bussey, H., Boone, C., Zhu, H., Vernet, T., Whiteway, M. and Thomas, D.Y.: Genetic and molecular approaches to synthesis and action of the killer toxin. *Experientia* 46: 193-200 (1990).
- 21- Ménard, R., Khouri, H.E., Plouffe, C., Dupras, R., Ripoll, D., Vernet, T., Tessier, D.C, Laliberté, F., Thomas, D.Y. and Storer, A.C. A protein engineering study of the role of aspartate 158 in papain's catalytic mechanism. (1990) *Biochemistry* 29, 6706-6713.
- 22- Vernet T., Tessier, D.C., Richardson, C., Laliberté, F., Khouri, H.E., Bell, A.W., Storer, A.C. and Thomas, D.Y. Secretion of functional papain precursor from insect cells: requirement for N-glycosylation of the pro-region. (1990) *J. Biol. Chem.* 265, 16661-16666.
- 23- Tessier, D.C., Thomas, D.Y., Laliberté, F. Khouri, H.E., and Vernet, T. Enhanced secretion from insect cells of a foreign protein fused to the honey bee melittin signal peptide. (1991) *Gene* 98 177-183.
- 24- Ménard, R., Plouffe, C., Khouri, H.E., , Dupras, D., Tessier, D.C, Vernet, T., Thomas, D.Y. and Storer, A.C. Removal of an inter-domain H-bound through site-directed mutagenesis: role of Ser176 in the mechanism of papain. (1991) *Protein Engineering* 4, 307-311.
- 25- Ménard, R., Khouri, H.E., Plouffe, C., Laflamme, P., Dupras, D., Vernet, T., Tessier, D.C, Thomas, D.Y. and Storer, A.C. Importance of hydrogen bonding interaction involving the side chain of Asp158 in the catalytic mechanism of papain. (1991) *Biochemistry* 30, 5531-5538.
- 26- Vernet, T. Khouri, H.E., Laflamme, P., Tessier, D.C, Gour-Salin, B., Storer, A.C. and Thomas, D.Y. Processing of the papain precursor: purification of the zymogen and characterization of its mechanism of processing. (1991) *J. Biol. Chem.* 266 , 21451-21457.

- 27- Khouri, H.E., Vernet, T., Ménard, R., Parlati, F., Laflamme, P., Tessier, D.C, Thomas, D.Y. and Storer, A.C. Engineering of papain: selective alteration of substrate specificity by site-directed mutagenesis. (1991) *Biochemistry* 30, 8929-8936.
- 28- Ménard, R., Carrière, J., Laflamme, P., Plouffe, C., Khouri, H.E., Vernet, T., Tessier, D.C, Thomas, D.Y. and Storer, A.C. Contribution of the glutamine 19 side chain to transition state stabilization in the oxyanion hole of papain. (1991) *Biochemistry* 30, 8924-8928.
- 29- Germain, D., Vernet, T., Boileau, G and Thomas, D.Y. Expression of the *Saccharomyces cerevisiae* Kex2 endoprotease in insect cells: evidence for a C-terminal autoprocessing event. (1992) *Eur. J. Biochem* 204, 121-126.
- 30- Vernet, T., Tessier, D., Khouri, H. and Altschuh, D. Correlation between coordinated amino acid changes at the two domain interface of cysteine proteases and protein stability. (1992) *J. Mol. Biol.* 224, 501-509.
- 31- Germain, D., Dumas, F., Vernet, T., Bourbonnais, Y., Thomas, D.Y. and Boileau, G. The pro-region of the Kex2 endoprotease of *Saccharomyces cerevisiae* is removed by self-processing (1992) *FEBS Letters* 299, 283-286.
- 32- Brömme, D., Bonneau, P.R., Lachance, P., Wiederanders, B., Kirschke, H., Peters, C., Thomas, D.Y., Storer, A. C. and Vernet, T. Functional expression of human cathepsin S in *Saccharomyces cerevisiae*: purification and characterization of the recombinant enzyme. (1993) *J. Biol. Chem.* 268, 4832-4838.
- 33- Vernet, T., Chatellier, J., Tessier, D.C. and Thomas, D.Y. Expression of functional papain precursor in *Saccharomyces cerevisiae* rapid screening of mutants. (1993) *Protein Engineering* 6, 213-219.
- 34- Vernet, T., Ziomek, E., Recktenwald, A. Scharg, J. D., Tessier, D. C., Thomas, D. Y. and Cygler, M. Cloning and Expression of *Geotrichum candidum* Lipase II in Yeast: Probing Of The Enzyme Active Site by Site-Directed Mutagenesis (1993) *J. Biol. Chem.* 268 26212-26219.

- 35- Bertolini, M. C., Laramée, L., Thomas, D. Y. Cygler, M., Schrag, J. D. and Vernet, T. Polymorphism in the lipase genes of *Geotrichum candidum* strains (1993) *Eur. J. Biochem.* (sous presse).
- 36- Whiteway, M., Dignard, D., Vernet, T. and Thomas, D.Y. Comparison of the *Saccharomyces cerevisiae* K1 and K2 cDNA's toxins. (1994) In *Proceedings of the EMBO conference on the viruses of fungi and simple eukaryotes* ed M.J. Leibowitz. University of Delaware Press (sous presse).
- 37- Vernet, T. and Brousseau, R. In *Handbook of protein and peptide design*. In vitro mutagenesis: recent advances and applications. P.R. Carey ed. Academic Press Inc. (1994) (sous presse).

ARTICLES SOUMIS POUR PUBLICATION

- 1- Altschuh, D. Tessier, D.C. and Vernet, T. Modulation of the enzymatic activity of papain by inter-domain residues remote from the active site (sous révision).
- 2- Lima, A.P., Sharfstein, J., Tessier, D.C., Thomas, D.Y. and Vernet, T. Identification of new cysteine protease gene isoforms in *T. cruzi*.
- 3- Vernet, T., Tessier, D.C., Chatellier, J. Plouffe, C., Lee, T.S., Thomas, D.Y., Storer, A.C. and Ménard, R. Structural and functional roles of asparagine 175 in papain.



SYNTHESE DE L'ACTIVITE SCIENTIFIQUE



1- Préambule

Le travail de recherche décrit dans ce document s'étend de la période de formation prédoctorale (3^{ième} cycle) jusqu'à présent et couvre une période de 14 années. J'ai effectué ma formation de chercheur à l'Institut de Biologie Moléculaire et Cellulaire (IBMC) du CNRS de Strasbourg sous la supervision du Dr. J. Fleck. Après l'obtention de mon doctorat de troisième cycle en septembre 1981 j'ai été engagé en novembre 1981 comme associé de recherche au Conseil National de Recherches du Canada à Ottawa (Canada) dans le laboratoire du Dr. L. Visentin. Ce poste, équivalent à un post-doc, m'a permis non seulement d'étudier les plasmides colicinogènes d'*E. coli* mais aussi de me familiariser avec les méthodes de travail en Amérique du Nord. J'ai eu rapidement la responsabilité d'un technicien et de plusieurs étudiants.

En juin 1984 peu après l'annonce de la création de l'Institut de Recherche en Biotechnologie de Montréal (IRB) j'ai été engagé, d'abord comme associé de recherche jusqu'en août 1985, puis comme agent de recherche adjoint (poste permanent, jusqu'en juillet 1988). J'ai été ensuite promu au rang d'agent de recherche associé (équivalent chargé de recherche) dans le laboratoire de génétique des levures du Dr. D.Y. Thomas.

Les quatre étapes de la carrière décrites ci-dessous correspondent à des réorientations de la thématique de recherche découlant d'un changement de laboratoire et, plus récemment, d'un choix délibéré. Ces thèmes sont:

"Différenciation chez les plantes supérieures" (page 26)

"Les plasmides colicinogènes" (page 27)

"Le système toxine "killer" de la levure" (page 33)

"L'ingénierie des hydrolases" (page 42)

Ces quatre étapes, présentées dans l'ordre chronologique, ne sont pas, bien entendu, isolées les unes des autres. Les concepts et méthodes acquis antérieurement sont intégrés dans les travaux ultérieurs. Par exemple le développement de systèmes d'expression dans plusieurs types de cellules et la bonne connaissance de la génétique des levures ont été essentiels aux recherches en ingénierie des protéines.

La bibliographie présentée à la fin du document est organisée de façon à faire ressortir les relations entre les différents thèmes de recherche abordés.

2- Période 1978-1981: Différenciation chez les plantes supérieures (publications 1 et 2)

2.1- Contexte professionnel

Durant cette période de préparation du doctorat de troisième cycle, j'ai travaillé sous la direction du Dr. Jacqueline Fleck à l'IBMC, Strasbourg. Les travaux du laboratoire étaient orientés vers l'étude de la différenciation cellulaire chez les plantes supérieures avec comme modèle d'étude la perte de la fonction photosynthétique chez les protoplastes, isolés des feuilles de tabac et mis en culture *in vitro*.

Nous avons démontré que ce passage d'un état différencié à un état non différencié s'accompagne d'une très forte diminution de la concentration d'une enzyme clé de la photosynthèse, la ribulose 1,5 bisphosphate carboxylase (rubisco). Lors d'études préliminaires de la synthèse des protéines dans les protoplastes nous avons établi le profil des polypeptides spécifiquement induits par le stress osmotique. L'augmentation de la pression osmotique est en effet nécessaire à la conservation des cellules dépourvues de paroi cellulaire.

La rubisco qui catalyse les étapes initiales de la photosynthèse est constituée de huit grandes sous-unités L de 55 kDa de poids moléculaire et de huit petites sous-unités S (12 kDa). Le gène codant pour S est localisé dans le noyau de la cellule alors que ceux codant pour L sont d'origine chloroplastique.

2.2- Clonage d'une sonde génétique pour la petite sous-unité de la rubisco

Le gène codant pour la sous-unité S a été choisi comme marqueur de la différenciation. Mes objectifs étaient de cloner une sonde d'acide nucléique et de l'utiliser pour suivre l'expression de la sous-unité S au cours de la différenciation cellulaire. Le principal défi, outre l'inexpérience du laboratoire en biologie moléculaire, tenait aux difficultés de préparer des ARN messagers (ARNm) de plantes en raison des polyphénols contaminants les préparations. L'ARNm codant pour la protéine S a finalement été purifié par électrophorèse préparative à partir de polysomes de feuilles de tabac. Les différentes étapes de la purification ont été suivies par analyse des produits de traduction *in vitro*. L'ADN complémentaire (ADNc) de l'ARNm codant pour la sous-unité S a été cloné dans *E. coli* puis identifié par hybridation différentielle. L'apprentissage des techniques du génie génétique et le clonage de l'ADNc ont été réalisés dans le laboratoire du Prof. P. Chambon (LGME, Strasbourg) sous la supervision du Dr. F. Ganon. La purification de l'ARNm de la sous-

unité S par hybridation avec le fragment d'ADNc a permis la confirmation du clonage de l'ADNc recherché, celui-ci étant traduit *in vitro* en une protéine unique co-migrant avec la protéine S (Fig. 1).

2.3- Régulation de l'expression de la petite sous-unité

L'ADNc codant pour la sous-unité S a été utilisé comme sonde de différenciation. Ainsi nous avons confirmé l'arrêt précoce de la synthèse de l'ARNm codant pour S dans les protoplastes fraîchement placés en culture. Cet arrêt n'est pas accompagné d'une disparition immédiate de la petite sous-unité en raison de la stabilité de la molécule.

Par contre la sensibilité de la méthode a révélé la présence de l'ARNm codant pour la sous-unité S dans les cultures de cellules non photosynthétiques. Ce faible niveau d'expression, observé pour la première fois, indique une transcription de base de ce gène dans les cellules indifférenciées. Cette expression de base pourrait être en relation avec une fonction secondaire inconnue de la sous-unité S. Aucune activité correspondante à la rubisco n'a été détectée dans les cultures de cellules non différenciées probablement en raison de l'absence de la grande sous-unité chloroplastique.

Ce travail, dont une des conséquences a été l'introduction de la biologie moléculaire dans le laboratoire, a été poursuivi après mon départ par le séquençage du gène et par l'étude de la régulation de son expression.

3- Période 1981-1984: Etude des plasmides colicinogènes et applications (publications 3, à 7)

3.1- Contexte professionnel

Mon choix d'effectuer un stage post-doctoral en Amérique du Nord a été facilité par le réseau de contacts canadiens de mes collègues de l'IBMC. J'ai été mis en relation avec différents laboratoires au Québec et en Ontario. Le Dr. L. Visentin au Conseil National de Recherches du Canada m'a offert un contrat renouvelable de 2 ans pour étudier la biologie des plasmides colicinogènes d'*E. coli* avec comme objectif le développement de systèmes d'expression applicables, entre autre, à la synthèse de l'insuline humaine. Ce dernier aspect a été réalisé en collaboration avec le Dr. S. Narang dont l'équipe venait de réaliser la synthèse complète du gène codant pour le précurseur de cette l'hormone.

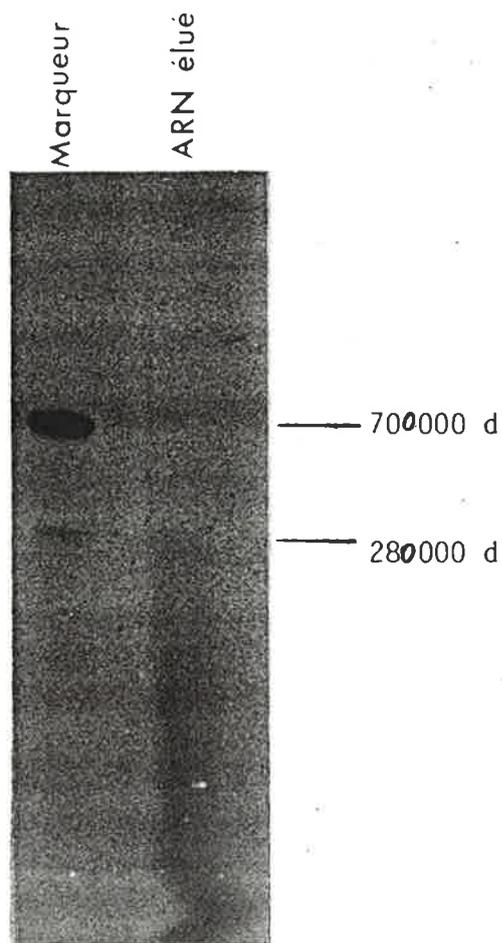


Fig. 1 a

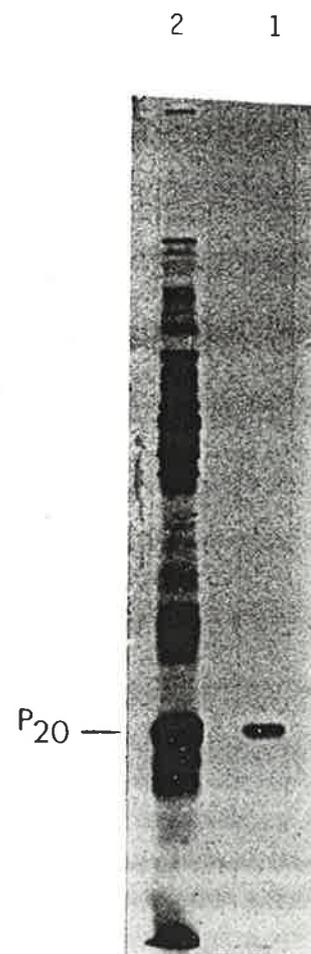


Fig. 1 b

Figure 1 : Identification du plasmide n° 34

Le plasmide n° 34, fixé sur papier DBM, a été utilisé pour purifier spécifiquement par hybridation homologue, l'ARN messager polysomique qui lui correspond.

(a) Analyse de l'ARN purifié par électrophorèse sur gel de polyacrylamide/agarose en conditions dénaturantes

Marqueur de PM : ARN 17S₁ et 12S de l'AMV marqué en 5' au pcp³²P

(b) l'ARN purifié (1) et l'ARN poly A⁺ polysomique total (2) ont induit la synthèse protéique *in vitro* (réticulocyte de lapin). Les protéines traduites sont analysées par électrophorèse sur gel de polyacrylamide (12,5%) en présence de SDS. Cette figure représente l'autoradiogramme du gel.

3.2- Caractérisation des plasmides colicinogènes d'*E. coli* (publications 4, 6)

De nombreux plasmides naturels présents chez *E. coli* codent pour des toxines, les colicines (Col), aux modes d'action variés. Ces toxines sécrétées détruisent très efficacement les souches d'*E. coli* ne possédant pas un plasmide identique sans affecter les souches productrices. Chaque plasmide code donc non seulement pour une toxine spécifique mais aussi pour une protéine d'immunité et un facteur, codé par le gène *hic*, favorisant le relachage de la toxine. L'expression du facteur d'immunité est constitutif, celle du gène *col* et *hic* étant inductible par le système SOS de la souche hôte. Au laboratoire l'induction de l'opéron est réalisée en présence de mitomycine.

Dans un premier temps, afin de me familiariser avec ce système, j'ai participé à la cartographie physique et génétique d'une série de sept plasmides de la famille Col E (Fig. 2). Ce travail, initié par R. Watson au laboratoire, nous a permis d'identifier la région C-terminale des toxines comme étant responsable de la spécificité du mode d'action des molécules.

Le facteur de relargage *hic*, dont la séquence nucléotidique a été déterminée par le Dr. P. Lau, a été étudié en détail. Nous avons démontré que le gène *hic* était sous le contrôle du promoteur de la toxine. En fait, l'induction de ce facteur provoque aussi la lyse de la cellule. L'avantage évolutif d'un système causant la mort de la cellule productrice peut être interprété dans le cadre de la dynamique d'une population de cellules. Quelques membres de cette population produisent de larges quantités de toxine, ineffectives sur les clones voisins, mais assurant la destruction de souches d'*E. coli* dépourvues du plasmide.

3.3- Développement d'un vecteur d'expression régulé à sélection directe (publications 3, 5)

Les particularités des plasmides colicinogènes ont été exploitées pour la construction et l'utilisation d'une nouvelle classe de vecteur de clonage, permettant la sélection directe de plasmides recombinants. Le principe et les avantages de ce système sont les suivants.

Lors de la ligation *in vitro* entre un vecteur de clonage et un fragment ou une population de fragments d'ADN, une proportion non négligeable de vecteur se circularise sans intégrer de fragment. Cette

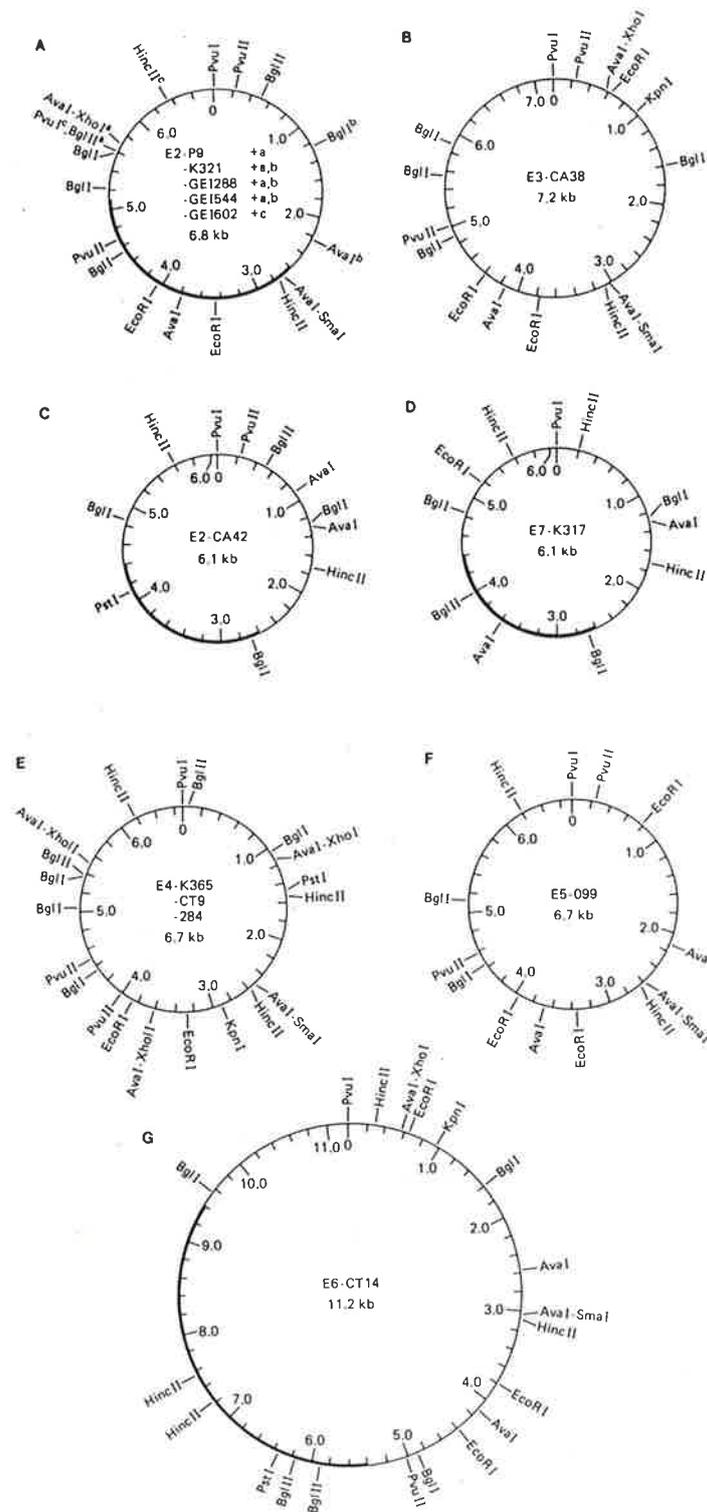


Fig. 2. Restriction maps of the ColE2-E7 plasmids. Plasmid purification and restriction mapping techniques have been described previously (4). (A) ColE2-P9 type plasmids. This is an amalgamation of the ColE2-P9 map (4) with those of ColE2-K321, ColE2-GEI288, ColE2-GEI544 and ColE2-GEI602. They differ in the sites with superscripts a, b, and c as indicated within the circle. The darker segment is the plasmid region of ColE2-P9 which is not homologous to ColE2-CA42. (B) ColE3-CA38 (4). (C) ColE2-CA42 (6). The 1.8-kb segment of this plasmid which is not homologous to ColE2-P9 is darkened. (D) ColE7-K317. The darker portion of the circle is the plasmid segment comparable to the darker segment of the ColE2-CA42 map. (E) ColE4 type plasmids. ColE4-K365, ColE4-CT9, and ColE4-284 have the same restriction map. (F) ColE5-099. (G) ColE6-CT14. The segment of this plasmid which is darkened is the 5.0-kb segment by which it differs from ColE3-CA38.

proportion varie, entre autres, en fonction de la nature des extrémités à lier, des conditions de réaction employées et des tailles respectives des fragments et du vecteur. Par conséquent il est souvent nécessaire de trier les colonies après transformation des bactéries réceptrices.

Le système mis au point évite ce triage par la sélection directe des clones contenant les vecteurs recombinés, en éliminant les clones porteurs de vecteur circularisé. Le vecteur pVT25, contient le gène codant pour la colicine E3 mais non le gène d'immunité correspondant (Fig. 3). Ce vecteur ne peut être maintenu dans *E. coli* qu'en présence d'un second plasmide fournissant, en trans, le facteur d'immunité. La toxicité a lieu même en absence d'induction du gène ColE3 probablement en raison de l'expression constitutive à un bas niveau du gène.

Lors de la ligation *in vitro*, le fragment d'ADN choisi est inséré à l'intérieur du gène ColE3 interrompant ainsi la phase de lecture du gène. En conséquence les plasmides recombinants perdent leur toxicité vis à vis des souches de laboratoire dépourvues du facteur d'immunité. Nous avons démontré que le nombre de colonies récupérées après transformation était en proportion directe avec la quantité de fragment d'ADN présent dans le milieu de ligation, une indication de l'efficacité de la sélection directe.

Un avantage supplémentaire de ce système de clonage réside dans sa capacité de produire, après induction du promoteur ColE3 par la mitomycine, une protéine de fusion avec ColE3. Nous avons exploité cet aspect du système en exprimant le gène synthétique codant pour le précurseur de l'insuline humaine (Fig. 4).

3.4- Etude de la maintenance des plasmides en absence de pression de sélection (publication 7)

Un des problèmes associés à l'utilisation de souches d'*E. coli* transformées par des plasmides artificiels provient de leur instabilité dans les populations de cellules en croissance rapide. Ce problème est relativement mineur dans un laboratoire de recherche car la pression de sélection peut être maintenue par l'utilisation d'antibiotique en excès. D'autre part, la durée des cultures y est limitée le plus souvent à quelques heures. La situation est différente en milieu industriel pour les cultures en gros volume et à longue durée où le nombre de générations cellulaires est important. Les antibiotiques utilisés pour forcer la maintenance des plasmides sont dégradés relativement rapidement. Par exemple, dans le cas de la sélection par l'ampiciline, la production de β -lactamase extra-cellulaire dégrade

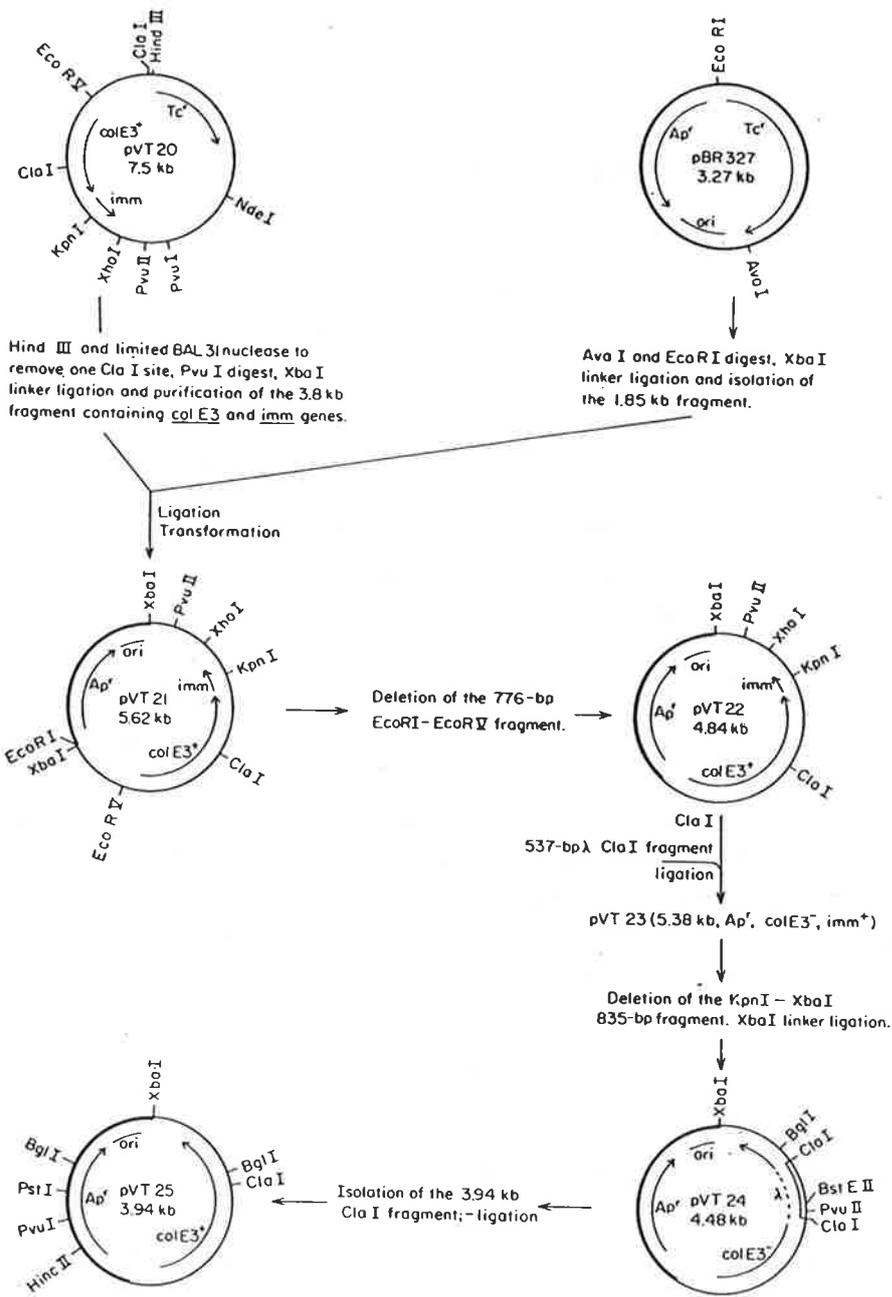


Fig. 3. Construction of the cloning vector pVT25. The portions of plasmids drawn in thick lines are from pBR327. The direction of transcription of *colE3*, *imm*, Ap^R and Tc^R genes are indicated by arrows; *ori* is the origin of replication.

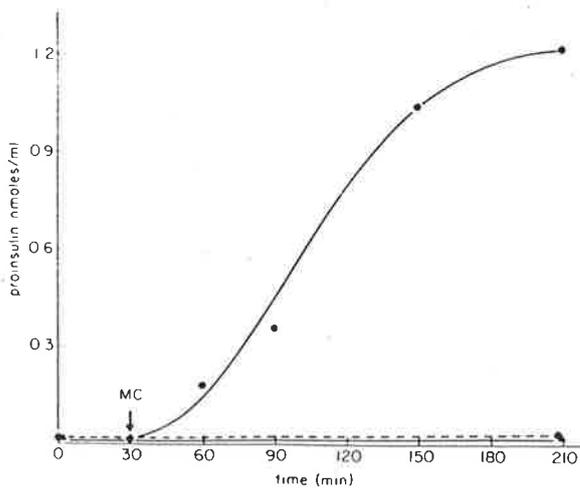


Fig. 4 Time course of proinsulin production in *E. coli* upon induction of plasmid pVT27 carrying the synthetic human proinsulin gene. Solid circles: induced cultures; Solid triangles: non-induced cultures; Solid line: WA802[pVT27]; dotted line: WA802[pVT27-1]. The downward arrow indicates the time addition of 0.5 μ g MC/ml. BW270[pVT27] and BW270 [pVT27-1] were grown in L-broth supplemented with 100 μ g Ap/ml to an $A_{620\text{nm}}$ of 0.5. Samples of 80 ml from a volume of 500 ml were removed at different times and the proteins extracted as follows: the cells were centrifuged for 5 min at 5000 rev./min, resuspended in 2 ml of 6 M urea, 10 mM Tris · HCl, pH 8.0, disrupted by sonication and centrifuged for 15 min at 15000 rev./min. Human C-peptide of proinsulin in the supernatant was quantified by RIA using a kit supplied by Novo Research Institute (Denmark). The supplier estimated that the cross-reaction of proinsulin with anti-C-peptide antibodies is 10%. Accordingly, the proinsulin concentration was corrected by a multiplication factor of 10.

l'antibiotique. L'ajout en cours de fermentation d'agents de sélection n'est pas désirable en raison de leur présence en fin de production et des coût additionnels.

En collaboration avec le Dr. I. McDonald du Laboratoire de Microbiologie de la Division des Sciences Biologiques à Ottawa, nous avons entrepris d'étudier les mécanismes moléculaires responsables de cette instabilité. Nous avons utilisé les cultures d'*E. coli* en chémostat, fermenteur asservi à un système de régulation de la croissance de cellules. Cet appareil maintient une densité cellulaire constante, fixée au préalable, par l'addition en continu d'éléments nutritifs. Les cellules excédentaires sont éliminées par un trop plein. De cette façon le taux de croissance des cellules et leurs conditions environnementales sont parfaitement déterminés.

Les expériences consistaient à mesurer la présence des plasmides dans la population bactérienne en fonction du nombre de générations, et ceci en absence totale de pression de sélection. Une série de plasmides couramment utilisée au laboratoire a été choisie pour ces études. Après environ 40 générations, moins de 5% des cellules contenaient ces plasmides (Fig. 5). Nous avons montré que la stabilité des plasmides augmentait de façon remarquable lorsqu'une région de 185 paires de base, en amont du gène codant pour la résistance à la tétracycline (*tet*), était délétée. En fait, dans la plupart des cas, la totalité des cellules contenait le plasmide délété, et ce après plus de cent générations. Ceci constitue une augmentation impressionnante de la stabilité des plasmides (Fig. 6).

Cette région du gène *tet* contient deux promoteurs en opposition, l'un d'entre eux étant utilisé pour la transcription du gène *tet*. Nous avons émis l'hypothèse que l'activité du second promoteur perturbait la répllication des plasmides par un phénomène de lecture abérante ("read-through). L'effet observé semble plus important que celui causé par le gène *rop*, facteur assurant la distribution des plasmides dans les cellules filles. Ce travail a établi qu'il était relativement facile de modifier les plasmides de laboratoire pour assurer leur maintenance en culture en continu.

4- Période 1984-1987: Etude du système "killer" chez *Saccharomyces cerevisiae* (publications 8, 10 à 14, 16, 20, 36)

4.1- Contexte professionnel

A partir de août 1984 j'ai été engagé comme associé de recherche dans le nouvel institut de recherche en biotechnologie de Montréal dans le laboratoire de génétique des levures du Dr. David Y. Thomas. Durant la construction du bâtiment et la préparation des

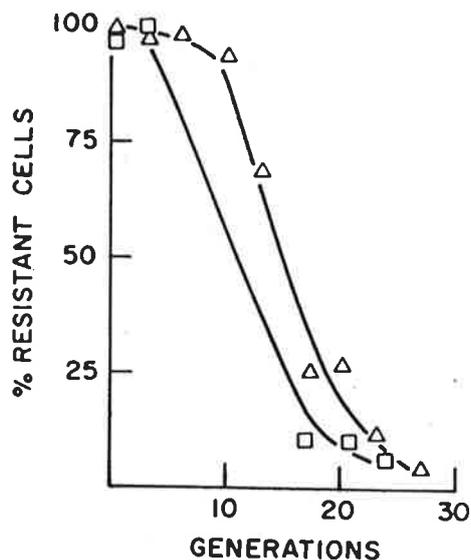


Fig.5 Maintenance of plasmid-determined antibiotic resistance in phosphate-limited chemostat cultures of *E. coli* BW270. Δ , pBR322; \square , pACYC184.

