

THESE

de DOCTORAT D'UNIVERSITE

pour l'obtention du grade de

DOCTEUR EN SCIENCES DE LA VIE ET DE LA SANTE

présentée par

Christophe FLAHAUT



Modifications post-traductionnelles et conformation des chaînes lourdes de l'inter- α -inhibiteur

Directeurs de thèse : Professeur Mizon J.
Docteur Michalski J.C.

Présentée le 25 janvier 2000 devant la commission d'examen

Jury

Président :	Professeur Verbert A.
Rapporteurs :	Professeur Guéant J.L. Professeur Salier J.P.
Examineurs :	Docteur Michalski J.C. Professeur Mizon J. Professeur Roussel P.

*A ma femme,
à mes parents,
à toute ma famille,
à mes amis.*

*Je tiens à exprimer ici ma profonde reconnaissance à
tous ceux qui m'ont soutenu, aidé et qui m'ont accordé leur
confiance tout au long de ces 4 années de recherches.*

Je tiens à remercier tout particulièrement :

Monsieur le Professeur Jacques Mizon ; vous m'avez accueilli au sein de votre laboratoire pour le DEA puis pour préparer cette thèse de sciences. Vos connaissances et nos discussions scientifiques ainsi que vos remarques constructives sur mon travail m'ont beaucoup apporté.

Monsieur le Docteur Jean-Claude Michalski ; tu m'as accueilli pour préparer mon DEA puis cette thèse. Tu m'as toujours fait confiance pour mener à bien mes travaux. J'espère avoir su tirer parti de tes connaissances scientifiques et avoir donné le maximum de moi-même à ton groupe.

Madame le Docteur Calliope Capon ; tu as accepté d'être mon tuteur de DEA et de m'accompagner dans mes premiers pas en recherche. Travailler avec toi fut un réel plaisir. J'admire ta disponibilité de tout instant, ta bonne humeur de tous les jours et surtout ta rigueur scientifique.

Monsieur Le Docteur Pierre Sautière ; vous m'avez toujours accueilli avec beaucoup de gentillesse et d'humour dans votre laboratoire. Vos connaissances et vos remarques constructives ont fortement contribué au succès de ces recherches.

Je remercie les équipes des deux laboratoires dans lesquels j'ai travaillé : le laboratoire de Biochimie de la Faculté des Sciences Pharmaceutiques et Biologiques de Lille et l'UMR n°8576 du CNRS.

Je remercie également le laboratoire de spectrométrie de masse (centre commun de mesure, bât C4, USTL) ainsi que le laboratoire du Professeur Sergheraert (UMR 8525 CNRS, Institut de Biologie et Institut Pasteur de Lille).

Ce travail a abouti grâce à la collaboration active et efficace des uns et aux encouragements des autres.

Je remercie pour l'honneur qu'ils me font :

Monsieur le Professeur André Verbert

en acceptant de présider ce jury ;

Monsieur le Professeur Jean-Louis Guéant

Monsieur le Professeur Jean-Philippe Salier

en acceptant d'être les rapporteurs de cette thèse ;

Monsieur le Docteur Jean-Claude Michalski

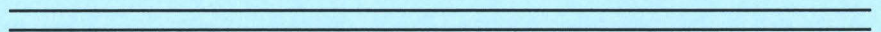
Monsieur le Professeur Jacques Mizon

Monsieur le Professeur Philippe Roussel

qui ont bien voulu juger ce travail.

Ce travail de recherches a été effectué sous la direction scientifique du Professeur Jacques Mizon au sein du laboratoire de Biochimie, Faculté des Sciences Pharmaceutiques et Biologiques de Lille en étroite collaboration avec le Docteur Jean-Claude Michalski de l'UMR n°8576 du CNRS.

Sommaire



SOMMAIRE

SOMMAIRE	7
PUBLICATIONS ET COMMUNICATIONS	9
ANNEXE I : ABREVIATIONS UTILISEES	11
ANNEXE II : SYMBOLES UTILISES POUR LES ACIDES AMINES	13
AVANT-PROPOS	14
GENERALITES	16
I Les membres de la famille de l'IαI	16
1) Chez l'homme.....	16
a) L'inter-alpha-inhibiteur.....	16
b) Les autres membres de la famille de l'IαI.....	19
2) Chez les mammifères.....	21
a) Les complexes sériques homologues à ceux de l'homme.....	21
b) Les complexes sériques différents de ceux de l'homme.....	22
II Une famille de molécules originales	24
1) Des activités biologiques variées.....	24
a) Activité anti-inflammatoire.....	24
- <i>Antiprotéase</i>	24
- <i>Stabilisation de la membrane des lysosomes</i>	27
- <i>Inhibition de la production des cytokines pro-inflammatoires</i>	
- <i>rôle de la régulation du flux calcique cytosolique par la bikunine</i>	28
- <i>La bikunine inhibe la production de prométalloprotéinases</i>	30
- <i>Modulation de la croissance cellulaire</i>	30
b) Activités anti-métastatiques.....	31
- <i>Récepteurs cellulaires</i>	31
- <i>Une protéine composite (uPA-UTI) comme agent anti-métastatique</i>	33
c) Stabilisation de la matrice extracellulaire.....	33
- <i>L'IαI et l'AH : une vieille histoire</i>	34
- <i>L'IαI et la protéine TSG-6</i>	37
- <i>Chaînes lourdes et protéoglycannes</i>	39

2) Une originalité structurale et caractéristique.....	42
a) Diversité des glycosaminoglycannes.....	42
b) Les protéoglycannes.....	45
c) Les complexes de la famille de l'I α I : des protéoglycannes ?.....	49
3) Une biosynthèse originale.....	53
a) Gènes et précurseurs.....	53
b) Assemblage moléculaire.....	55
c) Régulation de la biosynthèse au cours de pathologies inflammatoires	56
d) P α I et I α I comme protéines recombinantes.....	58
III Structures des différentes chaînes polypeptidiques constituant les protéines de la famille de l'IαI.....	60
1) Structure de la bikunine	60
2) Structure des chaînes lourdes	61
a) Ossatures peptidiques.....	63
b) Domaines structuraux.....	64
c) Modifications post-traductionnelles : ponts disulfure et glycosylation.....	66
3) Homologies inter espèces.....	67
OBJECTIFS.....	71
RESULTATS.....	73
I Topologie des ponts disulfure des chaînes lourdes de l'IαI humain:	
implications structurales et fonctionnelles.....	73
1) Publication	74
2) Discussion et résultats complémentaires.....	84
II Structures glycaniques des chaînes lourdes de l'IαI humain.....	89
1) Publication	91
2) Discussion et résultats complémentaires.....	100
III Schéma de glycosylation des chaînes lourdes de l'IαI porcine.....	103
1) Résultats	104
DISCUSSIONS ET PERSPECTIVES.....	137
BIBLIOGRAPHIE.....	140

Ce travail a fait l'objet de publications et de communications :

Publications :

Disulphide bonds assignment in the inter- α -inhibitor heavy chains : Structural and functional implications. (1998)

Flahaut, C., Mizon, C., Aumercier-Maes, P., Colson, P., Bailly, C., Sautière, P. and Mizon, J.

Eur. J. Biochem., **255**, 107-115.

Glycosylation pattern of human inter- α -inhibitor heavy chains. (1998)

Flahaut, C., Capon, C., Balduyck, M., Ricart, G., Sautière, P. and Mizon, J.

Biochem. J., **333**, 749-756.

Pig inter- α -inhibitor heavy chains are N- and O-glycosylated as their human counterparts.

Flahaut, C., Capon, C., Drobecq, H., Michalski, J.C., Mizon, J. (en préparation)

Communications orales et posters :

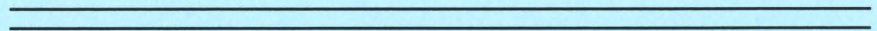
Structure des glycanes des chaînes lourdes de l'inter- α -inhibiteur (I α I)

Flahaut, C., Capon, C., Balduyck, M., Ricart, G., Sautière, P. and Mizon, J.
Congrès de Printemps de la Société Française de Biochimie et Biologie Moléculaire (24-26 mars 1997, Paris).

Structure des glycanes des chaînes lourdes de l'inter- α -inhibiteur (I α I) (apport de la spectrométrie de masse mode MALDI)

Flahaut, C., Capon, C., Balduyck, M., Ricart, G., Sautière, P. and Mizon, J. *14^{ème} Journée de la Société Française de Spectrométrie de Masse* (16-18 septembre 1997, Lille).

Annexes



ANNEXE I : ABREVIATIONS UTILISEES

ADNc	Acide désoxyribonucléique complémentaire à l'ARNm
α 1-m	α 1-microglobuline
α 1-PI	α 1-protéinase inhibiteur
ADP	Adénosine diphosphate
AMBP	α 1-microglobuline/bikunin precursor
ARNm	Acide ribonucléique messenger
B	Bikunine
BrCN	Bromure de cyanogène
CAM	Carboxamidométhylé
CS	Chondroïtine sulfate
DS	Dermatane sulfate
FACE	Fluorophore assisted carbohydrate electrophoresis
GAG	Glycosaminoglycane
HA	Acide hyaluronique
HABP	Hyaluronic acid binding protein
HS	Héparane sulfate
Hx	Heavy chain x
HxP	Heavy chain x precursor
I α I ou I α I _{H1,2,B}	Inter- α -inhibiteur
I α I _{H1,B}	Hétérodimère constitué de H1 et de B
I α I _{H4}	Chaîne lourde H4 mature
I α LI ou I α I _{H2,B}	Inhibiteur de type inter- α
IL	Interleukine
KS	Kératane sulfate
LPS	Lipopolysaccharides
MALDI	Matrix assisted laser desorption ionization
MMP	Matrix metalloproteinase
NAG	N-acétyl- β -D-glucosaminidase
Pr α I ou I α I _{H3,B}	Pré- α -inhibiteur
PARP	Poly (ADP-ribose) polymérase

PG	Protéoglycane
PGP	Protein-glycosaminoglycan-protein
PMN	Polymorphonucléaires neutrophiles
PNGase F	Peptide-N-glycosidase F
PPA	Protéines de la phase aiguë de l'inflammation
PTR	Proteoglycan tandem repeat
RHAMM	Receptor for HA-mediated motility
RP-HPLC	Reverse phase-high performance liquid chromatography
RT-PCR	Reverse transcriptase-polymerase chain reaction
SDS/PAGE	Sodium dodecyl sulfate/polyacrylamide gel electrophoresis
SHAPs	Serum-derived hyaluronan-associated proteins
TIMP	Tissue inhibitor of metalloprotéinases
TNF- α	Tumor necrosis factor α
TSG-6	TNF-stimulated gene 6 protein
UAP	Uronic-acid-rich protein
uPA	Urokinase-type plasminogen activator
uPAR	Urokinase-type plasminogen activator receptor
UTI	Urinary trypsin inhibitor
vWA	Domaine von Willebrand de type A
Gal	Galactose
GalNAc	N-acétylgalactosamine
GlcNAc	N-acétylglucosamine
GlcUA	Acide glucuronique
IdoUA	Acide L-iduronique
Man	Mannose
NeuAc	Acide N-acétylneuraminique
Xyl	Xylose

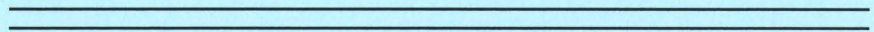
ANNEXE II : SYMBOLES UTILISES POUR LES ACIDES AMINES

Les acides aminés et leurs symboles

A	Ala	alanine
C	Cys	cystéine
D	Asp	acide aspartique
E	Glu	acide glutamique
F	Phe	phénylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	méthionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	sérine
T	Thr	thréonine
V	Val	valine
W	Trp	tryptophane
Y	Tyr	tyrosine

Hse	Homosérine
-----	------------

Avant-Propos



AVANT-PROPOS

Les antiprotéases présentes dans le plasma régulent l'activité des diverses protéinases impliquées dans de nombreux processus métaboliques concourant à la protection de l'organisme tels que l'inflammation, la coagulation ou l'activation du complément...

L'inter- α -inhibiteur (I α I) est le chef de file d'une famille de glycoprotéines présentant des nombreuses homologies immunologiques et structurales. Les complexes multiprotéiques de cette famille sont des inhibiteurs de sérine-protéinases de type Kunitz. Depuis bientôt 40 ans, différentes équipes de recherche ont contribué à la caractérisation génétique, métabolique, structurale et fonctionnelle des membres de la famille de l'I α I. Les résultats de ces années de recherche, succinctement passés en revue dans la première partie de ce mémoire, montrent que la structure très particulière de ces composés est désormais assez bien comprise. L'I α I se compose de trois chaînes polypeptidiques distinctes : deux chaînes lourdes dénommées H1 et H2 et une chaîne légère, la bikunine. Toutes trois sont covalentiellement unies par l'intermédiaire d'une chaîne de glycosaminoglycane de type chondroïtine sulfate.

La bikunine est le support de l'activité antiprotéasique de l'I α I. Cependant, parce que cette activité vis-à-vis des protéinases étudiées est relativement faible par rapport à celle développée par les autres inhibiteurs de protéinases du plasma, le rôle physiologique réel joué par cet inhibiteur reste à déterminer. De même, la véritable fonction des chaînes lourdes n'est pas clairement élucidée ; ne servent-elles qu'à cibler la bikunine vers les sites appropriés ou, au contraire, sont-elles le support d'activités biologiques qui leur sont propres ? Récemment, leur rôle stabilisateur vis-à-vis des matrices extracellulaires a été mis en évidence.

Les séquences en acides aminés des chaînes lourdes ont été pour l'essentiel déterminées par déduction à partir de la structure des acides désoxyribonucléiques complémentaires correspondants. Présentant de fortes homologies de séquence, elles diffèrent cependant par leur réactivité. Dès lors, la compréhension du rôle

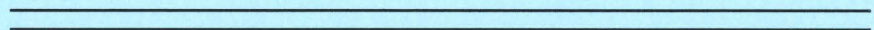
particulier joué par chaque chaîne lourde suppose une meilleure connaissance de leur conformation tridimensionnelle. Notre objectif a donc été de progresser dans ce domaine.

Les chaînes lourdes H1 et H2 comportent respectivement 5 et 4 résidus de cystéine. Nous avons dans un premier temps établi la topologie des ponts disulfure des chaînes lourdes H1 et H2 de l'I α I humain. Au cours de la biosynthèse de chaque chaîne lourde, deux ponts disulfure se forment. Nous avons ensuite montré que les chaînes lourdes humaines sont équivalentes en terme de structure secondaire. Chacune renferme 25 % d'hélicité de type α . Libres ou associées au sein de l'I α I, la sensibilité à la protéolyse partielle des extrémités C-terminales des chaînes lourdes témoigne d'une conformation désordonnée. Au contraire, leurs moitiés N-terminales apparaissent comme des structures ordonnées.

La présence de glycanes peut influencer de manière notable la réactivité et le métabolisme d'une protéine, nous démontrons que les chaînes H1 et H2 humaines présentent à la fois des N- et des O-glycanes. Ces derniers sont localisés à proximité de leurs extrémités C-terminales.

Des inhibiteurs de sérine-protéinases appartenant à la famille de l'I α I ont également été caractérisés dans de nombreuses espèces animales. Compte tenu de la stricte conservation du nombre et de la position des résidus de cystéine des chaînes lourdes des différentes espèces pour lesquelles la séquence en acides aminés est établie, nous proposons que la topologie des ponts disulfure des chaînes lourdes de l'I α I est **une donnée structurale conservée**. Profitant des techniques de purification de l'I α I porcin récemment mises au point au laboratoire, nous démontrons que la N- et la O-glycosylation des chaînes lourdes sont des **données structurales conservées** entre ces deux espèces. Dès lors, nous proposons que la sélectivité d'assemblage des membres de la famille de l'I α I repose sur la O-glycosylation des extrémités C-terminales des chaînes lourdes H1, H2 et H3.

Généralités



GENERALITES

I Les membres de la famille de l'I α I

1) Chez l'homme

a) L'inter-alpha-inhibiteur

L'Inter- α -Inhibiteur (I α I ou I α I_{H1,2,B}), identifié pour la première fois par Steinbuch et Loeb en 1961 [1], est une glycoprotéine plasmatique de masse moléculaire apparente 220 kDa. L'I α I est principalement synthétisé par le foie et est présent dans le plasma humain à une concentration voisine de 0,3 g/l [2]. Primitivement dénommé Inter- α -Trypsine Inhibiteur, cette glycoprotéine doit son nom à son activité antiprotéasique dirigée contre les sérine-protéinases et à sa mobilité électrophorétique, située entre celles des α 1 et α 2-globulines sériques, sur acétate de cellulose à pH alcalin [3].

L'I α I a été longtemps considéré comme constitué d'une chaîne polypeptidique monocaténaire [4] en raison de sa stabilité particulière vis-à-vis de divers agents chaotropiques ou réducteurs. En effet, en électrophorèse en gel de polyacrylamide en présence de dodécyl sulfate de sodium, l'I α I migre sous forme d'une bande unique, qu'il ait été ou non réduit par le β -mercaptoéthanol. Cependant, dès 1983, les travaux de biologie moléculaire réalisés par Bourguignon et coll [5], apportaient déjà la preuve de la pluricaténarité de cette protéine. Ils démontraient que l'I α I est synthétisé dans l'hépatocyte à partir de deux populations distinctes d'acides ribonucléiques messagers (ARNm). Celles-ci sont traduites *in vitro* respectivement en un précurseur polypeptidique léger **L** de 40 kDa et en un précurseur polypeptidique lourd **H** de 90 kDa, tous deux ne présentant aucune communauté antigénique. Par la suite, le clonage de plusieurs acides désoxyribonucléiques complémentaires (ADNc) distincts, à partir de la population d'ARNm codant pour le polypeptide lourd de l'I α I, a montré qu'il existe en fait plusieurs chaînes lourdes homologues [6, 7].

Il est à présent clairement établi que l'I α I est une molécule tricaténaire stable constituée de deux chaînes polypeptidiques dénommées chaînes lourdes **H1** et **H2** covalentiellement associées à une chaîne légère la **bikunine**.

La bikunine ou UTI (*Urinary Trypsin Inhibitor*), HI-30 ou encore EDC1 a d'abord été mise en évidence dans l'urine [8]. Facilement isolée grâce à sa stabilité en milieu acide [9], elle est à l'heure actuelle la chaîne la mieux connue d'un point de vue fonctionnel et structural. La bikunine possède dans sa partie N-terminale, une seule chaîne de glycosaminoglycane de type chondroïtine sulfate [10] constituée d'environ 15 à 17 unités disaccharidiques. La bikunine est donc **un protéoglycane**. Cette découverte en 1986 est à l'origine des travaux qui ont contribué à une meilleure compréhension de la structure très particulière de l'I α I.

Dans les années 1988 et 1989, la nature de la liaison covalente unissant les différentes chaînes polypeptidiques des membres de la famille de l'I α I fut partiellement élucidée. Différentes équipes démontrèrent alors que, sous l'action d'enzymes tels que la hyaluronidase (non spécifique), la chondroïtinase (AC ou ABC) ou sous l'action d'agents chimiques de déglycosylation (acide trifluorométhane sulfonique), la molécule d'I α I est clivée et libère des chaînes polypeptidiques distinctes [11-13]. Par conséquent, les différentes chaînes composant l'I α I sont unies covalentiellement par la chaîne de glycosaminoglycane de type chondroïtine sulfate portée par la bikunine. Jamais mise en évidence à ce jour dans la structure d'autres glycoprotéines, la nature de cette liaison covalente dénommée « *Protein-Glycosaminoglycan-Protein cross-link* » (liaison PGP) est le trait structural le plus caractéristique des membres de la famille de l'I α I [14].

La molécule d'I α I est donc composée de 3 chaînes polypeptidiques : deux chaînes lourdes (H1 et H2) et une chaîne protéoglycannique légère (la bikunine) **covalentiellement associées** par l'intermédiaire de la chaîne de chondroïtine sulfate (Figure 1).