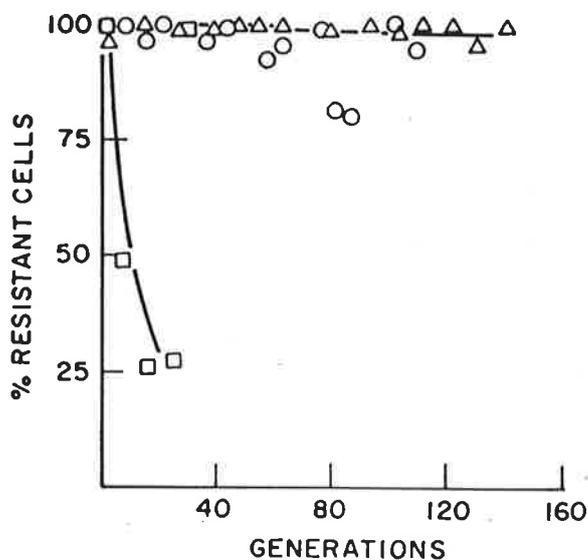


Fig.6 Maintenance of plasmid-determined antibiotic resistance in phosphate-limited chemostat cultures of *E. coli* BW270. Effect of insertion or deletion around the Tc^r gene promoter of pACYC184. Δ , pVT32; \circ , pACYC184-d1; \square , pACYC184-d2.

locaux à Montréal, j'ai continué à travailler dans les laboratoires de la division des sciences biologiques à Ottawa jusqu'en avril 1985, puis dans des locaux temporaires à l'hôpital Royal Victoria à Montréal. Le déménagement dans les locaux définitifs de l'avenue Royalmount a eu lieu en mars 1987. Ces multiples déménagements n'ont, en fait, que peu perturbé le déroulement des travaux de l'équipe.

4.2- Construction d'une série de vecteurs d'expression pour la levure (publication 12)

Vers 1984, les vecteurs disponibles pour le clonage génétique chez la levure étaient nettement moins avancés que ceux disponibles pour le clonage chez *E. coli*. Par conséquent, en parallèle avec les premières expériences sur le système "killer" nous avons développé une série de huit vecteurs de clonages adaptés pour la levure *Saccharomyces cerevisiae*. Ces vecteurs ont les caractéristiques et propriétés suivantes (Fig. 7).

1) multi-copies de type épisomal, basé sur le plasmide naturel de la levure 2 μ

2) de type navette, amplifiable dans *E. coli* (origine de replication ColE1)

3) expression des gènes d'intérêts sous le contrôle du promoteur fort constitutif ADHI

4) disponible avec le marqueur de sélection Ura3 ou Leu2d (ce dernier promoteur est défectif et assure un accroissement du nombre de copies du vecteur dans la levure par sélection forcée)

5) choix de 2 orientations du site de clonage multiple (7 sites uniques)

6) possibilité de synthèse d'ADN simple brin (ADNss) dans *E. coli* grâce à la présence de l'origine de replication du phage f1 (vecteur de type phagemide). Disponible avec deux orientations de l'origine du phage f1 permettant le choix du brin pour mutagenèse dirigée *in vitro* et séquençage.

7) grande flexibilité d'utilisation en raison de la disponibilité des différentes combinaisons de marqueurs de sélection, d'orientation du site de clonage multiple et de l'origine du phage f1.

Ces vecteurs ont grandement facilité nos travaux depuis leur construction, plus particulièrement pour l'expression et l'ingénierie des protéases à cystéine et des lipases (voir 4). De surcroît, jusqu'à présent, ces vecteurs ont été distribués dans environ 200 équipes de recherche en milieu académique et industriel dans le monde. La publication décrivant leur construction et leurs propriétés

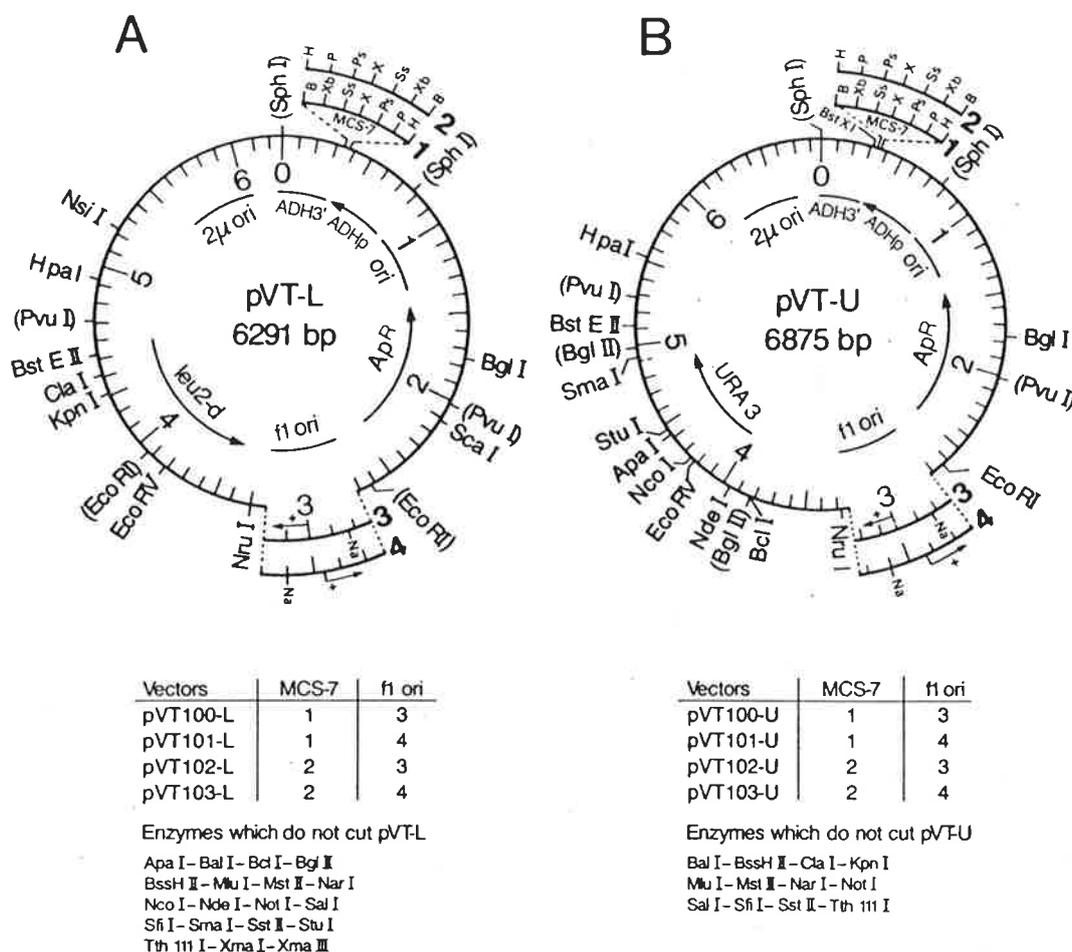


Fig. 7 Physical and genetic maps of pVT-L and pVT-U vectors. (A) Physical and genetic map of pVT-L vectors; (B) physical and genetic map of pVT-U vectors. The *E. coli* ori, the yeast origin (2μ ori) and the ori and intergenic region of the phage $\phi 1$ ($f1$ ori) are represented. The position and direction of transcription of the *leu2-d* gene, the *URA3* gene, the β -lactamase gene (Ap^R) and of the *ADH1* gene promoter (*ADHP*) and the position of the *ADH1* 3' sequence (*ADH3'*) are shown. Unique restriction sites within the plasmids are shown except for some of those present twice (in parentheses). Both orientations of MCS-7 and $f1$ ori are numbered and identify the plasmids in the tables below the maps. Restriction enzymes which do not cut the vectors are listed at the bottom of the figure. The sequence GCATGC of the *Sph*I site upstream from the *ADH3'* region has been chosen as the origin of the vector sequences (G is nucleotide No. 1). *Ps*, *Pst*I; *Ss*, *Sst*I; *Xb*, *Xba*I

a accumulé depuis sa parution environ 100 citations dans la littérature scientifique (source ISI Inc.).

4.3- Etude du système toxine "killer" de la levure

4.3.1- Introduction au système "killer"

Les bactéries ne sont pas les seuls microorganismes capables de sécréter des toxines dans leur environnement. En effet de nombreuses souches de la levure *Saccharomyces cerevisiae* sécrètent des toxines protéiques létales pour les cellules cibles. Les souches productrices sont insensibles à l'action de leur propre toxine. Parmi ces différentes toxines, la mieux connue est la toxine "killer" type 1 ou K1. Cette toxine et le facteur d'immunité correspondant sont codés par des éléments extrachromosomiques, des particules pseudo-virales.

La toxine active est composée de deux sous-unité, α et β , reliées par des ponts disulfures. Le mécanisme d'action de la toxine n'est pas encore complètement élucidé. Il est néanmoins établi que l'interaction de la toxine avec la membrane cellulaire provoque une fuite massive, et létale, des ions de la cellule par un mécanisme nécessitant de l'énergie. En fait la toxine est synthétisée sous forme d'un précurseur de haut poids moléculaire clivé lors de son passage dans la voie de sécrétion de la cellule (Fig. 8). Ce précurseur contient un peptide signal, région N-terminal éliminé lors de l'entrée dans la voie de sécrétion. Ce peptide signal est suivi de la région β puis d'une région intermédiaire γ et enfin de la région β en position C-terminal.

Lors de mon arrivée dans le groupe de génétique des levures, le clonage de l'ADN complémentaire (ADNc) complet codant pour le précurseur de la toxine venait d'être réalisé par D. Y. Thomas en collaboration avec le Prof. H. Bussey du département de biologie de l'université McGill. Cet ADNc placé dans un vecteur d'expression conférait le phénotype K1 aux souches de laboratoire transformées. Cette capacité de manipuler directement la structure du précurseur autorisait donc, entre autres, l'étude des mécanismes moléculaires de l'immunité de la toxine K1. La seule donnée certaine disponible à ce moment était que le facteur d'immunité était contenu dans le précurseur. Aucun résultat génétique ou biochimique n'avait encore permis de l'identifier. Notre objectif a donc été de définir par les méthodes de la biologie moléculaire, les rôles des différentes régions du précurseur de la toxine avec l'espoir d'identifier le ou les déterminants de l'immunité.

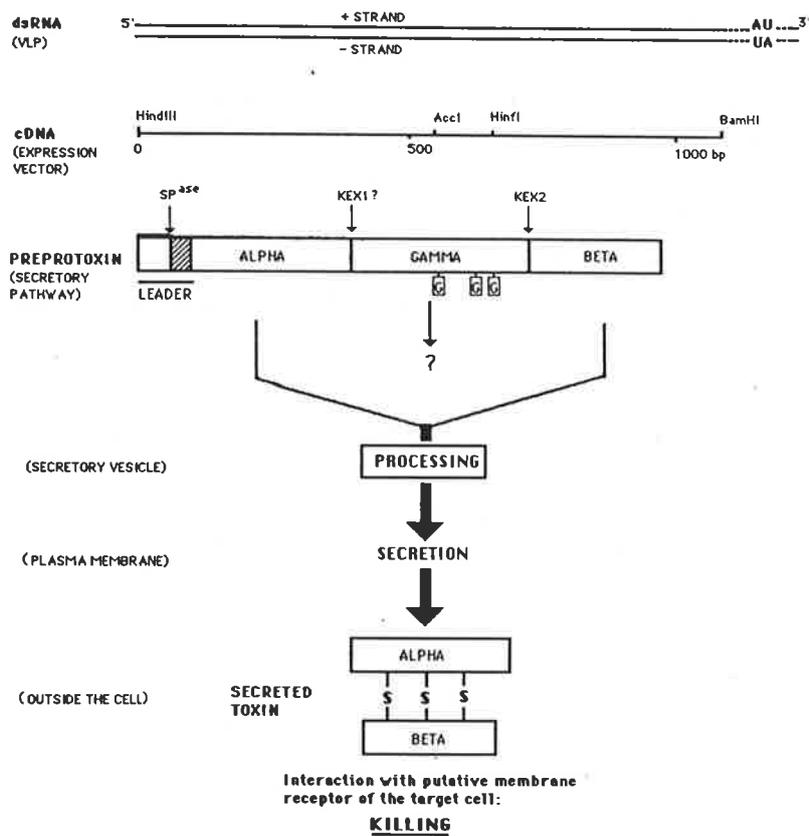


Fig. 8 : Schematic outline of the yeast K1 killer system. Cellular location of the molecules are given in parentheses on the left side of the figure. The A+U rich region of the dsRNA has been proposed to act as a transcription terminator. Only those restriction sites of the cDNA mentioned in this paper are represented. The processing protease cleavage sites are shown above the preprotoxin diagram. The exact location of the KEX1 cleavage site has not yet been defined. The function and fate of the hatched boxed peptide in the leader region and of the mature gamma peptide are not known. The three glycosylation sites of the gamma region are indicated by boxed G.

Ce travail a été réalisé en collaboration avec le Prof H. Bussey dont j'ai dirigé un des étudiants, Charles Boone, durant une partie de ces travaux.

4.3.2- *Elucidation du facteur d'immunité de la toxine K1 (publications 8, 10, 11, 13, 14, 20, 32)*

Certains précurseurs de protéine, notamment les précurseurs de neuropeptides, libèrent des polypeptides à activité biologique différente. Par analogie, la région γ du précurseur de la toxine K1 avait été proposée comme peptide responsable de l'immunité. En fait des études génétiques ont mis en doute cette hypothèse, les levures mutantes *kex2* étant immunes même en absence de maturation du précurseur.

Notre approche expérimentale a consisté à perturber, par mutagénèse dirigée, la structure des différentes régions du précurseur et à tester les phénotypes "killer" et immunité. Des analyses de la toxine sécrétée sur gel d'électrophorèse ont complété ce travail. La description des mutants est présentée par le Tableau I. Les résultats de mutagénèse par insertion d'acides aminés sont compilés dans la Fig. 9A. Ceux de la mutagénèse par délétion et substitution d'acides aminés sont reproduit dans la Fig. 9B.

Les résultats majeurs peuvent être brièvement résumés comme suit:

1- l'immunité est compromise par l'insertion d'acides aminés dans la partie C-terminale hydrophobe de la sous-unité α et dans la partie N-terminale de la région γ .

2- les délétions du précurseur confirment la localisation de la zone du précurseur responsable de l'immunité.

L'analyse des protéines mutantes marquées *in vivo* puis immunoprécipitées a montré que certaines mutations perturbaient la sécrétion de la toxine sans affecter l'entrée du précurseur dans la voie de sécrétion.

Nos résultats sont donc compatibles avec l'hypothèse selon laquelle le précurseur non mûré confère l'immunité. Cette observation a été la première démontrant une fonction spécifique d'un précurseur protéique en relation avec l'activité biologique de la protéine mûre correspondante.

Le modèle que nous avons proposé pour expliquer le mécanisme de l'immunité implique qu'une certaine proportion du précurseur n'est pas mûré en toxine active lors du transfert dans la voie de sécrétion. Cette fraction résiduelle du précurseur interagirait avec les récepteurs membranaires de la toxine lors de leur transport commun dans la voie de sécrétion. En conséquence, la saturation de

Table 1. Mutations in the Toxin-Immunity Precursor Gene

Plasmid	Restriction Site	Protein Modification
PADH 8A		Delete 14 N-terminal amino acids of preprotoxin leader. Insert Met-Gly-Asp-Ser-Ser prior to Leu 15 (Lohle et al., 1984)
PL7	Small at BglII	Insert Asn-Pro-Gly-Leu-Ile following Ile 52
PL8	Small at NcoI	Insert Pro-Pro-Gly-Gly-Met following Met 255
PL11	XhoI at StuI	Insert Pro-Arg-Gly-Ala following Ala 245
PL13	XhoI at Scal	Insert Pro-Leu-Glu-Gly following Ser 287
PL301-1	XhoI at HinfI	Insert Gly-Leu-Glu-Gly-Ile following Ile 212
PL312	EcoRI at AccI	Insert Gly-Ile-Pro-Ile following Ser 196
PL313	XhoI at AccI	Insert Glu-Asp-Thr-Pro-Ser-ochre following Leu 178
PL316	XhoI after deletion	Delete 34 amino acids following Ile 177 (to Ile 212); Insert Leu-Glu-Gly
PL317	5' deletion to BglII and addition of BamHI linker	Delete 51 N-terminal amino acids including the preprotoxin leader and 7 amino acids of α ; replace with Met-Gly-Asp-Pro-Gly
PL318	XhoI insertion	Insert Leu-Glu following Val 85
PL319	XhoI insertion	Insert Leu-Glu following Val 116
PL335		Change Val 85 to Thr 85
PL336		Change Asp 101 to Arg 101
PL338	XhoI insertion	Insert Leu-Glu following Asp 65
PL340		Change Val 116 to Thr 116
PL351	BamHI insertion	Stop codon following Gly 132
PL352	XhoI insertion	Insert Leu-Glu following Ala 147
NLP-1		Probable nonsense mutation approximately halfway into B-coding region (Busssey et al., 1982)

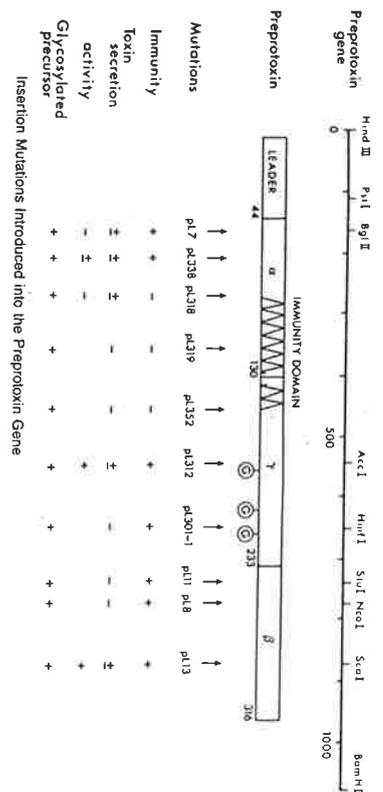


Fig. 9A The restriction map below shows the preprotoxin gene as it is cloned into the HindIII-BamHI sites of pYT760 ADH1; only sites used in this work are shown. Each interval below the line represents 100 bp. The joined open boxes depict the four peptides of the preprotoxin; the amino acid residue numbers are marked below, and the three circled Gs mark the locations of the glycosylated asparagine residues. The position of the immunity domain as it was identified in the present work is indicated by the cross-hatching of the open boxes in the preprotoxin diagram. Insertion 3 (for quantitation) is summarized by: (+) equivalent to that of the wild type, or (-), less than that of the wild type. pL308. The level of toxin secretion was determined following SDS-polyacrylamide gel electrophoresis of secreted protein (see Figure 4); (+), equivalent to pL308; (\pm), reduced but detectable toxin on gel; (-), no detectable toxin on gel. The biological activity of the toxin was measured by plate tests (see Experimental Procedures); since the amount of secreted toxin was known, the specific activity was estimated; (+), equivalent to pL308; (\pm), less than pL308; (-), no activity detectable. The 42 kd glycosylated precursor was identified following immunoprecipitation with antitoxin; (+) indicates the presence of the precursor.

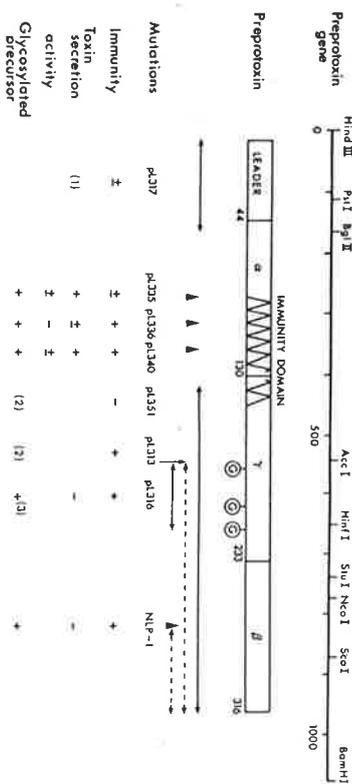


Fig. 9B Deletion and Single Amino Acid Substitution Mutations Introduced into the Preprotoxin Gene
 The preprotoxin gene and protein diagrams are as described for Figure 9A. The mutations are indicated here by the designations of plasmids containing them. The continuous horizontal arrows indicate the extent of the gene deletions; the dashed horizontal arrows the extent of the precursor protein deletions. The positions of the single amino acid substitution mutations are indicated by the solid triangles. The insertions leading to frameshift mutations are indicated by upward-pointing arrows. Other symbols: (1), toxin secretion reduced 90% (S. J. Lohle, unpublished data); (2), predicted precursors contain no methionine residues, and attempts to detect mutant precursors by immunoprecipitation following ³⁵S-labeling of cysteine residues have been unsuccessful; (3), precursor size consistent with glycosylation at a single asparagine residue (see Figure 4).

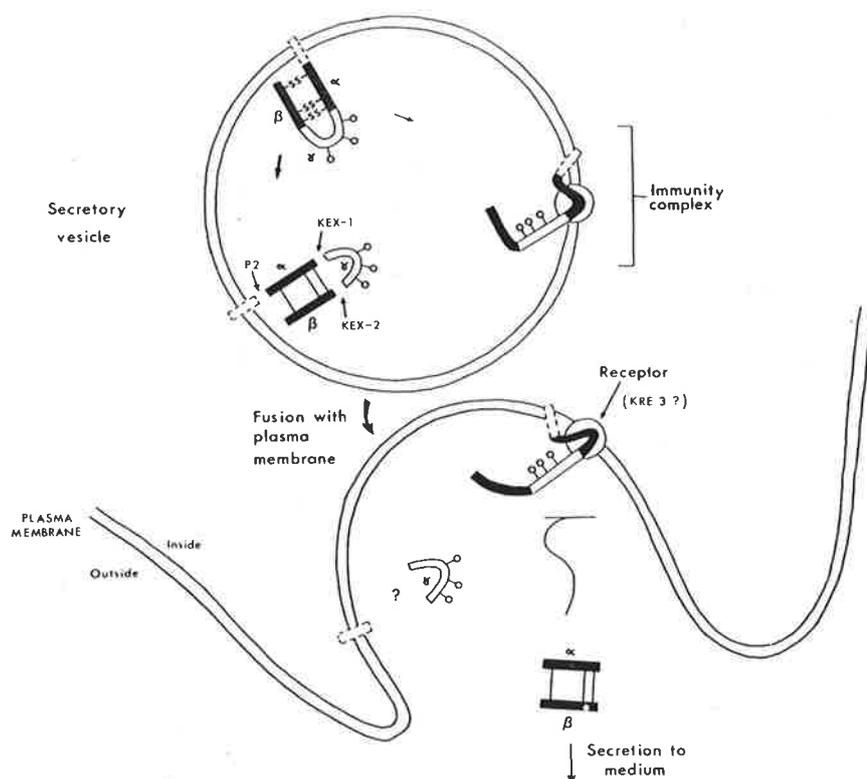


Fig. 10 Schematic Model for Immunity Action

The picture is modified from Tipper and Bostian (1984) to show that the receptor-binding domain of the immunity component includes the proposed toxin receptor-binding domain of α . Above, the toxin-immunity precursor is shown in the final stages of processing in a yeast secretory vesicle; below, secretion of mature toxin is shown after the vesicle has fused with the plasma membrane. In the secretory vesicle, most precursor is proteolytically cleaved by the KEX proteins and other gene products to generate mature toxin, indicated by the large arrow, above. A small fraction of the precursor, indicated following the smaller arrow, interacts via the immunity domain with a receptor, shown here as an integral membrane protein; the two form a stable immunity complex. This immunity complex enters the plasma membrane when the secretory vesicle fuses with it, and mature toxin is secreted. The fates of the leader (dotted line) and the γ region are unknown. In sensitive yeast cells the receptor is unoccupied; toxin can bind to it via the site on the α subunit and can subsequently form the lethal channel. In an immune cell, as shown here, the receptor is occupied by the precursor, and mature toxin is prevented from binding (indicated by the wavy line with a terminal bar).

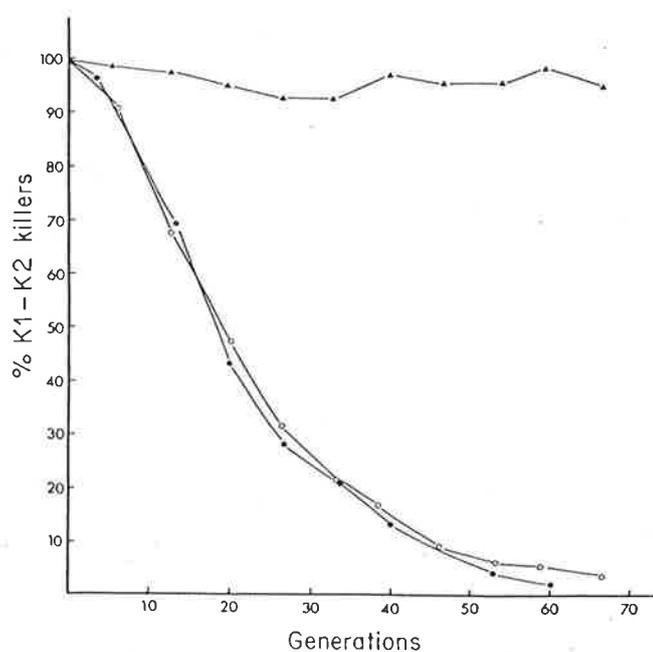


Fig. 11. K1 killer plasmid stability in a transformed K2 killer strain grown in grape juices. Strain M471 transformed with the K1 plasmid pL342 (AMS301) was serially subcultured in sterile grape juice, and K1-K2 killer measured as described in Methods. ▲, White grape juice, pH 3.4, at 30°C; ○, red grape juice, pH 2.8, at 30°C; ●, red grape juice, pH 2.8, at 18°C.

ces récepteurs au niveau de la membrane plasmique interdirait la liaison létale avec la toxine mature chez les cellules productrices (Fig. 10).

4.3.3- Application du système "killer" (publication 16)

La transformation de souches industrielles par des plasmides est une opération techniquement difficile du fait de l'absence de marqueurs d'auxotrophie. Le gène codant pour le précurseur de la toxine "killer" est un excellent marqueur de sélection des souches transformées. De plus la maintenance du plasmide dans la population de cellules est assurée par la présence de la toxine dans le milieu de culture. Les cellules filles qui n'héritent pas d'au moins une copie du plasmide sont éliminées car elle n'expriment pas le phénotype immunité.

En collaboration avec H. Bussey, nous avons exploité ces propriétés dans la construction d'une souche capable d'exprimer simultanément les toxines K1 et K2. Cette souche, obtenue par transformation d'une souche naturelle K2 par l'ADN complémentaire codant pour K1, exprime les deux phénotypes. Une telle souche n'existe pas dans la nature en raison de phénomènes d'exclusion génétique des génomes "killer" lors de la replication de l'ARN double-brin des particules pseudo-virales. Nous avons démontré que la stabilité du phénotype K1 dépendait des conditions de température et de pH du milieu de culture (Fig. 11). Ce travail, financé par une compagnie de production de levure, devrait pouvoir être exploité en milieu industriel pour éviter ou réduire la contamination par des levures "killer" indésirables lors des fermentations en grand volume.

5- Période 1987-1993: Ingénierie des hydrolases, étude des relations structure/fonction des protéines (publications 17 à 19, 21 à 28, 30, 32 à 35)

5.1- Contexte professionnel

Après la phase d'installation et de développement initial de l'IRB, la direction de l'institut a progressivement mis l'emphase sur le soutien de projets multidisciplinaires. Cette approche a l'avantage de permettre d'entreprendre des projets plus ambitieux, de les aborder par plusieurs angles méthodologiques complémentaires et de les faire progresser plus rapidement. Le défi consiste à coordonner efficacement les efforts des différentes équipes impliquées.

Dans ce contexte, j'ai été contacté par le Dr. Andrew Storer, enzymologiste, pour prendre en charge la partie biologie moléculaire d'un projet d'étude des protéases à cystéine. A travers cette collaboration j'ai rapidement développé un intérêt marqué pour

l'étude des relations structure/fonction des protéines, études auxquelles j'ai choisi depuis de consacrer l'essentiel de mes activités scientifiques. Cet intérêt s'est manifesté par la suite par le développement de projets plus personnels reliés à cette thématique et par 2 autres collaborations avec les Drs. M. Cygler (IRB) et D. Altschuh (IBMC). Bien que détaché de la thématique du groupe de génétique des eucaryotes auquel j'appartiens toujours, j'ai souvent bénéficié du soutien de mes collègues du groupe. Ce soutien s'est avéré critique lorsqu'il a été nécessaire de choisir ou de construire des souches de levure adaptées à des problèmes d'expression particuliers.

Une partie du travail de recherche sur les protéases à cystéine est intégrée depuis 1990 dans le réseau des centres d'excellence en génie protéique. Ce réseau transcanadien est axé sur le développement d'inhibiteurs spécifiques de protéases impliquées dans des pathologies, développement s'appuyant sur l'exploitation des résultats des recherches fondamentales. Le réseau regroupe plusieurs universités, compagnies pharmaceutiques et organismes gouvernementaux de recherche.

5.2- Ingénierie de la papaïne

5.2.1- Mise au point de systèmes d'expression (publications 18, 19, 23, 22, 33, 34)

La papaïne a été sélectionnée pour nos travaux. Cette enzyme abondante chez le *Carica papaya* (arbre à papaye), de 24 kDa (212 acides aminés), avait été abondamment étudiée par les enzymologistes. La structure cristallographique à haute résolution a confirmé la présence de deux acides aminés essentiels du site actif, la cystéine en position 25 (Cys25) et l'histidine159 (His159), placées de part et d'autre d'une dépression située à la surface de la protéine. Cette protéine monomérique est composée de deux domaines en étroite interaction (Fig. 12). Etrangement, au milieu des années 80, la classe des protéases à cystéines était la seule classe de protéase pour laquelle la littérature scientifique ne contenait aucune référence d'étude de ces enzymes par les méthodes de l'ingénierie des protéines. Cette lacune était en fait due aux difficultés rencontrées par plusieurs équipes lors de tentative d'expression de ces protéases dans des systèmes hétérologues. Par conséquent, le rôle de plusieurs acides aminés à proximité des deux résidus du site actif restait à déterminer. Existe-il une cavité oxianionique intervenant dans la stabilisation de l'intermédiaire tétrahédrique lors de la catalyse, analogue à celle présente chez les protéases à serine ? Quels sont les déterminants moléculaires régissant la préférence pour certaines

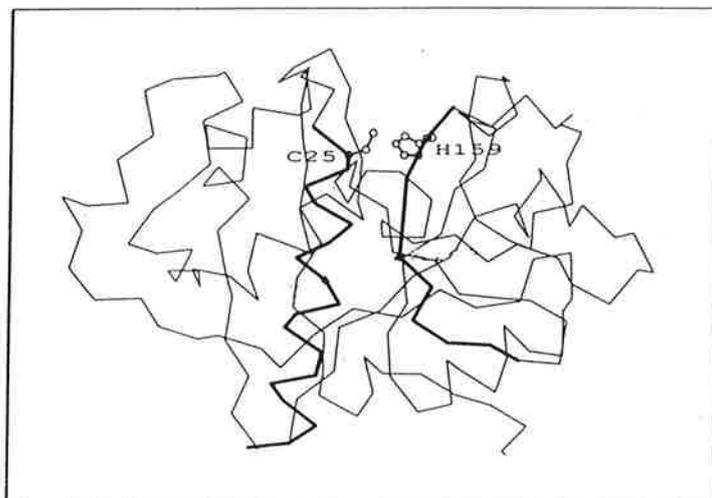


Fig. 12: structure tridimensionnelle de la papaïne mature. La chaîne des carbones alpha est représentée. La cystéine en position 25 et l'histidine en position 159 sont les deux acides aminés du site actif, localisés chacun dans l'un des deux domaines structuraux de la molécule. La longue hélice alpha porteuse de la Cys25 et une partie d'un feuillet bêta porteur de l'His159 sont représentés en trait gras.

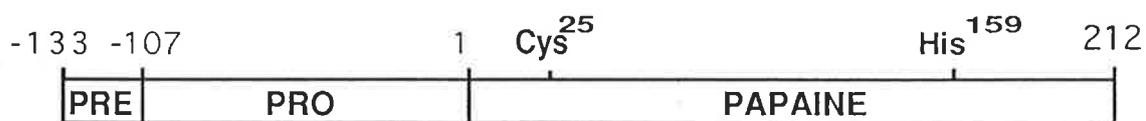


Fig. 13: structure du précurseur de la papaïne. Les acides aminés du site actif sont représentés. Par convention le résidu en position 1 du précurseur est l'Ile 1 de la papaïne mature.

chaines latérales du substrat? L'organisation en deux domaines est-elle impliquée dans la catalyse?

Ignorant à cette époque les problèmes rencontrés par d'autres, nous avons entrepris d'exprimer la papaïne. En fait la papaïne est traduite dans la plante sous forme d'un précurseur, la propapaïne, ayant l'organisation suivante: une région N-terminale de 26 acides aminés dont nous avons démontré ultérieurement le rôle comme peptide signal, une région pro de 107 acides aminés et la région de la protéine mature porteuse des résidus du site actif (Fig. 13). Nous nous sommes intéressés par la suite aux propriétés du précurseur et plus particulièrement au rôle de la région pro (voir 4.2.2).

Le gène codant pour le précurseur de la papaïne a été synthétisé *de novo* en biaisant le choix des codons vers ceux des gènes fortement exprimés dans la levure. Des sites de restrictions judicieusement placés ont facilité les manipulations ultérieures du gène.

Dans un premier temps, nous avons tenté d'exprimer ce gène chez *E. coli*, le système d'expression le plus simple à mettre en œuvre, soit sous forme libre, soit après fusion avec la β -galactosidase. En résumé, nous avons montré que le peptide signal était responsable de la toxicité de la molécule dans *E. coli*. L'expression de la protéine dépourvue du peptide signal ou fusionnée avec la β -galactosidase conduit à l'accumulation de quantités importantes de corps d'inclusion dans la bactérie. Dans les deux cas, les molécules ne sont pas fonctionnelles car non repliées de façon native. Nous n'avons pas réussi à induire un repliement correct *in vitro* de ces protéines.