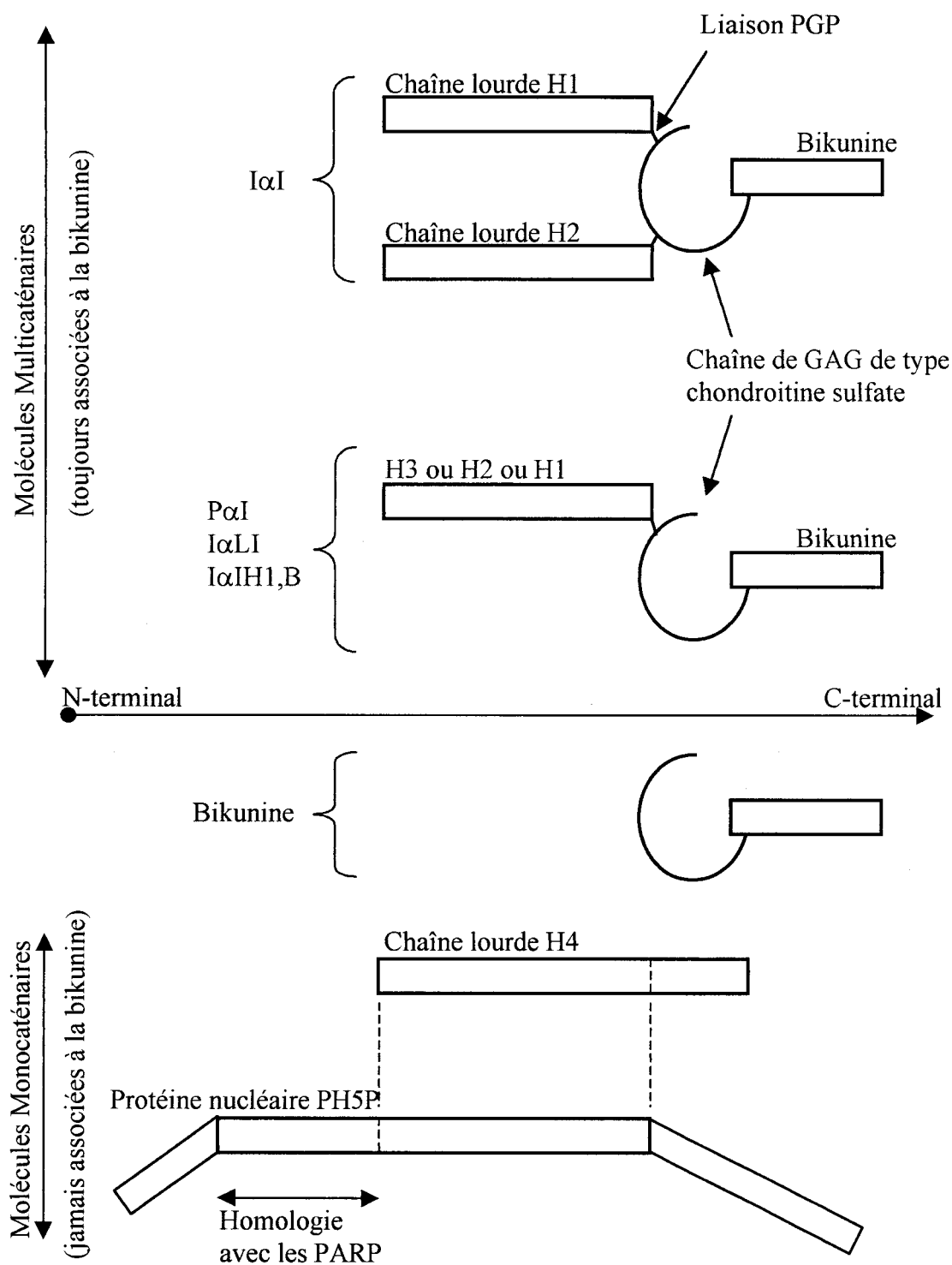


Figure 1: Représentation schématique des membres de la famille de l' $I\alpha I$ humain (PARP: poly (ADP-ribose) polymérase)

b) Les autres membres de la famille de l'I α I

A l'heure actuelle, la famille de l'I α I [15] peut être divisée en deux catégories. D'une part, celle des molécules **multicaténaires** toujours constituées de la bikunine covalentiellement associée à une ou deux chaînes lourdes distinctes et d'autre part, la catégorie des molécules **monocaténaires** présentant de nombreuses homologies immunologiques et structurales avec les chaînes lourdes précédemment citées et qui ne sont jamais associées à la bikunine. Le tableau 1 résume les caractéristiques des membres de cette famille dont la bikunine fait partie intégrante.

Dans le plasma humain, outre l'I α I, plusieurs molécules constituées d'une seule chaîne lourde associée par la liaison PGP à la bikunine ont été caractérisées : Le pré- α -inhibiteur (P α I ou I α IH3,B) constitué de la chaîne lourde H3 et de la bikunine [13, 14, 16] et l'inhibiteur de type Inter- α (I α LI ou I α IH2,B) composé de la chaîne lourde H2 et de la bikunine [17]. L'existence d'un complexe de même nature comportant la chaîne lourde H1 et la bikunine (I α IH1,B) n'est pas exclue [17]. Immunologiquement détecté par la suite, celui-ci n'a toutefois pas encore été isolé. Par contre, des composés tricaténaires constitués de deux chaînes lourdes identiques ou d'un mélange équimolaire de chaînes H1 et H3 associées à la bikunine n'existent pas dans le plasma humain. Les membres de la famille de l'I α I semblent donc présenter une **sélectivité d'assemblage**.

A côté de ces complexes sériques stables, une autre glycoprotéine sérique n'ayant aucune parenté antigénique avec la bikunine mais présentant de nombreuses homologies avec les trois premières chaînes lourdes de la famille de l'I α I a été récemment décrite. La chaîne H4, IHRP (*Inter- α Heavy chain Related Protein*) ou encore PK-120 (*Plasma Kallikrein-sensitive protein 120*) a été isolée pour la première fois en 1989 par Hammer et coll [18]. Néanmoins, sa parenté structurale n'a été mise en évidence que récemment grâce à l'analyse, par dégradation d'Edman, de fragments peptidiques [19]. Par la suite, ce lien de parenté avec les chaînes lourdes fut confirmé par détermination de la structure primaire complète déduite de celle de l'ADNc [20].

Très récemment, Jean et coll [21] viennent d'élargir la notion de superfamille de l'I α I. Grâce à la recherche informatisée des homologies de séquences peptidiques, ces

	ABREVIATIONS	NOMS	COMPOSITIONS	MASSE MOLECULAIRE APPARENTE (kDa)	LOCALISATION	REFERENCES
Complexes multicaténaires	I α I ou I α IH1,2,B	Inter- α - Inhibiteur	H1 / H2 / Bikunine	220	plasma	1
	P α I ou I α IH3,B	Pré- α - Inhibiteur	H3 / Bikunine	130	plasma	13, 14, 16
	I α LI ou I α IH2,B	Inhibiteur de type inter- α	H2 / Bikunine	125	plasma	17
	I α IH1,B	I α I-related H1/ bikunin	H1 / Bikunine	125	plasma	17
Molécules monocaténaires	I α IH4	IHRP, PK 120	H4	120	plasma	18,19
	PH5P	PARP-related/I α I-related H5/prolin-rich	PH5P	?	nucléaire	21
	Bikunine	UTI, EDC1, HI-30	Bikunine avec GAG	42-46	urine, plasma mucus bronchique	8
			Bikunine sans GAG	26		

Tableau 1: Nomenclature et caractéristiques des membres de la famille de l'I α I selon Salier et coll [15] et modifiée selon l'hypothèse formulée par Jean et coll [21].

auteurs viennent de présenter l'hypothèse d'une parenté structurale entre une protéine nucléaire dénommée PH5P et la chaîne H4 de la famille de l' $\text{I}\alpha\text{I}$. Cette protéine PH5P (*Poly(ADP-ribose)polymerase-related/ I αI -related H5/Prolin-rich*) présente également des homologies de structure avec la famille des poly(ADP-ribose)polymerases. Par ses différents domaines constitutifs, celle-ci pourrait faciliter les interactions moléculaires aboutissant à la réparation de l'ADN. Néanmoins, sa localisation nucléaire contraste avec la localisation plasmatique caractérisant les membres de la famille de l' $\text{I}\alpha\text{I}$.

2) Chez les mammifères

Bien que la présence de la bikunine ait été précédemment caractérisée chez divers mammifères, celle de protéines homologues a été établie par l'obtention des ADNc qui, actuellement sont séquencés chez 8 espèces animales qui vont de l'homme au poisson. En effet, la connaissance de l'ADNc pour une espèce donnée, permet souvent de caractériser chez d'autres espèces des séquences homologues avant même que la protéine ait pu être identifiée. En ce qui concerne les chaînes lourdes, la famille de l' $\text{I}\alpha\text{I}$ n'échappe pas à cette règle. Chez certains vertébrés supérieurs tels que la souris, la littérature ne rapporte que l'étude de l'expression des ARN messagers et le clonage des ADNc correspondant aux chaînes lourdes. Dans des modèles protéiques complexes comme ceux des membres de la famille de l' $\text{I}\alpha\text{I}$ humain, la caractérisation des ADNc et celle des séquences peptidiques qui en découlent ne peuvent servir de preuve quant à l'existence de complexes identiques chez telle ou telle espèce. Par conséquent, nous nous proposons de présenter succinctement les complexes sériques apparentés à l' $\text{I}\alpha\text{I}$ retrouvés chez les mammifères étudiés jusqu'à présent.

a) Les complexes sériques homologues à ceux de l'homme

Chez le babouin, l'analyse du sérum par immunodétection avec des anticorps spécifiquement dirigés contre des fragments des chaînes lourdes recombinantes humaines (H1, H2 et H3) révèle une composition identique pour les complexes $\text{I}\alpha\text{I}$, $\text{P}\alpha\text{I}$ et $\text{I}\alpha\text{LI}$ du babouin et de l'homme. Toutefois, la présence de l'hétérodimère $\text{I}\alpha\text{IH1,B}$ n'est pas

détectée dans le sérum de babouin contrairement à l'homme [22]. Le sérum de rat renferme également deux protéines similaires à l'I α I et au P α I humain par leur masse moléculaire (230 et 125 kDa) et leur réactivité vis-à-vis d'un anticorps anti-bikunine. Comme démontré par dégradation d'Edman, ces composés présentent la même composition peptidique que les protéines humaines [23]. Néanmoins, la teneur en P α I est 9 à 10 fois plus importante que celle de l'I α I dans le sérum de rat, ce qui n'est absolument pas le cas chez l'homme où l'I α I est quantitativement prédominant. En ce qui concerne le porc, Carrette et coll [24], isolant pour la première fois l'I α I porcin, ont démontré par séquençage peptidique qu'il possède une composition peptidique en tout point similaire à son homologue humain. Les hétérodimères porcins de la famille de l'I α I n'ont malheureusement pas été recherchés dans cette étude.

b) Les complexes sériques différents de ceux de l'homme

Bien que parfaitement connue, l'importance d'une caractérisation peptidique couplée à une étude nucléique a été mise en évidence par les travaux de Castillo et coll [25]. En effet ces auteurs, ne disposant d'aucune donnée relative aux ADNc des chaînes lourdes bovines, ont démontré que l'I α I bovin est constitué des chaînes H3, H2 et de la bikunine au lieu des chaînes H1, H2 et de la bikunine pour l'I α I humain. De même, le P α I bovin, constitué de l'assemblage de la chaîne lourde H2 et de la bikunine, diffère de celui de l'homme (I α IH3,B). Ces résultats surprenants ont ultérieurement été confirmés par détection immunologique dans le sérum de veau fœtal [22]. Cependant, les sérums de l'homme et du babouin renferment également l'hétérodimère I α IH2,B. Par conséquent, compte tenu de l'inversion des taux sériques d'I α I et de P α I observée entre l'homme et le rat [23], le complexe I α IH3,B bovin existe peut être, mais en trop faible quantité pour être détecté à l'époque. L'identification plus aisée de ces chaînes lourdes au niveau nucléique confirmerait peut-être l'absence de synthèse de la chaîne H1. Malheureusement, aucune information concernant soit l'expression des ARN messagers (ARNm) soit la caractérisation des ADNc des chaînes lourdes bovines qui permette d'étayer ce modèle d'assemblage original par rapport à l'homme, n'est disponible à ce jour.

Le bœuf n'est pas le seul mammifère à présenter ce type de composé : $I\alpha IH_{3,2,B}$. En 1996, Yamamoto et coll [26] présentaient les résultats concernant les différents complexes sériques apparentés à la famille de l' $I\alpha I$ chez le hamster atteint d'un carcinome pancréatique induit par le N-nitrosobis(2-oxopropyl)amine (BOP). Deux types d' $I\alpha I$ étaient identifiés en quantité similaire : le type humain $I\alpha IH_{1,2,B}$ et le type bovin $I\alpha IH_{3,2,B}$. Cependant, comme souligné par les auteurs, la présence de ces deux types d' $I\alpha I$ chez le hamster sain demande confirmation. Un seul type d'hétérodimère homologue à celui de l'homme pouvait être identifié, le $P\alpha I$ ($I\alpha IH_{3,B}$). Néanmoins, l'existence d'autres hétérodimères, probablement en quantité minime n'était pas écartée par les auteurs. La caractérisation des ADNc des trois chaînes lourdes du hamster corrobore en partie l'existence de ces différents complexes [27].

A ce jour, aucune déficience congénitale des protéines de la famille de l' $I\alpha I$ n'a été rapportée, suggérant que l'absence de cette famille de gènes constitue une délétion létale. La localisation et l'organisation des gènes responsables de la synthèse des membres de la famille de l' $I\alpha I$ sont homologues chez la souris et chez l'homme [28]. Ceci laisse supposer une organisation génétique similaire chez tous les mammifères terrestres, aucune étude n'ayant été menée chez les mammifères marins. Par conséquent, la signification biologique des différences de composition protéique entre espèces ne semble pas explicable d'un point de vue nucléique. Finalement, quelque soit l'espèce, son sérum contient toujours un **hétérotrimère** et au moins un **hétérodimère** de la famille de l' $I\alpha I$. Par ailleurs, le complexe tricaténaire n'est jamais constitué de deux chaînes lourdes identiques associées à la bikunine. Chez une même espèce, il peut exister jusqu'à trois types d'hétérodimères. De ces faits, il semble évident que les membres de la famille de l' $I\alpha I$ subissent au cours de leur biosynthèse un assemblage sélectif aboutissant à la formation de complexes protéiques bien définis. Au sein des précurseurs polypeptidiques ou des chaînes lourdes elles-mêmes, un signal de nature conformationnelle ou structurale doit donc exister.

II Une famille de molécules originales

1) Des activités biologiques variées

a) Activité anti-inflammatoire

-Antiprotéase

Les antiprotéases exercent dans de nombreux processus métaboliques (coagulation, fibrinolyse, inflammation, activation du complément) un rôle de régulation et de contrôle indispensable à l'organisme. Dans le plasma humain, celles-ci représentent environ 10% des protéines circulantes. L'activité antiprotéasique des membres de la famille de l' α_1 I est exclusivement portée par la bikunine. La conformation tertiaire de la chaîne peptidique de la bikunine forme deux domaines inhibiteurs de type Kunitz [29] arrangés en tandem (d'où son nom : bi-Kunitz). Ils sont responsables de l'activité inhibitrice de la bikunine vis-à-vis des sérine-protéinases. Les complexes sériques de la famille de l' α_1 I, par l'intermédiaire de la bikunine inhibent la trypsine, la chymotrypsine mais surtout des protéinases impliquées dans les mécanismes de la réaction inflammatoire telles que l'élastase et la cathepsine G leucocytaires ou la plasmine [30-32]. Cependant, selon ces auteurs, les constantes d'inhibition de l' α_1 I et de la bikunine pour ces différentes sérine-protéinases sont trop faibles pour traduire un effet physiologique.

Une des propriétés inhabituelles des membres de la famille de l' α_1 I est leur facilité relative à être dégradée par les enzymes qu'ils inhibent. En cas de syndrome inflammatoire, l'activation des protéinases leucocytaires, notamment l'élastase et la cathepsine G, induit une protéolyse de l' α_1 I [31-34] en dérivés de plus faibles masses moléculaires conservant l'activité inhibitrice. Parmi eux, se forme la bikunine. Celle-ci pourrait, par diffusion à travers les espaces intercellulaires, participer à la protection des tissus car elle est plus stable (acido-stable) et conserve son activité antiprotéasique. Dès lors, l' α_1 I apparaît comme le **précurseur physiologique** d'inhibiteurs de protéinases diffusibles et très stables. Cette propriété est dorénavant parfaitement comprise compte tenu de la structure

particulière des membres de la famille de l'I α I. Néanmoins, considérer que l'I α I, le P α I ou les autres complexes sériques ont pour seule fonction de cibler la bikunine vers des foyers inflammatoires est excessif. Comme nous le verrons ultérieurement, les chaînes lourdes participent également à la réparation tissulaire.

La réaction entre inhibiteurs et sérine-protéinases est généralement caractérisée par la formation d'un complexe **irréversible** entre l'inhibiteur et la protéinase [35] induisant des changements de conformation de l'inhibiteur qui permettent sa reconnaissance par son récepteur cellulaire en vue de l'élimination plasmatique du complexe [inhibiteur-protéinase]. Contrairement à ces antiprotéases telles que l' α 1-protéinase inhibiteur (α 1-PI), l'I α I ou la bikunine sont considérés comme des inhibiteurs transitoires. En effet, ils forment avec les protéinases des complexes **réversibles** qui sont, par ailleurs, rapidement dissociés en présence d'autres inhibiteurs de protéinases plus efficaces comme l' α 1-PI. Mis en évidence pour l'élastase leucocytaire, la plasmine et la trypsine [36, 37], cette propriété a abouti à la notion d'inhibiteurs "navettes" capables d'inhiber transitoirement certaines sérine-protéinases puis de les transférer vers les inhibiteurs responsables de leur clairance plasmatique. Les protéinases, associées en complexe avec l'I α I, sont d'ailleurs plus rapidement inhibées par les inhibiteurs secondaires que ces mêmes protéinases à l'état libre.

Finalement, l'I α I peut-il être considéré comme un agent biologique impliqué dans les mécanismes limitant l'inflammation ? (i) L'I α I inhibe de façon transitoire certaines sérine-protéinases. (ii) Malgré sa dégradation protéolytique, la fonction inhibitrice n'est pas altérée mais au contraire augmentée. (iii) L'I α I améliore la vitesse d'élimination de ces protéinases par des inhibiteurs plus efficaces. Une représentation schématique du mécanisme d'action de l'I α I vous est présentée dans la figure 2.

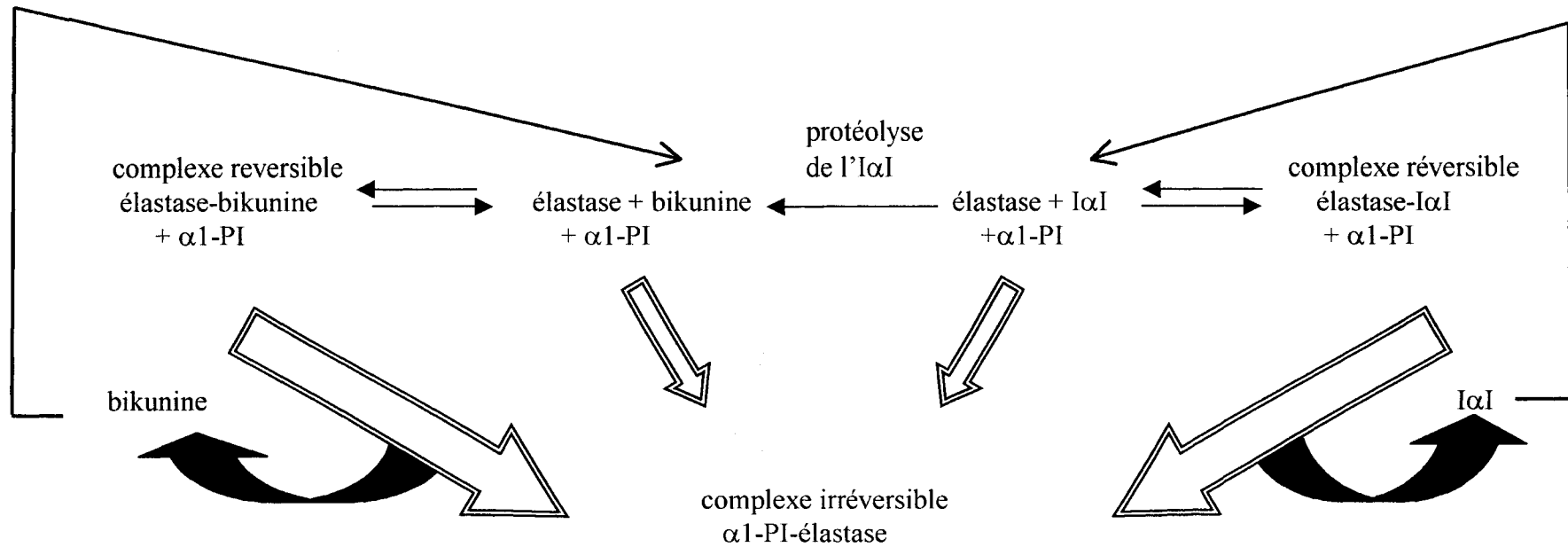


Figure 2 : Représentation schématique du mécanisme d'action de l'I α I

-Stabilisation de la membrane des lysosomes

Facilement isolable de l'urine humaine, la bikunine est préparée industriellement au Japon et est commercialisée sous le nom d'Ulinastatine (MiraclidTM, Mochida Pharmaceutical Co. Ltd.). Sa disponibilité a favorisé le développement d'un grand nombre d'études destinées à mettre en évidence ses effets thérapeutiques dans différentes affections telles que pancréatites aiguës, chocs septiques ou post-opératoires, insuffisance rénale aiguë...La pancréatite aiguë consiste en une réaction inflammatoire localisée, due à l'autodigestion du pancréas, induite par l'activation intracellulaire des enzymes qu'il produit, notamment la trypsine. Chez le rat, quelque soit le mode d'induction, l'administration d'ulinastatine améliore de manière significative les taux de survie. Elle s'avère plus efficace que l'aprotinine (inhibiteur pancréatique bovin de type Kunitz). Biochimiquement, la bikunine inhibe en partie l'activité trypsique dans le pancréas. Histologiquement, sa présence semble diminuer d'une part le phénomène de co-localisation des enzymes digestives et lysosomales responsables de l'activation prématurée du trypsinogène et d'autre part diminuer l'infiltration des polymorphonucléaires neutrophiles (PMN) limitant ainsi la réaction inflammatoire [38, 39].