Nos tentatives d'expression du précurseur de la papaïne dans *Saccharomyces cerevisiae* se sont aussi soldées, initialement, par un échec, la molécule étant aussi toxique pour les levures. Ce n'est que vers 1990 que nous avons résolu ce problème grâce à l'utilisation d'une souche de levure mutante *pep4* dépourvue de protéases vacuolaires. En effet, nous avons compris depuis que le précurseur inactif de la papaïne s'accumule dans la vacuole des levures où il est activé par les protéases résidentes. La papaïne mature active est probablement toxique pour les cellules. Le système d'expression, dans la souche *pep4*, sous le contrôle du promoteur régulé du facteur α de la levure, est employé depuis, notamment pour nos travaux utilisant la mutagenèse aléatoire, pour l'expression d'une autre protéase à cystéine humaine, la cathepsine S et pour l'expression des lipases de *Géotrichum candidum*. (voir 4.3).

Entre temps, nos études sur la papaïne mature et sur son précurseur ont été effectuées en utilisant le système des cellules

d'insectes en culture infectées par des baculovirus recombinants. Le précurseur est sécrété par les cellules d'insecte mais nous avons mis en évidence que le clivage du peptide signal et une ou plusieurs autres étapes en amont dans la voie de sécrétion étaient limitantes. D'autre part, nous avons démontré que deux asparagines de la région pro étaient glycosylées et que cette glycosylation était requise pour la sécrétion de la molécule. Les niveaux de production obtenus par ce système sont compatibles avec des expériences d'enzymologie et de biochimie mais sont insuffisants pour assurer le succès d'études structurales par cristallographie.

Afin d'améliorer le niveau de production, nous avons eu recours à une modification du système d'expression en remplaçant le peptide signal du précurseur de la papaïne par celui d'une protéine d'insecte: la mélitine d'abeille. Cette modification, intégrée dans un vecteur d'expression possédant d'autres caractéristiques pratiques (pVT-Bac, Fig. 14), accroît de façon significative le niveau de production et de sécrétion de la molécule. Un vecteur dérivé du pVT-Bac est disponible dans le catalogue de la compagnie Stratagène® depuis fin 1993.

5.2.2- Caractérisation du précurseur de la papaïne (publications 22, 26)

En parallèle avec les travaux d'enzymologie sur les mutants de la papaïne mature, nous avons entrepris de mieux comprendre les mécanismes moléculaires mis en jeu lors de la conversion de la propapaïne en papaïne active mature. Ces recherches ont aussi révélé plusieurs des fonctions spécifiques de la région pro (voir 4.2.3). Une partie de ces travaux a été publiée et sera résumé ici. Les résultats, en cours de rédaction, décrivant le rôle d'un motif structural conservé dans les régions pro seront présentés avec plus de détails.

Dans un premier temps nous avons suivi la conversion du précurseur en papaïne par analyse des produits de la réaction sur transfert de Western et par la mesure de l'activité protéolytique relâchée par spectrofluorométrie. La vitesse maximale de conversion *in vitro* de la propapaïne inactive (38 kDa) en papaïne active (24 kDa) est obtenue dans les conditions suivantes:

- pH proche de 4
- relativement inefficace à 37°C, maximale à 60°C
- facilitée par la présence d'agents réducteurs

Lorsque ces conditions optimales sont combinées la conversion complète a lieu en 20 min. environ. En fait, cette protéolyse ménagée

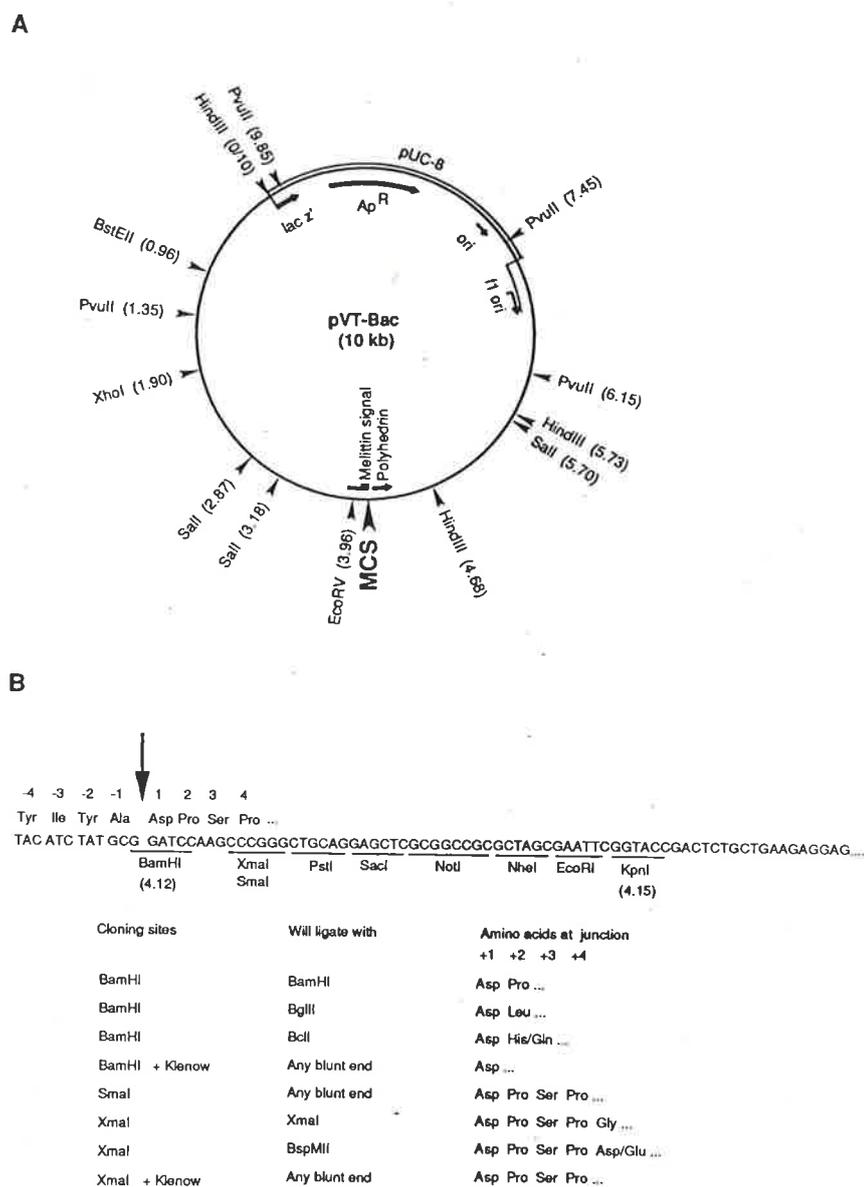


Fig. 14 Schematic structure and relevant sequences of pVT-Bac. (A) Physical and genetic map of pVT-Bac vector. The secretion vector pVT-Bac derives from IpDC176 following deletion of the propapain gene and insertion of a MCS using a 76-mer oligo. The *E. coli ori*, the *ori* and intergenic region of the phage $\phi 1$, the Ap^R gene, the α -complementation fragment of the *lacZ* gene (*lacZ'*) and the region containing pUCS (double line) are represented. Some restriction sites of the vector are indicated and their positions (kb) within the vector are given. The portion of the 3' coding region of the polyhedrin gene following the *KpnI* site is indicated by an arrow downstream from the melittin signal sequence and the MCS. Restriction sites and their position (kb), DNA sequence and deduced aa sequence in the region of the MCS is given. (B) The aa sequence context generated by fusion of the melittin signal peptide of pVT-Bac. The aa sequence at the junction between the melittin signal peptide is given for different cloning options into the MCS. The large downward arrow indicates the signal peptidase cleavage site predicted and shown to occur in the melittin-propapain fusion.

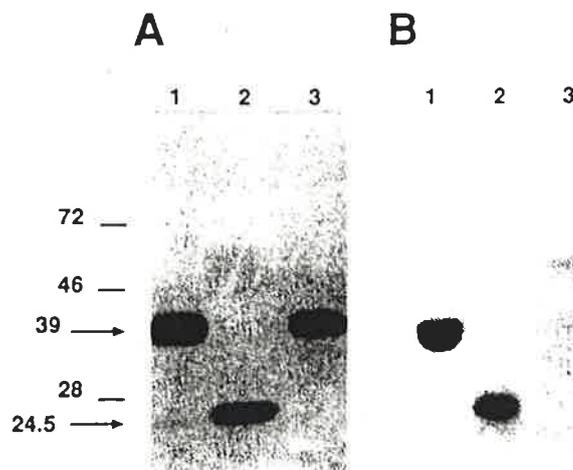


Fig. 15 Binding of ^{125}I -EpsTyrOBzl to the precursor of papain. Molecular masses of standards (kDa) are indicated in the *left margin*, and the *arrows* designate the molecular masses (kDa) of the detected proteins. *A*, Western blot of recombinant wild type papain precursor (*lane 1*), recombinant mature papain (*lane 2*), and recombinant C25S papain precursor (*lane 3*) incubated with ^{125}I -Eps-TyrOBzl. *B*, autoradiography of the Western blot from *panel A*.

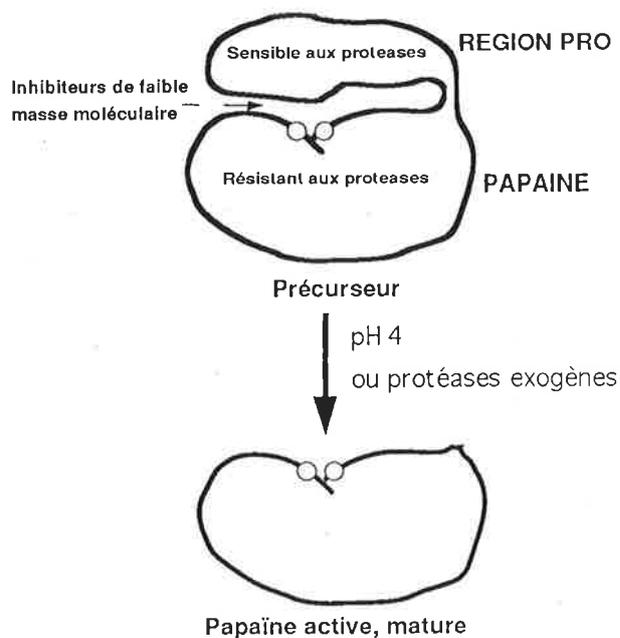


Fig. 16: modèle de maturation du précurseur de la papaine. Les acides aminés du site actif sont représentés (O).

est dépendante du site actif de la papaïne. Un mutant de la propapaïne dans lequel la cystéine en position 25 a été remplacée par une sérine (Cys25Ser) est stable dans les conditions de maturation définies précédemment. Cette observation est confirmée par l'utilisation d'une série d'inhibiteurs des protéases: seuls les inhibiteurs spécifiques des protéases à cystéine peuvent empêcher la maturation du précurseur. Un dérivé de petite taille de l'un de ces inhibiteurs spécifiques irréversibles a été utilisé pour marquer le site actif du précurseur (Fig. 15).

Des expériences de cinétique d'activation à différentes concentrations de précurseur nous ont permis de conclure que dans un premier temps les molécules de précurseur s'auto-clivent (réaction intra-moléculaire ou en cis). Par la suite, les molécules de papaïne matures libérées reconnaissent les molécules de précurseur résiduelles comme substrat (réaction inter-moléculaire ou en trans) et libèrent la papaïne dans une réaction en cascade. Cette deuxième voie de maturation n'est pas spécifique à la papaïne, d'autres protéases ayant le même effet. Autrement dit la région pro est sensible aux protéases mais non la papaïne mature.

De ces expériences nous avons tirés les conclusions suivantes:

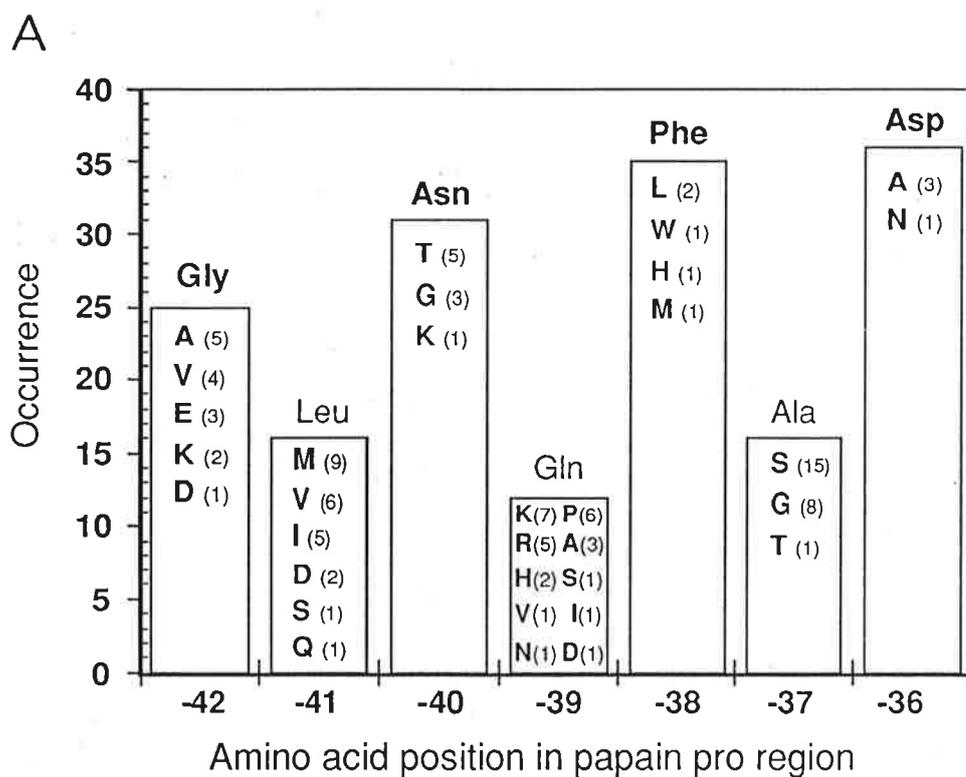
- le site actif de la papaïne est dans une conformation native dans le précurseur.
- le site actif de la propapaïne est accessible à des molécules de petites tailles.
- le précurseur est capable d'auto-maturation.

Le modèle du précurseur présenté en Fig. 16 intègre ces informations. Certains aspects de ce modèle pourraient être confirmés par la détermination de la structure tridimensionnelle de la propapaïne.

5.2.3- Définition des rôles de la région pro (publications 22, 26)

D'après les résultats précédents il est clair que la région pro est un inhibiteur de la papaïne. Lors du clivage intra-moléculaire du précurseur cet inhibiteur devient un substrat pour la protéase. Comment s'opère cette conversion? La réponse à cette question provient de travaux initiés à la suite de la comparaison des séquences de région pro de protéases homologues à la papaïne.

L'alignement de 40 séquences de précurseur de protéases à cystéine démontre clairement que la partie mature de ces molécules est homologue, les similarités de séquences étant évidentes. La situation est très différente pour les région pro ou seul un heptapeptide est reconnu comme motif partiellement conservé (Fig.



B

<i>Human cystatin C</i>	Gly	Val	Asn	Tyr	Phe	Leu	Asp
<i>Human cystatin S</i>	Gly	Val	Asn	Tyr	Phe	Phe	Asp
<i>Human cystatin SN</i>	Gly	Val	Asn	Tyr	Phe	Phe	Asp
<i>Human cystatin S5</i>	Gly	Val	Asn	Tyr	Phe	Phe	Asp
<i>Bovine colostrum</i>	Gly	Met	Asn	Tyr	Phe	Leu	Asp
Consensus	Gly	Xaa	Asn	Tyr	Phe	Xaa	Asp

Fig. 17: Presence of a common conserved motif within cysteine protease pro region and mammalian cystatin C sequences. A, Occurrence of amino acids in cysteine proteases between positions -36 and -42 (papain pro region numbering) from a set of 40 sequences of cysteine protease pro regions. The most frequent occurrence is given above each open box. Amino acids present in more than 65% of the sequences are given in bold characters. Other amino acids found at each position are listed using the one letter code and the number of occurrence is given between brackets. B, alignment of mammalian cystatin C and definition of a consensus sequence with the pro region of cysteine proteases (amino acids in bold character).

17A). Les positions conservées dans plus de 65% des séquences forment un motif d'acide aminé ayant la séquence suivante: **GlyXaaAsnXaaPheXaaAspXaa**. Ce motif localisé autour de la position -40 de la propapaïne a été recherché dans les banques de séquences de protéine. Aucune des rares protéines identifiées par ces comparaisons de séquences n'a de rapport évident avec les protéases à cystéine à une exception remarquable près: le motif est présent dans la structure d'inhibiteurs de protéases à cystéine, les cystatines de mammifères (Fig. 17B).

Les cystatines sont des protéines d'environ 12 kDa ayant une très grande affinité pour les protéases à cystéine. La relation entre les deux inhibiteurs, région pro et cystatine, n'est pas aussi évidente qu'il semble à priori. En effet les structures tridimensionnelles et modèles de complexes entre des protéases à cystéine et les cystatines ont montré que le motif conservé n'est pas en contact direct avec le site actif de l'enzyme.

Nous avons choisi de muter indépendamment chacune des positions de ce motif afin de tenter d'en déduire la ou les fonctions. Quatre bibliothèques de mutants ont donc été construites dans *E. coli*, amplifiées et exprimées dans la levure. Les résidus tolérés ou non tolérés à chaque position ont été déterminés par deux cycles successifs de mutagenèse aléatoire. Le criblage rapide des bibliothèques a été réalisé grâce à une méthode permettant la caractérisation biochimique d'environ 1000 mutants/jour. Lors du premier cycle, nous avons criblé les bibliothèques de levure pour les mutations non fonctionnelles. Ensuite, une de ces mutations a été choisie pour chacune des positions du motif et soumise à un deuxième cycle de mutagenèse aléatoire. De ces quatre nouvelles bibliothèques nous avons isolé les mutations capables de supprimer la mutation initiale, autrement dit nous avons identifié quelques un des acides aminés tolérés à chacune des positions du motif.

Ces collections de mutants de la propapaïne ont été étudiées en détail. Les résultats sont compilés dans la Fig. 18. Les mutants non fonctionnels sont peu exprimés et sont complètement dégradés par la subtilisine. Ceci indique que le repliement des mutants est compromis par les mutations. Les mutants fonctionnels appartiennent à deux classes. La première classe inclut les mutants produits sous forme de précurseur dans la cellule et pouvant être convertis en papaïne mature active, soit par une réaction en cis soit par une réaction en trans *in vitro* dans les conditions définies pour la molécule de type sauvage. Le niveau de production de certains mutants fonctionnels est inférieur à celui du type sauvage indiquant que ces mutations ne

Non functional mutants	Leu (1 TTG, 1 CTC)					Asn (1 AAT)
	Thr (1 ACC)					Tyr (1 TAC)
	Val (1 GTG)	Ile (1 ATC)		Pro (1 CCC)		Met (2 ATG)
	His (1 CAC)	Phe (1 TTC)		Gly (1 GGC)		Val (1 GTG)
	Ile (1 ATC)	Stop (1 TAG)		Ser (1 AGC)		Glu (1 GAG)
Gly Xaa Asn Xaa Phe Xaa Asp						
-42 -41 -40 -39 -38 -37 -36						
Functional mutants	Ser (1 TCG)	Gly (1 GGG)		Trp (1 TGG)		Gly (1 GGG, 1GGC)
	Ala (1 GCG)	Thr (1 ACC)		Met (1 ATG)		Cys (2 TGC)
	Cys (1 TGC)	Ser (2 TCG)		Leu (1 TTG, 1 CTC)		
				His (1 CAC)		

Fig. 18: Compilation of the four independent random mutagenesis experiments. Amino acid replacements that completely prevent the conversion of propapain into papain are given above the boxed conserved amino acid motif. Amino acids replacement that are tolerated, other than amino acids found in the WT sequence, are given below the motif. Mutation Asp-36Asn was constructed by site directed mutagenesis together with the mutation Ser-34Ala such as to avoid the creation of a potential N-linked glycosylation site. Codons and the number of time they were isolated are given between brackets.

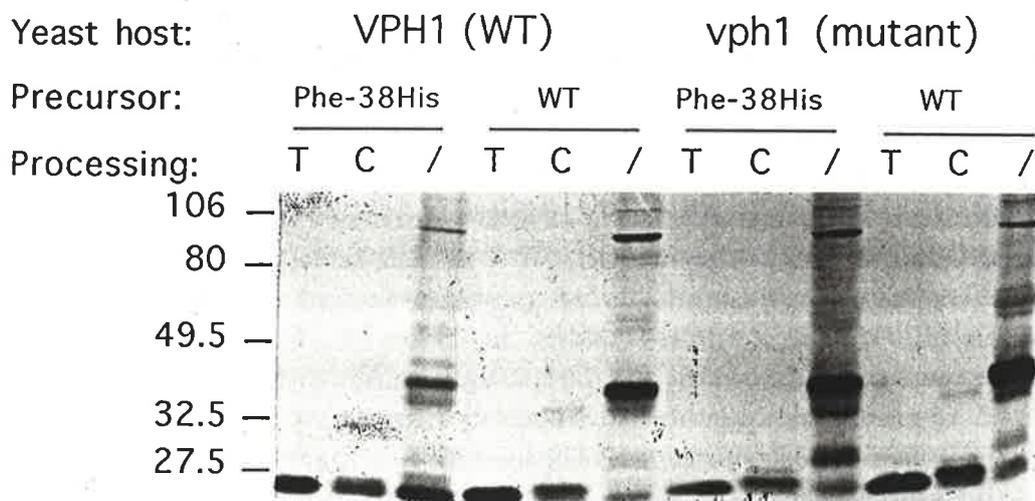


FIG. 19: rescue of the Phe-38His propapain mutant in a vph1 yeast strain. Western blot analysis. Total cellular extracts after treatment with endo H (/). Strain BJ3501 is wild type with respect to the VPH 1 gene. Propapain was processed in vitro using a trans (T) or a cis (C) reaction. Molecular masses of standards (kDa) are indicated in the left margin.

sont pas toutes neutres vis à vis du repliage et/ou de la stabilité des protéines.

La deuxième classe de mutants fonctionnels comprend trois membres, Phe-38His, Asp-36Gly et Asp-36Cys, qui sont en partie maturés en papaïne active dans la vacuole de la cellule. Le dénominateur commun entre ces trois mutations est l'élimination d'une charge négative (mutant -36) ou l'addition potentielle d'une charge positive (mutant Phe-38His). Le pH de la vacuole de la levure est d'environ 6, valeur à laquelle l'histidine du mutant Phe-38His possède probablement une charge positive. Si la présence de cette charge positive est responsable de la maturation intra-cellulaire de la papaïne, il devrait être possible de stabiliser le précurseur à un pH plus élevé où l'histidine ne serait pas protonnée. Une souche de levure ayant un pH vacuolaire plus élevé nous a été fournie par Elizabeth Jones (Pittsburg, Pa, USA). Dans cette souche, le mutant Phe-38His est effectivement produit sous forme de propapaïne (Fig. 19). L'accès à la forme stable du précurseur de ce mutant nous a permis de comparer son profil d'activation en fonction du pH avec celui de la molécule de type sauvage. Le pH auquel la vitesse d'activation du mutant est maximale est d'environ une unité de pH supérieure à celle du précurseur de type sauvage. La charge du motif conservé influence donc clairement l'efficacité du clivage intra-moléculaire.

Nous avons proposé comme site initial de clivage du précurseur la liaison peptidique entre les positions -37 et -38. Cette hypothèse est basée sur les quatre arguments suivants: accumulation transitoire d'un intermédiaire de maturation d'environ 30 kDa, clivage préférentiel d'un peptide synthétique entre les positions -37 et -38 à pH 4 mais non à pH 6, présence favorable d'une Phe en position P2 du substrat et clivage identique rapporté dans la littérature de la cystatine C par la cathepsine D lysosomale.

Le modèle du mécanisme de maturation dérivé de ces travaux est présenté en Fig. 20. A un pH supérieur à 4 la région pro est dans une conformation qui lui confère le rôle d'un inhibiteur de la papaïne, probablement par masquage conformationnel du site actif. Dans cet état l'Asp-36 est chargée négativement. Lorsque le pH décroît, la perte de charge en position -36 par protonation de l'Asp-36 provoque une transition structurale qui place la région pro dans le site actif de la papaïne. La région pro devient donc un substrat clivé entre les positions Leu-37 et Asp-36. Ce clivage déstabilise la région pro et libère la papaïne active capable de dégrader complètement les 36 résidus exédentaires en N-terminal.

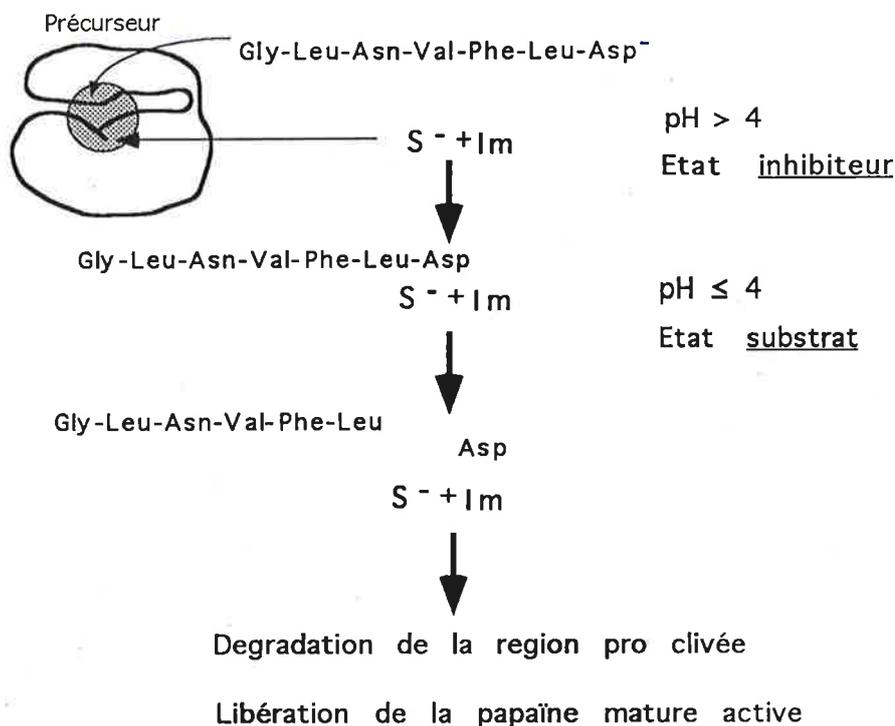


Fig. 20: modèle du mécanisme de la maturation du précurseur de la papaine. A pH supérieur à 4 la région pro joue le rôle d'inhibiteur de la papaine par masquage stérique du site actif représenté par la paire d'ions Cys25 (S^-) His159 (Im^+). A pH inférieur ou égal à 4 la région pro devient un substrat pour le site actif. La rupture de la liaison peptidique, probablement entre l'asp-36 et la Leu-37, destabilise la région pro qui est ensuite complètement hydrolysée.

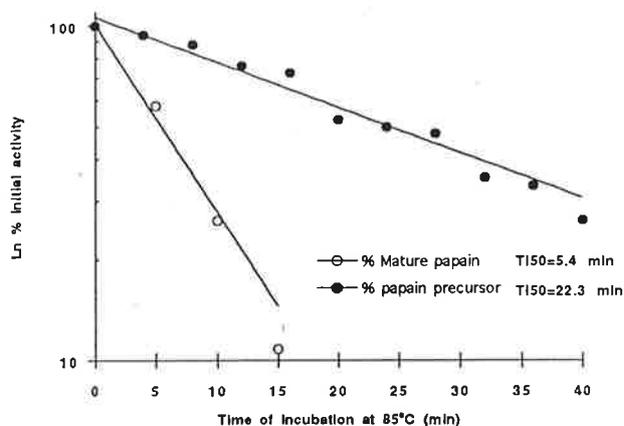


Fig. 21: stabilité thermique du précurseur de la papaine et du précurseur. La demi-vie des molécules à 85°C (T150) est indiquée.

Des travaux complémentaires ont aussi révélé que la région pro stabilisait la forme précurseur de la molécule (Fig. 21), la demi-vie du précurseur étant plus élevée que celle de la papaïne mature.

Des résultats préliminaires indiquent que la région pro est essentielle au repliage du précurseur. En effet, l'expression directe de la papaïne mature dépourvue de la région pro produit une molécule non fonctionnelle. Nous avons démontré que cette molécule n'était pas repliée de façon native. Une expérience de "sauvetage" en trans de la papaïne non pliée a démontré que le rôle de la région pro s'apparentait à celui des chaperons moléculaires. La co-expression, dans les cellules d'insectes, du mutant non auto-maturable Cys25Ser du précurseur de la papaïne et de la papaïne dépourvue de région pro permet la récupération d'une activité de protéases à cystéine.

Tout ou partie des fonctions distinctes de la région pro que nous avons mis en évidence sont spécifiques à la papaïne. Le remplacement de la région pro de la papaïne par celle d'une molécule homologue, la cathepsine S humaine, produit une molécule hybride non fonctionnelle car non repliée de façon native.

En résumé les 107 acides aminés de la région pro du précurseur de la papaïne combinent les fonctions spécifiques suivantes:

- inhibition de la papaïne
- régulation de la maturation
- chaperon moléculaire (?)
- stabilisation de la molécule

5.2.4- Etude du site actif (publications 21, 24, 25, 28, 3 soumise)

Notre capacité à manipuler le précurseur de la papaïne est à la base de nos études de la papaïne par les méthodes de l'ingénierie des protéines. Les mutants de l'enzyme ont été produits, soit dans les cellules d'insectes soit dans la levure, sous forme de précurseurs, convertis *in vitro* en papaïne mature puis purifiés. Les propriétés des mutants de l'enzyme sont ensuite déterminées par spectrofluorométrie en utilisant des substrats synthétiques de type CBZ-Phe-Arg-MCA. Les caractéristiques enzymatiques de l'enzyme recombinant sont identiques à celles de l'enzyme naturel. Les travaux publiés seront résumés ici. L'étude de la position Asn175, dont une partie des résultats seulement a été publiée, sera présentée avec plus de détails en raison de son intérêt particulier.

La papaïne est formée, à pH neutre, d'une paire d'ions entre la Cys25 et l'His 159 au pKas respectifs de 4.4 et 8.5. Le thiolate de la chaîne latérale de la cystéine joue le rôle d'un nucléophile lors de la

catalyse, et le noyau imidazole de l'histidine, celui d'une base générale. Ces valeurs inhabituelles des pK_a de ces deux résidus impliquent probablement l'intervention d'autres chaînes latérales pour la stabilisation de la paire d'ions. Le dipole formé par l'hélice α porteuse de la Cys25 pourrait aussi participer à cette stabilisation. Ce dernier aspect est difficile à tester expérimentalement et nous avons concentré nos études sur les chaînes latérales présentes à proximité du site actif.

L'Asp158 a souvent été proposée dans la littérature, suite à des études théoriques et expérimentales, comme jouant un rôle important, peut-être même essentiel, dans la catalyse. C'est en fait la seule chaîne latérale chargée proche de la diade Cys25-His159. Nous avons testé son rôle par l'étude d'une série de mutants en position 158. Nous avons démontré que l'influence de l'Asp158 dans la catalyse est minime et indépendante de la charge du résidu, le mutant Asp158Asn ayant une valeur de K_{cat}/K_M de $2 \times 10^4 \text{ M}^{-1}\text{S}^{-1}$ contre $1 \times 10^5 \text{ M}^{-1}\text{S}^{-1}$ pour l'enzyme de type sauvage. L'analyse fine d'autres mutants en position 158 a néanmoins permis d'établir le rôle des liaisons hydrogènes fournies par la chaîne latérale 158 dans la stabilisation de la paire d'ions du site actif.

Lors de la catalyse chez les protéases à cystéine la formation d'un complexe enzyme-substrat transitoire avait été établie. Cet état est hautement instable dû à l'apparition d'une charge négative au niveau de l'oxygène du carbonyle du substrat. Chez les protéases à sérine, un état de transition analogue est stabilisé par des interactions avec une cavité oxyanionique. L'observation de structures de la papaïne complexée avec plusieurs substrats a révélé qu'une telle cavité existait et que la Gln19 était en position adéquate pour interagir avec l'état de transition. Nous avons confirmé l'existence d'une cavité oxyanionique fonctionnelle dans la papaïne par mutagenèse de la Gln19.

La Ser176 forme une liaison hydrogène avec la Gln19 localisée dans l'autre domaine de la protéine. De plus la stricte conservation de cette chaîne latérale chez la famille des protéases à cystéine, suggère qu'elle possède un rôle important. Le remplacement de la Ser176 par une Ala n'a pourtant que des conséquences mineures sur l'activité de la papaïne. L'analyse des profils d'activité de l'enzyme en fonction du pH, et ce à plusieurs forces ioniques, a permis d'établir que la Ser176 contribue à stabiliser la paire d'ions du site actif.

L'Asn175 a été proposée par plusieurs équipes comme étant requise pour le contrôle de la rotation du noyau imidazole de

l'His159, une propriété jugée essentielle dans le fonctionnement de la papaïne. Cette proposition repose sur les observations suivantes:

- Asn175 est strictement conservée dans la superfamille des protéases à cystéine de la papaïne.

- Asn175 est en position analogue à l'Asp essentielle de la triade catalytique des protéases à serine.

- Asn175 forme une liaison hydrogène avec l'His159 du site actif.

Lors d'études préliminaires, nous avons montré que seul l'acide aminé analogue Gln pouvait partiellement remplacer l'Asn175. Ce travail ne nous avait toutefois pas permis de distinguer l'influence des mutations sur l'activité, la stabilité ou d'autres propriétés de l'enzyme. Deux mutants, Asn175Gln et Asn175Ala, ont été étudiés depuis de façon détaillée. Nous avons été surpris de constater que l'activité du mutant Asn175Gln, au pH optimum, n'était que 3,4 fois moins moindre que la molécule de type sauvage. De même l'élimination complète de toute possibilité pour la chaîne latérale de former des liaisons hydrogènes (mutant Asn175Ala) n'abolit pas totalement l'activité enzymatique. Dans cette dernière situation le rapport k_{cat}/K_M du mutant au pH optimum est de $3 \times 10^3 \text{ M}^{-1}\text{S}^{-1}$ soit 151 fois moindre que pour le type sauvage. Pour les deux mutants le paramètre affecté est le k_{cat} . Les profils d'activité en fonction du pH ont montré que, dans le cas du mutant Asn175Gln, la paire d'ion du site actif était déstabilisée par un facteur 7,6 ce qui correspond à une diminution de l'activité intrinsèque de l'enzyme de 1.5. Cette valeur remarquablement faible indique que la conservation stricte de l'Asn175 n'est probablement pas uniquement reliée au fonctionnement de la machinerie catalytique.