Des effets bénéfiques similaires sur la membrane des lysosomes ont également été rapportés lors d'insuffisance rénale aiguë induite chez le rat par l'intermédiaire de doses massives de gentamicine [40] (antibiotique utilisé pour combattre les infections bactériennes à Gram -). L'excrétion urinaire de la N-acétyl- β -D-glucosaminidase (NAG) est un témoin de l'altération des tubules rénaux. En cas d'insuffisance rénale, la présence de la bikunine limite l'excrétion de NAG et les anomalies histologiques. *In vitro*, celle-ci diminue la libération de NAG par les lysosomes isolés, ce qui suggère qu'elle pourrait agir *in vivo* en stabilisant les membranes des lysosomes. Récemment, Kato et coll [41] viennent de préciser la partie de la bikunine responsable de la stabilisation membranaire. La bikunine intacte supprime la libération de NAG, tandis que d'autres inhibiteurs de sérine-protéinases (aprotinine, gabexate mesilate (FOY), domaine II recombinant de UTI) n'ont aucun effet, indiquant que l'activité inhibitrice vis-à-vis des sérine-protéinases n'est pas impliquée. En fait, testant différents dérivés de l'UTI, ces

auteurs ont démontré que l'association de l'accepteur peptidique et de la chaîne de GAG constitue la structure biochimiquement active dans ce processus de stabilisation de la membrane des lysosomes. Néanmoins, aucun mécanisme physiologique n'est défini pour l'instant.

-Inhibition de la production des cytokines pro-inflammatoires -Rôle de la régulation du flux calcique cytosolique par la bikunine

Lors de processus inflammatoires et particulièrement au cours de certains cancers et d'infections bactériennes, l'activité anti-trypsique urinaire est fortement augmentée. Celle-ci est due à des taux urinaires de bikunine très élevés. Par conséquent, le dosage (aisé, automatisable et spécifique) de l'activité anti-trypsique urinaire due à la bikunine a été proposé comme marqueur fiable de la survenue de syndromes infectieux [42]. Les effets bénéfiques de la bikunine dans le traitement des chocs septiques induits par l'inoculation d'endotoxine ou de bactéries vivantes ont donc été étudiés dans des modèles animaux. Chez le chien, l'administration de bikunine induit une diminution du nombre de bactéries et du taux d'endotoxines libérées dans le flux sanguin [43] qui aboutit à une amélioration des constantes hémodynamiques et une augmentation significative des taux de survie.

La membrane des bactéries à Gram – est constituée de lipopolysaccharides (LPS ou endotoxine). La présence de LPS dans le sang induit l'activation des monocytes et macrophages. Ceux-ci produisent alors des cytokines pro-inflammatoires TNF- α (tumor necrosis factor α) et IL-1 (interleukine-1) qui potentialisent le choc septique. Dans les années 90, Endo et coll [44] rapportaient que la bikunine inhibe de manière dose-dépendante la production du TNF- α , de IL-1 α et de IL-1 β limitant ainsi les taux sanguins de ces médiateurs pro-inflammatoires considérés comme responsables de la survenue du choc septique.

Le LPS, le TNF- α ou encore IL-1 régulent, dans diverses cellules, la synthèse d'une autre cytokine inflammatoire l'IL-8, responsable du chimiotactisme des PMN et de la libération de leurs enzymes lysosomales, notamment l'élastase

[45]. Maehara et coll [46] ont démontré que dans une lignée cellulaire de leucémie humaine (HL60) stimulée par le LPS, la bikunine inhibe l'expression des ARNm de l'IL-8. Par ce processus, la bikunine pourrait exercer un contrôle négatif sur sa production puisqu'elle limiterait l'infiltration des PMN et la production d'élastase responsable de la protéolyse de l'I α I. Ces auteurs suggèrent que cette inhibition de la synthèse de IL-8 induite par le LPS est liée à la régulation des concentrations de calcium (Ca^{2+}) cytosolique puisque la bikunine n'inhibe pas la production ou la libération de l'IL-8 induite par un ionophore calcique (A23187). Par la suite, la même équipe de recherche rapportait également que la préincubation de neutrophiles ou de cellules endothéliales en présence de bikunine inhibe l'augmentation du Ca^{2+} libre cytosolique induite par le LPS [47]. Cette activité est attribuée à une séquence peptidique présente dans le deuxième domaine inhibiteur de type Kunitz de la bikunine. Outre l'activité d'inhibiteur de protéase, ce domaine serait donc impliqué dans la régulation du calcium cytosolique. Cependant, contrairement à la bikunine native, la bikunine ayant perdu son caractère protéoglycannique soit par déglycosylation (chondroitinase ACII) soit par l'utilisation d'anticorps monoclonaux dirigés contre la chaîne de chondroïtine sulfate, voit son pouvoir d'inhibition sur l'augmentation calcique intracellulaire induit par le LPS, partiellement diminuer [48]. Les GAG, par leur caractère fortement anionique à pH physiologique, sont des chélateurs potentiels de cations. Dans le cas de la bikunine, Kanayama et coll [48] ont établi qu'une molécule de bikunine fixe, de manière dose-dépendante jusqu'à 70 molécules de calcium par l'intermédiaire de sites de faible et de forte affinité. Cette capacité de liaison du calcium est fortement réduite pour la bikunine déglycosylée de même que pour l'I α I. Par conséquent, la bikunine est un **chélateur physiologique du calcium**. D'ailleurs, dans l'urine humaine, l'UAP (uronic-acid-rich protein) qui semble correspondre à une partie ou à la bikunine elle-même, inhibe la formation de cristaux d'oxalate de calcium [49]. Néanmoins, dans ce processus, la partie peptidique de la bikunine joue un rôle majoritaire par rapport à la partie protéoglycannique.

-La bikunine inhibe la production de prométalloprotéinases

Les métalloprotéinases (MMP : matrix metalloproteinase) sont les principales protéines responsables de la dégradation de la matrice extracellulaire. Sous l'action des cytokines pro-inflammatoires (IL-1 α), des cellules humaines telles que les fibroblastes utérins ou des cellules chorioniques produisent et libèrent dans l'espace interstitiel les précurseurs de métalloprotéinases (pro-MMP) en même temps que leurs inhibiteurs endogènes (TIMP : tissue inhibitor of metalloproteinases). Ces pro-MMP sont ensuite protéolytiquement activées par les protéinases (ex : la plasmine). *In vitro*, lorsque ces cellules sont incubées 24h en présence de l'IL-1 α et de la bikunine, l'excrétion extracellulaire ainsi que les taux d'ARNm de deux précurseurs de métalloprotéinases, la procollagénase (pro-MMP-1) et la prostromelysin 1 (pro-MMP-3), diminue [50]. Ces auteurs suggèrent que la bikunine contrôle la production de la pro-MMP1 et de la pro-MMP-3 au niveau transcriptionnel et post-transcriptionnel. De ces faits, il apparaît que la bikunine participerait à la protection de l'intégrité de la matrice extracellulaire. Cependant, d'autres études sont nécessaires pour confirmer et éclaircir les mécanismes mis en jeu.

-Modulation de la croissance cellulaire

Les cellules d'hépatome humain produisent deux polypeptides capables de maintenir ou d'augmenter la croissance des cellules endothéliales [51]. Il s'agit de deux inhibiteurs de sérine-protéases : l'inhibiteur sécrétoire pancréatique de la trypsine (PSTI) et la bikunine. Par contre, ceux-ci n'affectent pas la croissance des hépatomes qui les produisent ni celle des cellules musculaires lisses ou encore des fibroblastes. Inversement, la bikunine exerce un effet antimitotique vis-à-vis d'une lignée cellulaire de lymphome de Burkitt [52]. Néanmoins, puisqu'elle n'inhibe pas la croissance d'autres lignées cellulaires cancéreuses, ces auteurs ont suggéré l'existence de récepteurs cellulaires spécifiques présents ou non à la surface des cellules. En 1994, Perry et coll [53] ont rapporté, dans le cas des fibroblastes, un effet régulateur de croissance dose-dépendant (stimulant à faible dose, inhibant à forte dose). Ceci les a conduit à émettre l'hypothèse que ces effets cellulaires

opposés pourraient être médiés par des systèmes distincts (récepteurs d'affinité différente). Peu de connaissances sont disponibles à propos des mécanismes biochimiques par lesquels les inhibiteurs de protéinases stimulent ou inhibent la croissance cellulaire. Toutefois, ils sont souvent reliés métaboliquement aux facteurs de croissance autocrine ou non. Par exemple, une bikunine capable d'inhiber une protéinase activant l'HGF (hepatocyte growth factor) a été isolée du placenta et de plusieurs tissus cancéreux [54]. La chaîne de chondroïtine sulfate peut s'avérer également primordiale dans les processus de prolifération cellulaire. En effet, il est désormais couramment admis que les protéoglycannes, par l'intermédiaire de leur GAG contrôlent en partie la croissance cellulaire par les interactions moléculaires qu'ils forment avec les facteurs de croissance [55, 56].

L'ensemble de ces travaux témoigne de la diversité des activités biologiques exercées *in vitro* ou *in vivo* par la bikunine. Les mécanismes proposés sont disparates et rendent compte de la **double nature biochimique** de la bikunine : protéique et glycosaminoglycannique.

b) Activités anti-métastatiques

-Récepteurs cellulaires

Comme nous l'avons précédemment signalé, la bikunine inhibe la plasmine et est massivement présente (200 à 500mg/l) dans les urines de patients atteints de cancer, principalement métastasant. Lors de rémission chimiothérapeutique, cette concentration urinaire diminue fortement [57, 58].

Le pouvoir invasif des cellules tumorales dépend majoritairement des enzymes suivantes : uPA (urokinase-type plasminogen activator), plasmine, cathepsine B et D ou encore MMP. Ces cellules produisent l'uPA sous une forme inactive (pro-uPa). Le pro-uPA libre ou lié au plasminogène se fixe alors spécifiquement sur son récepteur cellulaire (uPAR : urokinase-type plasminogen activator receptor) où il protéolyse le plasminogène en plasmine. Celle-ci active ou

induit alors la cascade protéolytique (ex : pro-MMP en MMP) responsable de la dégradation de la matrice extracellulaire et de la membrane basale.

Dans des modèles de cellules cancéreuses ovariennes humaines (*in vitro*) ou pulmonaires de souris (*in vivo*), la bikunine inhibe de façon dose-dépendante l'invasion tumorale [59, 60]. Cette inhibition est corrélée à l'activité inhibitrice de la bikunine vis-à-vis de la plasmine [59, 60]. En fait, la bikunine possède deux récepteurs présents à la surface des cellules tumorales ou même des neutrophiles. Purifiés à partir de cellules de choriocarcinome humain (SMT-CC1), la caractérisation biochimique de ces récepteurs révèle une parenté structurale, aussi bien peptidique qu'immunologique, avec une forme tronquée de la protéine de liaison du cartilage humain (human cartilage link protein) [61]. Une fois fixée, la bikunine est toujours capable d'inhiber la plasmine libre, voir même liée à son récepteur cellulaire, contrairement à d'autres inhibiteurs plus efficaces (α 2-antiplasmine et α 2-macroglobuline) mais qui n'inhibent que la plasmine en solution [60, 62]. Kobayashi et coll [63] ont démontré que la bikunine se fixe par l'intermédiaire de son premier domaine inhibiteur de type-Kunitz sur ses récepteurs, où elle pourrait encore inhiber la plasmine grâce à son second domaine inhibiteur (domaine II). Cependant, le domaine II seul, ainsi que différents peptides de synthèse issus de ce domaine, exercent essentiellement la même inhibition que la bikunine native, vis-à-vis de l'invasion tumorale. Par conséquent, l'hypothèse selon laquelle la bikunine fixée sur son récepteur est impliquée dans la prévention de l'invasion des cellules cancéreuses *in vitro* est à exclure [63]. En ce qui concerne l'I α I, la fixation à la surface des cellules tumorales n'est pas médiée par la bikunine mais, comme nous le verrons ultérieurement, dépend des chaînes lourdes capables de se fixer à l'acide hyaluronique (HA). Ainsi fixé, l'I α I est clivé et libère d'une part les chaînes lourdes toujours associées à l'HA et d'autre part la bikunine apte à se fixer sur ses récepteurs cellulaires [64].

-Une protéine composite (uPA-UTI) comme agent anti-métastatique

La bikunine inhibe l'invasion tumorale mais n'exerce aucun effet sur la prolifération de la tumeur. Par conséquent, comme rapporté par Kobayashi [65], l'utilisation simultanée d'un agent anti-tumoral (etoposide) et de la bikunine pourrait constituer un traitement anti-métastatique prometteur. D'une autre manière, puisque l'uPA est enzymatiquement responsable de la transformation du plasminogène en plasmine, ces auteurs ont développé deux modèles de protéines composites (uPA-UTI) destinées d'une part à améliorer le ciblage cellulaire et d'autre part à diminuer la fixation du pro-uPA endogène sur son récepteur espérant ainsi augmenter le pouvoir anti-invasif. Pour ce faire, ils ont covalentiellement associé la séquence en acides aminés responsable de la fixation de uPA sur son récepteur au domaine II de la bikunine, soit par pontage entre les résidus de cystéines des deux molécules [66], soit par modification génétique et production d'une protéine recombinante (uPA-UTI) [67]. Dans les deux cas, l'activité inhibitrice sur l'invasion tumorale est augmentée, laissant entrevoir une utilisation thérapeutique de cette protéine recombinante en pathologie humaine.

c) Stabilisation de la matrice extracellulaire

A côté des propriétés biologiques de la bikunine, les chaînes lourdes des membres de la famille de l'I α I sont impliquées dans les processus de stabilisation de la matrice extracellulaire.

La matrice extracellulaire est composée de plusieurs macromolécules associées en un réseau complexe qui assure la cohésion cellulaire au sein d'un tissu donné. Par l'intermédiaire de ses différents constituants, elle intervient dans des mécanismes aussi divers que l'adhésion, l'agrégation, la migration et la prolifération cellulaires et module l'activité métabolique des cellules qui sont à son contact. Cette matrice extracellulaire est principalement constituée de collagène, de glycoprotéines, de protéoglycannes et d'acide hyaluronique.

L'acide hyaluronique (AH) est un glycosaminoglycane ubiquitaire, linéaire, de haut poids moléculaire, constitué de la répétition de l'unité disaccharidique [N-acétyl-D-glucosamine (β 1-4) acide D-glucuronique (β 1-3)]_n qui n'est jamais sulfatée. Les activités physiologiques relatives à l'AH sont très souvent corrélées aux interactions spécifiques qu'il est susceptible d'établir avec les autres molécules telles que les protéines et les protéoglycannes. Parmi les nombreuses protéines capables de se lier à l'AH, telles que la protéine de liaison du cartilage, figurent les protéines de la famille de l'I α I.

- l'I α I et l'AH : une vieille histoire

La présence d'I α I au sein du liquide synovial pathologique de patients atteints de polyarthrite rhumatoïde a été primitivement rapportée par Sandson et coll [68]. Par la suite, Jessen et coll [69] ont identifié, dans le liquide synovial pathologique mais également dans le liquide folliculaire humain normal, des complexes entre l'AH et les chaînes lourdes des membres de la famille de l'I α I. Partant d'observations relatives au milieu de culture contenant du sérum de veau fœtal, Yoneda et coll [70], quant à eux, purifiaient sans l'identifier une protéine sérique bovine de 85 kDa formant des interactions **covalentes** avec l'AH.

Chen et coll [71] ont été les premiers à assigner une fonction physiologique aux interactions entre l'AH et les membres de la famille de l'I α I. *In vitro*, ces auteurs ont mis en évidence la présence, dans le sérum de veau fœtal, d'un facteur (apparenté à l'I α I) indispensable pour l'organisation et la stabilisation de l'expansion du cumulus oophorus entourant les follicules ovariens de souris. *In vivo*, cette expansion est un pré-requis obligatoire avant l'ovulation et s'accompagne de la synthèse d'AH par les cellules folliculaires (mucification). Stimulée *in vitro*, cette synthèse d'AH participe, en présence de sérum de veau fœtal, à l'édification de la matrice extracellulaire. Au contraire, en l'absence de sérum, l'AH est libéré dans le milieu de culture et aucune expansion de la matrice extracellulaire n'est observable, faisant de ce facteur un composé structural **prédominant**. Un an plus tard, ce facteur du sérum de veau fœtal était identifié

comme étant le P α I bovin (I α I H2,B) [25]. En 1994, Chen et coll [72] rapportaient un rôle similaire pour la molécule d'I α I bovine ou humaine. Ils démontraient également l'existence d'interactions ioniques médiées par les résidus de lysine entre l'AH et le P α I bovin. Dans le même temps, Camaioni et coll [73] caractérisaient la structure glycosaminoglycannique minimale (5 unités disaccharidiques) capable de s'associer à ces facteurs sériques.

Lors de travaux menés en parallèle, Huang et coll [74], quant à eux, isolaient du sérum de veau fœtal et du sérum humain, un complexe covalent formé de l'AH et de protéines dénommé SHAPs (serum-derived hyaluronan-associated proteins). Après incubation de ces complexes en présence de hyaluronidase puis séquençage, deux protéines étaient identifiées : H1 et H2. La région de la chaîne H1 liant l'AH était localisée dans la région C-terminale de la chaîne peptidique. Par conséquent, selon ces auteurs, le même pontage PGP existant entre les chaînes lourdes et la bikunine pouvait exister au sein du complexe entre l'AH et les chaînes lourdes. Isolant ce complexe de liquides synoviaux pathologiques, Zhao et coll [75] ont identifié, par spectrométrie de masse, la nature covalente de cette liaison. Il s'agit effectivement d'une liaison ester, analogue à celle du pontage PGP caractérisant les membres de la famille de l'I α I, entre l'un des groupements carboxyliques de l'acide aspartique C-terminal des chaînes lourdes et la fonction hydroxyle du carbone 6 d'un résidu de N-acétylglucosamine de l'AH (Figure 3).

Selon certains auteurs [76], l'incubation de l'I α I ou de P α I purifiés avec des quantités variables d'AH seul, ne génère aucun complexe covalent détectable (chaînes lourdes et AH), ni de bikunine libre. Cependant, la présence de cellules de la granuloza de souris ou de leur milieu de culture, induit la formation des complexes covalents et la libération, dans le milieu de culture, de bikunine libre. Par contre, ce milieu de culture chauffé n'induit pas l'association des chaînes lourdes et de l'AH, laissant supposer l'existence d'une enzyme, synthétisée et excrétée par les cellules de la granuloza, capable de catalyser l'estérification de l'AH par les chaînes lourdes. La présence de chélateur de calcium (EDTA : ethylene diamine tetraacetic acid) dans les milieux d'incubation empêche également la formation des complexes covalents [74, 77].