Nous avons donc recherché à mettre en évidence d'autres effets des mutations en position 175 sur la protéine. Nous avons remarqué que le rendement de purification des mutants était très inférieur à celui de la papaïne de type sauvage. Notre protocole de purification incluait plusieurs étapes potentiellement traumatisantes pour les protéines mutantes (incubation avec de la subtilisine, précipitation acide à température élevée). Nous avons mesuré la stabilité thermique des mutants (Fig. 22). Il est clair que le remplacement de la position 175 par une Gln décroît la stabilité thermique de l'enzyme. Cette décroissance est encore plus marquée pour le mutant Asn175Ala.

Nous avons aussi développé une méthode de double Western quantitatif afin d'évaluer l'impact des mutations sur la capacité des molécules de se replier de façon native (Fig. 23). Une très grande proportion des précurseurs mutants s'agrègent ce qui indique que

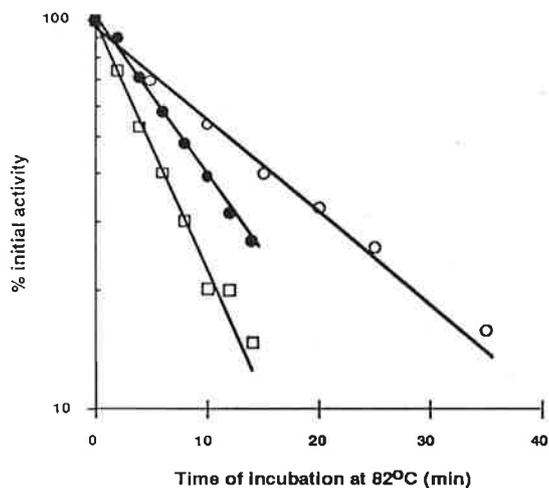
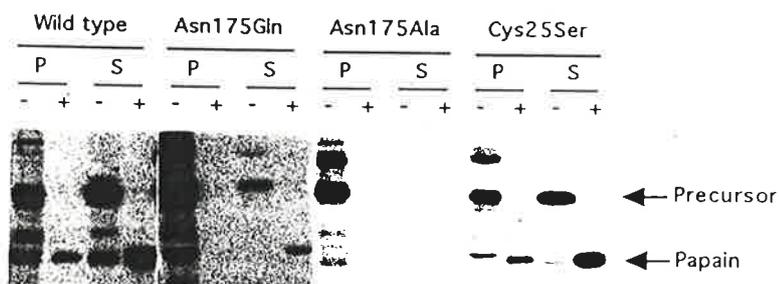


Fig. 22: stabilité thermique des mutants en position 175 de la papaïne.

La demi-vie des molécules à 82°C est de 12.5 min pour la papaïne de type sauvage, de 7.3 min pour le mutant Asn175Gln et de 4.6 min pour le mutant Asn175Ala.

A



B

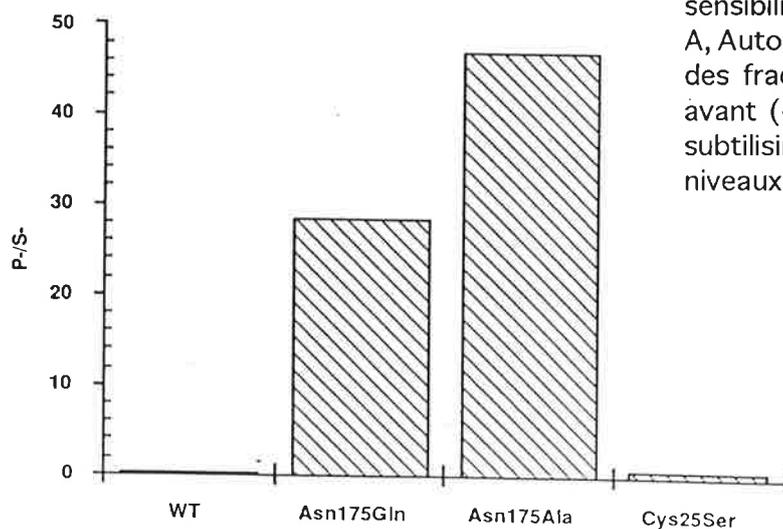


Fig. 23: ségrégation des mutants de la propapaine entre les fractions solubles et insolubles d'extraits cellulaires et sensibilité à la subtilisine.

A, Autoradiogramme d'un "Western blot" des fractions insolubles (P) et solubles (S), avant (-) et après (+) traitement par la subtilisine. B, histogramme représentant les niveaux relatifs des échantillons (P-) et (S-).

ces molécule ne sont pas repliées de façon native. La dégradation complète des molécules agrégées confirme leur état non natif. La fraction soluble est par contre résistante à la subtilisine. Environ 100 fois moins de molécules du mutant Asn175Gln sont repliées de façon native, par rapport aux molécules de type sauvage.

En conclusion:

- contrairement à de nombreuses prédictions basées sur l'analyse des structures tridimensionnelles et des calculs théoriques, l'Asn n'est pas essentielle à l'activité de la papaïne.

- le mécanisme catalytique de la papaïne est différent de celui des protéases à sérine.

- le site actif de la papaïne est composé d'une diade essentielle (Cys25 et His159).

- l'Asn 175 semble jouer un rôle critique lors du repliement de la molécule dans la cellule.

- L'Asn175 est impliquée dans la stabilité de la protéine native.

5.2.5- Etude de la spécificité de la papaïne (publications 27)

L'efficacité avec laquelle différentes protéases à cystéine hydrolysent les substrats dépend de la nature de ces substrats. Par exemple, la papaïne catalyse l'hydrolyse rapide de la liaison peptidique de substrats possédant en position P2 des chaînes latérales aromatiques de grosse taille. Les substrats ayant une Arg en P2 ne sont pas clivés avec autant d'efficacité. Par contre, la cathepsine B clive les substrats ayant une Arg en position P2 plus efficacement que la papaïne. La comparaison des structures et séquences disponibles de protéases à cystéine a révélé la nature des interactions entre les chaînes latérales des substrats et des enzymes. Ces interactions, qui diffèrent d'une protéase à cystéine à l'autre, sont probablement responsables de la sélectivité des substrats. Le sous-site S2 de la papaïne qui est en interaction avec la chaîne latérale en position P2 du substrat, est le mieux défini. L'alignement de séquences expose les différences de structure de ce sous-site entre la papaïne et les cathepsines B. Ces différences sont compilées dans le tableau II:

Enzymes	Position des résidus				
	68	133	157	160	205
Papaïne	Pro	Val	Val	Ala	Ser
Cat. B humaine	Pro	Ala	Gly	Ala	Glu
Cat. B de souris	Pro	Ala	Gly	Ala	Glu
Cat. B de rat	Pro	Ala	Gly	Ala	Glu
Cat. B bovine	Pro	Ala	Gly	Ala	Glu

Nous avons testé le rôle des résidus formant le sous-site S2 en batissant par étapes dans la papaïne le site correspondant de la cathepsine B. La cathepsin B est près de 4 fois plus active vis à vis d'un substrat possédant une Phe en P2 que vis à vis d'un substrat possédant une Arg à la même position. Cette valeur est d'environ 900 pour la papaïne. Le rapport d'activité équivalent pour le double mutant de la papaïne Val133Ala/Ser205Glu est de huit. Il a donc suffi de modifier deux des 212 acides aminés de la papaïne pour obtenir une modification importante de la reconnaissance du substrat et ce avec une perte d'activité de l'enzyme relativement modeste. Cette série d'expériences a non seulement confirmé certaines hypothèses reliées au mécanisme de sélectivité du substrat mais a aussi mis en évidence notre capacité à altérer de manière rationnelle les propriétés d'une enzyme.

En résumé, notre exploration de la région du site actif de la papaïne par mutagénèse a grandement amélioré notre connaissance du fonctionnement de l'enzyme en définissant la contribution propre à plusieurs chaînes latérales. Nous avons éliminé certaines des ambiguïtés et incertitudes présentes dans la littérature avant ce travail. Le site actif de la papaïne peut être décrit comme étant composé d'une diade ionique essentielle à l'activité. La formation et le maintien de cette paire d'ions sont assurés par l'influence de chaînes latérales d'acides aminés proches.

5.3- Ingénierie des lipases (publications 36, 37)

Au cours de l'année 1991 un de mes collègues de l'IRB, le DR. Mirosław Cygler m'a proposé une collaboration. Son équipe venait de déterminer la structure tridimensionnelle de la lipase II de *Géotrichum candidum* (GCLII). Le modèle de la protéine permettait de poser des hypothèses quant à la localisation du site actif et la fonction de plusieurs boucles d'acides aminés. Nous avons convenu de travailler ensemble, en collaboration avec le Dr. Edmund Ziomek enzymologiste à l'institut, afin de compléter les études structurales par des études fonctionnelles.

La protéine a donc été exprimée dans la levure, purifiée et caractérisée. La lipase GCLII recombinante a des propriétés enzymatiques similaires à celles de l'enzyme naturelle (Tableau III). Le rôle des acides aminés de la triade putative du site actif a été étudié par mutagénèse dirigée. Le remplacement de la Ser217 par Cys ou Thr ou de l'His463 par Ala abolit complètement l'activité enzymatique. Ces résidus sont donc essentiels pour le fonctionnement de l'enzyme et interviennent très probablement dans la catalyse. Le troisième membre de la triade catalytique, Glu354, tolère la

substitution par un autre acide aminé chargé, Asp, mais non la substitution par Ser, Ala ou Gln. En fait, la triade fonctionnelle Ser-His-Asp est présente naturellement dans plusieurs autres hydrolases (lipases, protéases et estérases). Nos résultats montrent que le site actif de la lipase GCLII est suffisamment flexible pour accepter une réduction d'un groupement méthylène en position 354. Dans cette configuration mutée, l'activité de l'enzyme vis à vis de la trioléine est environ la moitié de celle de l'enzyme de type sauvage (Tableau III). Ces résultats rappellent ceux que nous avons obtenu lors du remplacement conservateur de l'Asn175 de la papaïne.

Tableau III: Comparaison des activités spécifiques de la lipase GCLII naturelle, recombinante et du mutant Glu354Asp.

Substrat	Methode	Activité spécifique de GCLII (U/mg)			mutant/type sauvage
		naturelle	recombinant type sauvage	mutant Glu354Asp	
TO	HPLC	491±65	1017±213	542±598	0.53
TO	Titrimétrie	958±180	952±190	422±126	0.44
TB		54±25	40±14	5.3±1.6	0.13
PNPA	Spectrophotometrie	0.79±0.2	0.83±0.12	0.43±0.19	0.52
PNPB		2.05±0.36	2.95±0.78	1.08±0.25	0.37

Le champignon *Géotrichum candidum* sécrete au moins une autre lipase (GCLI) à laquelle certains auteurs attribuent des propriétés enzymatiques différentes notamment en ce qui concerne la sélectivité des substrats. La comparaison des structures et des fonctions des deux types de lipase pourrait accroître nos connaissances des mécanismes de la spécificité des substrats. Les résultats présentés dans la littérature ont été jusqu'à présent non concluants en raison des difficultés à purifier des enzymes aux caractéristiques physico-chimiques proches. Nous avons donc décidé dans un premier temps de dresser le répertoire des gènes codant pour les lipases homologues à GCLII dans plusieurs souches de *Géotrichum candidum*. Nous avons cloné, séquencé puis analysé quatre gènes codant pour GCLII et quatre codant pour GCLI. Le degré de polymorphisme de ces gènes pourrait expliquer les différences de spécificité des protéines correspondantes. Aucune autre séquence divergente de ces deux lipases n'a été mise en évidence. Une des GCLI recombinantes exprimée dans la levure est maintenant caractérisée au laboratoire.

5.4- Mutations coordonnées des protéines monomériques (publications 17, 30)

A la suite de la présentation à Montréal d'une méthode de recherche de mutations coordonnées chez la protéine de coque du virus de la mosaïque du tabac, le Dr. Danièle Altschuh et moi-même avons décidé de collaborer pour étendre cette méthode à d'autres protéines. Par la suite nous avons vérifié expérimentalement la validité de la théorie en utilisant la papaïne comme modèle. Cet effort a été soutenu par un programme d'échange entre le NRC et le CNRS.

L'étude détaillée des motifs de substitution dans les alignements de protéines a permis d'établir une relation entre leur appartenance à un même groupe de substitution et la proximité des résidus dans l'espace. Ces résidus mutent donc de manière non indépendante, autrement dit de manière coordonnée. Nous avons démontré l'existence de changements coordonnés dans plusieurs classes de protéines: les globines, les protéases à serine et les protéases à cystéine. Ces changements sont présents le plus souvent dans des régions contraintes de la protéine (site actif, amas hydrophobes). Nous avons aussi relaxé les paramètres de recherche de groupes de substitution de façon à retirer plus d'informations des alignements de séquences.

Lors de la recherche de mutations coordonnées chez les protéases à cystéine, le contact inter-domaine en position Val32-Ala162 de la papaïne a retenu notre attention: la séquence correspondante est Ala32-Val162 dans l'actinidine et Gly32-Leu162 dans une série de sept autres protéases homologues. Nous avons montré que les substitutions uniques, correspondant aux chaînes latérales présentes dans les protéases homologues, perturbaient la stabilité thermique de la molécule, lorsque introduites dans la papaïne. Cet effet néfaste est supprimé par l'introduction de secondes mutations reproduisant les contacts naturels. Nous avons donc démontré qu'il est possible de manipuler les régions contraintes des protéines de manière rationnelle en utilisant les informations présentes dans les séquences de protéines homologues.

Ce travail a eu comme conséquence imprévue de démontrer que la perturbation d'un contact entre les deux domaines de la papaïne n'avait aucune conséquence mesurable sur l'activité de l'enzyme. Comme le site actif est localisé à l'interface des deux domaines, nous avons émis l'hypothèse que l'agencement précis des deux domaines devait être important pour une catalyse efficace, puisque la configuration du couple Cys25/His 159 en dépend. En fait, ce résultat intrigant a stimulé notre intérêt vis à vis du rôle des

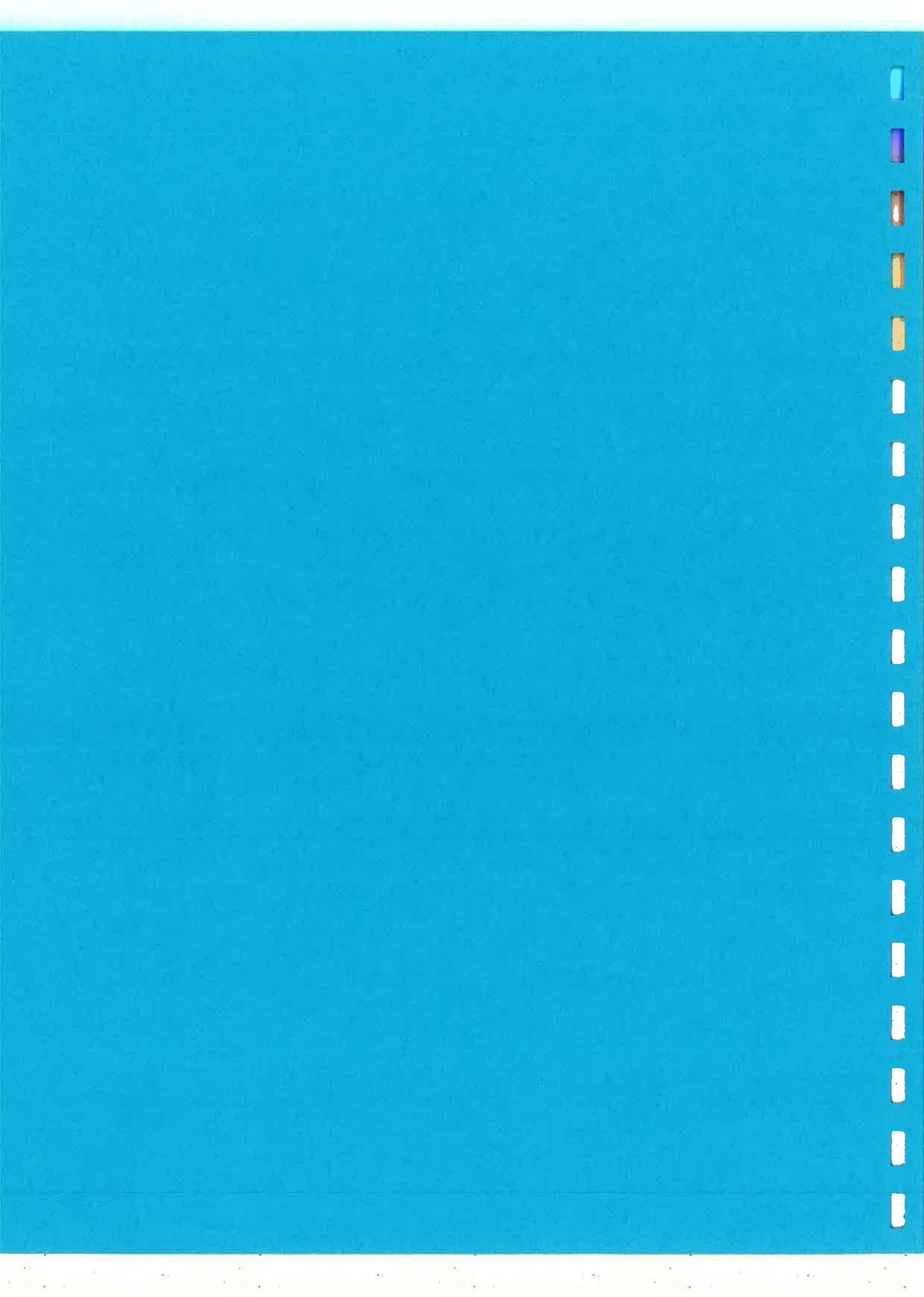
contacts inter-domaines. Nous avons donc décidé de perturber quelques autres contacts inter-domaines afin d'en mesurer les conséquences fonctionnelles.

En bref, l'étude du contact inter-domaine entre la Val13 et la Val164 a produit les résultats les plus inattendus. Lorsque les deux acides aminés sont simultanément remplacés par des Ala, le k_{cat} du double mutant diminué de 30 fois par rapport à celui du type sauvage. Ceci est d'autant plus surprenant que ce contact est à 15Å du site actif. La modification d'autres contacts, plus proches du site actif et impliquant le même nombre d'atomes, est neutre par rapport à l'activité de la papaïne.

Une relation entre les pKas de la papaïne et la distance entre la Cys25 et l'His159 a été établie. Les pKas du double mutant sont similaires à ceux de l'enzyme de type sauvage. Ces données impliquent que la géométrie du site actif n'a pas été perturbée de façon importante par les mutations. Nous avons par conséquent émis l'hypothèse que la réduction de l'efficacité de la catalyse provient de changements causés au comportement dynamique de la protéine durant la catalyse. Nous essayons actuellement de soutenir cette hypothèse par des analyses de dynamique moléculaire en collaboration avec le Dr. M. Karplus (Harvard).

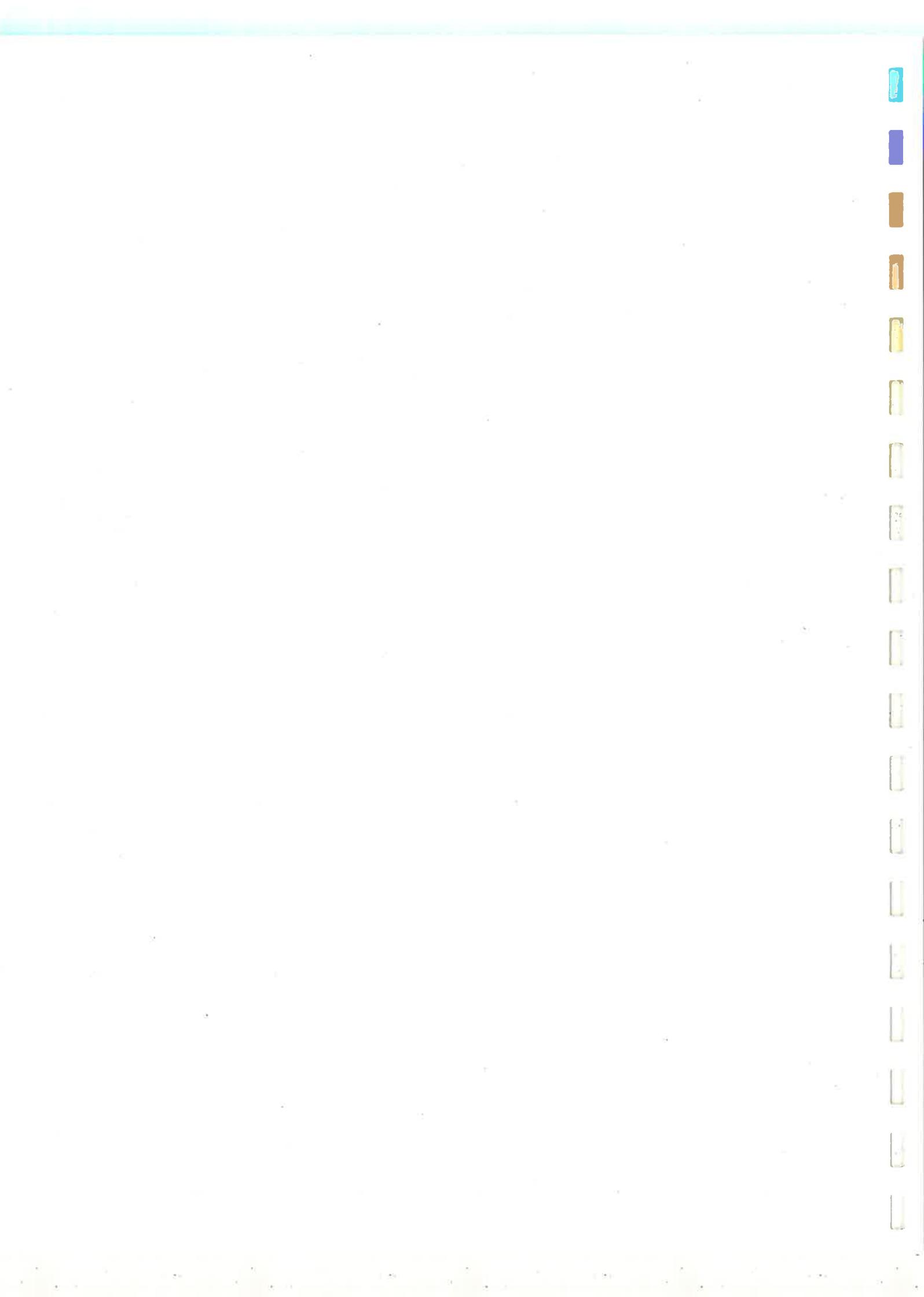


Deuxième partie: présentation des travaux publiés

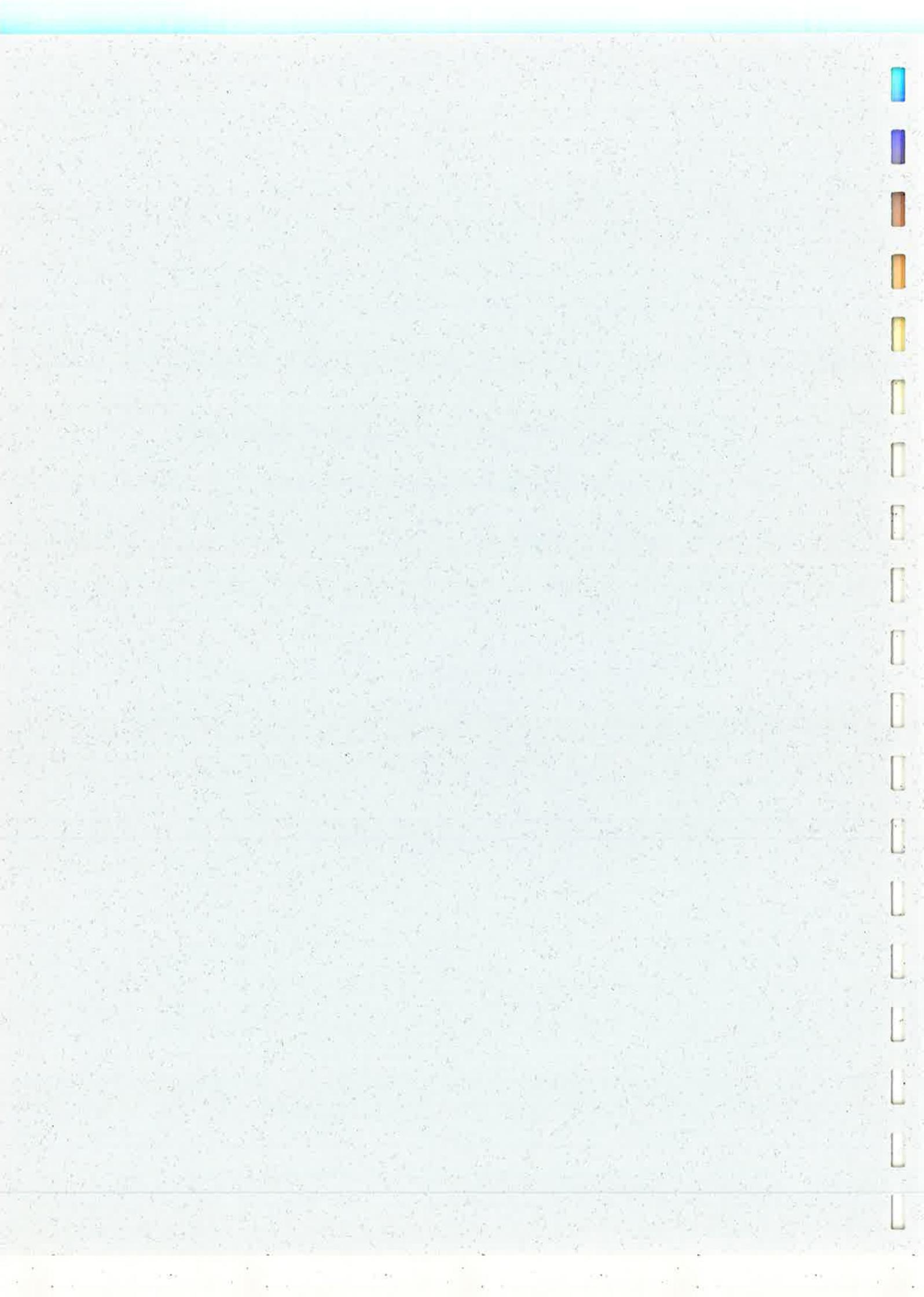


Les articles sous presse et publiés dans des revues à comité de lecture sont présentés par l'intermédiaire de la page contenant le résumé.

Les publications sont regroupées afin de faire ressortir les dénominateurs communs entre les différents thèmes de recherche abordés. Par conséquent elles ne sont pas toujours présentées dans l'ordre chronologique.



1- Différenciation chez les plantes supérieures



Expression of the Gene Coding for the Small Subunit of Ribulosebisphosphate Carboxylase during Differentiation of Tobacco Plant Protoplasts

Thierry VERNET, Jacqueline FLECK, Andrée DURR, Christiane FRITSCH, Monique PINCK, and Léon HIRTH

Institut de Biologie Moléculaire et Cellulaire du Centre National de la Recherche Scientifique,
Université Louis Pasteur, Strashbourg

(Received April 1; June 2, 1982)

A hybridization probe was used to study the regulation of expression of the gene coding for the small subunit of ribulose 1,5-bisphosphate carboxylase, during functional differentiation of protoplasts. A library of cDNA from poly(A)-containing RNA extracted from specially treated tobacco leaves was constructed in the plasmid pBR322, by blunt-end ligation. This library was screened by colony hybridization with ^{32}P -labelled cDNA prepared from mRNA coding for the precursor of the small subunit. A positive colony was identified containing recombinant plasmids with a nucleotide sequence homologous to this mRNA. These plasmids, bound to diazobenzoyloxymethylated cellulose paper, were then used as a hybridization probe. The results showed unambiguously that the small subunit was not transcribed in protoplasts but was transcribed in undifferentiated white and chlorophyll-containing green callus cultures derived from protoplasts. The discrepancy between these results and those obtained with classical techniques is discussed.

The regulation of gene expression at the molecular level has been studied extensively in prokaryotes. In eucaryotes such study is mainly restricted to animal cells. The protoplasts of higher plants would be a good subject for such studies because whole plants, that are in every respect similar to the original, can be regenerated from protoplasts isolated from any cell (for review see [1]).

We observed earlier that protoplasts freshly isolated from mesophyll cells of *Nicotiana sylvestris* immediately show dramatic changes in the proteins they synthesize [2, 3]. Some of these changes appear in response to the high osmotic strength of the medium [4] necessary for obtained viable protoplasts, but the most surprising change is that, although the protoplasts retain their ribulose-1,5-bisphosphate carboxylase (for a review of this enzyme see [5] and [6]) activity, they have lost the ability to synthesize the two subunits of this enzyme: the large subunit, L, and the small subunit, S.

These protoplasts can regenerate a cell wall and can divide, giving rise to white or green calluses, depending on the medium used. This finding suggests that some physiological functions have been restored in one case but not in the other.

We have therefore used freshly isolated protoplasts, white calluses and green calluses to study the regulation of the synthesis of the two subunit of ribulose-1,5-bisphosphate carboxylase.

Because the large subunit, L, encoded by the chloroplast genome, has been extensively studied (for a review see [7]),

Abbreviations. S, small subunit; L, large subunit; cDNA, complementary DNA; dscDNA, double stranded complementary DNA; Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; DBM-paper, diazobenzoyloxymethylated cellulose paper; NBM-paper, nitrobenzyloxymethylated cellulose paper; oligo(dT), oligo(thymidylic acid).

Enzymes. Ribulose-1,5-bisphosphate carboxylase (EC 4.1.1.39); T₄ RNA ligase (EC 6.5.1.3); RNA-dependent DNA polymerase (reverse transcriptase) (EC 2.7.7.49); DNA polymerase I (EC 2.7.7.7); S₁ nuclease (EC 3.1.30.1); restriction endonuclease *Sall* (EC 3.1.23.37); T₄ DNA ligase (EC 6.5.1.1).

we have restricted our research to the expression of the gene for the small subunit, S; this subunit is encoded by the nuclear genome and synthesized in the cytoplasm [8–10].

To determine unambiguously if the expression of this gene is regulated at the level of transcription, it was necessary to have available a specific probe corresponding to the mRNA coding for the small subunit. The present paper reports the construction of such a probe and its use to elucidate the mechanism of the regulation of the gene coding for the small subunit in tobacco plant cells at various stages of functional differentiation.

MATERIALS AND METHODS

Plant Material

The tobacco plant, *Nicotiana sylvestris*, was used throughout. The seeds were a gift from SEITA (Institut Experimental du Tabac, Bergerac, France). In order to increase the amount of mRNA encoding for the small subunit, the plants were specially light-treated, as previously described [11].

The protoplasts were isolated, tested for viability and incubated as reported before [2, 3]. The resulting calluses were maintained by transferring them into fresh Nagy and Maliga medium [12] but with the sucrose concentration reduced to 0.1–0.15 M. In this case the cell line is white and stable during numerous transfers in the same medium. Transfer of these white calluses into a medium containing Murashige and Skoog's major elements [13], and Heller's minor elements [14], including FeEDTA and 3 mM sucrose, results in the appearance of green nodules. Transfer of these nodules into the same medium results in the isolation of a stable chlorophyll-containing cell line. The physiological properties of these two cell lines will be reported elsewhere (A. Durr et al., unpublished results).

RNA Preparation

Polysomes were extracted from leaves and from protoplasts by a method based essentially on a procedure described

OSMOTIC-SHOCK 'STRESS PROTEINS' IN PROTOPLASTS OF *NICOTIANA SYLVESTRIS**

J. FLECK, A. DURR, C. FRITSCH, T. VERNET and L. HIRTH

Laboratoire de Virologie, Institut de Biologie Moléculaire et Cellulaire, 15, rue Descartes,
67084 Strasbourg Cédex (France)

(Received July 30th, 1981)

(Accepted January 12th, 1982)

SUMMARY

Because protoplasts isolated from leaves of *Nicotiana sylvestris* show dramatic changes in their gene expression, the proteins they synthesized were compared to those synthesized by cells incubated in hypertonic or isotonic medium. These comparisons led to the conclusion that the changes are in part due to the high osmotic strength of the isolation medium necessary to obtain viable protoplasts. The protoplasts responded to this environmental change very much as other types of cells respond to stresses.

INTRODUCTION

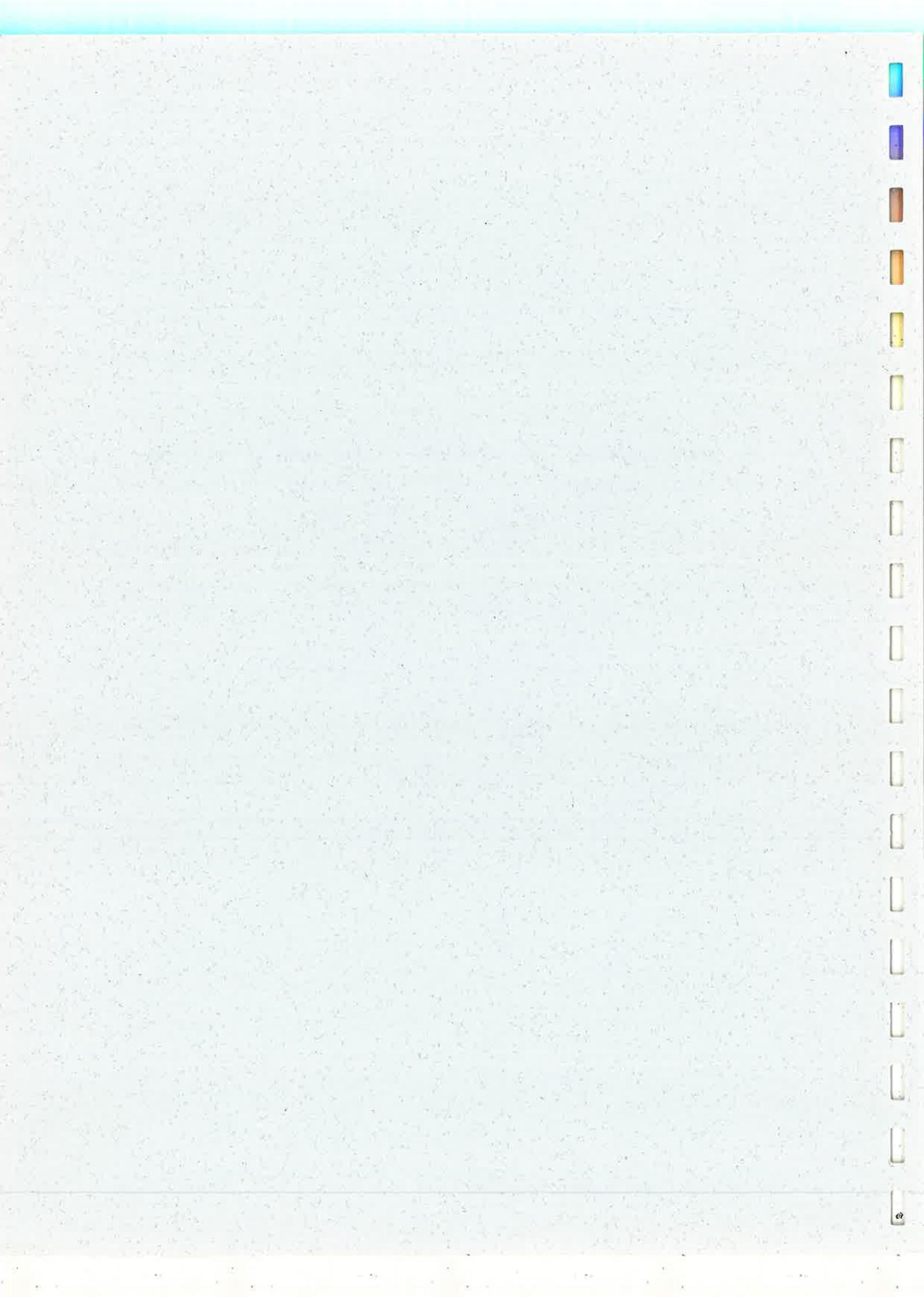
The isolation of protoplasts from fully differentiated, non-dividing mesophyll cells and their regeneration into mature, flowering plants is now a well established technique (for review, see Ref. 1). The number of species from which protoplasts can be cultured efficiently is increasing rapidly.

In previous experiments we observed dramatic changes in the proteins synthesized by freshly isolated protoplasts [2,3]. This is most evident with respect to ribulose biphosphate carboxylase (EC 4.1.1.39), a key enzyme in photosynthesis. In leaves, this enzyme (for review, see Ref. 4) constitutes about 50% of the total soluble proteins, but the synthesis of its two subunits, the large subunit L and the small subunit S is not detectable once the protoplasts are isolated and incubated in their culture medium. In addition, other proteins, not detected in leaves, are actively synthesized by protoplasts [2,3].

It seems that some genes active in the leaves are shut off during protoplast isolation while other genes become active — in other words — that protoplasts have a modified programme of gene expression.

*Part of this work was supported by Contract no. 0781 of the CNRS (ATP Biologie Moléculaire Végétale).

2- Développement de vecteurs, techniques et méthodes



GENE 1224

A direct-selection vector derived from pColE3-CA38 and adapted for foreign gene expression*

(Recombinant DNA; human proinsulin; colicin; immunity; pBR327; gene fusion)

Thierry Vernet**, Peter C.K. Lau, Saran A. Narang and Louis P. Visentin

Molecular Genetics Section, Division of Biological Sciences, National Research Council of Canada, Ottawa, Ontario K1A 0R6 (Canada) Tel. (613) 990-0853

(Received July 10th, 1984)

(Revision received November 4th, 1984)

(Accepted November 10th, 1984)

SUMMARY

The construction of a plasmid vector, pVT25, which allows an efficient and direct selection for transformed cells carrying recombinant plasmids is described. In this vector, the replicon and Ap^R gene from plasmid pBR327 are fused to the *colE3* gene of pColE3-CA38, whereby positive selection is based on the inactivation of the lethal colicin E3 by the insertion of a foreign DNA fragment. However, pVT25 can be maintained within the *Escherichia coli* cells when complemented with another plasmid, pVT26, which expresses the colicin E3 immunity (*imm*) and the Tc^R phenotypes. Furthermore, pVT25 was used to regulate the expression of the synthetic human proinsulin gene fused to the *colE3* gene at the single *Cla*I site. The production of the characteristic C-peptide of proinsulin, monitored by radioimmunoassay, was shown to be under the control of the inducible promoter of the *colE3* gene.

INTRODUCTION

Selection for cloning vehicles with DNA inserts is a time-consuming step in cloning experiments but it can be curtailed by using one of the direct or positive selection procedures which has been developed for

the *E. coli* system (Burns and Beacham, 1984 and references therein). In this report we describe the construction and use of a new direct selection cloning vector, pVT25, which is a hybrid between plasmid pBR327 (Soberon et al., 1980) and the *colE3* gene from bacteriocinogenic plasmid ColE3-CA38 (Watson and Visentin, 1980). Plasmid pBR327 was selected instead of the classic pBR322 (Bolivar et al., 1977) because it lacks the 1089-bp *Eco*RII fragment containing the *rop* gene (repressor of primer; Cesarini et al., 1982; Som and Tomizawa 1983). Since Twigg and Sherratt (1980) showed that deletion within the *rop* gene results in an increase of plasmid copy number, we consequently expect to benefit from this property. Furthermore the vector will be smaller in size. The *colE3* gene was chosen because

* Issued as NRCC publication No. 23828.

** To whom all correspondence and reprint requests should be sent at his present affiliation: Montreal Biotechnology Research Institute, c/o National Research Council of Canada, Ottawa, Ontario K1A 0R6 (Canada) Tel. (613) 990-0853.

Abbreviations: Ap, ampicillin; bp, base pairs; EtBr, ethidium bromide; kb, kilobase pairs; MC, mitomycin C; ^R, resistance; RIA, radioimmunoassay; ^S, sensitivity; Tc, tetracycline; [], indicates plasmid-carrier state.

GENE 01944

A family of yeast expression vectors containing the phage f1 intergenic region*

(Recombinant DNA; plasmid vector; filamentous phage; *URA3* gene; *Saccharomyces cerevisiae*; gene expression; killer toxin)

Thierry Vernet, Daniel Dignard and David Y. Thomas

Genetic Engineering Section, Biotechnology Research Institute, National Research Council of Canada, Montreal, Que.
(H4P 2R2 Canada)

(Received 2 October 1986)

(Revised 6 January 1987)

(Accepted 7 January 1987)

SUMMARY

The construction and characterization of a family of yeast expression vectors is described. They have the following features: (1) plasmid replication and selection (Ap^R) in *Escherichia coli*, (2) packaging of single-stranded (ss) DNA upon infection of *E. coli* with a filamentous helper phage, (3) replication in *Saccharomyces cerevisiae* based on the 2 μ plasmid origin of replication (*ori*), (4) selection in yeast by complementation of *LEU2* (pVT-L series, size 6.3 kb) or *URA3* gene (pVT-U series, size 6.9 kb) and (5) seven unique restriction sites for cloning within an 'expression cassette' which includes the promoter and 3' sequence of the *ADH1* gene. The multiple cloning site as well as the *ori* and intergenic region of the phage f1 have been cloned in two orientations for convenient gene cloning and ssDNA strand selection. As a result any of these eight vectors can be chosen for cloning, expressing genes in yeast, sequencing and mutagenesis without the need for recloning into specialized vectors.

INTRODUCTION

Progress in molecular genetics has relied partly upon the availability of a broad and increasingly sophisticated variety of cloning vectors. Indeed, shuttle vectors capable of replicating in both yeast

and *E. coli* have greatly simplified the isolation and characterization of yeast genes (for a review of yeast cloning vectors see Parent et al., 1985). However, these vectors are often large and lack both convenient restriction cloning sites and features needed to avoid recloning in other specialized vectors.

Correspondence to: Dr. T. Vernet, Genetic Engineering Section, Biotechnology Research Institute, National Research Council of Canada, 6100, Avenue Royalmount, Montreal, Que. (H4P 2R2 Canada) Tel. (514)496-6154.

* Issued as NRCC publication No. 26301.

Abbreviations: Ap, ampicillin; bp, base pair(s); Δ , deletion; ds, double-stranded; kb, kilobases or 1000 bp; Km, kanamycin; MCS, multiple cloning site; m.o.i., multiplicity of infection; *ori*, origin of replication; PEG, polyethylene glycol; PolIk, Klenow (large) fragment *E. coli* DNA polymerase I; R , resistance; ss, single stranded; Tc, tetracycline, 2YT medium, see MATERIALS AND METHODS, section a; [], designates plasmid-carrier state.

Synthesis of the Membrane Fusion and Hemagglutinin Proteins of Measles Virus, Using a Novel Baculovirus Vector Containing the β -Galactosidase Gene

JORGE VIALARD,¹ MANON LALUMIÈRE,¹ THIERRY VERNET,¹ DALIUS BRIEDIS,² GHALIB ALKHATIB,¹ DOMINIQUE HENNING,¹ DAVID LEVIN,^{1†} AND CHRISTOPHER RICHARDSON^{1*}

Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2,¹ and Department of Microbiology and Immunology, McGill University, Montreal, Quebec H3A 2B4,² Canada

Received 7 April 1989/Accepted 15 September 1989

An improved baculovirus expression vector was developed to expedite screening and facilitate oligonucleotide-directed mutagenesis. This vector contained twin promoters derived from the P10 and polyhedrin genes of *Autographica californica* nuclear polyhedrosis virus. The P10 promoter directed the synthesis of β -galactosidase, whereas the polyhedrin promoter controlled the synthesis of foreign gene products. These two genes recombined with wild-type virus genome to yield recombinants which were polyhedrin negative, produced the foreign gene product, and formed blue plaques when β -galactosidase indicator was present in the agarose overlay. An origin of replication derived from M13 or f1 bacteriophage was also included in the plasmid to permit the synthesis of single-stranded DNA. This template DNA was used to introduce or delete sequences through the process of site-specific mutagenesis. The measles virus virion possesses a membrane envelope which contains two glycoproteins: the hemagglutinin (H) and membrane fusion (F) proteins. The H polypeptide has receptor-binding and hemagglutinating activity, whereas the F protein mediates virus penetration of the host cell, formation of syncytia, and hemolysis of erythrocytes. Genes for these two glycoproteins were inserted into the *NheI* cloning site of the modified expression vector described above. The vector and purified wild-type viral DNA were introduced into Sf9 insect cells by calcium phosphate precipitation. A mixture of wild-type and recombinant virus was generated and used to infect Sf9 cells, which were subsequently overlaid with agarose. After 3 days, 0.1 to 1% of the plaques became blue in the presence of β -galactosidase indicator. At least 70% of these blue viral colonies contained the foreign gene of interest as determined by dot blot analysis. Recombinant virus was separated from contaminating wild-type virus through several rounds of plaque purification. Insect cells were then infected with the purified recombinants, and synthesis of H and F proteins was verified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis followed by immunoblot detection and Coomassie blue staining. Glycosylation of the proteins appeared to be impaired somewhat, and the precursor to the F protein was not completely cleaved by the proteases present in insect host cells. On the other hand, both proteins appeared to be active in hemagglutination, hemolysis, and cell fusion assays. Levels of synthesis were in the order of 50 to 150 μ g of protein per 10^8 cells.

The baculovirus-insect cell expression system has gained wide popularity as a means of expressing foreign genes for high-level production of relevant proteins (27, 36, 55). Baculovirus expression vectors use the strong, efficient promoter from the polyhedrin gene to direct transcription of the foreign gene. Polyhedrin protein is normally synthesized very late in infection (24 to 72 h postinfection) and can account for 20 to 50% of the total protein made in infected cells. This protein, which is not essential for virus replication, associates with virions in the nucleus to produce protective structures called occlusion bodies. These bodies can easily be discerned by light microscopy (59). The underlying principle behind this expression system relies upon vectors which direct recombination and substitution of the polyhedrin gene with the foreign gene of interest. Recombinant virus is produced, which forms plaques that are occlusion body negative and express the foreign gene product to various degrees.

The most widely used transfer vectors for introducing foreign genes into wild-type baculovirus are pAc373 (50, 51).

pE-55 (35), p89B310 (15, 31), pAcYM1 (32), and pVL941 (29). These vectors contain the 5'- and 3'-flanking regions of the polyhedrin gene, the polyhedrin promoter and polyadenylation site, and pUC8 Amp^r for growing the plasmid in *Escherichia coli* bacterial cultures. Insertion of the recombinant gene into the wild-type virus genome relies upon a process of homologous recombination between the flanking sequences of the vector and wild-type DNA. Other vectors consisting of fused polyhedrin and foreign gene-coding sequences have also been developed (28).

Screening recombinant viruses and purifying virus containing the foreign gene away from contaminating wild-type virus can be laborious and time-consuming. This process can involve several rounds of plaque purification by using visual screening or hybridization techniques to detect recombinant plaques. An approach whereby two different promoters controlling β -galactosidase and foreign gene expression recombine as a unit with wild-type viral DNA has proven to be very successful with vaccinia virus (5). Recombinant virus was visualized as blue plaques when β -galactosidase indicator was included in the agar overlay of plaque assays. Recently Pennock et al. (40) have described a baculovirus vector which contains β -galactosidase under control of the polyhedrin promoter together with a unique *PstI* site for

* Corresponding author.

† Present address: Probetec Corporation, Downsview, Ontario, Canada M3J 1P3.

GENE 03871

Enhanced secretion from insect cells of a foreign protein fused to the honeybee melittin signal peptide(Baculovirus; *Spodoptera frugiperda* cells; recombinant DNA; cysteine protease; polymerase chain reaction; papain; gene expression vector)Daniel C. Tessier^a, David Y. Thomas^a, Henry E. Khouri^b, France Laliberté^a and Thierry Vernet^a^aGenetic Engineering Section and ^bProtein Engineering Section, Biotechnology Research Institute, National Research Council of Canada, Montréal, Québec H4P 2R2 (Canada)

Received by J. Davidson: 4 July 1990

Revised: 17 September 1990

Accepted: 20 September 1990

SUMMARY

The baculovirus/insect cell system has been remarkably successful in yielding high levels of synthesis of many proteins which have been difficult to synthesize in other host/vector systems. The system is also capable of secreting heterologous proteins, but with generally low efficiency. We have increased the efficiency of secretion of the system by using signal peptides of insect origin to direct the secretion of a foreign protein. The precursor of the plant cysteine protease papain (propapain) has been used as a report enzyme to compare secretion efficiency. Insect cells infected with a baculovirus recombined with the gene encoding propapain fused to the sequence encoding the honeybee melittin signal peptide secreted over five times more papain precursor than the wild-type prepropapain which used the plant signal peptide. Based on these results, we have assembled pVT-Bac, an *Autographa californica* nuclear polyhedrosis virus transfer vector that may enhance secretion of other foreign proteins from insect cells. The vector incorporates a number of features: phage fl *ori* to facilitate site-directed mutagenesis, the strong polyhedrin promoter upstream from the melittin signal peptide-encoding sequence, and eight unique restriction sites to facilitate fusion of heterologous genes.

INTRODUCTION

The insect cell-baculovirus expression system has specific characteristics which make it remarkably efficient for the production of a wide variety of recombinant proteins (Miller, 1988). Insect cells infected with a recombinant baculovirus have the potential to produce large amounts of

foreign proteins when gene expression is regulated by the very late polyhedrin promoter. This high level of production is not achieved at the expense of the function of the protein as sometimes observed with other host vector systems for instance the accumulation of protein inclusion bodies in *Escherichia coli*. Many mammalian proteins which require post-translational modifications have been successfully

Correspondence to: Dr. T. Vernet, Genetic Engineering Section, Biotechnology Research Institute, National Research Council of Canada, 5100 Royalmount Avenue, Montréal, Québec (Canada H4P 2R2) Tel. (514)496-6154; Fax (514)496-6213.

Abbreviations: aa, amino acid(s); AcNPV, *Autographa californica* nuclear polyhedrosis virus; Ap, ampicillin; bp, base pair(s); β Gal, β -galactosidase; E-64, L-trans-epoxysuccinyl-L-leucylamido(4-guanido)butane;

hpi, hours post-infection; kb, kilobase(s) or 1000 bp; *mel*, gene encoding melittin signal peptide; nt, nucleotide(s); MCS, multiple cloning site; moi, multiplicity of infection; oligo, oligodeoxyribonucleotide; *ori*, origin of DNA replication; *P*, papain-encoding gene; PAGE, polyacrylamide-gel electrophoresis; PCR, polymerase chain reaction; pre, pre-region; pro, pro-region; *proP*, propapain-encoding gene; ^R, resistance; SDS, sodium dodecyl sulfate; Sf, *Spodoptera frugiperda*; ss, single strand(ed); wt, wild type; XGal, 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside.

Expression of functional papain precursor in *Saccharomyces cerevisiae*: rapid screening of mutants

Thierry Vernet¹, Jean Chatellier, Daniel C. Tessier and David Y. Thomas

Eukaryotic Genetics Group, Biotechnology Research Institute, National Research Council of Canada, 6100 Ave. Royalmount, Montréal, Québec, Canada H4P 2R2

¹To whom correspondence should be addressed

A microbial expression system for the study of the cysteine protease papain has been developed as a more useful alternative to the insect cell/baculovirus expression system we have previously used. A synthetic papain precursor (propapain) gene was expressed in the yeast *Saccharomyces cerevisiae* under the control of the α -factor promoter. Efficient expression required fusion of the propapain sequence with the yeast α -factor prepro region and a yeast host cell defective in the synthesis of vacuolar proteases. Surprisingly, the glycosylated form of the inactive papain precursor is not secreted, but accumulates within the yeast cell. Complete conversion of the intracellular zymogen into active mature papain could be achieved *in vitro*. Purified recombinant papain produced by the yeast system has kinetic characteristics similar to those of the natural enzyme. An advantage of the yeast expression system over the baculovirus/insect cell system is that we can perform mutagenesis and screening of papain mutants very efficiently. We have set up a 'one-tube' screening procedure for the simultaneous characterization of numerous mutants of the papain precursor. Yeast cells are grown and lysed in microtiter plate wells and the released papain precursor is then activated to mature papain. This assay allows easy discrimination between proteins with close to wild type properties and proteins that are not functional. We have applied this assay to investigate the spectrum of amino acids which are tolerated at Asn175 of papain using two independently derived libraries of mutants at this position. Many amino acid substitutions at this position are not accepted; only the reintroduction of Asn restored normal function.

Key words: cysteine protease/mutants screening/protein precursor/recombinant DNA/yeast expression system

Introduction

Proteases have been extensively studied as model enzymes using combinations of kinetic measurements and chemical modifications (Hartley, 1960). More recently, manipulations of the protein structure using genetic engineering techniques have allowed detailed investigations of the enzymatic mechanism of serine proteases through the alteration of specific amino acid residues (Craik *et al.*, 1985; Bryan *et al.*, 1986; Wells and Estell, 1988).

Our own studies of cysteine proteases have so far relied on the expression and secretion of mutants of the papain precursor from insect cells infected with recombinant baculoviruses (Vernet *et al.*, 1990). We have shown that the 39 kDa glycosylated

precursor of papain is an inactive protease (zymogen) that can be activated by heat and low pH treatment *in vitro*. We have also demonstrated that this activation involves limited proteolysis through an autocatalytic mechanism that releases active mature 24 kDa papain (Vernet *et al.*, 1991). Using the baculovirus expression system and applying the known structural information (Baker and Drenth, 1987), we have made a number of amino acid replacements designed to answer questions about the enzymatic mechanism (Ménard *et al.*, 1990, 1991a,b; Khouri *et al.*, 1991) and structural properties (Vernet *et al.*, 1992) of cysteine proteases.

However, the baculovirus system is not suited to investigate the structure–function relationships of proteins when the studies involve the replacement of many amino acid residues, nor is it suitable for random mutagenesis strategies which are combined with selection or screening procedures (Reidhaar and Sauer, 1988; Copley *et al.*, 1990). Often, it is not possible to predict the impact of mutations on the structure and function of a protein and efficient experimental means to establish the spectrum of accepted side chains at various positions of the molecule are needed. The information originating from these types of stochastic approaches when combined with more rational changes, provides more insight on protein structure and function. We have searched for an alternative expression system for efficient characterization of numerous mutants of the papain precursor to enable us to apply these approaches and procedures to papain.

To avoid the complex and time-consuming *in vitro* protein refolding steps required following expression of cysteine proteases in *Escherichia coli* (Smith and Gottesman, 1989; Eakin *et al.*, 1992; Taylor *et al.*, 1992) we have selected an eukaryotic expression system more likely to produce the native, correctly folded molecule. The yeast *Saccharomyces cerevisiae* combines most of the sophisticated post-translational modifications of eukaryotic cells with the convenience of a microbial system. Many recombinant proteins have been expressed in yeast under the control of the α -factor promoter and secreted from the cell when fused to the leader peptide of the α -factor precursor (Brake, 1989).

In this paper we described the expression of the papain precursor in a vacuolar protease deficient *S. cerevisiae* mutant. We have used this expression system to develop a simple screening system for papain mutants. To test the screening system we have defined the spectrum of side chains accepted at position Asn175 of the putative active site of papain.

Materials and methods

Yeast and E. coli strains

Saccharomyces cerevisiae strains BJ3501 (MAT α , pep4::HIS3 prb1- Δ 1.6R his3- Δ 200 ura3-52 can1 gal2) (Jones, 1991), YE409 (MAT α , pep4-3 ura3-52 leu2 mnn9) (a gift from J. Ernst) and A232-34B (MAT α , ura3-52, leu2, trp1, can1) (a gift from M. Whiteway) were used in this study. *Escherichia coli* strain CJ236 (Bio-Rad, Richmond, CA) was used for the preparation of ssDNA for *in vitro* site-directed mutagenesis. The *E. coli* strain

Ligation of Single-Stranded Oligodeoxyribonucleotides by T4 RNA Ligase¹

DANIEL C. TESSIER, ROLAND BROUSSEAU, AND THIERRY VERNET

Montreal Biotechnology Research Institute, Royal Victoria Hospital, Pavillon Hersey,
687 Avenue des Pins ouest, Montreal, Quebec H3A 1A1, Canada

Received March 24, 1986

Despite its unique ability to ligate single-stranded DNA molecules, T4 RNA ligase has so far seen little use in molecular biology due to long reaction times, modest yields, and apparent inability to promote ligation of long oligodeoxyribonucleotides. We describe here a set of reaction conditions which dramatically shorten the reaction time and give reproducible 40 to 60% ligation of DNA fragments of up to 40 bases in length. These improvements open promising new fields of application to T4 RNA ligase. © 1986 Academic Press, Inc.

KEY WORDS: ligases; cloning enzymes; recombinant technology; DNA modification enzymes; DNA; oligonucleotide synthesis.

Since the initial discovery of T4 RNA ligase's ability to catalyze the joining of single-stranded DNA molecules (1), several studies exploring the mechanism of this reaction have been published (2,3). Potential applications such as the 5'-tailing of DNA with RNA (4) and the synthesis of oligodeoxyribonucleotides (5,6) have also been published. Until now the enzyme has not found widespread use, probably because the joining reactions are often extremely slow and inefficient. Overcoming these limitations would open new areas of application for this potentially very useful enzyme; in particular the assembly of synthetic genes from oligodeoxynucleotides using only one DNA strand would be of great value.

Reaction conditions for this enzyme have already been extensively studied (7,8); however, recent publications describing improved yields in DNA ligation using polyethylene glycol (9) and hexamine cobalt chloride (10) in conjunction with T4 DNA ligase prompted us to initiate similar studies on T4 RNA ligase. Our initial results made us reexamine the other components of the RNA ligase reaction mixture and attempt ligation of oligodeoxyribo-

nucleotides of a larger size than hitherto reported.

MATERIALS AND METHODS

Materials

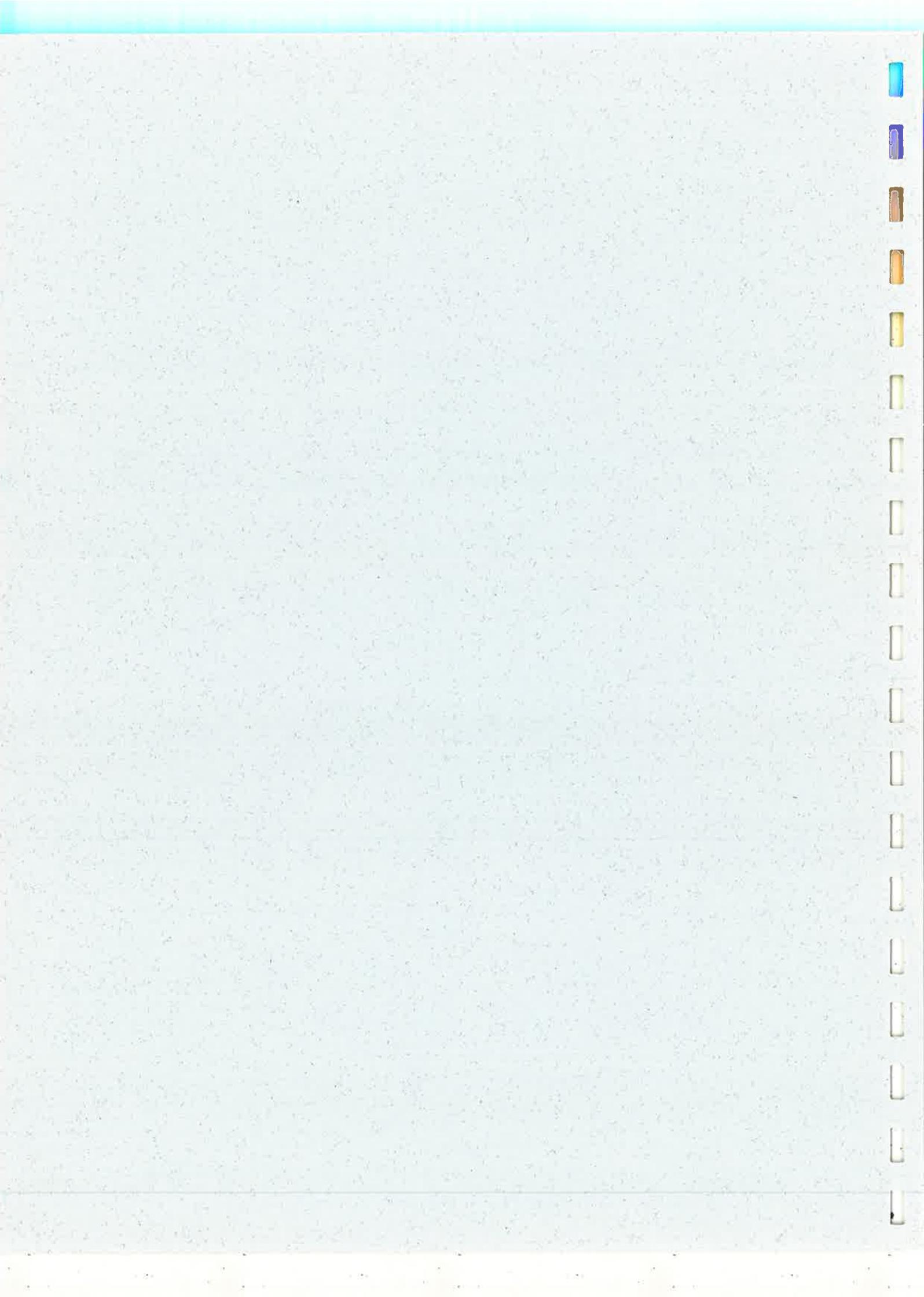
T4 RNA ligase was purchased from Pharmacia (Lot No. 881-12). The enzyme activity was determined by the supplier was 6000 units per milliliter, according to the method of Snopce *et al.* (1). Our determination of the enzyme activity on a 4-month-old sample of this lot gave a value of 5930 units per milliliter (as the average of four experiments). Our ligase assays gave reproducible results over a 3-month period thus establishing the excellent stability of this commercial preparation of enzyme.

Polynucleotide kinase and terminal deoxynucleotidyl transferase were obtained from Pharmacia. DNA polymerase I (Klenow fragment) was purchased from Bethesda Research Labs, as were the restriction enzymes *Hinf* I, *Hae* III, and *Alu* I. PEG² 8000 was obtained

² Abbreviations used: acceptor, acceptor oligodeoxyribonucleotide; donor, donor oligodeoxyribonucleotide; HCC, hexamine cobalt chloride; PEG, polyethylene glycol; BSA, bovine serum albumin; BPB, bromophenol blue; XCFE, xylene cyanol FF.

¹ Issued as NRCC publication No. 25961.

**3- Etude des systèmes toxine/immunité chez *E. coli* et chez
*Saccharomyces cerevisiae***



GENE 1030

Characterization and nucleotide sequence of a colicin-release gene in the *hic* region of plasmid ColE3-CA38*

(Recombinant DNA; lacuna; immunity; promoter; *E. coli*; lysis; induction; inverted repeat; CloDF13)

Robert J. Watson**, Peter C.K. Lau, Thierry Vernet and Louis P. Visentin

Molecular Genetics Group, Division of Biological Sciences, National Research Council of Canada, Ottawa, Ontario K1A 0R6, and **Chemistry and Biology Research Institute, Agriculture Canada, Ottawa, Ontario, K1A 0C6 (Canada) **Tel. (613) 995-3700

(Received January 24th, 1984)

(Revision received March 14th, 1984)

(Accepted March 27th, 1984)

SUMMARY

Downstream from its colicin and immunity genes (*col*, *imm*), *Escherichia coli* plasmid ColE3-CA38 contains a 0.81-kb DNA segment, the *hic* region, which is required for high colicin production. Characterization of derived plasmids, carrying the *col-imm* operon but varying in the *hic* region, showed that the latter functions in lacuna production, colicin release, cell death, and lysis. The *hic* gene expression after induction was shown to be dependent on the *col* gene promoter. The nucleotide sequence of the 0.81-kb region was determined and the *hic* gene localized to its *imm*-distal portion following an open reading frame (ORF) with no known function. There are two overlapping ORFs in that portion of the sequence, one of which was identified as the *hic* gene by its partial homology to lysis gene *H* of CloDF13. The 3' half of the *hic* gene is non-essential and contains a terminator-like DNA sequence. Preceding the gene, there are also inverted repeats which may attenuate its transcription.

INTRODUCTION

ColE3-CA38 is a 7.2-kb plasmid coding for the production of a bactericidal protein, colicin E3, and an immunity protein which protects the host cell from the lethal activity of its own colicin. Though

immunity protein is produced constitutively, colicin production occurs only after activation of the host cell "SOS" system, probably as a direct effect of RecA-mediated cleavage of the LexA protein (Horii et al., 1981; Little and Mount, 1982). Induced cells may be detected by their ability to produce clearings, known as lacuna, in a lawn of colicin-sensitive cells in top agar. Colicin induction is accompanied by increased synthesis of immunity protein (Jakes, 1982), a marked reduction of host-cell protein synthesis (Tyler and Sherratt, 1975), cell death, and release of large amounts of colicin-immunity protein complex into the media.

The inducible nature of colicin production by

* Issued as NRCC No. 23007; CBRI No. 1438.

** To whom all correspondence and reprint requests should be sent.

Abbreviations: bp, base pairs; cfu, colony-forming units; IR, inverted repeat; kb, 1000 bp; lfu, lacuna-forming unit; MC, mitomycin C; ORF, open reading frame.

Relationships of the Col Plasmids E2, E3, E4, E5, E6, and E7: Restriction Mapping and Colicin Gene Fusions¹

ROBERT J. WATSON,*² THIERRY VERNET,³ AND LOUIS P. VISENTIN

*Molecular Genetics Group, Division of Biological Sciences, National Research Council of Canada, and *Chemistry and Biology Research Institute, Agriculture Canada, Ottawa, Ontario, K1A 0C6, Canada*

Received October 9, 1984; revised February 5, 1985

Thirteen ColE plasmids representing the E2-E7 types have been compared by restriction mapping. Over 80% of their restriction sites were found to be similarly positioned, indicating that these plasmids share a common structure. Three variants are ColE2-CA42 and ColE7-K317, both of which contain 1.8-kb DNA segments in place of a 2.5-kb segment common to the other plasmids, and ColE6-CT14, which has an additional 5.0-kb DNA segment compared to the other plasmids. The colicin (*col*), immunity (*imm*), and colicin release (*hic*) genes of these plasmids have been localized to regions corresponding to those known for ColE3-CA38 and ColE2-P9, with the *imm* and *hic* genes adjacent to the 3' end of the *col* gene. Active colicin is produced from hybrid *col* genes containing 5' and 3' ends from different E-type plasmids. The 3'-termini of the fused *col* genes specify the colicin type. © 1985 Academic Press, Inc.

ColE plasmids encode bacteriocidal proteins, known as colicins, which are active against wild-type *Escherichia coli* strains but not those lacking the outer membrane protein specified by the *btuB* gene. They also encode immunity proteins which can complex with their own or closely related colicins and thus neutralize their killing effect. The specificity of the colicin/immunity protein interaction has been used to classify the ColE plasmids into types E1 to E9, such that strains carrying the same type of ColE plasmid are immune to each other's colicins. Of these, the ColE2-E7 plasmids share several characteristics which suggest they are related to each other, but not to the E1 types (1-3). However, this has been demonstrated at the molecular level only for ColE3-CA38, ColE2-P9 and ColE2-CA42 (4-6). In this communication, we describe the characterization of ten other plasmids in the E2-E7 series by restriction mapping, localization of their colicin (*col*),

immunity (*imm*), and colicin release (*hic*) genes and fusions of their *col* genes. Recently, ColE8-J has also been characterized and its physical and genetic organization found to be similar to the E2-E7 plasmid series described here (7).

Figure 1 shows the restriction maps of 10 ColE plasmids of types E2, E4, E5, E6, and E7 together with the known maps of ColE3-CA38, ColE2-P9, and ColE2-CA42. ColE2-P9 and ColE3-CA38 have many restriction sites in common due to large regions of homology (8,4). Similarly over 80% of the sites shown in Fig. 1 are common to two or more plasmids of different immunity types (Fig. 2), indicating that they comprise a family of plasmids related by extensive homologies. Comparisons of these maps also indicates regions where size differences are apparent.