L'un des groupements carboxyliques de l'acide aspartique C-terminal estérifie la fonction hydroxyle du carbone 6 d'un résidu de N-acétylglucosamine interne de l'AH

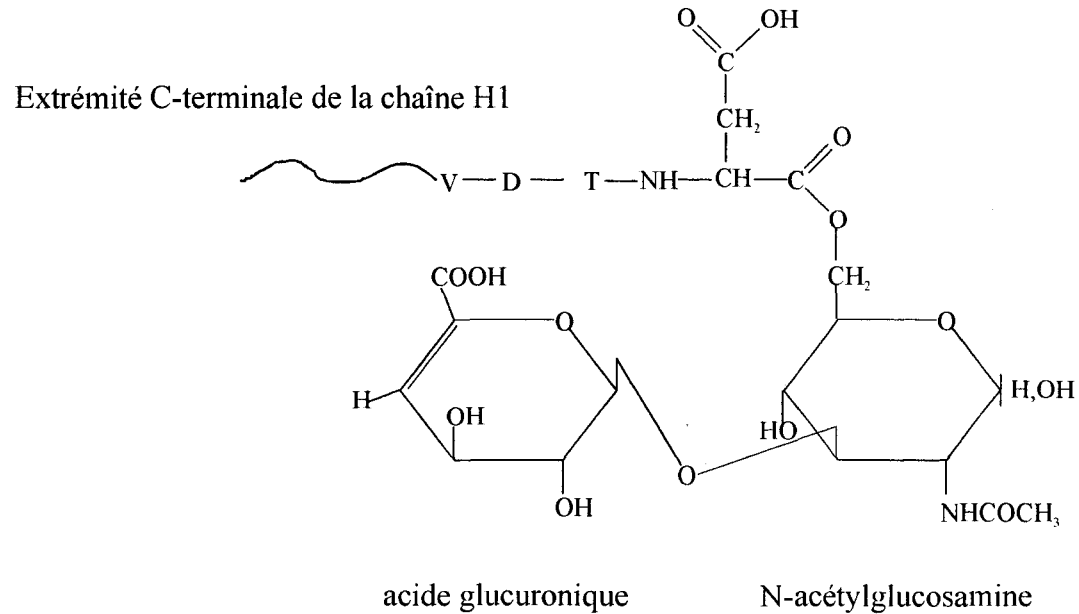


Figure 3 : Représentation schématique de la liaison ester unissant les chaînes lourdes H1, H2 et H3 des membres de la famille de l' α I et l'acide hyaluronique (AH). La liaison insaturée résulte de la digestion de l'AH par la hyaluronidase.

Thomas et coll [77] caractérisaient en 1998, la présence de bikunine libre dans le sérum et la présence de la bikunine libre, d'I α I et de P α I dans le liquide péritonéal de patients soumis à une dialyse péritonéale continue. Ce milieu contient des protéines sériques, en raison de la fuite du sérum dans la cavité péritonéale, et de l'AH synthétisé par les cellules mésothéliales du péritoine. Trois années plus tôt, Blom et coll [78] avaient démontré la participation de l'I α I dans la stabilisation du manteau péricellulaire des fibroblastes et des cellules mésothéliales en culture. Par conséquent, Thomas et coll [77] ont émis l'hypothèse selon laquelle, dans la cavité péritonéale de ces patients, l'I α I, interagissant avec l'AH, libère la bikunine (retrouvée dans le liquide péritonéal et dans le sérum) et forme parallèlement des complexes chaînes lourdes-AH qui stabilisent le manteau péricellulaire des cellules mésothéliales du péritoine. Une fois encore, l'I α I et l'AH seraient donc impliqués dans le maintien de la cohésion tissulaire.

De ces faits, ressort un mécanisme (Figure 4) où, dans un premier temps, l'I α I ou le P α I interagissent avec l'AH à la surface des cellules par des liaisons ioniques réversibles. Par la suite, une enzyme semble catalyser la réaction de transestérification permettant l'incorporation de l'AH en lieu et place de la bikunine [76]. Dans toutes ces étapes, la présence de calcium apparaît primordiale.

D'un point de vue structural, les membres de la famille de l'I α I ne présentent aucune communauté avec les deux classes de protéines liant l'AH (HABP : hyaluronic acid binding protein) dont les leaders sont d'une part la protéine de liaison du cartilage (présence de séquence PTR : proteoglycan tandem repeat) et d'autre part le RHAMM (receptor for HA-mediated motility). Par ces différences structurales et de mécanismes d'interaction, les membres de la famille de l'I α I pourraient constituer une **nouvelle catégorie** d'HABP [79].

- L'I α I et la protéine TSG-6

La stimulation des fibroblastes, des chondrocytes et des cellules synoviales par des cytokines pro-inflammatoires telles que le TNF- α et l'IL-1 induit la

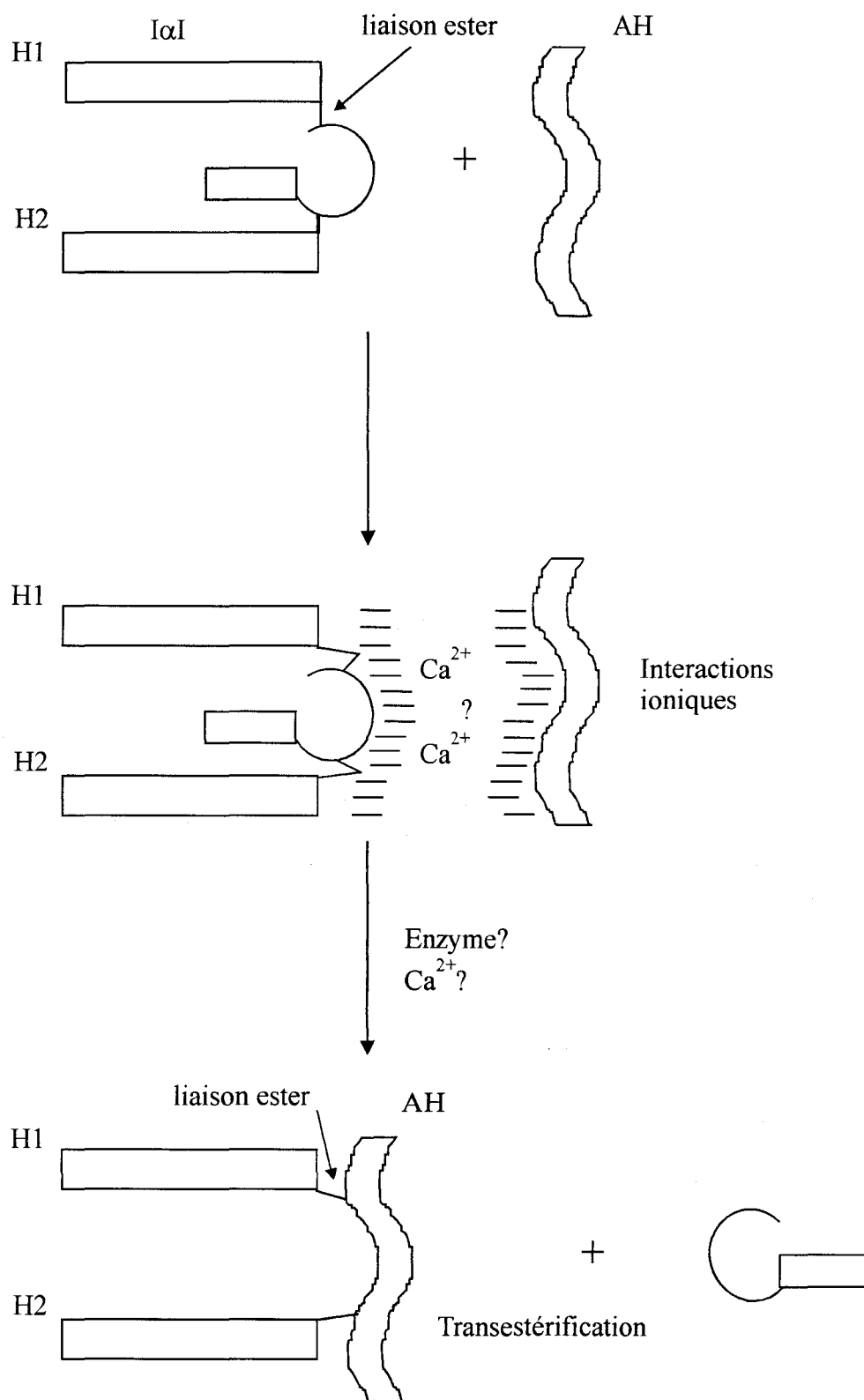


Figure 4: Représentation schématique du mécanisme d'interaction entre les membres de la famille de l'I α I et l'AH

synthèse d'une protéine dénommée TSG-6 (TNF-stimulated gene 6 protein). Cette glycoprotéine comme le CD44 ou la protéine de liaison du cartilage, appartient à la famille des protéines capables de s'associer à l'AH : les hyaladhérines.

Cette protéine TSG-6 (35 kDa) est retrouvée en grande quantité dans les liquides synoviaux de patients atteints de maladies articulaires. Dans ces milieux, la protéine TSG-6 forme également un complexe stable (120 kDa) avec une protéine provenant de l'I α I. Par microséquençage, Wisniewski et coll [80] ont démontré que ce complexe renfermait la TSG-6, la chaîne H2 et la bikunine. Sous l'action de la chondroitinase ABC, ce complexe stable se dissocie en ses différents composés suggérant que la chaîne de chondroïtine sulfate portée par la bikunine est impliquée dans sa formation.

Dans un modèle d'inflammation aiguë chez la souris, l'injection de la protéine TSG-6 au niveau du foyer inflammatoire inhibe l'infiltration des PMN. L'analyse des protéines de l'exsudat inflammatoire révèle l'existence du complexe TSG-6/H2/B. Wisniewski et coll [81] ont démontré qu'ainsi assemblée, la protéine TSG-6 potentialise l'activité antiplasmine de la bikunine. Par l'intermédiaire des nombreuses propriétés de la bikunine, ce complexe semble donc être le facteur responsable de l'activité anti-inflammatoire observée.

-Chaînes lourdes et protéoglycannes

Très récemment, deux protéoglycannes différant par leur taille et leur composition ont été isolés du liquide folliculaire humain. L'un d'eux présente une masse moléculaire estimée à 3000 kDa et possède 15 à 20 chaînes de GAG de type chondroïtine sulfate. Ce protéoglycane traité par la chondroitinase ABC, libère 2 protéines : l'une de 600 à 700 kDa et l'autre de 70 kDa. La plus imposante réagit avec un anticorps dirigé contre le versicanne humain tandis que la protéine de 70 kDa est immunologiquement détectée par un anticorps anti-I α I humain [82]. Par conséquent, les auteurs suggèrent que cette dernière protéine correspond à une des

chaînes lourdes de l'I α I et qu'elle pourrait être liée aux chaînes de chondroïtine sulfate du protéoglycane par l'intermédiaire d'un pont PGP.

Finalement, les connaissances actuelles s'accordent sur le fait que les protéines de la famille de l'I α I possèdent la faculté d'interagir directement ou indirectement avec l'acide hyaluronique au sein de la matrice extracellulaire. Induit par différents processus encore inconnus, la dissociation moléculaire de l'I α I libère d'une part la bikunine, toujours à même de limiter l'intensité des phénomènes inflammatoires et de contribuer à la régénération tissulaire, et d'autre part les chaînes lourdes qui associées de manière covalente à l'AH deviennent des composants primordiaux dans le maintien de l'intégrité des matrices extracellulaires (Figure 5).

Bikunine:

- inhibition de protéinases
- stabilisation de la membrane des lysosomes
- inhibition de la production de cytokines pro-inflammatoires
- modulation de la croissance cellulaire
- inhibition de la synthèse de prométalloprotéinases
- activités anti-métastatiques

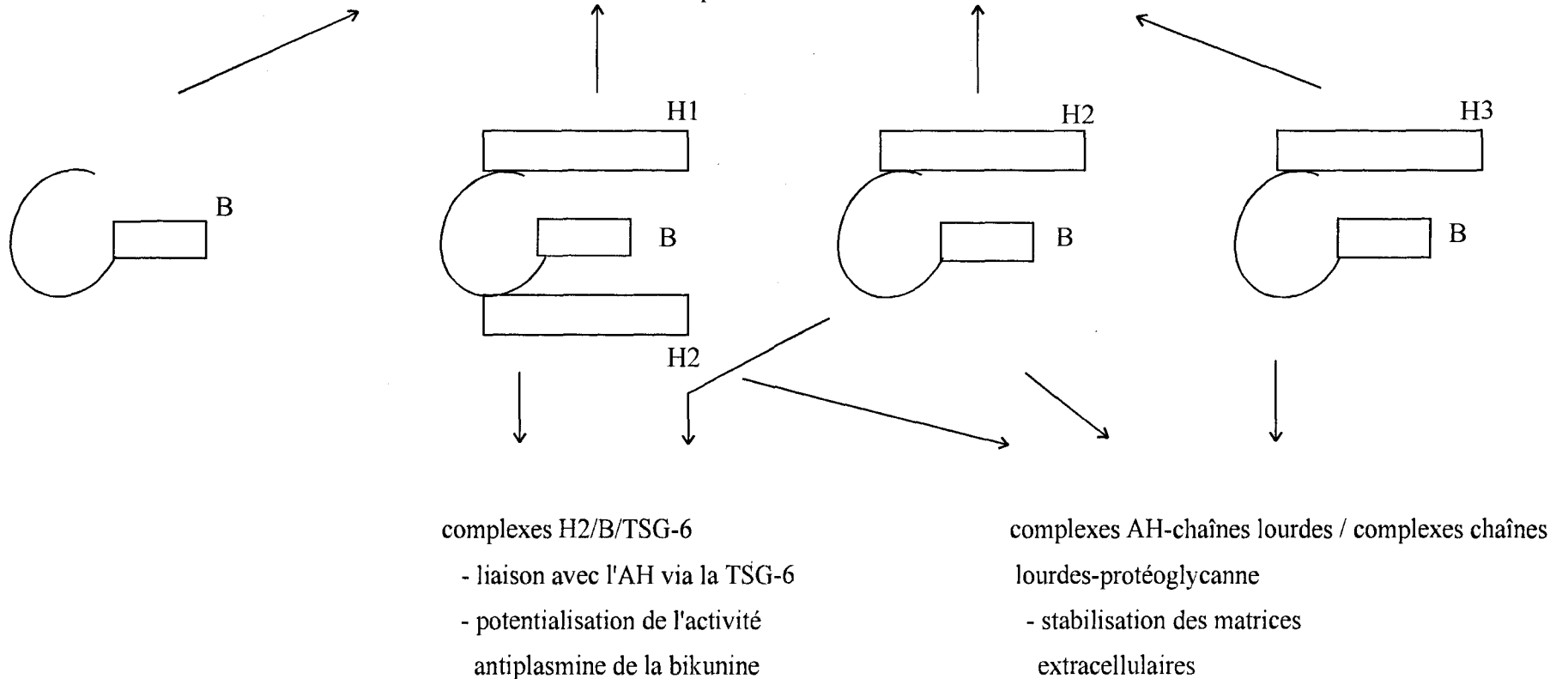


Figure 5: Fonctions physiologiques de la bikunine et des complexes sériques appartenant à la famille de l' α I.
(schéma selon Bost et coll [79] modifié)

2) Une structure originale et caractéristique

Les protéoglycannes (PG) sont présents dans les tissus animaux et sont synthétisés par la majorité des types cellulaires. Ce sont des molécules très complexes à localisation extracellulaire, membranaire ou intracellulaire constituées d'une à une centaine de chaînes de glycosaminoglycannes (GAG) liées de manière covalente à un axe protéique unique. Comme nous le verrons en détail ultérieurement, les complexes de la famille de l'I α I sont les exceptions qui confirment la règle. La double nature biochimique des PG est à la base de leur très grande capacité à interagir avec d'autres constituants matriciels, des composants membranaires et des médiateurs solubles. Par le biais de ces multiples interactions, les PG jouent des rôles structuraux et fonctionnels au sein des matrices extracellulaires [56].

a) Diversité des glycosaminoglycannes

Les GAG sont des polysides linéaires composés de la répétition d'unités disaccharidiques de structures définies qui contiennent toujours une hexosamine et soit un acide hexuronique, soit un galactose. Il existe quatre catégories de GAG : (i) l'acide hyaluronique, (ii) les chondroitines et dermatanes sulfates, (iii) les kératanes sulfates et (iv) les héparanes sulfates et l'héparine.

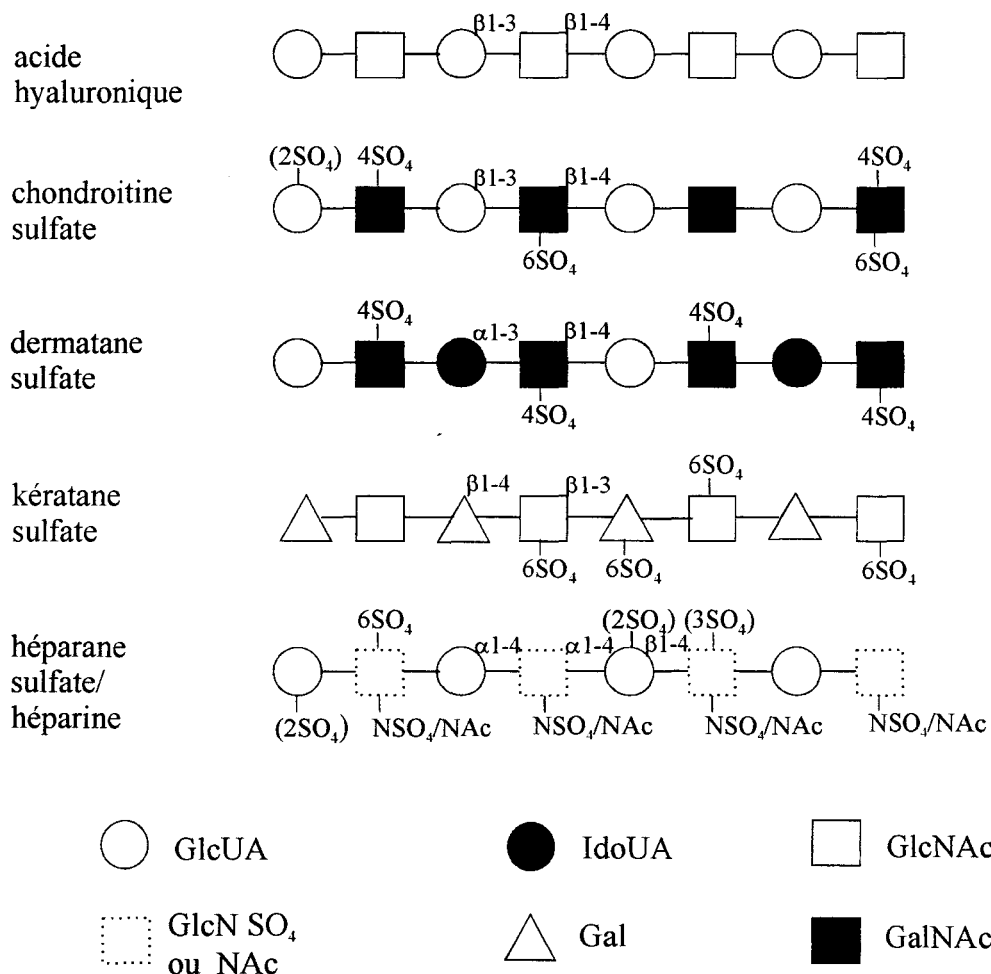
Comme nous l'avons signalé précédemment, l'AH est un GAG non-sulfaté, constitué de la répétition de l'unité disaccharidique suivante : [N-acétylglucosamine (β 1-4) acide D-glucuronique (β 1-3)]_n. Les chondroitines sulfates (CS) possèdent la même ossature osidique que l'AH excepté que le résidu de N-acétylglucosamine (GlcNAc) est remplacé par un résidu de N-acétylgalactosamine (GalNAc). Les CS se caractérisent donc par la répétition du chaînon osidique : [GalNAc (β 1-4) GlcUA (β 1-3)]_n. Les carbones 4 et/ou 6 des résidus de GalNAc et/ou le carbone 2 des résidus de GlcUA peuvent être substitués par un groupement sulfate. Par rapport aux CS, les dermatanes sulfates (DS) résultent de l'épimérisation de certaines molécules d'acide glucuronique (GlcUA)

en acide L-iduronique (IdoUA). L'unité répétée est alors : [GalNAc (β 1-4)-IdoUA (α 1-3)]_n. La O-sulfatation des DS a lieu respectivement sur les carbones 4 et/ou 2 de la GalNAc et de l'IdoUA. Les kératanes sulfates (KS), quant à eux se composent de la répétition de GlcNAc et de galactose (Gal) unis respectivement par une liaison β (1-3) et β (1-4). Sur ces chaînes, la O-sulfatation s'effectue généralement sur les carbones 6 de l'une ou de l'ensemble des unités osidiques. Il s'agit donc d'une structure polyN-acétyllactosaminique O-sulfatée. Les héparanes sulfates et l'héparine ont une structure disaccharidique répétée distincte des autres GAG. La GlcNAc est liée par une liaison α 1-4 à un résidu de GlcUA, lui-même étant associé au résidu de GlcNAc suivant par une liaison (β 1-4) : [GlcNAc (α 1-4) GlcUA (β 1-4)]_n. Après l'élongation, cette ossature polyosidique subit de nombreuses modifications telles que l'épimérisation de certains GlcUA en IdoUA, des O-sulfatations mais surtout la N-sulfatation en 2 de la glucosamine [83, 84]. Une représentation schématique de ces structures est présentée figure 6A.

Toutes ces substitutions forment, le long des chaînes de GAG, des motifs précis capables d'interagir avec d'autres protéines. Parmi ceux-ci, l'exemple le mieux caractérisé est celui de l'unité pentasaccharidique minimale conférant à l'héparine son activité anticoagulante grâce à son affinité pour un site spécifique porté par l'anti-thrombine III [85].

Chaque chaîne de GAG de type CS, DS, HS et héparine est covalentiellement fixée, par l'intermédiaire d'une structure tétrasaccharidique caractéristique GlcUA β 1-3 Gal β 1-3 Gal β 1-4 xylose β 1-O, sur un résidu de sérine appartenant à une séquence primaire semi-consensus Ser-Gly-X-Gly [86]. Cependant, compte tenu du nombre de séquences primaires caractérisées à l'heure actuelle, la xylosyltransférase responsable de l'étape d'initiation de la biosynthèse de ces types de GAG semble plus probablement influencée par une conformation peptidique plutôt que par une séquence consensus universelle [84]. Ce tétrasaccharide de liaison peut également être substitué : présence de phosphoester sur le carbone 2 du xylose ou O-sulfatation des résidus de galactose. Cette dernière modification semble d'ailleurs être restreinte aux chaînes de CS [84]. L'originalité de liaison revient aux chaînes de KS. Dans la cornée, celles-ci sont N-liées à un

A



B

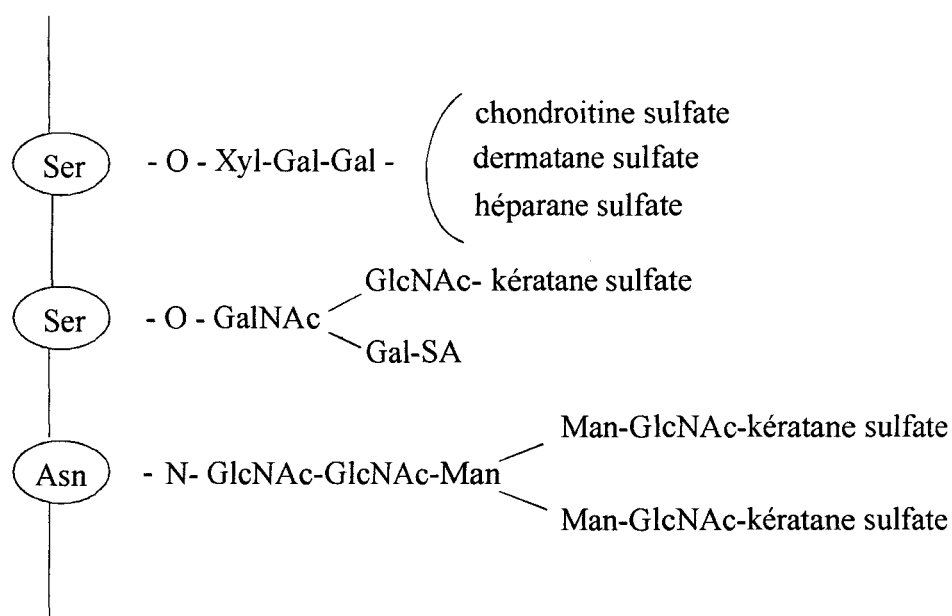


Figure 6 : Représentation schématique des structures osidiques des GAG (A) et de leur point d'ancrage sur la protéine (B) selon Hardingham et coll [88].

résidu d'asparagine via la structure pentasaccharidique (Mannose)₃-(GlcNAc)₂ communes à tous les N-glycannes. Au contraire, dans le cartilage, les chaînes de KS sont, par l'intermédiaire d'un résidu de GalNAc, O-liées sur un résidu de sérine ou de thréonine (Figure 6B). Parmi les GAG, l'AH est le seul GAG synthétisé sous une forme polysidique stricte par une enzyme membranaire [87]. Par conséquent, les interactions moléculaires (covalentes ou non) qu'il développe, n'interviennent que dans l'espace extracellulaire.

b) Les protéoglycannes

Compte tenu de l'hétérogénéité glycosaminoglycannique des PG, leur classification repose sur leur distribution cellulaire. Trois familles sont classiquement admises : les PG extracellulaires (agrécane, versicane, décorine), les PG membranaires (syndécane, CD44) et les PG intracellulaires généralement localisés dans les granules de sécrétion (serglycine). Dans le paragraphe suivant, nous développerons quelques exemples de PG destinés d'une part à démontrer la diversité structurale caractérisant cette famille de molécules et d'autre part à souligner l'originalité structurale des complexes de la famille de l'IαI.

Parmi les PG extracellulaires, l'agrécane est un protéoglycane de haute masse moléculaire présent dans le cartilage. Il est constitué d'une chaîne polypeptidique de 225 à 250 kDa sur laquelle sont fixées, en plus des N et O-glycannes, plus de 100 chaînes de GAG de type CS ou DS (Figure 7). L'ossature peptidique de l'agrécane est divisée en trois régions distinctes : une partie N-terminale composée de deux domaines globulaires G1 et G2, une zone centrale recevant les chaînes de GAG et une région C-terminale formant un domaine globulaire G3. Ces domaines globulaires renferment des motifs polypeptidiques précis, responsables des interactions moléculaires de l'agrécane. Par exemple, les interactions de l'agrécane, de la protéine de liaison du cartilage ou encore des CD44 avec l'AH, sont médiées par des séquences dénommées PTR (proteoglycan tandem repeat). Celles-ci s'associent à une séquence décасaccharidique de l'AH [84, 88, 89]. Les chaînes de GAG, quant à elles, jouent un rôle dans l'hydratation,

la viscosité et la résistance à la compression des milieux dans lesquels l'agrécanne est présent.

La décorine, le biglycane et la fibromoduline sont des PG de faible masse moléculaire appartenant également à la famille des PG extracellulaires. Excepté la région N-terminale, ces molécules présentent la même organisation protéique, faite d'une zone centrale riche en résidus de leucine et d'une partie C-terminale homologue, traduisant une origine ancestrale commune. Dans la région N-terminale, la décorine et le biglycane portent, respectivement, une ou deux chaînes de CS ou de DS, tandis que les chaînes de KS de la fibromoduline sont localisées dans sa région médiane (Figure 7). Malgré de fortes homologies de séquence, ces molécules ne présentent pas les mêmes réactivités. Le biglycane est incapable de se lier aux fibres de collagène de type I et II alors que la décorine et la fibromoduline ralentissent, par leur fixation, l'élongation de ces fibres. Néanmoins, la diversité des chaînes de GAG ne semble pas responsable des différentes propriétés de ces PG vis-à-vis des fibres de collagène [88].

La subdivision des protéoglycannes membranaires repose sur le type d'association de ces PG avec la membrane. Il peut s'agir d'une liaison de type glycosylphosphatidylinositol (glypicane, cérebroglycane) ou d'une intercalation hydrophobe. Ainsi, les syndécannes, le CD44 ou la thrombomoduline se caractérisent par un court domaine C-terminal cytoplasmique, une région transmembranaire et une partie extracellulaire portant les chaînes de GAG (Figure 8). Ces molécules (syndécannes, CD44) possèdent un nombre variable de chaînes de GAG de type HS et/ou CS. Certaines, comme la thrombomoduline, sont également associées aux membranes sous une forme non-glycosylée [88]. De part leur position membranaire, ces PG peuvent jouer un rôle de transducteur de signaux extracellulaires. Ils sont également impliqués dans les processus d'adhésion et de migration cellulaire par l'intermédiaire d'une part de leur chaîne d'HS et d'autre part de motifs peptidiques définis [90]. Par exemple, les chaînes d'HS des syndécannes s'associent à diverses molécules de la matrice extracellulaire (collagène, fibronectine...). Inversement, le CD44, grâce à son motif peptidique PTR, participe au mécanisme d'adressage lymphocytaire. Principalement démontré dans le cas des CS et HS, les chaînes de GAG des PG extracellulaires et membranaires

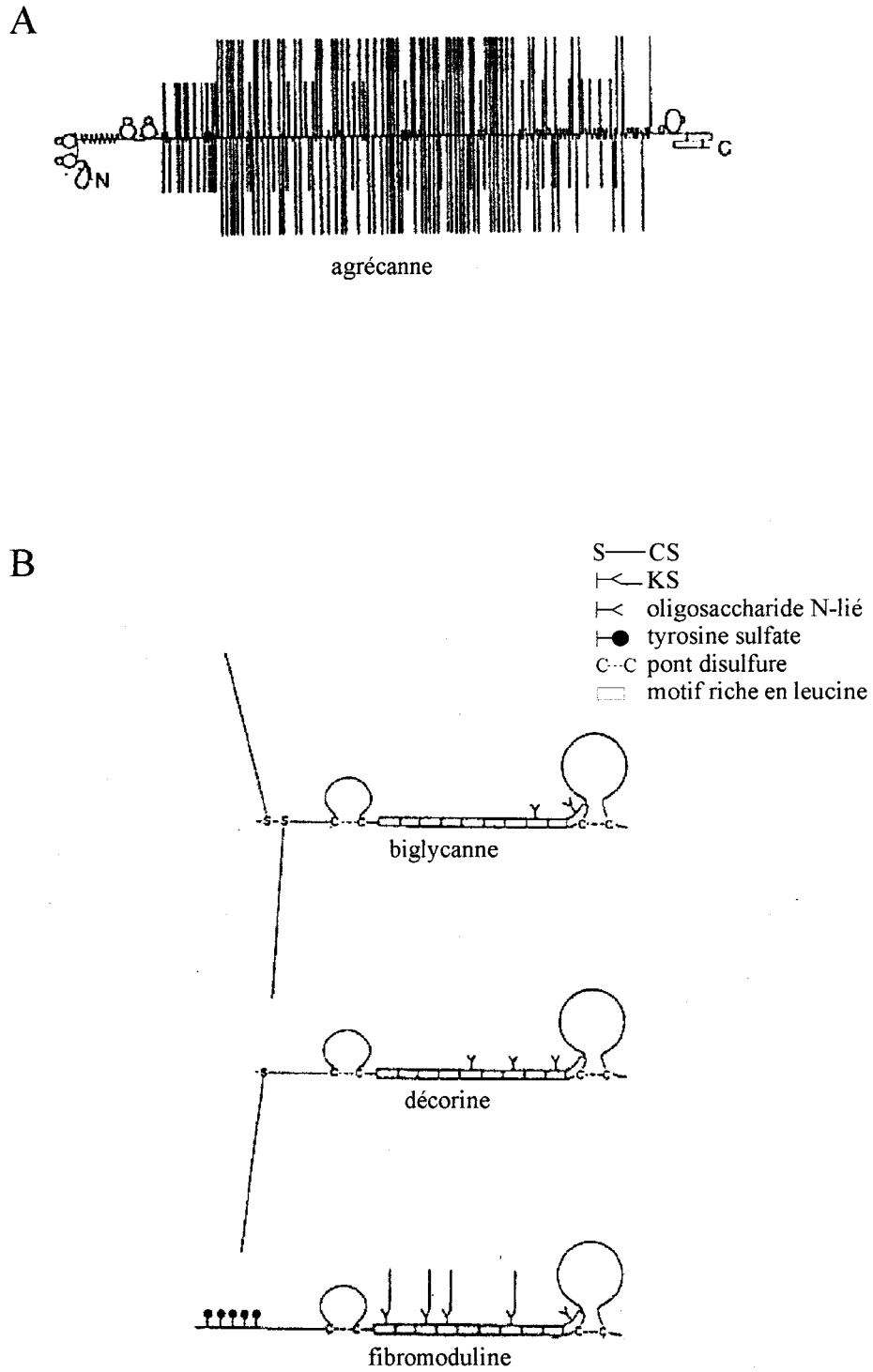
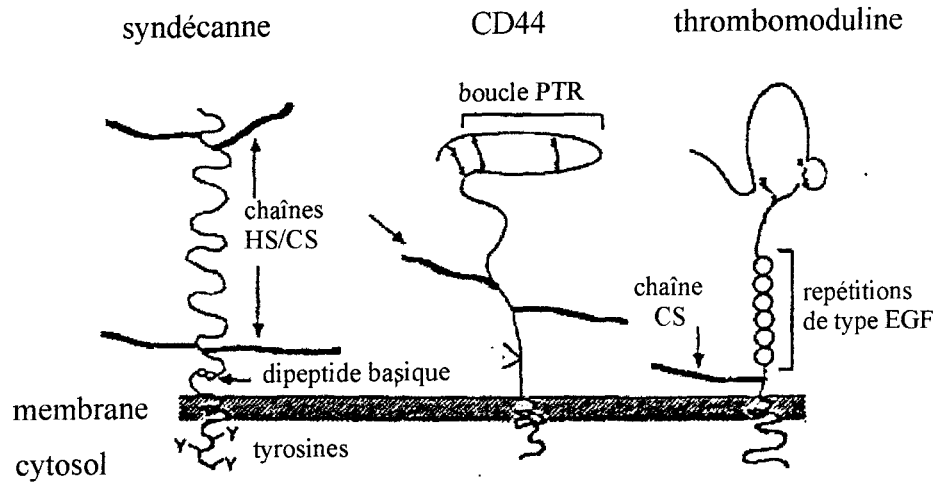


Figure 7 : Variété des protéoglycannes extracellulaires [83].
 (A) PG de haut poids moléculaire; (B) PG de faible poids moléculaire.

PG membranaires



PG intracellulaires

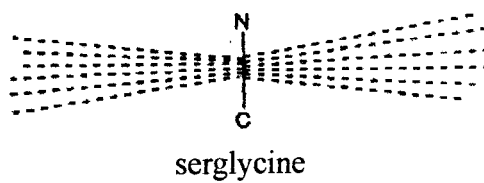


Figure 8 : Modèles de protéoglycannes (PG) membranaires et intracellulaires [83, 88].

participent également au contrôle de l'activité, de la biodisponibilité et de la stabilité des cytokines telles que le TGF- β (transforming growth factor β) et le bFGF (basic fibroblast growth factor) [56].

Les PG intracellulaires sont localisés dans les granules de sécrétion cellulaire. La serglycine, synthétisée par les mastocytes et les cellules hématopoïétiques est structurellement le PG intracellulaire le mieux connu. Il est composé de 10 à 15 chaînes de GAG de type héparine ou CS hypersulfatées fixées dans une région centrale bien définie de la protéine. Cet agencement glycosaminoglycannique fournit une charge négative importante responsable des interactions moléculaires de la serglycine avec les composants des granules de sécrétion. Par le biais de ces interactions, la serglycine stockerait et régulerait la libération de ces composés [83, 91].

Finalement, les PG présentent une variété structurale importante en étroite relation, la plupart du temps, avec leurs fonctions cellulaires. Cependant, ils sont toujours constitués d'un axe protéique unique sur lequel sont greffées des chaînes de GAG en nombre variable.

c) Les complexes de la famille de l'I α I : des protéoglycannes ?

Contrairement aux divers protéoglycannes décrits précédemment, les complexes de la famille de l'I α I résultent de **l'association covalente de 2 ou 3 chaînes peptidiques distinctes sur un axe polyosidique unique**, en l'occurrence une chaîne de GAG. Cette chaîne polyosidique est une chaîne de CS faiblement sulfatée [10,11]. Elle prend naissance sur la bikunine au niveau de la sérine¹⁰ [92]. Celle-ci est incluse dans une séquence peptidique Ser¹⁰-Gly-Gly-Gly correspondant au site consensus de fixation des GAG.

Les analyses structurales du tétrasaccharide de liaison permettant l'ancrage protéique de cette chaîne de CS ont conduit à des résultats discordants. Aussi bien pour l'I α I [93] que pour le P α I [14], la structure, caractérisée par spectrométrie de

masse, correspond à la séquence osidique de base : Δ HexUA-Gal-Gal-Xyl (Δ HexUA : acide hexuronique insaturé en position 4,5). Cependant, dans le cas de l'UTI et de l' α I, Yamada et coll [94, 95] ont identifié, par résonance magnétique nucléaire du proton, une structure hexasaccharidique disulfatée : Δ HexUA-GalNAc(4S)-GlcUA-Gal(4S)-Gal-Xyl (4S : O-sulfatation du résidu en 4). Cette structure correspond au tétrasaccharide de liaison additionné d'un groupement sulfate et à la première unité de répétition disaccharidique sulfatée en 4 sur la GalNAc. Selon ces auteurs [94, 95], la O-sulfatation en 4 du résidu de galactose appartenant au tétrasaccharide de liaison pourrait orienter la synthèse du glycosaminoglycane vers la formation d'une chaîne de CS. Cependant, cette hypothèse n'a pas pu être confirmée en raison de la présence de structures non sulfatées aux points d'ancrage des chaînes de CS de divers protéoglycannes.

La chaîne de CS faiblement sulfatée est constituée de 15 ± 3 unités disaccharidiques [14, 96]. La 4-O-sulfatation de cette chaîne intervient sur environ 5 résidus consécutifs de GalNAc ; un motif fortement anionique est ainsi localisé au voisinage du tétrasaccharide de liaison [97]. Le reste de la chaîne (approximativement 10 unités disaccharidiques) est au contraire non sulfaté. Selon ces auteurs [97], cette région semble être le lieu d'addition des chaînes lourdes en raison de l'incompatibilité stérique existant entre les groupements sulfates et la présence de la liaison unissant les chaînes lourdes à la chaîne de CS. Très récemment, Enghild et coll [96] ont démontré que l'addition des chaînes lourdes a effectivement lieu dans la région non sulfatée (extrémité non réductrice) de la chaîne de CS. Le fait que les chaînes lourdes soient affines pour l'AH qui est un GAG non sulfaté, suggère aussi que la sulfatation de la chaîne de GAG influence négativement le mécanisme d'assemblage de l' α I.

La dégradation enzymatique ménagée (chondroïtinase ABC) de la chaîne de CS de l' α I [11, 12, 96] conduit à la libération d'un composant comportant les deux chaînes lourdes covalentiellement associées par la chaîne de CS. Ce fait suggère que les points d'ancrage de ces deux chaînes sur le GAG sont relativement proches. D'autre part, le site de fixation de la chaîne H2 sur le GAG est plus proche de la bikunine [96].