Of the six ColE2 plasmids included in this study, ColE2-P9 (Fig. 1A) and ColE2-CA42 (Fig. 1C) are known to be homologous except for 1.8 and 2.5-kb segments, respectively (6). In ColE2-P9 this segment is within the 2.6-5.1-kb coordinates on its restriction map and includes a series of eight restriction sites (*Sma*I - *Hinc*II - *Eco*RI - *Ava*I - *Eco*RI - *Bgl*I -

¹ Issued as NRCC No. 24058 (CBRI No. 1493).

² To whom all correspondence and reprint requests should be addressed.

³ Present affiliation: Montreal Biotechnology Research Institute, c/o National Research Council of Canada; address as above.

Stable maintenance in chemostat-grown *Escherichia coli*
of pBR322 and pACYC184 by disruption of the
tetracycline resistance gene*

Thierry VERNET^{1**}, Ian J. McDONALD², Dave R. CAMERON³
and Louis P. VISENTIN¹

¹Molecular Genetics Section and ²Microbiology Section,
Division of Biological Sciences, National Research Council
of Canada, Ottawa, Ontario, Canada, K1A 0R6; and

³Department of Microbiology and Immunology, McGill University,
3775 University Street, Montreal, Quebec, Canada, H3A 2B4

(Received 3 December 1984)

Plasmid stability was studied in antibiotic-free chemostat cultures. Disruption, either by deletion or insertion, of the tetracycline resistance gene in the EcoRI/EcoRV region of the cloning vector pBR322 or in the HindIII/BamHI region of pACYC184 yields plasmids markedly more stable than the parent plasmids. Thus, at least for these two instances, cloning of a partitioning (*par*) locus is not prerequisite for plasmid maintenance.

The usefulness of plasmid cloning vectors could be questioned if they show inherent instability. At the laboratory level, it is generally easy to maintain a stable population of plasmid carrying cells by the means of antibiotic selection. However, selection is often difficult and expensive to apply to large scale industrial fermentations. Therefore, it is of major interest to study plasmid stability, i.e. the maintenance of plasmids in their host cells after numerous generations in the absence of selective pressure.

Continuous cell growth in chemostat culture provides an excellent way to study plasmid cloning vector stability (Jones & Melling, 1984; Wouters et al., 1980; Noack et al., 1981) since many parameters including the environmental conditions during growth and the growth rate of host cells are strictly controlled.

The identification of DNA regions involved in the maintenance of plasmids has often been realized by sub-cloning these DNA fragments into unstable plasmids and then testing for increased stability. This approach allowed the identification of the *par* locus of pSC101 (Meacock & Cohen, 1980), the *cer* determinant of pColE1 (Summer

*Issued as NRCC publication No. 23992.

**To whom reprint requests should be sent. Present affiliation: Montreal Biotechnology Research Institute, c/o National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada.

Yeast Killer Toxin: Site-Directed Mutations Implicate the Precursor Protein As the Immunity Component

Charles Boone,* Howard Bussey,* Deirdre Greene,* David Y. Thomas,† and Thierry Vernet†

* Department of Biology

McGill University

Montreal, Quebec, Canada H3A 1B1

† National Research Council of Canada

Biotechnology Research Institute

c/o Royal Victoria Hospital

Montreal, Quebec, Canada H3A 1A1

Summary

Yeast killer toxin and a component giving immunity to it are both encoded by a gene specifying a single 35 kd precursor polypeptide. This precursor is composed of a leader peptide, the α and β subunits of the secreted toxin, and a glycosylated γ peptide separating the latter. The toxin subunits are proteolytically processed from the precursor during toxin secretion. Using site-directed mutagenesis, we have identified a region of the precursor gene necessary for expression of the immunity phenotype. This immunity-coding region extends through the C-terminal half of the α subunit into the N-terminal part of the γ glycopeptide. Mutations in other parts of the gene allow full immunity but produce precursors that fail to be processed. The precursor can therefore confer immunity, and we propose that it does so in the wild type by competing with mature toxin for binding to a membrane receptor.

Introduction

K1 killer yeast cells secrete a heterodimeric, $\alpha\beta$ -subunit protein toxin. The toxin is lethal to sensitive yeasts, but cells secreting the toxin are immune to its action. Protein toxins are produced by many species, and these species must have strategies to prevent self-killing. The molecular mechanism of host immunity to toxin action is understood in outline for some bacteriocins (Weaver et al., 1981; Bishop et al., 1985), but it is at present unclear for eukaryotic toxins. Here we present an initial study of the molecular basis of immunity to killer toxin in yeast.

The genome of the K1 killer in *Saccharomyces cerevisiae* is a linear double-stranded RNA species that is encapsidated in a virus-like particle (for reviews, see Tipper and Bostian, 1984; Wickner, 1981). A cDNA clone of the killer precursor gene has been sequenced (Bostian et al., 1984; Skipper et al., 1984) and expressed in yeast (Lolle et al., 1984). The precursor has a preprotoxin structure: it consists of an N-terminal leader, followed by the two toxin subunits, α and β , which are separated by a central glycosylated γ peptide. The expression of the precursor gene cDNA on a multicopy plasmid has shown that both the immunity and toxin activities are encoded on the precursor gene.

The preprotoxin has similarities with other precursor

systems, including other toxins; examples include preproinsulin, preprovasopressin, and preproricin (Steiner et al., 1980; Richter, 1983; Lamb et al., 1985). Secretion of mature, active toxin involves cleavage of the disulfide-linked $\alpha\beta$ dimer during processing along the yeast secretory pathway (Bussey et al., 1983). We have undertaken a study of the processing, cellular localization, and action of this precursor to learn more about the cell biology of this event sequence.

An immediate aim in this work was to find, through site-directed mutagenesis, the region of the precursor involved in the immunity process. Little is known about immunity to killer toxin, but current ideas about it are derived from studies of toxin action. The initial step of the killing process involves binding of the toxin to a (1 \rightarrow 6)- β -D-glucan receptor on the yeast cell wall (Hutchins and Bussey, 1983). Beyond this first step the mechanism of toxin action is unknown; however, biochemical and genetic evidence suggest that the toxin binds to a receptor on the plasma membrane and disrupts the membrane potential by the formation of ion channels (Kagan, 1983; de la Pena et al., 1981). Immunity appears to be conferred at the plasma membrane level; toxin-producing cells bind toxin normally to the cell wall receptor, and they retain immunity as spheroplasts (Al-Aidroos and Bussey, 1978). By analogy with some polyprotein precursors containing peptides of different function, it was suggested that the γ glycoprotein, when cleaved from the precursor, acted as the immunity component by masking a toxin plasma membrane receptor (Bostian et al., 1984). However, other results have shown that there is no need to process the K1 precursor to confer immunity (Bussey et al., 1982, 1983; Wickner, 1974). Neutral mutants of the dsRNA as well as *kex* mutants, which are defective in precursor processing, fail to secrete toxin or secrete greatly reduced levels of toxin, but they do make glycosylated toxin precursor and also retain full immunity. The phenotype of these mutants suggests that the precursor can functionally express immunity, and it raises the possibility that, in the wild type, residual unprocessed precursor may also express immunity. Results presented here are inconsistent with the hypothesis that the γ glycoprotein is the immunity component. We have found that mutations in the α region of the precursor gene, resulting in a product that can be proteolytically processed and express unaltered γ glycoprotein, can perturb immunity function. An interpretation of this unanticipated finding is that immunity is conferred by the precursor, which can act as a competitive inhibitor of mature toxin, perhaps by saturating a cell membrane receptor that normally mediates toxin action.

Results

Our initial approach was to construct a set of in-frame insertion mutations throughout the precursor gene. We reasoned that such mutations would lead to local disruption of the native precursor structure and thus allow a func-

Genetic and molecular approaches to synthesis and action of the yeast killer toxin

H. Bussey, C. Boone, H. Zhu, T. Vernet, M. Whiteway and D. Y. Thomas

Dept of Biology, McGill University, Montreal, Quebec (Canada H3A 1B1), and Genetic Engineering Section, National Research Council of Canada, Biotechnology Research Institute, Montreal, Quebec (Canada H4P 2R2)

Summary. The K1 killer toxin of *Saccharomyces cerevisiae* is a secreted, virally-coded protein lethal to sensitive yeasts. Killer yeasts are immune to the toxin they produce. This killer system has been extensively examined from genetic and molecular perspectives. Here we review the biology of killer yeasts, and examine the synthesis and action of the protein toxin and the immunity component. We summarise the structure of the toxin precursor gene and its protein products, outline the proteolytic processing of the toxin subunits from the precursor, and their passage through the yeast secretory pathway. We then discuss the mode of action of the toxin, its lectin-like interaction with a cell wall glucan, and its probable role in forming channels in the yeast plasma membrane. In addition we describe models of how a toxin precursor species functions as the immunity component, probably by interfering with channel formation. We conclude with a review of the functional domains of the toxin structural gene as determined by site-directed mutagenesis. This work has identified regions associated with glucan binding, toxin activity, and immunity.

Key words. *Saccharomyces cerevisiae*; protein toxin; yeast toxin precursor; protease processing; lectin; (1 → 6)- β -D-glucan; receptor; resistant mutants; spheroplasts; ion-permeable channels; site-directed mutagenesis; toxin functional domains.

Introduction

The *Saccharomyces cerevisiae* K1 killer system is one of many described among the yeasts (see review by Young⁴⁷), although the structure and action of the protein toxins produced by most of these killers have not been examined in any detail. Because *Saccharomyces cerevisiae* is a convenient organism to use genetically, and because the killer phenotype is easy to detect, the system has been exploited to pursue the genetic basis of maintenance of the killer virus in host cells and the expression of its toxic gene product. Some comprehensive reviews cover these topics^{8, 41, 44}. The problem of toxin action has been primarily studied by genetic approaches, and this biased approach could profit from more detailed biochemical work. One purpose of this review is to convey the message that genetic and mutational dissection of the toxin structural gene is now at a point where precise biochemical and biophysical work could proceed on the pore-forming mechanism of the protein.

Killer biology

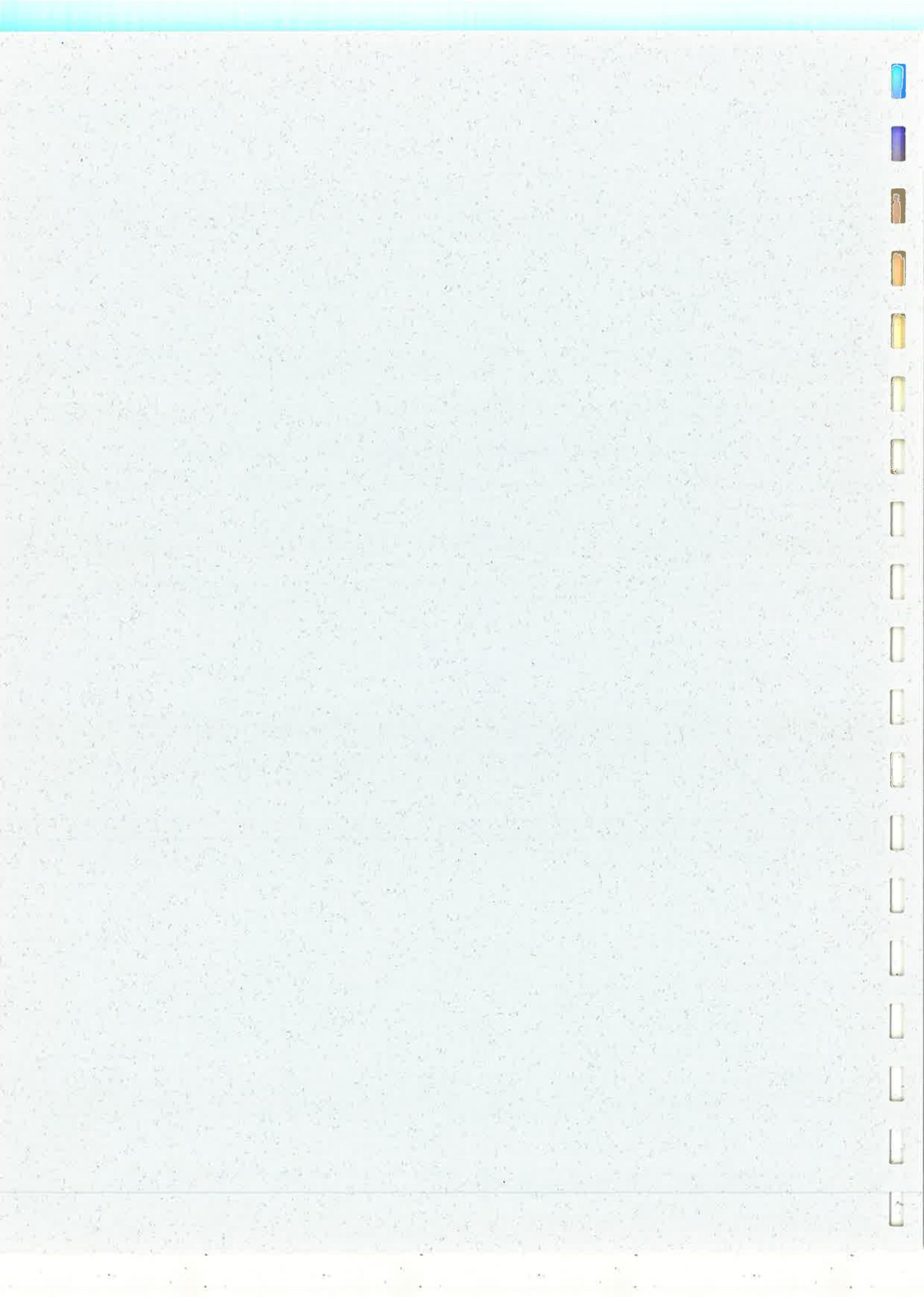
The presence of killer strains of *Saccharomyces cerevisiae* was discovered by Makower and Bevan²⁹. The K1 killer they found has been the archetype in subsequent work. Such killers contain a virus and secrete a virally-coded protein toxin that kills yeasts of the same species lacking the virus. Killer strains are not killed by the toxin they produce and are said to be immune. At least three different viruses can confer a killer phenotype in *Saccharomyces cerevisiae*. All produce different proteins, the so-called K1, K2 and KT28 toxins, and each immunity system is specific for only one toxin type. The presence of an antagonistic killer phenotype in natural yeast populations may confer a competitive advantage. Some support

for this view comes from a survey of killers among yeasts colonizing cactus fruits, where the results suggest that the killer yeasts can exclude non-killers from this habitat³⁹. Killer yeasts present a problem in the contamination of industrial fermentations³⁰ and have been shown in the laboratory to be efficient competitors with non-killers in some situations⁴⁸. The killer system can be used as a dominant selectable marker for genetic transformation and plasmid maintenance and in the latter application serves as an anticontaminant trait for yeasts used in industrial fermentations^{16, 43}.

Viral nature. The detailed virology of the killer system is beyond the scope of this article and will just be summarised here (see Wickner⁴⁴ for a review). The *Saccharomyces cerevisiae* K1 killer virus system is composed of two distinct double-stranded RNA species: L (large) and M (medium), that are separately packaged by a common major capsid protein into the virus-like particles of the mycovirus. In laboratory situations the virus is normally vertically transmitted by vegetative cell division or through sexual fusion where it behaves as a classic cytoplasmically transmitted genetic element³⁸. Recently conditions have been developed where it is possible to transmit the virus in a quasi-infectious way to yeast cells using modified DNA transformation procedures²⁶.

The killer and immunity phenotypes conferred by the virus have allowed extensive genetic analysis of nuclear genes necessary for replication and maintenance of the virions and of others that determine virion copy number. Other studies have explored the molecular biology of replication of the dsRNA genome and its packaging into capsids. In many ways this system is a paradigm for the interaction of a stably maintained virus with a eukaryotic host cell.

4- Relation structure/fonction des hydrolases



GEN 02951

The expression in *Escherichia coli* of a synthetic gene coding for the precursor of papain is prevented by its own putative signal sequence

(Recombinant DNA; β -galactosidase fusion; plasmid vector; endopeptidase; cysteine protease; insoluble granules)

Thierry Vernet, Daniel C. Tessier, France Laliberté, Daniel Dignard and David Y. Thomas

Genetic Engineering Section, Biotechnology Research Institute, National Research Council of Canada, Montréal, Québec (Canada)

Received by J. Davison: 17 October 1988

Accepted: 30 December 1988

SUMMARY

A 1048-bp gene coding for prepropapain was assembled from chemically synthesized oligodeoxyribonucleotides and cloned into a variety of *Escherichia coli* expression plasmids. We observed loss of plasmid when the preproP gene was expressed in *E. coli* either as the native precursor or fused at the C terminus of the first 592 amino acids (aa) of β -galactosidase (β Gal). Deletion of the putative 26-aa signal peptide (pre-region) increased plasmid stability. The level of maintenance for the different plasmid constructs correlated with the level of expression detected by immunoblotting. Constitutive expression of the β Gal-propapain fusion generated insoluble granules in a protease-deficient *E. coli* host. The fusion protein was easily purified to near homogeneity by differential solubilization of the granules.

INTRODUCTION

Thiol proteases constitute a large group of endopeptidases which are characterized by the presence of a cysteine residue essential for catalytic activity. This group includes the archetypical 23-kDa enzyme papain (EC 3.4.22.2) for which a large body of

enzymatic and structural data is available (for reviews see Baker and Drenth, 1987). Recombinant DNA technology provides the opportunity to study further the catalytic mechanism of papain and to increase our knowledge of the structure-function relationships within this class of proteases. Engineering of this enzyme could also lead to alteration of the

Correspondence to: Dr. T. Vernet, Genetic Engineering Section, Biotechnology Research Institute, National Research Council of Canada, 6100, Avenue Royalmount, Montréal, Québec (Canada H4P 2R2) Tel. (514)496-6154; Fax (514)496-6232.

Abbreviations: aa, amino acid(s); Ap, ampicillin; bp, base pair(s); β Gal, β -galactosidase; cDNA, DNA complementary to mRNA; Δ , deletion; EtdBr, ethidium bromide; IPTG, isopropyl-

β -D-thiogalactopyranoside; kb, kilobase(s) or 1000 bp; MCS, multiple cloning sites; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; *ori*, origin of DNA replication; *P* gene, papain-coding gene; PAGE, polyacrylamide-gel electrophoresis; pre, pre-region; pro, pro-region; ^R, resistance; SDS, sodium dodecyl sulfate; ss, single stranded; 2YT medium, see MATERIALS AND METHODS, section a; [], designates plasmid-carrier state; ::, novel joint (insertion, fusion).



Secretion of Functional Papain Precursor from Insect Cells

REQUIREMENT FOR *N*-GLYCOSYLATION OF THE PRO-REGION*

(Received for publication, January 31, 1990)

Thierry Vernet^{‡§}, Daniel C. Tessier[‡], Christopher Richardson[‡], France Laliberté[‡],
Henry E. Khouri[¶], Alexander W. Bell[¶], Andrew C. Storer[¶], and David Y. Thomas[‡]

From the [‡]Genetic Engineering Section and [¶]Protein Engineering Section, Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec H4P 2R2, Canada

The synthetic gene coding for the precursor of the cysteine protease papain (EC 3.4.22.2) has been expressed using the baculovirus/insect cell system. The prepropapain gene was cloned into the transfer vector IpDC125 behind the polyhedrin promoter. The recombinant construct was then incorporated by homologous recombination into the *Autographa californica* nuclear polyhedrosis virus genome. The host *Spodoptera frugiperda* Sf9 cells infected with the recombinant baculovirus secrete an enzymatically inactive *N*-glycosylated papain precursor. This zymogen could be activated *in vitro* to yield about 400 nmol of active papain per liter of culture. The recombinant active mature papain was enzymatically indistinguishable from natural papain but the precursor was not processed to the same amino acid residue. The insect cells also accumulated prepropapain and glycosylated propapain intracellularly. This accumulation was an indication that there are rate-limiting steps in the secretion of proteins from insect cells in this expression system. Characterization of mutants of the precursor has shown that entry into the secretory pathway and addition of carbohydrate are prerequisite conditions for the production and secretion of functional propapain.

Thiol proteases are endoproteases that share a common enzymatic mechanism which involves a thiol-imidazole contact. These proteases are found in many different animal and plant species (Sakanari *et al.*, 1989). The 24-kDa enzyme papain from *Carica papaya* (EC 3.4.22.2) is from a structural point of view the best characterized cysteine protease of plant origin. The molecular structure of papain has been refined to a resolution of 1.65 Å (Kamphuis *et al.*, 1985) and crystallographic data of enzyme-substrate complexes have also been reported (Varughese *et al.*, 1989). The enzymatic characteristics of papain have been studied up to now mostly by using chemical modification, by interaction with specific inhibitors and by spectroscopic methods (for reviews, see Brocklehurst *et al.* 1987; Baker and Drenth, 1987). Further studies of the enzymatic mechanism of papain and thiol proteases in general will be facilitated by recombinant DNA techniques.

The amino acid sequence of the active enzyme has been

* Issued as National Research Council of Canada publication No. 32383. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Genetic Engineering Section, Biotechnology Research Inst., NRCC, 6100 Royalmount Ave., Montreal, Quebec H4P 2R2, Canada. Tel.: 514-496-6154; Fax: 514-496-6232.

known for some time. Analysis of the sequence of a cDNA from *C. papaya* coding for the papain gene has shown that the protease is synthesized as a precursor with an NH₂-terminal extension of 133 amino acids (Cohen *et al.*, 1986). This precursor form is a feature that papain has in common with all the other protease genes so far isolated. We chose to express the precursor of papain rather than the mature form of the enzyme for two reasons. First, there is evidence that the precursor region of other proteases is necessary to produce a functional molecule through its participation in the folding of the mature protein (Silen *et al.*, 1989; Zhu *et al.*, 1989). Second, we are interested in assessing the role of the precursor region (prepro region) in the targeting of the protein in the cell, in the correct folding of the enzyme, and in the inactivation of the enzyme. In addition, this information is of interest since there is, as yet, no report on the characterization of the papain precursor from plants.

Our initial attempts to express a synthetic gene coding for the papain precursor failed because of the toxic effect of the hydrophobic signal sequence in *Escherichia coli* (Vernet *et al.*, 1989). We selected the insect cell-baculovirus expression system as an alternative to *E. coli*, since this system has demonstrated a remarkable capacity to produce a variety of different functional proteins including enzymes and storage proteins from plants (Andrews *et al.*, 1988; Bustos *et al.*, 1988).

The characterization of the papain precursor enables us to define the steps involved in intracellular processing of this enzyme in insect cells. We have also shown that processing of the signal peptide of prepropapain and subsequent step(s) in the transit of propapain through the insect cell secretory pathway are rate limiting. These findings have implication for the production of other proteins in this transient expression system.

EXPERIMENTAL PROCEDURES

Construction of Recombinant Baculoviruses—The 1048-base pair synthetic gene coding for prepropapain (Vernet *et al.*, 1989) was cloned by blunt-end ligation into the *Bam*HI site of IpDC125 (Vialard *et al.*, 1990) to create plasmid IpDC127. These plasmids are derivatives of the transfer vector pAc373 (Smith *et al.*, 1985) which have been modified for rapid sequencing and facile *in vitro* mutagenesis. Single-stranded IpDC127 DNA was used as template for *in vitro* site-directed mutagenesis (Kunkel, 1985). A *Hind*III-*Bam*HI 973-base pair DNA fragment from plasmid pWRI^a-proP (Vernet *et al.*, 1989) containing the papain precursor from which the signal peptide has been deleted (propapain) was inserted into the *Bam*HI site of IpDC125, yielding plasmid IpDC123. The two target sites for *N*-glycosylation within the proregion of papain were removed simultaneously by site-directed mutagenesis using the following oligonucleotides: 5'-GGTTCTATTGCCGGCCAGTACACTACAAC-3' and 5'-CCAAC-AAGAAACAGAAATTCTTACAGGCTAG-3' yielding the plasmid IpDC179. The entire sequence of the modified gene was confirmed (Sanger *et al.*, 1977). Purified double-stranded DNA from each of

Processing of the Papain Precursor

PURIFICATION OF THE ZYMOGEN AND CHARACTERIZATION OF ITS MECHANISM OF PROCESSING*

(Received for publication, May 10, 1991)

Thierry Vernet^{‡§}, Henry E. Khouri^{¶||}, Pierre Laflamme[¶], Daniel C. Tessier[‡], Roy Musil[‡],
Barbara J. Gour-Salin[¶], Andrew C. Storer[¶], and David Y. Thomas[§]

From the [‡]Genetic Engineering Section and [¶]Protein Engineering Section, Biotechnology Research Institute, National Research Council of Canada, Montréal, Québec H4P 2R2, Canada

The precursor of the cysteine protease papain has been expressed and secreted as propapain from insect cells infected with a recombinant baculovirus expressing a synthetic gene coding for prepropapain. This 39-kDa secreted propapain zymogen molecule is glycosylated and can be processed *in vitro* into an enzymatically active authentic papain molecule of 24.5 kDa (Vernet, T., Tessier, D. C., Richardson, C., Laliberté, F., Khouri, H. E., Bell, A. W., Storer, A. C., and Thomas, D. Y. (1990) *J. Biol. Chem.* 265, 16661-16666). Recombinant propapain was stabilized with Hg²⁺ and purified to homogeneity using affinity chromatography, gel filtration, and ion-exchange chromatographic procedures. The maximum rate of processing *in vitro* was achieved at approximately pH 4.0, at a temperature of 65 °C and under reducing conditions. Precursor processing is inhibited by a variety of reversible and irreversible cysteine protease inhibitors but not by specific inhibitors of serine, metallo or acid proteases. Replacement by site-directed mutagenesis of the active site cysteine with a serine at position 25 also prevents processing. The inhibitor ¹²⁵I-N-(2S,3S)-3-*trans*-hydroxycarbonyloxiran-2-carbonyl-L-tyrosine benzyl ester covalently labeled the wild type papain precursor, but not the C25S mutant, indicating that the active site is accessible to the inhibitor and is in a native conformation within the precursor. Based on biochemical and kinetic analyses of the activation and processing of propapain we have shown that the papain precursor is capable of autoproteolytic cleavage (intramolecular). Once free papain is released processing can then occur *in trans* (intermolecular).

proteases (Kassell and Kay, 1973).

Of the major classes of proteases, the best characterized precursors are trypsinogen of the serine protease family (Kunitz and Northrop, 1936) and pepsinogen (Herriott, 1938) of the aspartyl protease family (for a review see Desnuelle (1986)). The former is activated following limited proteolysis by other proteases, whereas the activation of the latter is by an autocatalytic mechanism, being intramolecular below pH 3 and mainly intermolecular at higher pH values (Al-Janabi *et al.*, 1972). The structural basis of inactivation-activation is known for these precursors. The three-dimensional structure of pepsinogen has been elucidated and has revealed that the N-terminal 44 residues of the pro region interact with the protease active site via salt bridges which are disrupted at low pH (James and Sielecki, 1986). A similar type of interaction between the activation segment (pro region) and the active site has been recently reported for procarboxypeptidase B (Coll *et al.*, 1991). Precursors of cysteine proteases are comparatively poorly understood and no three-dimensional structures of cysteine protease precursors have been reported so far. Autocatalytic conversion of precursors of cysteine proteases including those of cathepsin B (Felleisen and Klinkert, 1990) and cathepsin L (Smith and Gottesman, 1989) have been reported. Processing of the precursor of the vacuolar plant cysteine protease aleurain, on the other hand, requires the activity of another cysteine protease (Holwerda *et al.*, 1990).

The canonical member of the cysteine protease family, papain (EC 3.4.22.2) is a 24-kDa plant endoprotease which has a wide substrate specificity and which contains an essential thiol group in its active site. Despite extensive structural and enzymatic characterization of mature papain (for a review see Brocklehurst *et al.*, 1987; Baker and Drenth, 1987) very little is known about its precursor form. Cloning of a cDNA encoding papain has revealed that the enzyme is synthesized as a precursor that includes a 133-amino acid N-terminal extension¹ (Cohen *et al.*, 1986). Such a high molecular weight precursor form of papain has not been observed in *Carica papaya* latex despite earlier reports of the presence of a catalytically inactive form of papain (Brocklehurst and Kierstan, 1973; Kierstan *et al.*, 1982). This, in addition to the difficulties in expressing the functional recombinant molecule in *Escherichia coli* (Cohen *et al.*, 1990; Vernet *et al.*, 1989), has impaired further biochemical characterization of the precursor. We have recently overcome this obstacle by expressing the functional papain precursor using the baculovirus/insect cell system. The initial translation product of the prepropa-

Proteolytic enzymes are ubiquitous in biological systems. Their physiological roles range from protein degradation within specialized cellular compartments to precise cleavage of protein precursors (Neurath, 1984). Proteases are often synthesized as higher molecular weight precursors, zymogens or proproteases. There are a variety of mechanisms by which different protease precursors can be activated to functional

* This is publication 32781 of the National Research Council of Canada. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Genetic Engineering Section, Biotechnology Research Inst., NRCC, 6100 Royalmount Ave., Montréal, Québec H4P 2R2, Canada. Tel.: 514-496-6154; Fax: 514-496-6213.

¶ Present address: Merck Frosst Canada Inc., 16711 Trans-Canada Hwy., Kirkland, Québec, H9H 3L1, Canada.

¹ The numbering of amino acid residues of the precursor of papain has been described previously (Vernet *et al.*, 1990); the pro region amino acids are designated with negative numbers starting at the pro region papain junction (Asn⁻¹/Ile¹).

A Protein Engineering Study of the Role of Aspartate 158 in the Catalytic Mechanism of Papain[†]

Robert Ménard,[†] Henry E. Khouri,[†] Céline Plouffe,[†] Robert Dupras,[†] Daniel Ripoll,[†] Thierry Vernet,[§]
Daniel C. Tessier,[§] France Laliberté,[§] David Y. Thomas,[§] and Andrew C. Storer^{*.‡}

*Protein Engineering and Genetic Engineering Sections, Biotechnology Research Institute, National Research Council of Canada,
6100 Royalmount Avenue, Montréal, Québec H4P 2R2, Canada*

Received October 30, 1989; Revised Manuscript Received February 16, 1990

ABSTRACT: The controversy concerning the various suggested roles for the side chain of Asp158 in the active site of papain has been clarified by using site-directed mutagenesis. Both wild-type papain and an Asp158Asn variant were produced in a baculovirus-insect cell expression system, purified to homogeneity from the culture, and characterized kinetically. With CBZ-Phe-Arg-MCA as substrate, the k_{cat}/K_M and k_{cat} values obtained for the Asp158Asn papain are $20\,000\text{ M}^{-1}\cdot\text{s}^{-1}$ and 34 s^{-1} , respectively, as compared with values of $120\,000\text{ M}^{-1}\cdot\text{s}^{-1}$ and 51 s^{-1} obtained for the wild-type papain. In addition, the pH- (k_{cat}/K_M) profile for the Asp158Asn enzyme is shifted relative to that for the wild-type enzyme to lower values by approximately 0.3 pH unit. This shows clearly that Asp158 is not, as previously postulated, an essential catalytic residue. In addition, the pH dependency data are interpreted to indicate that, contrary to earlier suggestions, the negatively charged side chain of Asp158 does not significantly stabilize the active-site thiolate-imidazolium ion pair. However, its presence does influence the $\text{p}K_a$'s associated with ion-pair formation in a manner compatible with electrostatic considerations.

Since papain was first described in 1879 (Wurtz & Bouchut, 1879), a large database of information on the enzyme has been gathered. In particular, evidence has accumulated for the formation of an ion pair at neutral pH between residues Cys25 and His159, and it has been proposed that this is the active form of the enzyme [e.g., Polgar (1974), Sluyterman and

Wijdenes (1976), Halasz and Polgar (1977), Lewis et al. (1981), and Migliorini and Creighton (1986)]. For an in-depth review of cysteine proteases in general and papain in particular, see Brocklehurst et al. (1987). Mechanistically and kinetically, papain has been extensively characterized, and the structure of the native enzyme has been determined at 1.65-Å resolution (Kamphuis et al., 1984). In addition, the structures of several papain-inhibitor complexes are also available (Drenth et al., 1976; Varughese et al., 1989). These facts together with the recent construction of a synthetic gene coding for the papain precursor (Vernet et al., 1989) and its expression in a bacu-

[†] NRCC Publication No. 31445.

* Author to whom correspondence should be addressed.

[‡] Protein Engineering Section.

[§] Genetic Engineering Section.

Importance of Hydrogen-Bonding Interactions Involving the Side Chain of Asp158 in the Catalytic Mechanism of Papain[†]

Robert Ménard,[†] Henry E. Khouri,[†] Céline Plouffe,[†] Pierre Laflamme,[†] Robert Dupras,[†] Thierry Vernet,[‡]
Daniel C. Tessier,[§] David Y. Thomas,[§] and Andrew C. Storer^{*†}

*Protein Engineering and Genetic Engineering Sections, Biotechnology Research Institute, National Research Council of Canada,
6100 Avenue Royalmount, Montréal, Québec H4P 2R2, Canada*

Received January 15, 1991; Revised Manuscript Received March 15, 1991

ABSTRACT: In a previous study, it was shown that replacing Asp158 in papain by Asn had little effect on activity and that the negatively charged carboxylate of Asp158 does not significantly stabilize the active site thiolate-imidazolium ion pair of papain (Ménard et al., 1990). In this paper, we report the kinetic characterization of three more mutants at this position: Asp158Gly, Asp158Ala, and Asp158Glu. From the pH-activity profiles of these and other mutants of papain, it has been possible to develop a model that enables us to dissect out the contribution of the various mutations toward (i) intrinsic activity, (ii) ion pair stability, and (iii) the electrostatic potential at the active site. Results obtained with mutants that place either Gly or Ala at position 158 indicate that the hydrogen bonds involving the side chain of Asp158 in wild-type papain are indirectly important for enzyme activity. When CBZ-Phe-Arg-MCA is used as a substrate, the $(k_{\text{cat}}/K_{\text{M}})_{\text{obs}}$ values at pH 6.5 are 3650 and 494 $\text{M}^{-1} \text{s}^{-1}$ for Asp158Gly and Asp158Ala, respectively, as compared to 119 000 $\text{M}^{-1} \text{s}^{-1}$ for papain. Results with the Asp158Glu mutant suggest that the side chain of Glu moves closer to the active site and cannot form hydrogen bonds similar to those involving Asp158 in papain. From the four mutations introduced at position 158 in papain, we can conclude that it is not the charge but the hydrogen-bonding interactions involving the side chain of Asp158 that contribute the most to the stabilization of the thiolate-imidazolium ion pair in papain. However, the charge and the hydrogen bonds of Asp158 both contribute to the intrinsic activity of the enzyme.