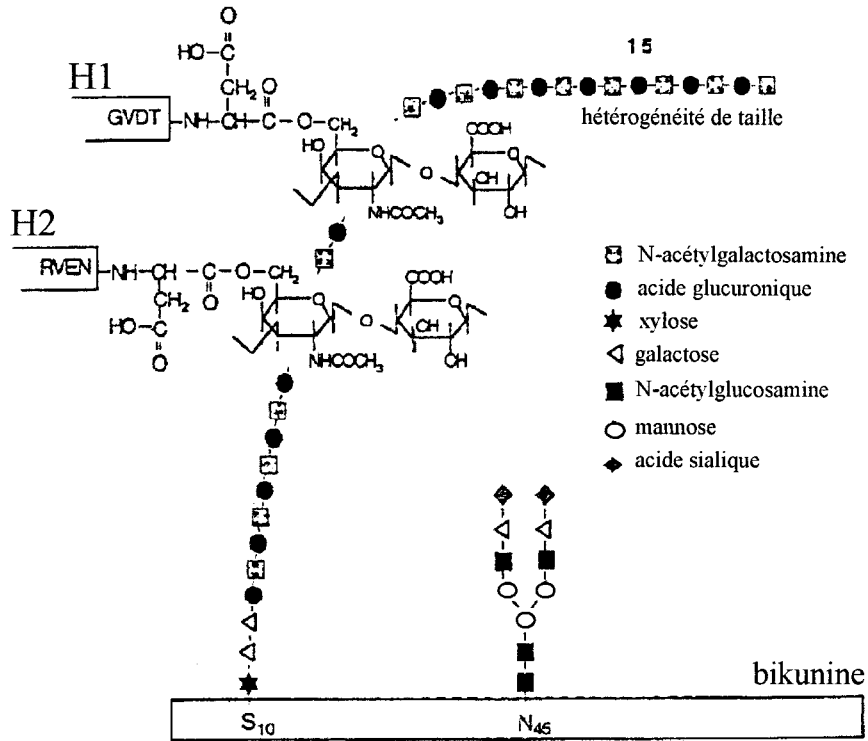
Jamais mise en évidence à ce jour dans d'autres molécules, la structure covalente permettant à une chaîne de GAG de porter plusieurs protéines est la caractéristique structurale des membres de la famille de l'I α I. Celle-ci a été primitivement décrite par Enghild et coll [14, 17] et a été dénommée *protein-glycosaminoglycan-protein* [14]. A l'aide de la spectrométrie de masse, ces auteurs ont démontré que le carboxyle α du résidu d'acide aspartique C-terminal des chaînes lourdes H2 et H3 estérifie en position 6 un résidu de GalNAc de la chaîne de CS (Figure 9A). Ce type de liaison a également été retrouvé au sein de la molécule d'I α I [93].

Quelle est la raison d'être d'un tel assemblage moléculaire au sein des membres de la famille de l'I α I ? En d'autres termes, quels avantages sont apportés par cette liaison qui est absente des complexes multicaténaires covalents circulant dans le sang tels les immunoglobulines ?

Il est connu que la bikunine de petite taille est une molécule très diffusible. Ainsi, elle est très rapidement éliminée dans l'urine [98]. Dans la mesure où elle est responsable d'une bonne part de l'activité biologique de l'I α I, celui-ci de haut poids moléculaire peut représenter une réserve mobile de bikunine. En cas de lésion tissulaire, l'I α I pourrait être retenu, dans la matrice extracellulaire devenue accessible, par les chaînes lourdes interagissant avec l'AH. L'I α I pourrait donc s'accumuler à l'endroit voulu. De plus, compte tenu de sa nature glucidique, la chaîne de CS apporte d'une part une flexibilité peut-être favorable au développement d'interactions ; d'autre part, elle crée, au sein des complexes, un pôle hydrophile et anionique dont le caractère électro-négatif est renforcé au voisinage de la bikunine par la présence de groupements sulfates (Figure 9B). Par conséquent, ces molécules peuvent présenter à ce niveau une fragilité particulière. Ainsi, sous l'action de protéinases, ces régions accessibles et hydrophiles peuvent permettre la libération des chaînes peptidiques pratiquement intactes et donc biologiquement actives.

Quelque soit son rôle, une telle structure suppose l'existence de mécanismes de reconnaissance et d'enzymes très particuliers pour être biosynthétisée.

A



B

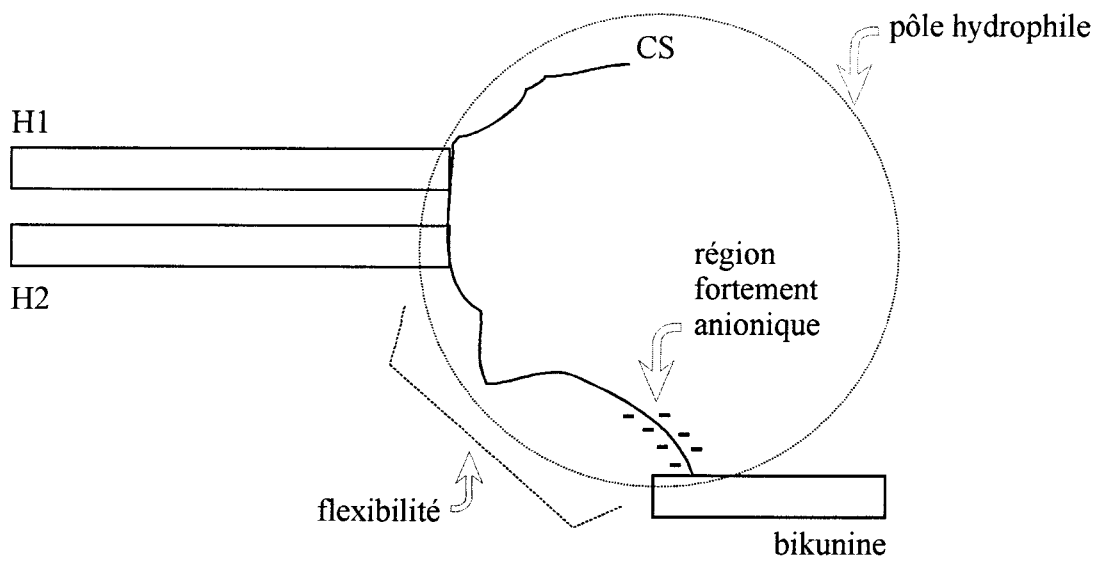


Figure 9 : (A) Agencement moléculaire de l'I α I selon Enghild et coll [96];
 (B) Propriétés conformationnelles potentielles d'une telle structure.

3) Une biosynthèse originale

a) Gènes et précurseurs

Les membres de la famille de l'I α I sont principalement produits par le foie. Quatre gènes situés sur trois chromosomes sont à l'origine des précurseurs polypeptidiques impliqués dans leur formation. Les gènes (*ITIH1* et *ITIH3*) codant pour les chaînes lourdes H1 et H3 sont situés sur le chromosome 3 tandis que celui de la chaîne H2 est localisé sur le chromosome 10 [99]. Seules, les structures des gènes *ITIH1* et *ITIH3* sont décrites à ce jour [100, 101]. L'ARNm codant pour la bikunine porte l'information nécessaire à la synthèse de deux protéines distinctes : l' α 1-microglobuline (α 1m) et la bikunine. Le gène correspondant, dénommé gène du précurseur de l' α 1m/bikunine ou AMBP (alpha-1-microglobulin/bikunin precursor) dont la structure est connue [102] est, quant à lui, localisé sur le chromosome 9 [99]. Il résulte de la fusion de deux gènes ancestraux et est très conservé au cours de l'évolution. Des ARNm homologues au modèle humain ont été caractérisés chez d'autres mammifères comme le porc [103] ou le bœuf [104].

Les précurseurs polypeptidiques résultant de l'expression de ces gènes, subissent d'importantes modifications post-traductionnelles (Figure 10). L'AMBP, précurseur de la bikunine, est composé d'un peptide signal suivi de la séquence de l' α 1m, d'un dipeptide basique (site de clivage enzymatique) et de la séquence de la bikunine [105]. Dans l'appareil de Golgi des cellules du foie, cette protéine (α 1m/bikunine) est glycosylée : N et O-glycosylation de l' α 1m et de la bikunine. Les précurseurs (HxP : heavy chain precursor) des chaînes lourdes H1, H2 et H3 présentent une organisation similaire. Tous, excepté le précurseur de la chaîne H1 qui ne renferme pas de propeptide N-terminal, contiennent un peptide signal suivi d'un propeptide N-terminal de fonction inconnue, de la séquence de la chaîne mature et d'un propeptide C-terminal. La maturation protéolytique du propeptide C-terminal des précurseurs H1P, H2P et H3P a lieu par clivage d'une liaison Asp-Pro et génère un résidu d'acide aspartique en position C-terminale. La fonction α -carboxylique de celui-ci devient alors disponible pour former une liaison ester avec l'hydroxyle en 6 d'un résidu de GalNAc de la chaîne de CS portée par la bikunine.

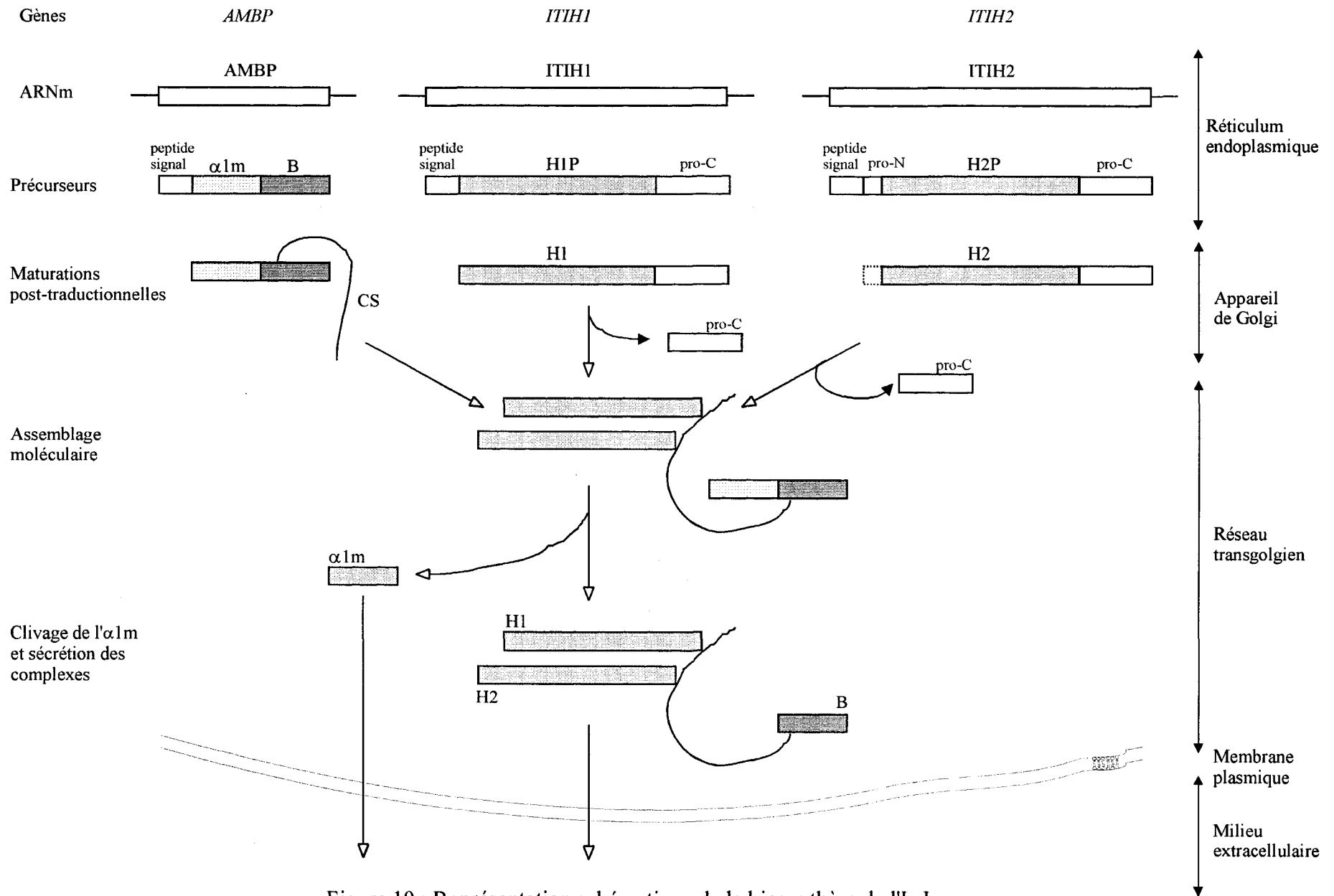


Figure 10 : Représentation schématique de la biosynthèse de l'I α I.

(AMBP, $\alpha 1$ -microglobuline/bikunine; HxP, heavy chain precursor; $\alpha 1m$, $\alpha 1$ -microglobuline; CS, chondroïtine sulfatée; B, bikunine; pro-N ou pro-C, propeptide N ou C-terminal).

Cet acide aspartique appartient à une séquence consensus Asp-Pro-His-Phe-Ile-Ile retrouvée dans la séquence primaire des précurseurs des chaînes lourdes H1, H2 et H3 de toutes les espèces étudiées jusqu'à présent. Travaillant à partir de cellules COS-1 transfectées par les ADNc de la chaîne H3 et de la bikunine de rat, Thuveson et coll [106] ont démontré par mutagenèse dirigée que 5 acides aminés proches ou appartenant à la séquence consensus Asp-Pro-His-Phe-Ile-Ile sont essentiels pour le clivage du propeptide C-terminal. L'absence de cette séquence consensus au niveau du précurseur H4P explique l'absence de protéolyse C-terminale de ce précurseur et par conséquent l'incapacité de la chaîne H4 à se lier à la chaîne de CS de la bikunine.

b) Assemblage moléculaire

Bourguigon et coll [107] ont été les premiers à étudier la biosynthèse des membres de la famille de l'I α I par des cellules d'hépatome humain (HepG2). Dans le milieu intracellulaire, la bikunine apparaît précocement sous la forme d'une protéine hybride : α 1m/bikunine [107, 108]. Aussi bien chez le rat que chez l'homme, l'addition de la chaîne de chondroïtine sulfate s'effectue sur cette protéine hybride [108, 109]. D'autre part, les précurseurs des chaînes lourdes sont clivés au niveau de la liaison Asp-Pro libérant d'un côté les chaînes lourdes matures à même de s'associer à la chaîne de CS et de l'autre les propeptides C-terminaux.

Certaines études [110, 111] ont par la suite démontré que les cellules HepG2 sont incapables de synthétiser, en quantité voulue, la chaîne H1 rendant, de fait, impossible la synthèse de l'I α I. Par contre, les hépatocytes primaires en culture sont capables de synthétiser l'I α I. Ainsi, les études suivantes réalisées sur des hépatocytes primaires humains [110] ou de rat [109], ont démontré que l'addition des chaînes lourdes matures se fait sur la chaîne de CS portée par la protéine hybride α 1m/bikunine. Cet assemblage moléculaire en deux étapes, maturation protéolytique C-terminale des chaînes lourdes puis création de la liaison ester, semble dépendre d'une protéinase et d'une activité enzymatique nouvelle [110]. Dans des

cellules COS-1 transfectées, le propeptide C-terminal de la chaîne H3 de rat semble être impliqué dans le mécanisme de formation de la structure PGP puisque, une fois clivé, il reste associé à la chaîne H3 mature (Figure 11) jusqu'à la création de la liaison ester comme récemment rapporté [106]. Dans la dernière étape de la biosynthèse, l' α 1m est enzymatiquement libérée des complexes et chaque entité est alors sécrétée.

Le foie n'est pas le seul organe capable d'exprimer les gènes des précurseurs des membres de la famille de l' α I. Par hybridation *in situ* et/ou RT-PCR (reverse transcriptase-polymerase chain reaction), deux équipes ont rapporté, chez le rat [112] comme chez l'homme [113], la présence des ARNm relatifs aux membres de la famille de l' α I dans divers organes tels que le cerveau, l'intestin, l'estomac, le pancréas ou le rein. Par conséquent, d'autres types cellulaires peuvent peut être synthétiser les protéines apparentées à l' α I.

c) Régulation de la biosynthèse au cours de pathologies inflammatoires

Sous l'influence de stimuli inflammatoires, la synthèse hépatique de certaines protéines, dénommées protéines de la phase aiguë de l'inflammation (PPA), est soit augmentée (PPA positive), soit diminuée (PPA négative). Chez des patients atteints de syndromes inflammatoires d'origine infectieuse [114], le taux d'ARNm (détecté par Northern blot) de la chaîne H1 reste constant, ceux de la chaîne H2 et de l'AMBP diminuent alors que celui de la chaîne H3 augmente. Ces variations quantitatives sont parfaitement corrélées aux quantités de protéines sériques observées par Western-blot puisque les auteurs montrent une diminution significative des concentrations de l' α I_{H1,2,B} et de l' α I_{H2,B} et au contraire une augmentation de la concentration de P α I (I α I_{H3,B}). De même, à partir d'un modèle inflammatoire chez le rat, Daveau et coll [112] ont détecté par RT-PCR des variations similaires des taux hépatiques des ARNm de l'AMBP et des chaînes H1, H2 et H3 ainsi qu'une augmentation du taux d'ARNm de la chaîne H4. Par conséquent, l' α I_{H1,2,B}, l' α I_{H2,B} et la bikunine sont des PPA négatives alors que le P α I (I α I_{H3,B}) et la chaîne H4, au moins chez le rat, sont des PPA positives.

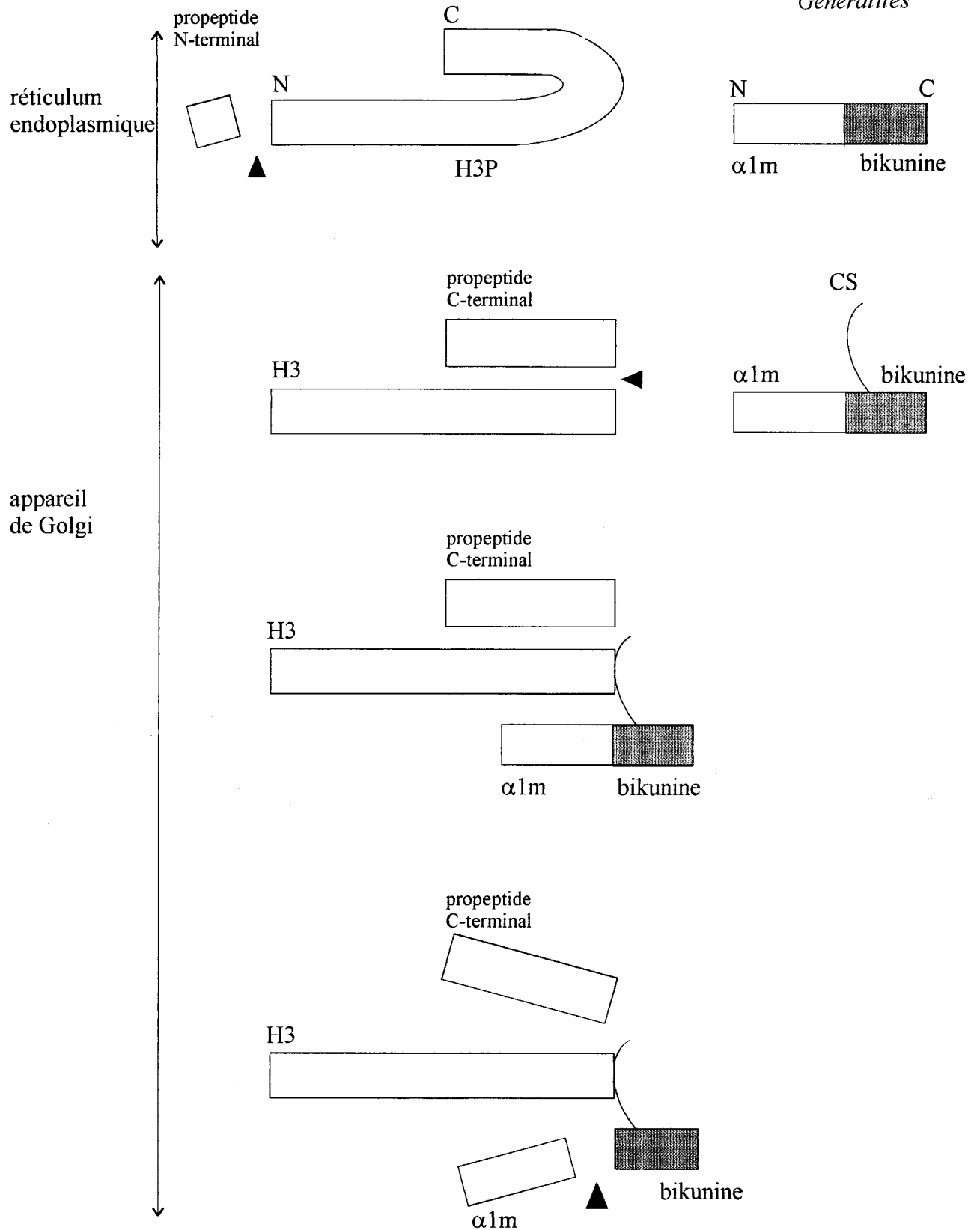


Figure 11 : Assemblage du PoI dans des cellules COS-1 transfectées par les ADNc de H3 et de l'AMBP de rat [106].