An understanding of the mode of action of cysteine proteases is of interest for many reasons. Cysteine proteases are members of a class of enzymes that are widely distributed among living organisms and participate in a number of physiological processes in plants, bacteria, and animals (Brocklehurst et al., 1987; Baker & Drenth, 1987). There is a great deal of evidence that implicates cysteine proteases in various disease states. In particular, lysosomal cathepsins are believed to be

involved in several pathological conditions such as muscular dystrophy, heart diseases, inflammatory diseases, and tumor invasiveness (Kar & Pearson, 1977; Poole et al., 1978, 1980; Mort et al., 1984; Sloane & Honn, 1984; Rich, 1986). These enzymes form a family of homologous proteins with comparable structures (Kamphuis et al., 1985) and functions, and the information obtained from a given enzyme can be very useful in understanding the behavior of related cysteine proteases. The most widely studied member of this group is the plant protease papain, and it is taken as the prototypic member of this class of enzymes.

The structure of papain has been solved at 1.65-Å resolution and is available (Kamphuis et al., 1984), and the general

[†]NRCC Publication No. 32763.

* Author to whom correspondence should be addressed.

[†]Protein Engineering Section.

[‡]Genetic Engineering Section.

Contribution of the Glutamine 19 Side Chain to Transition-State Stabilization in the Oxyanion Hole of Papain†

Robert Ménard,† Julie Carrière,‡§ Pierre Laflamme,† Céline Plouffe,† Henri E. Khouri,‡|| Thierry Vernet,‡ Daniel C. Tessier,‡ David Y. Thomas,‡ and Andrew C. Storer*·‡

Protein Engineering and Genetic Engineering Sections, Biotechnology Research Institute, National Research Council of Canada, 6100 Avenue Royalmount, Montréal, Québec, H4P 2R2 Canada

Received June 3, 1991; Revised Manuscript Received July 30, 1991

ABSTRACT: The existence of an oxyanion hole in cysteine proteases able to stabilize a transition-state complex in a manner analogous to that found with serine proteases has been the object of controversy for many years. In papain, the side chain of Gln19 forms one of the hydrogen-bond donors in the putative oxyanion hole, and its contribution to transition-state stabilization has been evaluated by site-directed mutagenesis. Mutation of Gln19 to Ala caused a decrease in k_{cat}/K_M for hydrolysis of CBZ-Phe-Arg-MCA, which is $7700 \text{ M}^{-1} \text{ s}^{-1}$ in the mutant enzyme as compared to $464000 \text{ M}^{-1} \text{ s}^{-1}$ in wild-type papain. With a Gln19Ser variant, the activity is even lower, with a k_{cat}/K_M value of $760 \text{ M}^{-1} \text{ s}^{-1}$. The 60- and 600-fold decreases in k_{cat}/K_M correspond to changes in free energy of catalysis of 2.4 and 3.8 kcal/mol for Gln19Ala and Gln19Ser, respectively. In both cases, the decrease in activity is in large part attributable to a decrease in k_{cat} , while K_M values are only slightly affected. These results indicate that the oxyanion hole is operational in the papain-catalyzed hydrolysis of CBZ-Phe-Arg-MCA and constitute the first direct evidence of a mechanistic requirement for oxyanion stabilization in the transition state of reactions catalyzed by cysteine proteases. The equilibrium constants K_i for inhibition of the papain mutants by the aldehyde Ac-Phe-Gly-CHO have also been determined. Contrary to the results with the substrate, mutation at position 19 of papain has a very small effect on binding of the inhibitor. This important finding indicates that different types of inhibitors and substrates utilize the catalytic machinery of papain to differing degrees. Implications for the role of oxyanion hole interactions in cysteine proteases are discussed.

As suggested by Pauling in the 1940s, enzymes can cause appreciable rate accelerations by preferentially binding the transition-state complex formed during the reaction (Pauling, 1946, 1948). Serine and cysteine protease catalyzed hydrolysis involves a transition-state complex that is inherently unstable due to the development of a negative charge on the carbonyl oxygen of the substrate. In serine proteases, stabilization of such a transition state is achieved by an arrangement of dipoles in a cavity that has been termed the oxyanion hole (Henderson, 1970; Henderson et al., 1971; Robertus et al., 1972; Matthews et al., 1975; Poulos et al., 1976; Asbóth & Polgár, 1983; Bryan et al., 1986; Wells et al., 1986; Carter & Wells, 1990). However, the existence of such a cavity in cysteine proteases remains controversial (Wolthers et al., 1971; Asbóth & Polgár, 1983; Asbóth et al., 1985; Mackenzie et al., 1986). Examination of X-ray structures of papain-inhibitor complexes (Drenth et al., 1975, 1976) indicate that the Gln19 side chain and the Cys25 amide proton are positioned to interact with the oxyanion of a papain-substrate transition state. On the other hand, the ability of papain to hydrolyze thionoesters has been considered as evidence against the contribution of oxyanion hole interactions in the catalytic mechanism of papain

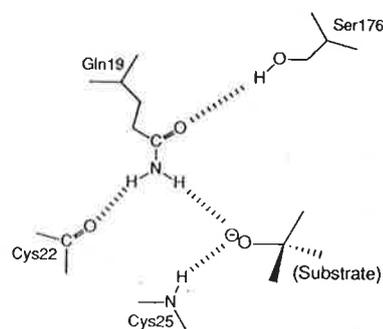


FIGURE 1: Schematic representation of the interactions involving the side chain of Gln19 and the amide proton of Cys25 in the oxyanion hole of papain.

(Asbóth & Polgár, 1983; Asbóth et al., 1985). Mackenzie et al. (1986) have also proposed that interactions between a tetrahedral intermediate and the oxyanion hole may not be possible in papain. Using ^{13}C NMR spectroscopy, they have shown that both *N*-acetyl-D- and *N*-acetyl-L-phenylalanyl-glycinals form thiohemiacetals with the same stereospecificity in the S_1 subsite of papain, and from computer modeling work, they have suggested that the hydroxyl function is not located in the oxyanion hole of papain.

Quantitative evaluation of transition-state stabilization in an enzyme-catalyzed reaction is an important aspect of enzymology. The advent of site-directed mutagenesis, however, has rendered such studies more accessible and allows structure-function correlations to be established. We have used this technique to investigate the contribution of interactions

† NRCC Publication No. 32776.

* Author to whom correspondence should be addressed.

‡ Protein Engineering Section.

§ Present address: Syntex Inc., 2100 Syntex Ct., Mississauga, Ontario, L5N 3X4 Canada.

|| Present address: Merck Frosst Canada Inc., 16711 Trans-Canada Hwy., Kirkland, Québec, H9H 3L1 Canada.

‡ Genetic Engineering Section.

Removal of an inter-domain hydrogen bond through site-directed mutagenesis: role of serine 176 in the mechanism of papain

Robert Ménard, Céline Plouffe, Henry E. Khouri,
Robert Dupras, Daniel C. Tessier¹, Thierry Vernet¹,
David Y. Thomas¹ and Andrew C. Storer²

Protein Engineering and ¹Genetic Engineering Sections, Biotechnology Research Institute, National Research Council Canada, 6100 Royalmount Avenue, Montréal, Canada, H4P 2R2

²To whom correspondence should be addressed

A mutant of papain, where an inter-domain hydrogen bond between the side chain hydroxyl group of a serine residue at position 176 and the side chain carbonyl oxygen of a glutamine residue at position 19 has been removed by site-directed mutagenesis, has been produced and characterized kinetically. The mutation of Ser176 to an alanine has only a small effect on the kinetic parameters, the k_{cat}/K_m for hydrolysis of CBZ-Phe-Arg-MCA by the Ser176Ala enzyme being of 8.1×10^4 /M/s compared with 1.2×10^5 /M/s for papain. Serine 176 is therefore not essential for the catalytic functioning of papain, even though this residue is conserved in all cysteine proteases sequenced. The pH–activity profiles were shown to be narrower in the mutant enzyme by up to 1 pH unit at high ionic strength. This result is interpreted to indicate that replacing Ser176 by an alanine destabilizes the thiolate–imidazolium form of the catalytic site Cys25–His159 residues of papain. Possible explanations for that effect are given and the role of a serine residue at position 176 in papain is discussed.

Key words: mechanism/mutagenesis/papain/protein engineering

Introduction

Papain is the best known member of a series of homologous cysteine proteases. It is a globular protein consisting of a single polypeptide chain of 212 residues, folded to form two domains with a deep cleft between them (Drenth *et al.*, 1968; Kamphuis *et al.*, 1984). The active site consists of a cysteine residue (Cys25) and an histidine residue (His159), believed to exist as a thiolate–imidazolium ion pair (Polgar, 1974; Lewis *et al.*, 1976; Sluyterman and Wijdenes, 1976; Lewis *et al.*, 1981). Many other residues are considered to be of importance for the proper functioning of the enzyme based on comparison of amino acid sequences between members of this group of cysteine proteases and on examination of available X-ray structures for papain, actinidin and calotropin DI (Baker, 1980; Kamphuis *et al.*, 1984, 1985; Baker and Drenth, 1987; Brocklehurst *et al.*, 1987; Hilgenfeld, 1987). Such residues include Asp158, Asn175, Gln19 and Ser176 in papain. Recently, Ménard *et al.* (1990) have shown, using site-directed mutagenesis, that Asp158 is not essential for catalytic activity in papain. The presence of the negative charge of the Asp158 side chain in the vicinity of the active site merely affects the pK_a s associated with ion pair formation through electrostatic interactions. The asparagine at position 175 is hydrogen bonded to the active site histidine to

form a 'catalytic triad': Cys25–His159–Asn175. Glutamine 19 is believed to assist catalysis by interacting with the oxyanion of a tetrahedral intermediate formed during catalysis in a manner similar to serine proteases (Robertus *et al.*, 1972; Drenth *et al.*, 1975, 1976; Kraut, 1977). However, the fact that cysteine proteases like papain have been shown to hydrolyze specific thionoester substrates at rates comparable to oxygen esters has put some doubt on the existence of such an interaction for cysteine proteases (Asboth and Polgar, 1983; Asboth *et al.*, 1985).

Very little information on the role of the serine residue at position 176 in papain is available and very seldom is it mentioned in mechanistic considerations, even though it is conserved in all reported cysteine proteases sequences. The hydroxyl group of Ser176 is hydrogen bonded to the side chain carbonyl oxygen of Gln19 and this might have an orientational effect on the putative Gln19–substrate interaction in the oxyanion hole. This Ser176–Gln19 hydrogen bond also bridges the two domains of papain and since the active site Cys25 and His159 are located on different walls of the cleft, such an inter-domain interaction could play a significant role in catalysis by controlling active site geometry and mobility. To study the role of Ser176 in papain and in analogous cysteine proteases, the Ser176–Gln19 hydrogen bond linking the two domains of papain has been removed by replacing the serine residue by an alanine. This paper reports the kinetic characterization of the Ser176Ala mutant. The role of Ser176 and of inter-domain interactions in general in the mechanism of papain is discussed.

Materials and methods

Papain from Sigma Chemical Co. ($\times 2$ crystallized suspension in sodium acetate) was further purified by the method of Sluyterman and Wijdenes (1970) using a mercurial agarose column. The mercurial enzyme was converted to active papain by addition of β -mercaptoethanol followed by gel filtration on Sephadex G-15. DTNB (5,5'-dithiobis-2-nitrobenzoic acid) was purchased from Sigma Chemical Co., the inhibitor E-64 (1-[(L-trans-epoxysuccinyl)-L-leucyl]amino]-4-guanidino butane) and the substrate CBZ-Phe-Arg-MCA (carbobenzoxy-L-phenylalanyl-L-arginine-4-methylcoumarinyl-7-amide) were from IAF Biochem International Inc, Laval Québec. Thiopropyl Sepharose was obtained from Pharmacia.

Mutagenesis

A synthetic gene coding for the precursor of papain (Vernet *et al.*, 1989) has been inserted into the transfer vector IpDC125 (Vialard *et al.*, 1990). *In vitro* site-directed mutagenesis was carried out using uracil-containing ssDNA as template (Kunkel, 1985) and the following 23 mer oligonucleotide as primer: 5'-TGATCAAGAACGCGTGGGGTACC-3'. The resulting IpDC141 mutated plasmid was easily identified due to the creation of a *Mlu*I restriction site within the altered region of the gene. To ensure accuracy of the *in vitro* DNA repair during the mutagenesis, the entire papain precursor gene was resequenced.

Reprinted from *Biochemistry*, 1991, 30.
Copyright © 1991 by the American Chemical Society and reprinted by permission of the copyright owner.

Engineering of Papain: Selective Alteration of Substrate Specificity by Site-Directed Mutagenesis[†]

Henry E. Khouri,^{‡§} Thierry Vernet,^{||} Robert Ménard,[‡] Francesco Parlati,[‡] Pierre Laflamme,[‡] Daniel C. Tessier,^{||} Barbara Gour-Salin,[‡] David Y. Thomas,^{||} and Andrew C. Storer^{*;‡}

Protein Engineering and Genetic Engineering Sections, Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Avenue, Montréal, Québec H4P 2R2, Canada

Received April 8, 1991; Revised Manuscript Received June 25, 1991

ABSTRACT: The S₂ subsite specificity of the plant protease papain has been altered to resemble that of mammalian cathepsin B by site-directed mutagenesis. On the basis of amino acid sequence alignments for papain and cathepsin B, a double mutant (Val133Ala/Ser205Glu) was produced where Val133 and Ser205 are replaced by Ala and Glu, respectively, as well as a triple mutant (Val133Ala/Val157Gly/Ser205Glu), where Val157 is also replaced by Gly. Three synthetic substrates were used for the kinetic characterization of the mutants, as well as wild-type papain and cathepsin B: CBZ-Phe-Arg-MCA, CBZ-Arg-Arg-MCA, and CBZ-Cit-Arg-MCA. The ratio of k_{cat}/K_M obtained by using CBZ-Phe-Arg-MCA as substrate over that obtained with CBZ-Arg-Arg-MCA is 8.0 for the Val133Ala/Ser205Glu variant, while the equivalent values for wild-type papain and cathepsin B are 904 and 3.6, respectively. This change in specificity has been achieved by replacing only two amino acids out of a total of 212 in papain and with little loss in overall enzyme activity. However, further replacement of Val157 by Gly as in Val133Ala/Val157Gly/Ser205Glu causes an important decrease in activity, although the enzyme still displays a cathepsin B like substrate specificity. In addition, the pH dependence of activity for the Val133Ala/Ser205Glu variant compares well with that of cathepsin B. In particular, the activity toward CBZ-Arg-Arg-MCA is modulated by a group with a pK_a of 5.51, a behavior that is also encountered in the case of cathepsin B but is absent with papain. Results of this study suggest that sequence alignment of cysteine proteases coupled with the structural information that is available for papain can be used to achieve a better understanding of the molecular mechanism and specificity of structurally and functionally related cysteine proteases.

Papain is a cysteine protease isolated from the latex of tropical papaya fruit (*Carica papaya*). This enzyme has been extensively characterized (Baker & Drenth, 1987; Brocklehurst et al., 1987) and its three-dimensional structure has been determined at high resolution (Kamphuis et al., 1984). Cathepsin B, on the other hand, is a cysteine protease found in the tissues of many animals, including human liver (Barrett, 1977), for which no crystal structure is presently available. However, the complete or partial amino acid sequences of cathepsin B from many sources are available, and comparison of the aligned sequences for papain and cathepsin B suggests that both proteases have a similar folding pattern and three-dimensional structure (Kamphuis et al., 1985; Dufour, 1988). Furthermore, the mechanistic data gathered so far for cathepsin B suggests that it shares a common basic enzymatic mechanism with papain. Features of this mechanism include the formation of a thiolate-imidazolium ion pair involving the side chains of the catalytic-site cysteine and histidine residues (Polgár, 1974; Lewis et al., 1976, 1981) and the presence of an acyl-enzyme intermediate in the catalytic pathway (Bajkowski & Frankfater, 1983).

From comparisons of the available cysteine protease structures and sequences, it can be seen that groups required

for interacting with the backbone of the substrates are provided by the backbone atoms of the enzyme and are therefore conserved. However, differences exist between the regions of the enzyme that bind the substrates' side chains. If papain and cathepsin B do possess very similar three-dimensional structures, then from these differences it should be possible to predict the features that lead to their different substrate specificities. Schechter and Berger (1967) have shown that the active site of papain can be considered to consist of seven subsites (S₁-S₄ and S₁'-S₃'), each able to accommodate one amino acid residue of a substrate (P₁-P₄ and P₁'-P₃', respectively). The nature of the residue at position P₂ has been shown to be the most significant in terms of determining specificity, and it is well established that papain exhibits a strong catalytic preference for amino acids with a bulky nonpolar side chain (e.g., Phe, Tyr) at this position (Berger & Schechter, 1970). Examination of the crystal structures of chloromethyl ketone derivatives of papain published by Drenth et al. (1976) allows identification of the residues in the S₂ subsite whose side chains make the most intimate contacts with the P₂ side chain of the substrate. These residues are Pro68,¹ Val133, Val157, Asp158, and Ala160; they form a hydrophobic cavity in papain (Table I.) The corresponding residues in cathepsin B, obtained from the sequence alignment, remain hydrophobic and the nonpolar character of the S₂ subsite is therefore conserved. Accordingly, it has been shown that substrates with an aromatic group (i.e., Phe) at P₂ are

[†] NRCC Publication No. 32775.

* Author to whom correspondence should be addressed.

[‡] Protein Engineering Section.

[§] Present address: Merck Frosst Canada Inc., 16711 Trans-Canada Hwy., Kirkland, Québec H9H 3L1, Canada.

^{||} Genetic Engineering Section.

¹ The amino acid numbering of papain is used throughout the text.

Functional Expression of Human Cathepsin S in *Saccharomyces cerevisiae*

PURIFICATION AND CHARACTERIZATION OF THE RECOMBINANT ENZYME*

(Received for publication, July 8, 1992)

Dieter Brömme‡, Pierre R. Bonneau§, Paule Lachance, Bernd Wiederanders¶, Heidrun Kirschke¶, Christoph Peters||, David Y. Thomas, Andrew C. Storer, and Thierry Vernet

From the Molecular Biology Sector, Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec H4P 2R2, Canada, the ¶Institute of Biochemistry, Martin-Luther-University, Halle (Saale), D-O-4020, Germany, and the ||Institute of Biochemistry II, Georg-August-University, Göttingen D-W-3400, Germany

A cDNA encoding the human lysosomal cysteine proteinase cathepsin S precursor has been expressed in yeast using the pVT100-U expression vector containing the α -factor promoter. The procathepsin S gene was expressed either as a fusion protein with the pre-region or with the prepro-region of the yeast α -factor precursor gene. Following *in vitro* processing both constructs gave an identical active mature enzyme with a molecular weight of 24,000. After prolonged cultivation of the cells the recombinant protein is also found as an active proteinase in the culture supernatant. The precursor can be activated *in vitro* at pH 4.5 and 40 °C under reducing conditions. The *in vitro* activated enzyme has a 6-amino acid NH₂-terminal extension when compared with the native bovine enzyme.

The purified enzyme displays a bell-shaped pH activity profile with a pH optimum of 6.5 and pK values of 4.5 and 7.8. The isoelectric point of the recombinant human cathepsin S is between 8.3 and 8.6 and about 1.5 pH units higher than for the bovine enzyme. The kinetic data for several synthetic substrates and inhibitors reveal a preference for smaller amino acid residues in the binding subsites S₂ and S₃ of cathepsin S. Like the bovine enzyme, the recombinant human cathepsin S is characterized by a broader range of pH stability (pH 5–7.5) than cathepsins B and L.

Lysosomal cysteine proteinases (cathepsin B, EC 3.4.22.1; cathepsin L, EC 3.4.22.15; cathepsin H, EC 3.4.22.16) are considered to be involved both in the physiological protein breakdown (1) and in pathological degradative and invasive processes (2–5). Normally organisms are protected against accidental release or mistargeting of lysosomal cysteine proteinases by protein inhibitors of the cystatin superfamily (6) which are found both intra- and extracellularly. Further protection is afforded by the instability of most of these cysteine proteinases at neutral pH. In contrast to the cathepsins B and L, the more recently characterized cathepsin S (EC 3.4.22.27) is both stable and highly active at pH 7.0 (7) and

can be regarded as a new potential participant in proteinase related disorders.

Cathepsin S is a lysosomally located, single-chain, cysteine proteinase with high endopeptidase activity against proteins including elastin and collagen (7, 8). Its substrate specificity shows some similarities with cathepsin L but it is clearly different in its S₂ subsite specificity (9). The primary structures of bovine and human cathepsin S have been reported (10–12); they are 85% identical and show similarities with cathepsin L.

The expression of the cDNAs of mammalian proteolytic enzymes in microorganisms is a useful method to obtain sufficient quantities of these enzymes and to manipulate their sequences for studies of their mechanism and structure. To date expression has been obtained for the following cysteine proteinases: papain (13), cathepsin B (14), and cathepsin L (15). Here we report the expression of a functional human cathepsin S in *Saccharomyces cerevisiae* and the characterization of the recombinant enzyme.

EXPERIMENTAL PROCEDURES

Materials—Restriction endonucleases were obtained from New England Biolabs; DNA polymerase (Klenow fragment) from Bethesda Research Laboratories; T4 DNA polymerase, T4 DNA ligase, and thiopropyl-Sepharose 6B were purchased from Pharmacia LKB Biotechnology Inc. Oligonucleotides were synthesized using an Applied Biosystems DNA synthesizer. Endoglycosidase F/glycosidase F was purchased from Boehringer Mannheim and used according to the instructions of the manufacturer.

Z¹-VVR-MCA, Z-FR-MCA, Boc-FFR-MCA, Z-LLR-MCA, Z-FVR-MCA, and Z-VVK-NHO-Nbz² were synthesized using standard methods in peptide chemistry as reported elsewhere (9). Z-FA-CHN₂ was purchased from Enzyme System Products. Rat cathepsins L and B were purified as described previously (16, 17).

Plasmid Construction—The yeast expression vector used, Yp-DC222 (39), is based on a modification of the shuttle vector pVT100-U (18), which contains the α -factor promoter and the part of the gene which codes for the prepro- α -factor. The entire cDNA (11) coding for human cathepsin S (GenBank™ with accession number M90696) including a nontranslated part of the 3' end (1.2-kb *EcoRI*/blunt ended *XbaI* fragment) was cloned into an *EcoRI* and blunt-ended *BamHI* site of the vector to create plasmid YpCS1. By site-directed mutagenesis (19) the coding sequence of pro-cathepsin S was brought into the reading frame of the α -factor prepro-sequence to create

* Issued as National Research Council of Canada Publication No. 33660. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Molecular Biology Sector Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Ave., Montreal, Quebec H4P 2R2, Canada. Tel.: 514-496-6169; Fax: 514-496-5143.

§ Bio-Méga Industrial Research Fellow.

¹ The abbreviations used are: Z-, benzyloxycarbonyl; Boc-, *t*-butyloxycarbonyl; -MCA, 4-methyl-7-coumarylamide; -NBz, 4-nitrobenzoyl; -NHO- is the hydroxylamine moiety; E-64, L-3-carboxy-*trans*-2,3-epoxypropionyl-leucylamido-(4-guanidino)butane; PAGE, polyacrylamide gel electrophoresis. For the discussion of the interactions between proteinase and substrate the nomenclature of Schechter and Berger (38) was used.

² D. Brömme, U. Neumann, H. Kirschke, and H. U. Demuth, in preparation.

Coordinated amino acid changes in homologous protein families*

D. Altschuh, T. Vernet¹, P. Berti¹, D. Moras and K. Nagai²

Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue Descartes, 67084 Strasbourg Cédex, France, ¹Conseil National de Recherches Canada, Institut de Recherche en Biotechnologie, 6100 Avenue Royalmount, Montréal, Québec H4P 2R2, Canada and ²Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, UK

*Issued as NRCC BRI no. 26327

In the tobamovirus coat protein family, amino acid residues at some spatially close positions are found to be substituted in a coordinated manner [Altschuh *et al.* (1987) *J. Mol. Biol.*, 193, 693]. Therefore, these positions show an identical pattern of amino acid substitutions when amino acid sequences of these homologous proteins are aligned. Based on this principle, coordinated substitutions have been searched for in three additional protein families: serine proteases, cysteine proteases and the haemoglobins. Coordinated changes have been found in all three protein families mostly within structurally constrained regions. This method works with a varying degree of success depending on the function of the proteins, the range of sequence similarities and the number of sequences considered. By relaxing the criteria for residue selection, the method was adapted to cover a broader range of protein families and to study regions of the proteins having weaker structural constraints. The information derived by these methods provides a general guide for engineering of a large variety of proteins to analyse structure–function relationships.

Key words: protein structure/sequence similarities/protein engineering/proteases/haemoglobins

Introduction

Within the serine protease, cysteine protease and globin families, the tertiary structure is remarkably similar despite considerable variation in the amino acid sequences. Conserved residues or conservative changes alone cannot account for structural conservation. Thus the structures must be stabilized in different ways in related proteins. Knowing which alternative residues produce a given local structure would improve our understanding of protein folding and residue interactions. A modified function of a member of an enzyme family may arise from a small number of substitutions. The ability to predict which particular residues among a large number of substitutions determine functional differences would be particularly useful for protein engineering.

A recent analysis of tobamovirus coat proteins showed that some amino acid replacements in these related proteins are coordinated. Positions in the protein sequence that are replaced together, and therefore show the same replacement pattern in related sequences, are mostly in spatial proximity in the crystallographic structure of the coat protein disk of tobacco mosaic virus *vulgare*. This strongly suggests that mutations are not stabilized independently of each other in the tobamovirus

protein family (Altschuh *et al.*, 1987), and that these coordinated mutations may play important roles in stabilizing protein structure and/or in defining functional diversity.

A simple method of finding coordinated changes in homologous sequences and relating these changes with structure has previously been described (Altschuh *et al.*, 1987). In order to investigate whether this method can be applied successfully to protein families other than tobamoviruses, we analysed three additional families with various degrees of sequence similarities, and for which several sequences and at least one crystallographic structure are available: serine proteases (11 sequences with similarities from 22 to 44%), cysteine proteases (12 sequences with similarities ranging from 25 to 95%) and haemoglobin beta chain (12 sequences with similarities ranging from 40 to 97%).

Coordinated amino acid substitutions were found in all three protein families. Optimal conditions for detecting such changes are highly constrained structures and low sequence similarities. A modification of the original method is described for an extended analysis of the pattern of substitutions in less-constrained structures. Results are illustrated by describing some groups of residues that are substituted together in the two protease and the haemoglobin families.

Materials and methods

Source of sequences

Serine proteases. The sequences used were the same as those aligned by Greer (1981): chymotrypsin, trypsin, elastase, haemoglobin heavy chain, kallikrein, factor IX_a (Christmas factor), factor X_a (Stuart factor), plasmin B chain, group-specific protease, thrombin B chain and bacterial trypsin.

Cysteine proteases. The twelve cysteine protease sequences used in this study are: papain (Cohen *et al.*, 1986), actinidin (Praekelt *et al.*, 1988), *Dictyostelium discoideum* cysteine proteases 1 and 2 (Pears *et al.*, 1985), mouse cathepsin L (Troen *et al.*, 1987), rat cathepsin L (Ishidoh *et al.*, 1987a), chicken cathepsin L (Wada *et al.*, 1987), rat cathepsin H (Ishidoh *et al.*, 1987b), aleurain (Whittier *et al.*, 1987) and human, mouse and rat cathepsins B (Chan *et al.*, 1986).

Haemoglobins. Twelve haemoglobin beta chain sequences were selected among various animal groups: human, bovine, dog, rabbit, chicken, duck, starling, ostrich, carp, goldfish, crocodile and alligator. These sequences were extracted from the NBRF data bank where all references can be found.

Source of structures

The crystallographic structures used for this study are from the Brookhaven data bank. They are papain (Kamphuis *et al.*, 1984), chymotrypsin A (Birktoft and Blow, 1972) and human deoxy haemoglobin beta chain (Fermi *et al.*, 1984).

Alignment procedures

Since the purpose of this paper is methodological, the alignments are not presented here. Alignments and structure–function

Correlation of Co-ordinated Amino Acid Changes at the Two-domain Interface of Cysteine Proteases with Protein Stability†

Thierry Vernet, Daniel C. Tessier, Henry E. Khouri‡

*Biotechnology Research Institute, National Research Council of Canada
6100, Avenue Royalmount, Montréal, Québec, H4P 2R2, Canada*

and Danièle Altschuh

*Institut de Biologie Moléculaire et Cellulaire, Centre National de la Recherche Scientifique
15 rue Descartes, 67084 Strasbourg, Cedex, France*

(Received 1 August 1991; accepted 4 December 1991)

The engineering of a protein containing an alternative local residue packing for a set of side-chains has proven to be a major challenge because compositional, volumetric and steric constraints must be respected. Homologous proteins should provide examples of alternative groups of residues leading to a similar functional result. The functional significance of a pair of co-ordinated changes that are observed in the cysteine proteases family has been investigated by comparing the effect of individual or double changes on secretion, stability and activity of papain. The two changes are not independent. Detrimental effects of single mutations at one of the two positions can be partly suppressed by the co-ordinated mutation that reproduces naturally occurring contacts, indicating that these changes are concerted. Single mutations at the other position produce milder effects, suggesting a pathway for evolution.

Keywords: cysteine proteases; sequence similarities; packing; protein engineering; structure-function relationship

1. Introduction

Minor structural changes have been found to affect measurable properties of proteins such as catalytic activity or stability (for reviews, see Knowles, 1987; Alber, 1989). On the other hand related proteins show large sequence differences, while retaining structural and functional integrity. This apparent contradiction may be resolved if amino acid replacements in natural proteins are concerted and follow a directed pathway (Lim & Sauer, 1989), so that their acceptability is constrained by the local atomic environment. Homologous proteins are a reliable source of information for acceptable changes (Altschuh *et al.*, 1987; Nagai *et al.*, 1988; Malcolm *et al.*, 1990). To investigate the relationship between groups of changes observed in protein families and their individual acceptability, we have replaced an inter-

domain contact of papain by alternative residues observed in the cysteine proteases family.

In the crystallographic structure of papain (Kamphuis *et al.*, 1984) Val32 and Ala162 are buried in the protein and are in van der Waals' contact (average distance 3.16 Å between C^{γ2} of Val32 and C^β of Ala162; 1 Å = 0.1 nm) across the interface of the two domains of the protein. No other side-chain is found within 3.8 Å of these residue's side-chains. They are both located about 9 Å away from the active site residues Cys25 and His159 (Fig. 1(a)). Positions 32 and 162 are not conserved in the cysteine proteases family, but they are replaced in a non-random manner (Altschuh *et al.*, 1988; Fig. 1(b)). Three pairs of alternative contacts found in nine cysteine proteases conserve a hydrophobic character and a total number of four carbon atoms for the two side-chains: Val32/Ala162, Ala32/Val162 and Gly32/Leu162. In the fourth alternative pair found in cathepsins B, both positions are replaced by charged residues (Glu32/Arg162). These replacements will not be considered in the present study because cathepsins B show large variations in the

† Published as NRC publication number 32804.

‡ Present address: Merck Frosst (Canada Inc., 16711 Trans-Canada Hwy, Kirkland, Québec, H9H 3L1, Canada.

Expression of the *Saccharomyces cerevisiae* Kex2p endoprotease in insect cells

Evidence for a carboxy-terminal autoprocessing event

Doris GERMAIN^{1,2}, Thierry VERNET¹, Guy BOILEAU² and David Y. THOMAS¹

¹ Eukaryotic Genetics Group, National Research Council of Canada, Biotechnology Research Institute, Montréal, Canada

² Département de biochimie, Université de Montréal, Montréal, Canada

(Received August 9, 1991) – EJB 91 1079

The pheromone-processing Kex2p endoprotease of *Saccharomyces cerevisiae* has been difficult to characterize due to its low level of expression in yeast cells. To overcome this problem, we have overexpressed Kex2p using the baculovirus/insect cell expression system. *Spodoptera frugiperda* Sf9 insect cells infected with a recombinant baculovirus, containing the complete *KEX2* gene which encodes the Kex2p protease (814 amino acids), accumulate an 120-kDa functional form of the enzyme. The inhibition profile of the insect-cell-derived endoprotease is similar to that of the yeast enzyme. The recombinant infected insect cells also secrete into the medium about half of the total Kex2p activity produced. Deleting the carboxyl-terminal tail and the transmembrane domain of Kex2p (Kex2Δp, 666 amino acids) does not measurably interfere with the enzyme characteristics and results in the secretion of up to 90% of the total enzyme activity. The truncated form, Kex2Δp, of the endoprotease accumulates in the cell supernatant to 6.7×10^5 U/l. The molecular mass of the secreted forms for both the wild-type Kex2p and Kex2Δp is the same (70 kDa) and is 50-kDa lower than the intracellular form. This result implicates a processing event which gives rise to shorter extracellular forms of both the wild-type Kex2p and Kex2Δp and which trims their carboxy termini upstream of amino acid 666. This processing event requires the integrity of the Ser385 of the Kex2p active site.