H3P, précurseur de la chaîne H3; $\alpha 1m$, $\alpha 1$ -microglobuline; CS, chondroïtine sulfate; N et C, extrémités N et C -terminales

d) P α I et I α I comme protéines recombinantes

Les protéines de la famille de l'I α I sont les seules protéines connues dans lesquelles une chaîne de CS développe une liaison covalente avec trois chaînes polypeptidiques distinctes H1, H2 et H3. Cette liaison très particulière peut donc être le facteur limitant pour la production de protéines recombinantes. De plus, aucune des lignées d'hépatocytes étudiées n'est capable, à la différence des hépatocytes primaires, de synthétiser *in vitro* l'I α I. Cependant, Blom et coll [23], transfectant les ADNc de la chaîne H3 et de la bikunine de rat dans des cellules COS-1, ont mis en évidence la présence d'une protéine recombinante, similaire au P α I, composée de la chaîne H3 liée à la bikunine par l'intermédiaire de la chaîne de CS. La nature exacte de la liaison formée reste cependant à démontrer. Des cellules autres que les hépatocytes primaires semblent donc capables de former une liaison de type PGP.

Une fois transfectées [115], les cellules COS-7 sont également capables de synthétiser le protéoglycane hybride (α 1m/bikunine) ainsi que des protéines apparentées à l'I α I. Dans des modèles de doubles transfections (H1, H2, ou H3 et l'AMBP), seul un complexe protéique pouvant correspondre au P α I est synthétisé. Dans les milieux de culture des cellules COS-7 triplement transfectées par les ADNc humains de deux chaînes lourdes et de l'AMBP, un seul complexe tricaténaire associé via la chaîne de CS peut être caractérisé : le complexe H2/H3/B. Malheureusement, celui-ci n'existe pas chez l'homme. Ces cellules co-transfectées par les 4 ADNc sont également incapables de produire l'I α I humain (I α I_{H1,2,B}).

Finalement, *in vitro*, aucun type cellulaire autre que les hépatocytes primaires n'a été démontré être capable d'assembler correctement les différentes chaînes polypeptidiques composant l'I α I humain. Or, *in vivo*, cet assemblage moléculaire est conservé. L'I α I est toujours constitué d'une chaîne H1, d'une chaîne H2 et de la bikunine et présente une homogénéité structurale. Par conséquent, ce mécanisme d'assemblage peut dépendre de la séquence protéique, mais aussi de modifications post-traductionnelles ou de la conformation spatiale

des chaînes lourdes. Un **signal** permettant l'assemblage **sélectif** des complexes appartenant à la famille de l'I α I doit donc exister.

III Structure des différentes chaînes polypeptidiques constituant les protéines de la famille de l'I α I

1) Structure de la bikunine

A l'heure actuelle, les séquences correspondant à l'AMBP (α 1m/bikunine) de 8 espèces animales sont publiées. Leur numéro d'accès dans la base de données SWISS-PROT est présenté dans le tableau II ci-dessous. Ces séquences sont principalement déduites du clonage et du séquençage des ADNc relatifs au précurseur AMBP.

Tableau II : Numéros d'accès dans la base de données SWISS-PROT aux différentes séquences de l'AMBP

Espèce	Numéro d'accès
Homme	P 02760
Bœuf	P 00978
Porc	P 04366
Rat	Q 64240
Hamster	Q 60559
Souris	Q 07456
Gerbille de Mongolie	Q 62577
Carrelet	P 36992

La bikunine humaine est, de toutes celles caractérisées, la bikunine la mieux connue d'un point de vue structural. Comme nous l'avons signalé précédemment, la bikunine est un protéoglycane constitué d'une chaîne peptidique de 147 acides aminés (masse moléculaire théorique = 15974 Da) portant une chaîne de GAG de type chondroïtine sulfate (cf. chapitre II.2.c). Elle présente une masse moléculaire apparente de 67000 Da en chromatographie de gel filtration [116] et 42 à 46000 Da par électrophorèse en gel de polyacrylamide en présence de dodécyl sulfate de sodium (SDS/PAGE) [9].

La séquence primaire de la bikunine montre la présence de deux domaines (domaine I : Lys²²-Arg⁷⁷ ; domaine II : Thr⁷⁸-Arg¹⁴⁴) inhibiteurs de sérineprotéinases arrangés en tandem. Chaque domaine contient 3 ponts disulfure dont l'arrangement est caractéristique du domaine inhibiteur de type Kunitz trouvé dans l'inhibiteur pancréatique bovin de la trypsine (BPTI : bovin pancreatic trypsin inhibitor). La résolution de la structure cristalline [117], grâce à la diffraction des rayons X, d'une partie de la bikunine (Ser²⁵-Val¹³⁴) révèle que chaque domaine est constitué de 2 feuillets β -plissés antiparallèles et d'une hélice α . Le second domaine renferme également une courte hélice α dans sa partie N-terminale (Figure 12). Ces domaines présentent une orientation spatiale décalée d'un angle de 60° qui, néanmoins, ne permet pas l'inhibition simultanée de 2 molécules de sérineprotéinases en raison d'encombrements stériques dûs à la proximité des domaines. La bikunine possède également une région hydrophobe proche de l'extrémité N-terminale probablement impliquée dans la liaison avec ses récepteurs cellulaires et une région C-terminale basique qui peut être responsable de l'inhibition de la formation des cristaux d'oxalate de calcium urinaire.

En plus de la chaîne de chondroïtine sulfate, la bikunine possède, sur le résidu d'asparagine 45 [118], un N-glycane biantenné portant au maximum 2 acides N-acetylneuraminiques (NeuAc) [119]. Celui-ci, est responsable de l'affinité de la bikunine vis-à-vis d'une lectine : la Concanavalline A (*Concanavalia ensiformis agglutinin*) [120]. Ce protéoglycane digéré par la chondroïtinase ABC, AC ou la hyaluronidase (non spécifique) permet d'obtenir la bikunine dépourvue de GAG (26 kDA). Celle-ci conserve ses propriétés antiprotéasiques [121].

2) Structure des chaînes lourdes

Toutes les séquences des chaînes lourdes ont été déduites des structures des ADNc. Chez l'homme, les 4 précurseurs polypeptidiques sont connus contre 3 chez le porc, le hamster et la souris et uniquement 2 chez le rat. Leur numéro d'identification dans la base de données SWISS-PROT est rassemblé dans le tableau III. Au cours de leur biosynthèse, ces chaînes subissent des modifications post-traductionnelles telles que formation de ponts disulfure et glycosylation.

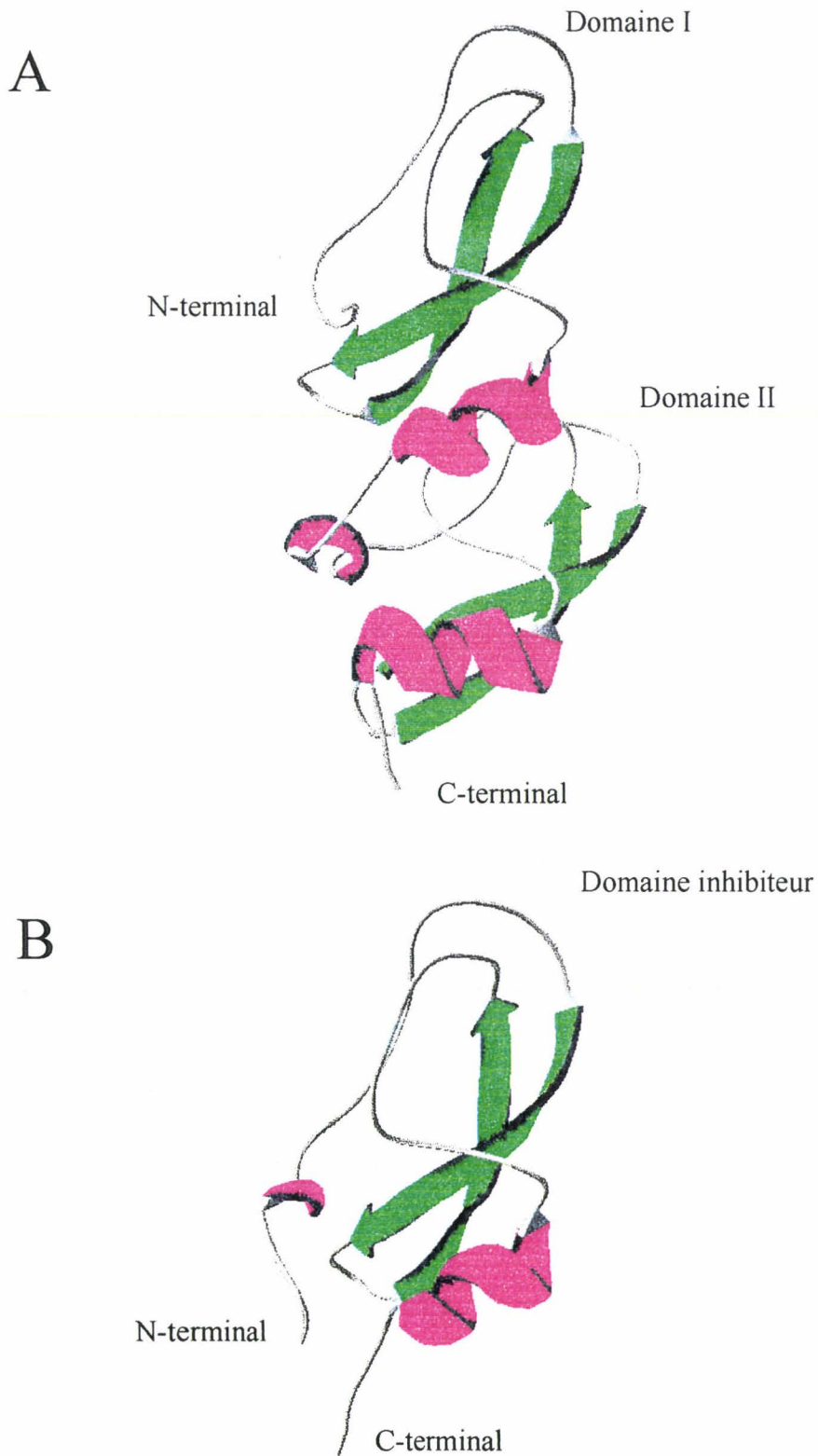


Figure 12: Structures tridimensionnelles de la bikunine (A) et de l'inhibiteur pancréatique bovin de la trypsine (B). Chaque domaine inhibiteur de type Kunitz renferme deux feuillets β -plissés antiparallèles (vert) et une hélice α (rouge). Le second domaine inhibiteur de la bikunine contient également une courte hélice α dans sa partie N-terminale.

Lorsque nous avons abordé ce travail, aucune étude destinée à les mettre en évidence n'avait été réalisée. Dans ce chapitre, nous nous proposons de résumer succinctement les caractéristiques structurales des chaînes lourdes humaines et de les comparer à celles des autres espèces.

Tableau III : Dénomination et numéro d'identification des précurseurs des chaînes lourdes dans la base de données SWISS-PROT.

ITH1-human	P 19827	ITH1-pig	Q 29052
ITH2-human	P 19823	ITH2-pig	O 02668
ITH3-human	Q 06033	ITH3-pig	-----
ITH4-human	Q 14624	ITH4-pig	P 79263
ITH1-meseau	P 97278	ITH1-mouse	Q 61702
ITH2-meseau	P 97279	ITH2-mouse	Q 61703
ITH3-meseau	P 97280	ITH3-mouse	Q 61704
ITH4-meseau	-----	ITH4-mouse	-----
ITH1-rat	-----		
ITH2-rat	-----		
ITH3-rat	X 83231		
ITH4-rat	Y 11283		

a) Ossatures peptidiques

L'électrophorèse en gel de polyacrylamide, en milieu réducteur et en présence de dodécyl sulfate de sodium (SDS/PAGE), des complexes de la famille de l' $\text{I}\alpha\text{I}$ [122,16] révèle des masses moléculaires de 220 kDa ($\text{I}\alpha\text{I}$) et 125 kDa ($\text{P}\alpha\text{I}$ et $\text{I}\alpha\text{IH2,B}$). Les masses moléculaires des chaînes lourdes H1, H2 et H3 sont respectivement estimées à 96, 86 et 85 kDa en milieu non réducteur [122, 16]. Dans des conditions réductrices, seule la masse moléculaire apparente de la chaîne H1 est modifiée (78 kDa). Par spectrométrie de masse de type MALDI (matrix assisted

laser desorption ionization), la masse moléculaire de la chaîne H3 a été plus précisément évaluée à 74100 Da [16]. Ces masses sont en accord avec les structures primaires de ces différentes chaînes constituées d'environ 650 acides aminés [123-126]. Par contre, la chaîne H4 se caractérise par une masse moléculaire plus élevée : 120 kDa [20]. En effet, elle ne subit pas de protéolyse C-terminale et est donc composée de 902 acides aminés pour une masse moléculaire théorique avoisinant les 100000 Da (Tableau IV) [20].

Les homologues de séquence des précurseurs (HxP) des chaînes lourdes sont de 54 % entre H1P et H3P, 41 % entre H2P et H3P et 38 % entre H1P et H2P [126, 127]. Les précurseurs qui ne sont jamais associés (quelque soit l'espèce) en complexe tricaténaire présentent donc une forte homologie de séquence. En ce qui concerne H4P, les pourcentages d'homologie sont de l'ordre de 31 à 38 % selon les précurseurs considérés [20]. Seuls les deux tiers de la séquence du précurseur de H4 possèdent des homologues avec les précurseurs précédemment cités. En fait, ces homologues reflètent la présence de domaines structuraux conservés.

b) Domaines structuraux

Selon Chan et coll [127], excepté les 60 premiers acides aminés très divergents, les séquences primaires des précurseurs H1P, H2P et H3P peuvent être grossièrement divisées en trois régions : (i) la première moitié des séquences (environ 500 acides aminés) est constituée de séquences conservées régulièrement espacées le long de la chaîne et contient un domaine von Willebrand de type A (vWA) détecté dans toutes les chaînes lourdes ; (ii) les 200 acides aminés suivant constituent une région faite de groupes d'acides aminés totalement identiques entrecoupés de larges zones hypervariables ; (iii) les parties C-terminales excisées présentent, quant à elles, une disposition des séquences conservées identique à celle du premier domaine. Certaines (H1P et H3P) renferment également un domaine potentiel de liaison du cuivre fortement similaire aux sites liant le cuivre de certaines oxydases. Comme les précurseurs précédemment évoqués, celui de la chaîne H4 renferme un domaine vWA localisé dans la même région. Sa séquence C-terminale qui n'est pas protéolysée, renferme une séquence proche de celle de

Tableau IV : Caractéristiques structurales des chaînes matures humaines de la famille de l'IαI

Chaînes	Nombre d'acides aminés	Masses moléculaires			Nombre de cystéines	Sites potentiels de N-glycosylation	N-glycannes détectés	Sites de O-glycosylation	Domaines structuraux
		Théoriques (Da)		Expérimentales					
		SDS/PAGE (kDa) Non réduite/ réduite							
H1	644	71887	96/78	Nd	5	N ²⁵⁶ -L-T N ⁵⁵⁹ -L-S	Oui	Absence de GalNAc	VWA Cu ²⁺
H2	647	72425	86/86	Nd	4	N ⁴² -N-S N ⁶⁴ -F-S N ³⁹¹ -I-S	Oui	Thr ⁶³⁷	VWA
H3	618	69447	85/85	74100	2	N ⁵⁸ -F-T N ⁵⁴⁷ -L-T	Oui 2	Absence de GalNAc	VWA Cu ²⁺
H4	902	100199	120/Nd	Nd	3	N ⁵³ -F-S N ¹⁷⁹ -K-T N ⁴⁸⁹ -I-T N ⁵⁴⁹ -L-T	Oui	Potentiel : Ser ⁶⁶⁸ Thr ⁶⁷³ Ser ⁶⁷⁴	VWA «Bradykinine» Protéinase dépendante de l'ATP

Nd, non déterminé ; vWA, domaine von Willebrand de type A ; « Bradykinine », domaine similaire à la bradykinine ; Cu²⁺, domaine de liaison du cuivre.

la bradykinine ainsi qu'un motif retrouvé dans les protéinases dépendantes de l'ATP [128].

c) Modifications post-traductionnelles : ponts disulfure et glycosylation.

La masse moléculaire de la seule chaîne H1 diffère si elle est estimée en SDS/PAGE dans des conditions réductrices ou non, ce qui témoigne de la présence de ponts disulfure. Les chaînes matures H1, H2 et H3 comportent respectivement 5, 4 et 2 résidus de cystéine [126]. En raison du nombre impair de cystéine de la chaîne H1, un de ces résidus peut donc être libre. Toutes les régions N-terminales des chaînes précédemment citées ont 2 cystéines, proches l'une de l'autre, qui forment un motif Cys-Pro-Xaa-Cys (CPXP). Les 2 autres cystéines de la chaîne H2 sont voisines contrairement à celles de la chaîne H1. D'autre part, les séquences C-terminales excisées contiennent également 2 cystéines qui, selon Bourguignon et coll [126], pourraient être pontées entre elles. Par contre, la chaîne H4 ne renferme que 3 cystéines [20] dont les 2 dernières, situées dans la partie C-terminale, sont pontées [19].

Selon Kornfeld et coll [129], toutes les séquences consensus Asn-Xaa-Ser/Thr (ou Xaa est différent de la proline) sont des sites potentiels de N-glycosylation. Par conséquent, selon ces règles, les chaînes matures H1, H2 et H3 renferment respectivement 2, 3 et 2 sites potentiels de N-glycosylation (Tableau IV). En 1988, Gebhard et coll [125] annonçaient les asparagines 64 et 617 comme glycosylées. Cependant, l'Asn⁶¹⁷ appartient à une séquence Asn-Pro-Ser connue pour n'être glycosylée que dans de très rares cas. Peu de temps après, Malki et coll [122], par la composition en monosaccharides et la réactivité des chaînes H1 et H2 vis-à-vis de la Concanavaleine A, apportaient la preuve de l'existence de N-glycannes de type complexe sur ces chaînes. D'autre part, Mizon et coll [16] dénombraient 2 N-glycannes biantennés sur la chaîne H3. Comme l'indique d'une part la différence de masse moléculaire entre la masse théorique et expérimentale de la chaîne H4, et d'autre part sa teneur en sucres de 22 %, cette chaîne subit des modifications post-traductionnelles de nature glycanique. Traitée par la peptidase N-glycosidase F (PNGase F), sa masse moléculaire décroît. Par conséquent, cette

chaîne porte des N-glycannes. D'ailleurs, elle renferme 4 sites potentiels de N-glycosylation (Tableau IV). Cependant, leur localisation n'a pas encore été rapportée à ce jour.

Dès 1979, la présence de O-glycannes était suspectée sur l'I α I en raison de sa réactivité positive vis-à-vis de lectines reconnaissant certains types de O-glycannes [130]. La chaîne H2 était démontrée comme responsable de cette réactivité puisque sa composition en monosaccharides permettait, contrairement à celle de la chaîne H1, d'identifier 2 résidus de GalNAc. Par la suite, Enghild et coll [17] caractérisaient, dans la région C-terminale de cette chaîne H2, un O-glycane trisaccharidique NeuAc-Gal-GalNAc situé sur la thréonine 637. Inversement, compte tenu de sa composition en monosaccharides, la chaîne H3 ne semble posséder aucun O-glycane [16]. Par contre, pour la chaîne H4, l'absence de phenylthiohydantoïnes pour les résidus Ser⁶⁶⁸, Thr⁶⁷³ et Ser⁶⁷⁴ tend à démontrer, selon Nishimura et coll [20], une substitution O-glycannique de ces résidus.