In mammals, most active peptide hormones or neuropeptides are released from precursor proteins of higher molecular mass (Cohen et al., 1989; Hutton, 1990). The specific endoproteases involved in the processing of these precursors have been difficult to isolate and characterize biochemically, primarily due to their low concentrations in the cell. In the yeast *Saccharomyces cerevisiae*, maturation of the α -factor pheromone precursor and of the K1 and K2 killer-toxin precursors requires the endoprotease Kex2p (Julius et al., 1984; Dignard et al., 1991). The *KEX2* gene has been cloned (Julius et al., 1984) and its primary structure determined (Mizuno et al., 1988). The protein structure deduced from the 2442-nucleotide coding sequence of the *KEX2* gene reveals a type-III transmembrane protein with a highly charged carboxy terminus. The protein contains sites for potential *N*-glycosylation and five Lys–Arg or Arg–Arg amino acid pairs which, by analogy with prohormone precursors, are potential proteolytic processing sites (Cohen et al., 1989; Hutton, 1990). Sequence comparisons (Mizuno et al., 1988) and the spectrum of sensitivity to specific inhibitors (Julius et al., 1984) indicate that the Kex2p enzyme belongs to the subtilisin family of serine proteases. Partial purification of the Kex2p enzyme from solubilized membranes of yeast overexpressing the endoprotease, has been reported. Kex2p was shown to be a Ca^{2+} -dependent endoprotease that cleaved at the carboxy

side of pairs of basic amino acids such as Lys–Arg or Arg–Arg (Fuller et al., 1989). Co-transfection of the yeast *KEX2* gene and of proopiomelanocortin (POMC) cDNA into mammalian cells that cannot normally process the POMC precursor showed that the yeast enzyme could process the POMC precursor in a cellular environment (Thomas et al., 1988; Zollinger et al., 1990; Germain et al., 1990). These recent reports have engendered an interest in further biochemical characterization of the Kex2p endoprotease. However, further characterization, including accurate determination of the structure of the active form of the protease and elucidation of the three-dimensional structure of Kex2p, requires a more abundant source of the protease.

In this study, we have used the Sf9 insect cell/baculovirus expression system to express the complete Kex2p protease and a truncated version lacking the transmembrane domain and the cytoplasmic tail (Kex2Δp). Insect cells have the potential to generate complex post-translational modifications of foreign proteins and large amounts of a variety of functional proteins have been successfully produced with this system (for a review, see Luckow and Summers, 1988). Our results indicate that both the full-length and the truncated form of the enzyme are active. Furthermore, we show that a post-translational maturation event occurs at the carboxy-terminal region of the Kex2p protein and that this event is dependent on the integrity of its active site.

Correspondence to D. Germain, Eukaryotic Genetics Group, National Research Council of Canada, Biotechnology Research Institute, Montréal, Québec, Canada H4P 2R2

Abbreviations. POMC, proopiomelanocortin; TPCK, L-1-tosyl-2-phenylethyl chloromethane; MCA, 7-amino-4-methylcoumarin; MNPV *Autographa californica* nuclear polyhedrosis virus.

EXPERIMENTAL PROCEDURES

Construction of the Ac[KEX2] recombinant virus

In order to express the Kex2p protease in the Sf9 insect cell/baculovirus system, we first cloned the coding region of

The pro-region of the Kex2 endoprotease of *Saccharomyces cerevisiae* is removed by self-processing

Doris Germain^{a,c}, France Dumas^a, Thierry Vernet^a, Yves Bourbonnais^a, David Y. Thomas^{a,b} and Guy Boileau^c

^aEukaryotic Genetics Group, National Research Council of Canada Biotechnology Research Institute, Montréal, H4P 2R2, Canada,

^bBiology Department, McGill University, Montréal, H3A 1B1, Canada and ^cDépartement de Biochimie, Université de Montréal Montréal, H3C 3J7, Canada

Received 24 January 1992; revised version received 10 February 1992

We have produced in the baculovirus/insect cells expression system a soluble secreted form of the *Saccharomyces cerevisiae* Kex2 endoprotease. This secreted enzyme was purified and its NH₂-terminal sequence determined. The NH₂-terminal sequence started at residue Leu¹⁰⁹ of the sequence deduced from the KEX2 gene nucleotide sequence, showing that the Kex2 enzyme is produced as a proenzyme. Residue Leu¹⁰⁹ is preceded by a pair of basic amino acid residues (Lys¹⁰⁷-Arg¹⁰⁸) which is a potential processing site for the Kex2 endopeptidase. Furthermore, expression of an inactive form of this truncated enzyme resulted in the production of a protein with a higher molecular weight. These observations suggest that the pro-region of Kex2 endoprotease is removed by a self-processing event.

Endoprotease; Prohormone processing; Insect cell; Protein sequence

1. INTRODUCTION

The yeast Kex2 endoprotease (Kex2p) is an integral membrane protein involved in the proteolytic processing of the α -pheromone precursor and of the K1 and K2 killer toxin precursors [1–3]. Cleavage of precursor proteins by Kex2p occurs on the COOH-terminal side of Arg residues which are found in pairs of basic amino acid residues such as Lys-Arg or Arg-Arg [1]. Doublets of basic amino acid residues are also preferred proteolytic processing sites in mammalian prohormone or pro-neuropeptide precursors [4,5]. Co-transfection of the KEX2 gene and of the proopiomelanocortin cDNA into mammalian cells, that cannot normally process prohormones, showed that the yeast Kex2 enzyme could process this precursor in a cellular environment and release authentic maturation products such as β -lipotropin and β -endorphin [6–8]. Thus, Kex2p can be used as a model for mammalian maturation endoproteases.

The Kex2 protein structure deduced from the 2848 nucleotide long coding sequence of the KEX2 gene [9] revealed a type I transmembrane protein with a luminal directed NH₂-terminus and a highly charged COOH-terminal cytoplasmic region. The protein also contains sites for *N*-glycosylation and five pairs of basic amino acid residues Lys-Arg or Arg-Arg. A Ser/Thr-rich

region which may be the site of *O*-glycosylation precedes the transmembrane domain. Sequence comparisons [9] and the spectrum of sensitivity to specific inhibitors [1] indicated that the Kex2 enzyme belongs to the subtilisin family of serine proteases.

In yeast, several proteases are synthesized as inactive precursors. Similarly, subtilisin contains an NH₂-terminal pro-region which is removed by an autoprocessing event [10]. Since two putative Lys-Arg dibasic cleavage sites are located in the NH₂-terminal region of Kex2p [9], we have undertaken the present work to investigate the possibility of NH₂-terminal maturation of the Kex2 endoprotease.

2. EXPERIMENTAL

2.1. Production of truncated forms of Kex2p in insect cells

Production of a truncated version of Kex2p (Kex2 Δ p) in insect cells has been described elsewhere [11]. We used the KEX2 transfer plasmid KEX2 Δ pDC [11] to substitute by site-directed mutagenesis the active site Ser³⁶⁵ residue with an Ala residue in order to generate an inactive form of Kex2 Δ p. Both KEX2 Δ pDC and the newly obtained KEX2 Δ 385pDC plasmids were transferred by homologous recombination into the AcMNPV viral genome (*Autographa californica* nuclear polyhedrosis virus) as described elsewhere [11,12]. The recombinant baculoviruses were respectively termed Ac[KEX2 Δ] and Ac[KEX2 Δ 385].

For purification of Kex2 Δ p, Ac[KEX2 Δ] baculovirus at a multiplicity of infection of 2 was used to infect a 4 l bioreactor of Sf9 cells grown to a density of 1.5×10^6 cells/ml in SF900 serum-free medium (Gibco). The supernatant was recovered 72 h after infection by centrifugation of the cells at $5,000 \times g$ for 10 min at 4°C and concentrated 16-fold by ultrafiltration using an Amicon Diaflo S1Y10 spiral cartridge. Protein concentration was determined by the method of Brad-

Correspondence address: D. Germain, Eukaryotic Genetics Group, National Research Council of Canada Biotechnology Research Institute, Montréal, H4P 2R2, Canada.

Cloning and Expression of *Geotrichum candidum* Lipase II Gene in Yeast

PROBING OF THE ENZYME ACTIVE SITE BY SITE-DIRECTED MUTAGENESIS*

(Received for publication, May 6, 1993, and in revised form, July 15, 1993)

Thierry Vernet‡, Edmund Ziomek, Achim Recktenwald, Joseph D. Schrag, Chantal de Montigny, Daniel C. Tessier, David Y. Thomas, and Miroslaw Cygler‡

From the Biotechnology Research Institute, National Research Council of Canada, 6100 Avenue Royalmount, Montréal, Québec H4P 2R2, Canada

The three-dimensional structure of lipase H of *Geotrichum candidum* strain ATCC34614 (GCL II) has provided insights with respect to the nature of the catalytic machinery of lipases. To support these structural observations, we have carried out an analysis of GCL II by mutagenesis. The gene encoding lipase II of *Geotrichum candidum* strain ATCC34614 (GCL II) was amplified using the polymerase chain reaction, cloned, and sequenced. The intronless lipase gene was expressed and secreted from *Saccharomyces cerevisiae* at approximately 5 mg/liter of culture. Recombinant GCL II was purified by immunoaffinity chromatography and characterized using a combination of substrates and independent analytical methods. The recombinant enzyme and the enzyme isolated from its natural source have comparable specific activities against triolein of about 1000 μmol of oleic acid released/min/mg of protein. The putative catalytic triad Ser²¹⁷-His⁴⁶³-Glu³⁵⁴ was probed by site-directed mutagenesis. The substitution of Ser²¹⁷ by either Cys or Thr and of His⁴⁶³ by Ala led to a complete elimination of the activity against both triolein and tributyrin. Substitution of Glu³⁵⁴ by either Ser, Ala or Gln renders the enzyme inactive and also perturbs the enzyme stability. However, the enzyme with the conservative replacement Glu³⁵⁴Asp is stable and displays only a small decrease of triolein activity but a 10-fold decrease in activity against tributyrin. There was no appreciable difference in esterase activity between the native, recombinant wild type, and Glu³⁵⁴Asp mutant. These results confirm that the triad formed by Ser²¹⁷-Glu³⁵⁴-His⁴⁶³ is essential for catalytic activity. They also show that the active site of GCL II is more tolerant to a conservative change of the carboxylic side chain within the triad than are other hydrolases with similar catalytic triads.

small, water-soluble esters. The hydrolytic activity of lipases is greatly enhanced by the presence of an interface, a phenomenon known as the interfacial activation (Desnuelle, 1972). This activation was long ago proposed to be caused by a conformational change (rearrangement) in the enzyme and that notion was confirmed when the first 3-D¹ structures of lipases were elucidated (Winkler *et al.*, 1990; Brady *et al.*, 1990; Schrag *et al.*, 1991). They showed that in the absence of the interface the active site is covered by parts of the polypeptide chain and is not accessible for the substrate. Rearrangement of some loops has to occur to open the access to the active site (Brzozowski *et al.*, 1991; Grochulski *et al.*, 1993).

The reaction mechanism of lipases has been suspected to be similar to that of serine proteases (Chapus and Sémériva, 1976), and the active site was thought to be a Ser-His-Asp catalytic triad. The 3-D structure of all lipases supported this notion as all three lipases which have been crystallized contain such a catalytic triad. Although the presence of Asp has been observed in the catalytic triad of HPL and Rhizomucor mieher lipase, the structure of GCL II revealed a Ser-His-Glu triad arrangement (Schrag *et al.*, 1991, 1993). This amino acid triad is common to a large family of esterases and lipases (Cygler *et al.*, 1993). The other members of this family for which the 3-D structure has been determined are AChE (Sussman *et al.*, 1991) and very recently *Candida cylindracea* lipase (Grochulski *et al.*, 1993). The 3-D structures confirmed the presence of the Ser-His-Glu configuration. Mutations of AChEs and butyrylcholinesterases from various organisms have been studied extensively (Gibney *et al.*, 1990; Neville *et al.*, 1992a, 1992b; Shafferman *et al.*, 1992; Fournier *et al.*, 1992) confirming the assignment of the active site and probing the substrate binding gorge. The mutation of the active site Ser and Glu in human AChE resulted in a total loss of activity toward acetylcholine (Schafferman *et al.*, 1992). A conservative replacement of Glu³⁵⁴ by Asp also led to a loss of biological activity, indicating the fine tuning of this enzyme structure for the glutamate containing triad. This is of special interest since some of the enzymes of this homologous family, notably cholesterol esterases, were predicted to have the traditional Ser-His-Asp catalytic triad (Cygler *et al.*, 1993). The identification by site-directed mutagenesis of Ser¹⁹⁴ (DiPersio *et al.*,

Lipases are triacylglycerol hydrolases (EC 3.1.1.3) whose substrates are insoluble lipids. These enzymes act at the water-lipid interface and show rather low activity toward

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

This work was issued as National Research Council of Canada publication No. 36152.

‡ To whom correspondence should be addressed: Biotechnology Research Institute, National Research Council of Canada, 6100 Ave. Royalmount, Montréal, Québec, H4P 2R2, Canada. Tel.: 514-496-6154; Fax: 514-496-6213; e-mail: Vernet@BRI.NRC.CA.

¹ The abbreviations used are: 3-D, three-dimensional; AChE, acetylcholinesterase; DO, diolein; GCL, *Geotrichum candidum* lipase; HL, hepatic lipase; HPL, human pancreatic lipase; MO, monoolein; PNPA, *p*-nitrophenyl-acetate; PNPB, *p*-nitrophenyl-butyrate; OA, oleic acid; oligo, oligodeoxyribonucleotide; TB, tributyrin; TO, triolein; WT, wild type; MES, 4-morpholinepropanesulfonic acid; HPLC, high performance liquid chromatography; ELISA, enzyme-linked immunosorbent assay.

Polymorphism in the lipase genes of *Geotrichum candidum* strains

¹ Maria Célia BERTOLINI, Louise LARAMÉE, David Y. THOMAS, Mirosław CYGLER, Joseph D. SCHRAG and Thierry VERNET
Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec, Canada

(Received September 7, 1993) – EJB 93 1355/1

The fungus *Geotrichum candidum* produces extracellular lipases. Purification and characterization of different lipase isoforms from various *G. candidum* strains is difficult due to the close physical and biochemical properties of the isoforms. Consequently, the characterization of these enzymes and their substrate specificities has been difficult. We have determined the lipase genes present in four strains of *G. candidum* (ATCC 34614, NRCC 205002, NRRL Y-552 and NRRL Y-553) by molecular cloning and DNA sequencing. Each strain contains two genes similar to the previously identified lipase I and lipase II cDNAs. Our data suggest that no other related lipase genes are present in these strains. Each lipase-gene family shows sequence variation (polymorphism) that is confirmed by Southern-blot analysis. This polymorphism and the sequence differences between lipase I and lipase II have been localized within the previously determined three-dimensional structure of lipase II. Although most of the amino acid substitutions are located on the protein surface, some are present in structural features possibly involved in determining substrate specificity.

Lipases (glycerol ester hydrolases) are rather unusual amongst soluble enzymes in that they exhibit increased activity at a water/lipid interface (Brockman, 1984), a phenomenon called interfacial activation. This activation process involves large movement of one surface loop (flap) which is sometimes associated with movements of other loops (Brzozowski et al., 1991; Grochulski et al., 1993; van Tilbeurgh et al., 1993). A great need to understand the molecular basis of lipase catalysis and substrate selectivity arises from their high industrial potential.

The lipase produced by the fungus *Geotrichum candidum* (GC) was one of the first lipases investigated and many reports on its production, isolation, activity and selectivity appeared over the years (for review see Charton, 1991). Crude enzyme preparations from different GC strains exhibited varying degrees of specificity for the kind and position of the fatty acids hydrolyzed from triacylglycerols and fatty esters. For example, using methyl and butyl esters of oleic and palmitic acids to examine the specificity of crude enzyme preparations from a series of GC strains, strains ATCC 34614 and NRRL Y-552 were reported to produce lipases which did not discriminate between 18:1 and 16:0 fatty acid esters. However, lipases from other strains (including NRRL Y-553) showed a preference for 18:1 methyl ester (Baillargeon et al., 1989; Baillargeon and Sonnet, 1991). That GC produces multiple isozymes was realized early, but their separation proved to be difficult. An added complication came from the fact that some of the observed heterogeneity was due to differences in glycosylation. Deglycosylation with endoglycosidases reduces the number of isoforms (Baillargeon,

1990; Baillargeon and McCarthy, 1991; Sidebottom et al., 1991; Jacobsen and Poulsen, 1992). The reports on substrate selectivities of apparently purified lipase isoforms published in recent years are contradictory. Shimada et al. (1989, 1990) isolated and sequenced two lipase cDNAs from GC strain ATCC 34614 and later purified the two isozymes, termed lipases I and II (Sugihara et al., 1990). They reported no significant differences in the substrate selectivities, although the two forms had slightly different stabilities and biochemical properties. Sidebottom et al. (1991) showed, in contrast, that two lipase isozymes from the strain CMICC 335426 (termed lipases A and B) had markedly different substrate specificities. While both isoforms displayed preference for long chain fatty acids, lipase B was, in addition, quite selective for unsaturated substrates with a *cis* double bond at position 9. These two isozymes could be distinguished by immunopurified antisera apparently specific for each type, although unpurified antiserum recognized both types, demonstrating that they share some epitopes (Charton et al., 1992). These investigators also isolated two isozymes from the ATCC 34614 strain which, in their hands, showed a pattern of selectivities similar to those of the isoforms isolated from the strain CMICC 335426. Furthermore, comparison of specificity profiles and partial peptide sequences revealed that lipase A is similar to lipase II and lipase B to lipase I (Charton, 1991). These contradictory results regarding the substrate selectivity of various isozymes may originate from the different growth conditions used by different investigators, as growth conditions are known to alter the relative yield of different lipases from GC (Charton et al., 1992). Alternatively, the presence of multiple differentially expressed lipase genes coding for products having similar chromatographic properties cannot be excluded. This has been observed in another organism, *Candida rugosa*, which has at least five genes which code for lipases that are closely related to those of GC (Lotti et al., 1993).

Correspondence to M. C. Bertolini, Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec H4P 2R2, Canada

Abbreviations. GC, *Geotrichum candidum*; PCR, polymerase chain reaction.

Enzyme. Glycerol ester hydrolase (lipase) (EC 3.1.1.3).

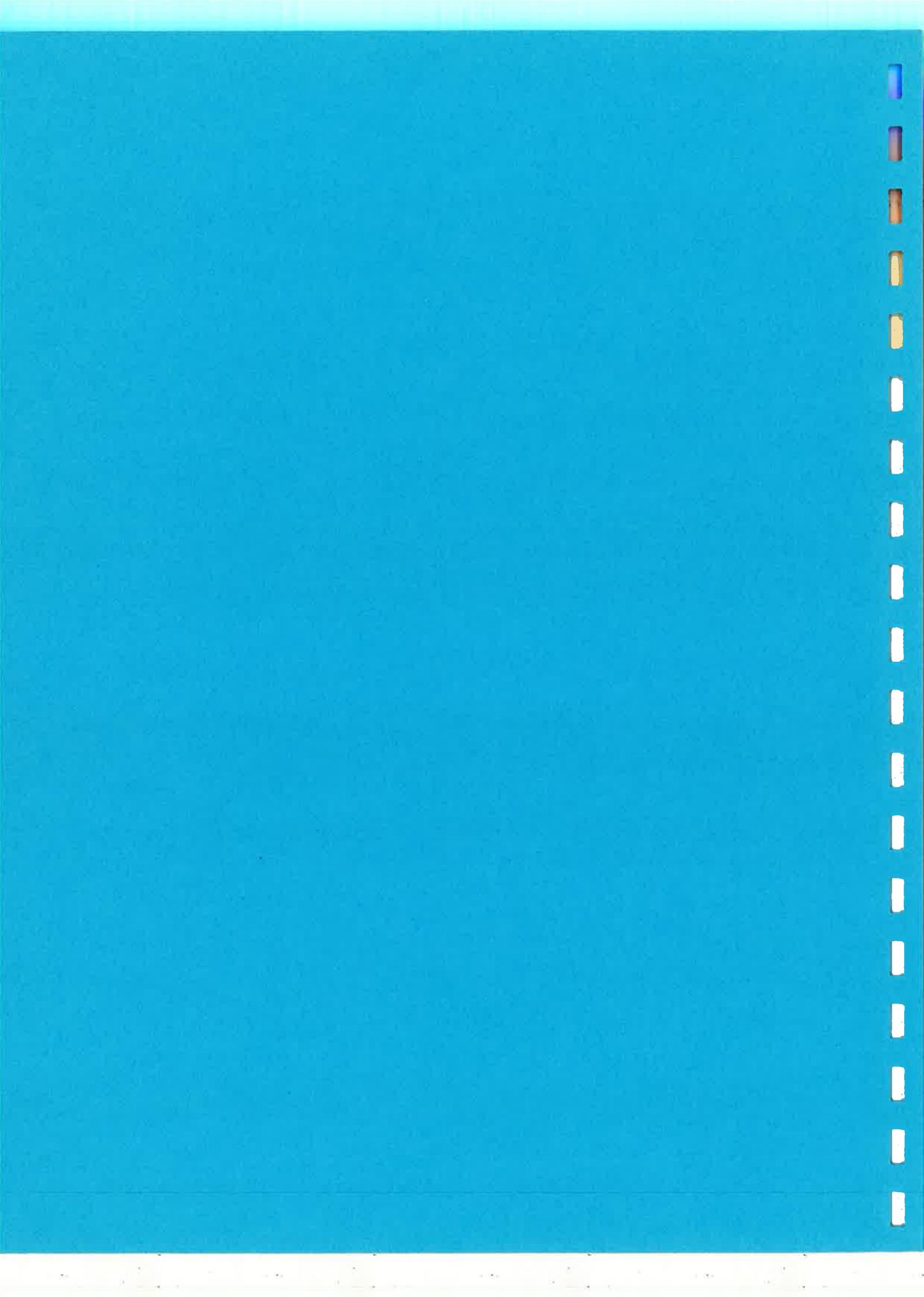
FEBS Eur. J. Biochem. Ms.-No. 378 (93 1355/1) Ms. 1-33 Pages 1-7

Springer-Verlag, 69042 Heidelberg 1 / Wiesbadener Graphische Betriebe GmbH, 65050 Wiesbaden

Provisorische Seltenzahl / Provisional page number 111sg35378

1. Korr. Date

Troisième partie: programme de recherche



1- *Préambule*

Les perspectives de recherche à moyen terme décrites ici s'inscrivent dans le cadre général de l'étude des relations structure/fonction des protéines. Les deux protéines modèles (papaïne, lipases) avec lesquelles nous sommes déjà familiers, seront utilisées pour l'étude de plusieurs aspects de ces relations structure/fonction. Une troisième protéine modèle, un Fab, sera aussi introduite. En résumé, nous désirons contribuer à répondre aux questions suivantes:

- comment la structure de la région pro d'une protéase à cystéine contribue-t-elle aux multiples fonctions de cette région pro?
- Quels sont les déterminants moléculaires de la reconnaissance du substrat par les lipases et quel est leur rôle? Quelles sont les différentes étapes de la catalyse enzymatique?
- Comment les contacts entre domaines et sous-unités des protéines modulent-ils la stabilité et les fonctions des protéines?

Différentes approches méthodologiques sont requises et certaines nécessiteront l'expertise de collaborateurs (cristallographie des protéines, mesure d'affinité, dynamique moléculaire). Nos efforts porteront particulièrement sur l'intégration de ces différentes approches complémentaires.

Un nouveau projet de recherche sera aussi présenté en rapport avec le pliage des protéines.

2- *Etude du précurseur de la papaïne*

Le comportement du précurseur de la papaïne lors de l'auto-maturation est maintenant bien compris. Si le mécanisme déclenchant le clivage de la région pro est clair, nous n'avons que des preuves indirectes quant à la localisation du site de coupure primaire. Nos efforts portent actuellement sur la détermination de la séquence N-terminale de l'intermédiaire principal de maturation afin d'établir sans ambiguïté ce site de clivage. Afin de purifier des quantités suffisantes de cet intermédiaire pour réaliser ces expériences, nous produirons une forme modifiée de la propapaïne dans les cellules d'insectes infectées par des baculovirus recombinants. Cette forme modifiée inclura une extension N-terminale (6 histidines) compatible avec la fixation sur une matrice d'affinité de type Ni^{++} -NTA. Nous avons vérifié préalablement que cette modification de la protéine n'affecte pas ses propriétés enzymatiques. Cette approche combinera donc un système

d'expression relativement performant et une méthode de purification par chromatographie d'affinité simple et efficace.

La disponibilité de quantités plus importantes de propapaïne pure sera aussi exploitée pour en déterminer la structure tridimensionnelle par diffraction des rayons X. Ce travail se fera en collaboration avec le Dr. M. Cygler (BRI). Actuellement, la classe des protéases à cystéine est la seule pour laquelle aucune structure de précurseur n'est connue. Cette structure nous permettra de vérifier certaines de nos prédictions basées sur les analyses fonctionnelles.

Les nouvelles méthodes de répertoires moléculaires seront mises à profit pour analyser les relations structure/fonction de la région pro de la papaine. La région pro sera exposée à la surface d'un phage filamentueux sous forme de fusion avec la protéine III du phage. Par analogie avec le fait que région pro de la cathepsin B sous forme libre possède une grande affinité pour la cathepsin B mature, nous testerons l'inhibition de la papaine par le phage recombinant. Dans l'hypothèse où ce test serait positif nous disposerons donc d'un système permettant de produire et de cribler un très grand nombre de variants de la région pro. Ce système sera utilisé pour identifier les acides aminés impliqués dans la reconnaissance avec la papaine mature. De plus, il sera possible de rechercher des mutants de haute affinité pour d'autres protéases à cystéine impliquées dans des pathologies humaines (cathepsine B, L, S).

3- Relations structure/fonction des lipases GCL

Nous allons concentrer notre attention sur les mécanismes d'activité de l'enzyme et sur l'étude des déterminants de la spécificité des lipases.

Nous avons récemment remodelé le site actif de GCLII par une double mutation (S249D/E354S). Ces remplacements de résidus, choisis d'après la structure de la molécule, n'affectent que peu l'activité de l'enzyme envers la trioléine. Le site actif d'un enzyme peut donc être reconstruit à l'intérieur d'un échafaudage moléculaire pré-établi. Le spectre d'activité de ce mutant envers une série de substrats sera déterminé afin d'étudier l'impact des mutations sur la reconnaissance du substrat. La détermination de la structure du double mutant est en cours et permettra de visualiser la nouvelle triade catalytique et de mesurer l'effet des changements sur l'ensemble de la structure.

Le site actif de GCLII n'est pas accessible au solvant car masqué par une boucle (positions à 68-74) occluant la dépression du site actif. Cette boucle a probablement un rôle dans la reconnaissance

du substrat et/ou dans le phénomène d'activation interfaciale, un phénomène unique aux lipases responsables de l'accroissement de l'activité de l'enzyme en présence d'une interface eau-lipide. Deux mutants de délétion de la boucle d'occlusion ont été construits et sont en cours de purification. Leur caractérisation enzymatique et structurale devrait permettre de vérifier nos hypothèses quant à la fonction de cette boucle.

Nous avons récemment obtenu l'expression de la lipase GCLI dans la levure. La comparaison des propriétés enzymatiques (activité spécifique envers une série de substrats) de GCLI avec celles de GCLII devrait permettre de lever les ambiguïtés décrites dans la littérature à ce propos. Dans le cas où les deux enzymes auraient des spectres de reconnaissance des substrats significativement différents, nous déterminerons la structure de GCLI de façon à corrélérer la structure des molécules avec leur spécificité pour les lipides. La construction d'hybrides entre les deux lipases et la construction de mutations ponctuelles seront aussi utilisées pour identifier les acides aminés responsables de la spécificité envers les différents substrats.

4- Rôles des contacts domaine-domaine et entre sous-unités des protéines

Ces études seront réalisées en collaboration avec le Dr. D. Altschuh (IBMC, Strasbourg). Certains de nos résultats sur le comportement de mutants à l'interface des deux domaines de la papaïne peuvent être interprétés comme témoignant du rôle des mouvements de la protéine lors de la catalyse. Cet aspect dynamique de la fonction des protéines est difficile à mettre en évidence expérimentalement. Nous essayons donc actuellement d'étayer cette hypothèse par des simulations du comportement dynamique de la papaïne sur ordinateur en collaboration avec le Dr. Annick Dejaegere (IBMC, Strasbourg) et le Prof. Martin Karplus (Strasbourg/Université Harvard, Boston, USA).

Les limitations imposées par les systèmes d'expression de la papaïne nous ont conduits à choisir un deuxième modèle d'étude des relations entre les sous-unités des protéines: les Fabs. Les Fabs constituent la partie des anticorps reconnaissant les antigènes. Les Fabs sont composés de l'association d'une chaîne légère et d'une chaîne lourde, le site de liaison avec l'antigène étant défini par un ensemble de boucles appartenant aux deux chaînes. Ce projet implique les étapes suivantes:

1- Etude théorique des résidus à l'interface des sous-unités des Fab. Choix des résidus à muter basé sur cette étude.

2- Expression dans *E. coli* de mutants d'un Fab dérivé d'un anticorps monoclonal reconnaissant un peptide d'origine virale de 12 acides aminés.

3- Etudes fonctionnelles fines des mutants (mesure d'affinité pour l'antigène et des sous-unités entre elles) sur BIAcore.

4- Détermination de la structure cristallographique de mutants sélectionnés pour leur intérêt (par exemple: mutants neutres au niveau de la reconnaissance des deux sous-unités mais perturbés dans la reconnaissance de l'antigène).

5- Simulation du comportement dynamique de mutants par le programme informatique CHARMM.

Ce projet fait actuellement l'objet d'une demande de fonds conjointe (International Human Frontier science program) entre notre laboratoire et ceux de M. Cygler (Montréal), D. Altschuh (Strasbourg) et Martin Karplus (Harward).

5- Etude du repliement des protéines par analyse de répertoires de protéines à séquence aléatoire

Il est maintenant établi que le répertoire naturel de repliement des protéines est limité. Quelle est l'origine de ce répertoire restreint? S'agit-il d'une propriété fondamentale découlant des propriétés des 20 acides aminés entrant dans la composition des protéines "modernes"? S'agit-il d'une restriction imposée par l'évolution, l'ensemble des propriétés fourni par les protéines étant satisfaites avec un nombre restreint de repliements? Les 20 acides aminés sont-ils tous nécessaires pour obtenir des repliements stables mais souples?

Les méthodes modernes de synthèse de gène, d'expression et de sélection des protéines sont utilisables pour recréer *de novo* un répertoire de protéines indépendamment du répertoire naturel.

Nous proposons donc de générer une large collection de protéines de séquence aléatoire, de rechercher parmi cette collection les protéines ayant des caractéristiques physico-chimiques proches de celles des protéines naturelles et de les étudier en détail. Le contrôle total exercé sur la fabrication de cette collection sera utilisé pour faciliter la production, la purification et la caractérisation des protéines par introduction d'étiquettes moléculaires ("tags" d'affinité, par exemple épitope ou séquence peptidique (HPQ) reconnue par la streptavidine). Plusieurs types de répertoires pourront être synthétisés. Par exemple, nous pourrons n'utiliser que certains acides aminés ou biaiser la proportion relative de certains acides aminés ou

type d'acides aminés (hydrophobe/hydrophile) afin d'en tester les conséquences sur l'efficacité de production de molécules repliées.

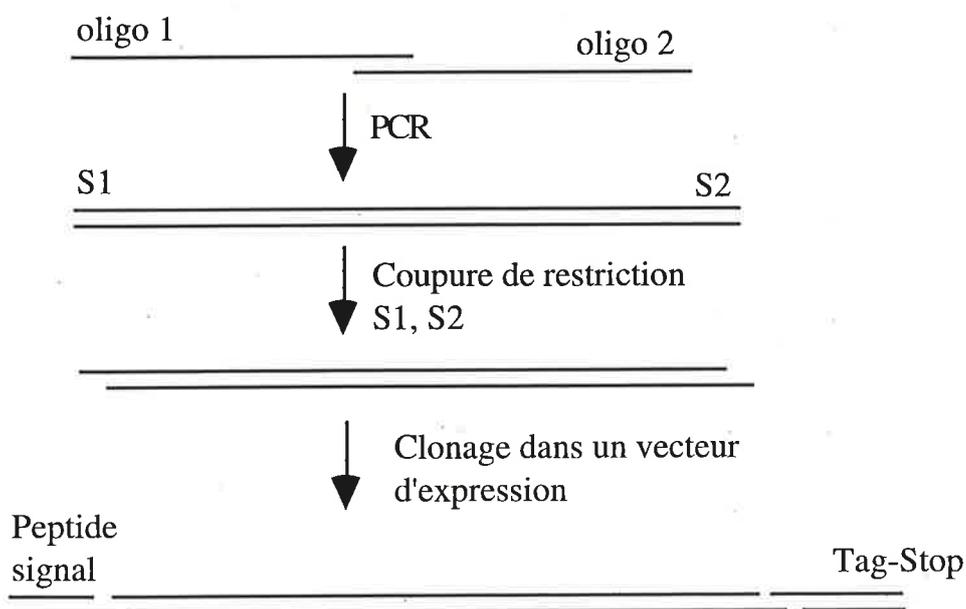
Le répertoire de protéines sera exprimé dans *E. coli*. sous forme fusionné avec un signal peptide. Les avantages de cette configuration sont les suivants:

- expression d'un vaste répertoire (10^7 - 10^8 , limité par l'efficacité de transformation).

- permet un criblage primaire des protéines "viables" c'est à dire des protéines sécrétées donc a priori solubles (non agrégées) et stables (non dégradées par les protéases cellulaires). Le criblage se fera grâce à la présence de l'étiquettes moléculaires soit par criblage classique sur boite de pétri soit par "panning" direct des cellules.

L'assemblage génétique du répertoire se fera par PCR. La taille des protéines sera d'environ 100 acides aminés soit une taille suffisante pour produire des molécules capables de se replier spontanément (Fig. 1).

Fig. 1: construction du répertoire. Oligo 1 et 2 sont de séquence aléatoire à l'exception des sites de restriction S1 et S2 et de la région d'hybridation centrale. En raison de la taille maximale des oligonucléotides qu'il est possible de synthétiser efficacement, au moins 4 oligo seront nécessaires. Dans cette configuration le "tag" est placé en C-terminal de la protéine.



La séquence des protéines ayant franchi le premier criblage sera déterminée après récupération des plasmides. Les propriétés de ces protéines seront déterminées par une batterie de tests incluant, entre autres, des mesures de sensibilité aux protéases *in vitro*, la détermination du point isoélectrique, de la stabilité thermique et la stabilité en présence d'agents dénaturants, la mesure du degrés de structures secondaires. Ces mesures nous permettront de comparer certaines des propriétés des protéines "artificielles" avec celles des protéines naturelles.

6- Remerciements

Je remercie ma collègue Danièle Altschuh pour avoir relu et commenté ce document. Je suis aussi très reconnaissant aux rapporteurs et membres du jury d'avoir accepté de réviser et de juger ma demande d'habilitation à diriger des recherches.