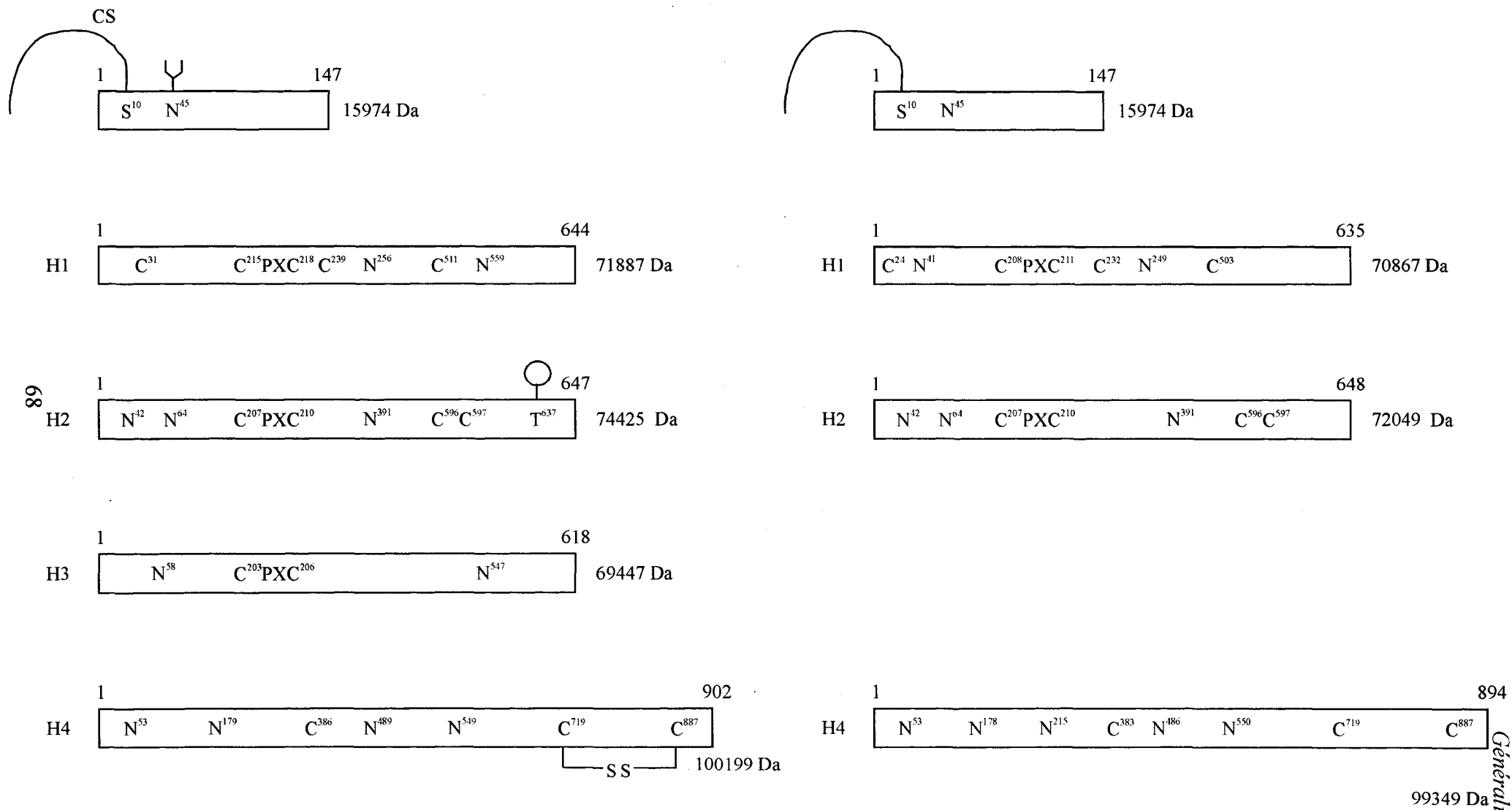
Quoique présentant une parenté structurale indéniable, les précurseurs H1P, H2P et H3P sont finalement clairement différenciés en terme de séquences, de nombre de résidus de cystéine ou de sites de glycosylation (Figure 13). Par conséquent, elles pourraient exercer chacune, une fonction physiologique propre. Cette notion est, d'ailleurs, appuyée par l'étude des homologies constatées pour une même chaîne provenant d'espèces différentes.

3) Homologies inter espèces

La détermination des séquences primaires des précurseurs des chaînes lourdes chez d'autres espèces que l'homme a permis d'opérer des comparaisons inter espèces. Ainsi, Chan et coll [127] ont été les premiers à rapporter une conservation inter espèces, pour le même précurseur, plus importante (de l'ordre de 83% et plus) qu'entre les précurseurs H1P, H2P et H3P à l'intérieur d'une même espèce. De même, comparés à la souris, l'homme ou le porc, les précurseurs H1P, H2P et H3P du hamster présentent des taux d'identité de séquence similaires [27].

Chaînes matures humaines

Chaînes matures porcines



Généralités

99349 Da

Figure 13 : Représentation schématique comparée des connaissances structurales acquises et des modifications post-traductionnelles potentielles concernant la structure des chaînes matures humaines et porcines de la famille de l'IαI. La schématisation des ponts disulfure de la bikunine n'est pas représentée dans un souci de clarté. (Y, N-glycane biantenné; O, O-glycannes; CS, chondroïtine sulfate; --S-S--, pont disulfure.)

Toutes les séquences des précurseurs des chaînes lourdes H1, H2 et H3 du hamster [27] et de la souris [127] peuvent être divisées en trois régions distinctes présentant une organisation et des domaines structuraux similaires à celles de l'homme. Par exemple, la première région contient un domaine vWA, la seconde se compose de courtes séquences strictement conservées espacées de larges régions hypervariables et la troisième renferme dans certains cas un domaine de liaison du cuivre. Une preuve de l'importance biologique des séquences conservées est apportée par la séquence consensus Asp-Pro-His-Phe-Ile-Ile. Celle-ci, comme nous l'avons signalé précédemment, est un élément majeur impliqué dans la formation des motifs PGP. Ainsi, toutes les chaînes lourdes capables de se lier de manière covalente à la chaîne de CS de la bikunine possèdent cette séquence consensus.

La répartition des résidus de cystéine s'avère également être une donnée structurale conservée. Parmi les 10 précurseurs (H1P à H3P) connus en 1997, 4 résidus de cystéine formant d'une part le motif CPXC et appartenant d'autre part à la région C-terminale excisée, sont strictement conservés d'une chaîne à l'autre mais également d'une espèce à l'autre [27]. De plus, pour un précurseur donné que ce soit H1P, H2P, H3P ou H4P, la position des cystéines est identique, à de rares exceptions près, quelque soit l'espèce considérée.

En terme de sites de N-glycosylation, les sites potentiels identifiés dans les précurseurs humains sont tous strictement conservés chez la souris. Cependant, les précurseurs de souris renferment également un site supplémentaire par rapport à l'homme [127]. En fait, de manière générale, quelque soit l'espèce, les sites potentiels de N-glycosylation, pour une chaîne donnée, sont très fortement conservés.

Finalement, ces homologies de séquence, de domaines structuraux, du nombre et de la localisation d'une part des résidus de cystéine et d'autre part des sites potentiels de N-glycosylation (Figure 13) indiquent que les différences inter espèces sont moins importantes que celles existant entre les différentes chaînes lourdes au sein d'une même espèce. Par conséquent, ces similitudes soutiennent l'hypothèse d'un métabolisme, du développement d'interactions moléculaires spécifiques ou de fonctions physiologiques propres à chaque chaîne lourde.

Cependant, l'importance biologique des domaines structuraux caractérisés et/ou des modifications post-traductionnelles potentielles reste à démontrer.

Objectifs

OBJECTIFS

La revue bibliographique que nous avons conduite montre à l'évidence que parmi les inhibiteurs de sérine-protéinases du plasma humain, l'I α I occupe une place bien particulière. Identifié initialement comme le précurseur possible de la bikunine responsable de l'essentiel de l'activité antitrypsique acido-stable présente dans la plasma ou l'urine, il apparaît aujourd'hui comme le chef de file d'un ensemble de composés qui diffèrent les uns des autres par le nombre et la nature des chaînes lourdes associées à la bikunine.

Bien que présentant entre elles de nombreuses homologues structurales et donc des propriétés communes en ce qui concerne leur réactivité comme leur métabolisme, les chaînes lourdes apparaissent aussi comme capables de présenter un comportement et un métabolisme propre à chacune d'elles.

La compréhension du rôle particulier joué par chaque chaîne lourde dépend d'une meilleure connaissance de leurs caractéristiques structurales. Notre objectif a donc été de progresser dans ce domaine. En effet, quand ce travail a démarré, seule la séquence en acides aminés déduite de la structure de l'ADNc était connue pour les chaînes lourdes alors identifiées chez l'homme.

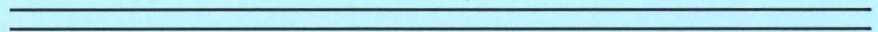
Nous avons choisi d'étudier les chaînes lourdes H1 et H2 qui entrent dans la composition de l'I α I et de nous engager dans trois directions :

- Les propriétés d'une protéine dépendent de sa conformation tridimensionnelle qui est elle-même liée à la présence des liaisons disulfures qui s'établissent entre les résidus de cystéine. Nous nous sommes donc attachés à démontrer et à identifier ces liaisons et à examiner les répercussions des particularités de la structure covalente de chaque chaîne sur leur conformation appréhendée par dichroïsme circulaire d'une part, par leur sensibilité à une protéolyse partielle d'autre part.

- La présence de glycanes peut influencer de manière notable la réactivité et le métabolisme d'une protéine. Quelques indications avaient été obtenues qui laissaient supposer que les chaînes H1 et H2 pouvaient différer par leur structure glycanique. Nous démontrons qu'effectivement la chaîne H2 humaine présente à proximité de son extrémité C-terminale une région fortement O-glycosylée.

- Des inhibiteurs de sérine-protéinases appartenant à la famille de l'I α I et en particulier la bikunine ont été caractérisés dans de nombreuses espèces animales. Cependant, les chaînes lourdes présentes ont rarement été isolées. Nous avons donc dans la dernière partie de ce travail analysé la structure glycanique des chaînes lourdes de l'I α I porcin. Cet animal présente en effet un intérêt évident à la fois comme modèle d'étude de nombreuses maladies humaines et comme source possible de tissus ou d'organes peut-être un jour utilisables en transplantation.

Résultats



RESULTATS

I Topologie des ponts disulfure des chaînes lourdes de l'I α I humain : implications structurales et fonctionnelles.

Avec l'aide du Laboratoire français du Fractionnement et des Biotechnologies de Lille, un procédé de préparation, à l'échelle industrielle, d'un concentré d'I α I humain de qualité thérapeutique a été mis au point au laboratoire [135]. Ainsi, disposant de quantités suffisantes de protéine pure, nous avons étudié la topologie des ponts disulfure des chaînes lourdes H1 et H2 de l'I α I humain.

Avant toute étude structurale concernant les chaînes lourdes de l'I α I, il est nécessaire, compte tenu de la structure particulière de l'I α I, de la dissocier [11-13, 24] puis de purifier [2] les chaînes lourdes.

La méthodologie employée repose sur la caractérisation par la dégradation d'Edman et la spectrométrie de masse de type MALDI, des peptides comportant un ou plusieurs résidus de cystéine susceptibles d'être impliqués dans la formation des ponts disulfure. Pour cela, après avoir dosé les groupements thiols libres et par conséquent estimé le nombre de ponts disulfure probables, nous avons clivé les chaînes lourdes par le bromure de cyanogène (BrCN), afin d'obtenir des fragments peptidiques facilement purifiables par chromatographie de gel filtration. Dès lors, l'analyse électrophorétique et le séquençage des fractions obtenues par gel filtration ont permis d'identifier celles contenant les peptides d'intérêt. Afin de caractériser les résidus de cystéine engagés dans les ponts disulfure, ces peptides purifiés ont ensuite été soumis à la dégradation d'Edman et/ou analysés par spectrométrie de masse avant et après réduction-alkylation.

Parce que la protéolyse des protéines s'exerce de préférence sur les régions non structurées de la molécule, nous avons soumis les chaînes lourdes à une digestion ménagée par la protéinase V8 afin d'étudier leur sensibilité à la protéolyse et d'évaluer l'influence exercée par les ponts disulfure. Cependant, la

conformation des chaînes lourdes, associées en complexe au sein de la molécule d'I α I, peut être stabilisée par des interactions inter chaînes. Leur dégradation peut donc s'avérer totalement différente de celle des chaînes isolées. Par conséquent, nous avons également analysé, par immunodétection spécifique des chaînes lourdes, la dégradation ménagée de l'I α I humain sous l'action de la protéinase V8.

Finalement, afin de mieux évaluer l'influence de la présence des ponts disulfure sur la conformation des chaînes lourdes et de l'I α I, il était indispensable d'analyser, par dichroïsme circulaire, les chaînes lourdes libres ou associées au sein de l'I α I.

1) Publication

Les résultats obtenus ont fait l'objet de la publication suivante : « Disulphide bonds assignment in the inter- α -inhibitor heavy chains : structural and functional implications ».

Disulphide bonds assignment in the inter- α -inhibitor heavy chains Structural and functional implications

Christophe FLAHAUT¹, Charlotte MIZON¹, Pierrette AUMERCIER-MAES¹, Pierre COLSON², Christian BAILLY³, Pierre SAUTIERE⁴ and Jacques MIZON¹

¹ Laboratoire de Biochimie, Faculté de Pharmacie, Lille, France

² Laboratoire de Chimie Macromoléculaire et Chimie Physique, Université de Liège, Liège, Belgium

³ INSERM U-124, IRCL, Lille, France

⁴ Département V, URA 1309 CNRS, Institut de Biologie et Institut Pasteur de Lille, Lille, France

(Received 24 December 1997/24 March 1998) – EJB 97 1822/3

Human inter- α -inhibitor (I α I) is a plasma serine-proteinase inhibitor. It consists of three polypeptide chains covalently linked by a glycosaminoglycan: a light one named bikunin, carrying the antiproteinase activity and two heavy chains H1 and H2. The amino acid sequences of these heavy chains are highly similar; however when I α I is digested by neutrophil proteinases, their proteolytic susceptibility strongly differs [Balduyck, M., Piva, F., Mizon, C., Maes, P., Malki, N., Gressier, B., Michalski, C. & Mizon, J. (1993) Human leucocyte elastase (HLE) preferentially cleaves the heavy chain H2 of inter- α -trypsin inhibitor (ITI), *Biol. Chem. Hoppe-Seyler* 374, 895–901].

We mapped the disulphide topology of the I α I heavy chains in order to investigate whether or not disulphide bonds might be responsible for their differential susceptibility to proteolysis. Using amino acid sequencing and mass spectrometry analysis, we demonstrate that the H1 heavy chain contains one free thiol group and two disulphide bridges of which one links two largely spaced cysteine residues (Cys239 and Cys511). Thus H1 is clearly different from H2 which contains two disulphide bonds between closely located cysteine residues. However, using immunoprint analysis, we show that, when I α I is subjected to a limited digestion by *Staphylococcus aureus* V-8 proteinase, the two polypeptide chains are similarly susceptible to proteolysis. This enzyme preferentially cleaves the I α I heavy chains from their N-terminal extremity. These results are consistent with the circular dichroism (CD) analysis, suggesting that the conformation of the polypeptide backbone of H1 is not very different from that of H2, with calculated α -helicities of 24% and 28%, respectively. The CD measurements reveal that the aromatic amino acids of H1 and H2 are in a different asymmetrical environment.

Inside the I α I molecule, the heavy chains are linked to the glycosaminoglycan chain via their C-terminal aspartic acid residue. Thus we suggest that the affinity of cationic neutrophil proteinases for the anionic glycosaminoglycan is responsible for the cleavage of the heavy chains (mainly H2) near their C-terminal end and the high susceptibility of I α I to these proteinases.

Keywords: inter- α -inhibitor; heavy chain; disulfide bond; proteolysis.

Inter- α -inhibitor (I α I) is a serine-proteinase inhibitor found in mammalian plasma. I α I is a complex of three different polypeptide chains: the first one named bikunin ($m = 26\,000$) is a double-headed Kunitz-type inhibitor carrying the antiproteinase activity of I α I. The two other polypeptide chains exhibit higher molecular masses, near 80 000 Da. For human I α I, they are designated H1 and H2 heavy chains. The three polypeptide chains of I α I are covalently cross-linked by a chondroitin sulphate-type glycosaminoglycan (recently reviewed in [1]).

During severe inflammatory processes such as septic shock [2], a large consumption of I α I has been reported, suggesting

that this molecule which is unusually sensitive to proteinases released by triggered neutrophils, would be partially cleaved in acquired diseases in which there is a systemic proteolysis. Thus I α I would be converted into smaller size derivatives able to reach various tissues where they exert their physiological activity: proteinase inhibition but also for instance modulation of cell proliferation [3] and stabilization of extracellular matrix [4].

We have previously determined the degradation mechanism of I α I incubated *in vitro* in the presence of human leucocyte elastase [5]. The H2 heavy chain appeared particularly sensitive to proteolysis and the high-molecular-mass components released from I α I were composed of H1 linked to bikunin. Thus they have lost their immunoreactivity towards anti-H2 antibodies. A similar pattern was obtained when I α I was incubated in the presence of triggered neutrophils [6]. Furthermore, we have shown that this process could be physiologically relevant because the addition of plasma proteinase inhibitors (notably α 1-proteinase inhibitor) blunted but did not completely block the I α I proteolysis.

We hypothesized that the preferential cleavage occurring in C-terminal regions of the I α I heavy chains could be explained

Correspondence to J. Mizon, Laboratoire de Biochimie, Faculté de Pharmacie, avenue du Pr Laguesse, B.P. 83, F-59006, Lille, France

Fax: +33 3 20 95 90 09.

E-mail: jmizon@phare.univ-lille2.fr

Abbreviations. I α I, human inter- α -inhibitor; H, heavy chain; vWA, von Willebrand type A domain; Hse, homoserine; CD, circular dichroism; MALDI-TOF-MS, matrix-assisted laser desorption ionization mass spectrometry; RP, reverse-phase.

Enzymes. Endoproteinase Lys-C (EC 3.44.99.30); V-8 proteinase, endoproteinase Glu-C (EC 3.4.21.19).

by affinity of cationic neutrophil proteinases for the anionic chondroitin sulphate which cross-links the three polypeptide chains of IαI.

However the protein digestion sites might be influenced by other characteristics of the target protein as well as by the amino acid specificity of the proteinase involved. With respect to the target protein, susceptibility to digestion is predominantly associated with unstructured regions. Disulphide bonds strongly contribute to the stability of the tertiary structure of proteins. Therefore, using amino acid sequencing and mass spectrometry of Cys-containing peptides, we have determined the disulphide bond topology in the IαI heavy chains.

SDS/PAGE analysis of the fragments released from IαI incubated with proteinases of various specificities showed that the degradation patterns were different [7, 8]. As reported by Potempa et al. [8], *Staphylococcus aureus* V-8 proteinase induces a degradation pattern clearly different from that produced by neutrophil proteinases (namely elastase and cathepsin G). Thus we have subjected IαI and isolated H1 and H2 heavy chains to a limited proteolysis by V-8 proteinase and comparatively analysed the digests by immunoprint, using anti-IαI immunoglobulins as well as specific anti-H1 and anti-H2 sera.

MATERIALS AND METHODS

Materials. Endoproteinase Lys-C and V-8 proteinase from *S. aureus* (endoproteinase Glu-C) were from Boehringer. Anti-IαI immunoglobulins were from Dako. Specific anti-H1 and anti-H2 antibodies were those described in [9]. External calibrants used for mass spectrometry were from Sigma. All the solvents and reagents used for gas-phase sequencing were from Perkin-Elmer. All the other reagents used were of analytical grade.

Isolation and purification of the heavy chains of IαI. IαI was isolated from a fraction of human plasma originating from an industrial preparation of clotting factors (LFB) as already reported [10]. IαI was dissociated by hydroxylamine treatment, as described in [11] and the heavy chains isolated according to Mizon et al. [9].

Disulphide topology study. Determination of free thiol groups of H1 and H2 was carried out according to Mantle et al. [12]. The protein content was determined from absorbance at 280 nm ($\epsilon^{1\% \cdot 1\text{cm}} = 6.41 \pm 0.27$ for H1 and 11.75 ± 0.35 for H2; $n = 6$). Each heavy chain (1.6 mg) was dissolved in 1.6 ml of a 200 mM sodium acetate, pH 5.4, containing 20 mM EDTA and 2% SDS and incubated for 15 min at 37°C before addition of 0.2 ml of 2 mM 4,4'-dipyridyl disulphide (Sigma) in a 200 mM sodium acetate, pH 5.0. After 30 min incubation at room temperature, the absorbance of the mixtures was measured at 324 nm against a blank containing all reagents except the protein. All assays were run in duplicate.

Prior to CNBr cleavage, the H1 heavy chain was carboxamidomethylated. For this purpose, H1 (150 nmol), dissolved in 10 ml of a 20 mM Tris/HCl, pH 7.6, containing 100 mM NaCl, 4 mM EDTA, was denatured by guanidine hydrochloride (6 M final concentration). The sample was flushed with nitrogen and incubated at 50°C for 30 min. The pH was then controlled and was adjusted to 8.0. Iodoacetamide was added in a 100-fold molar excess over the protein present. The reaction was carried out in the dark at 37°C for 30 min under nitrogen. The sample was then extensively dialyzed against a 20 mM ammonium bicarbonate, pH 7.4 and lyophilized.

CNBr cleavage was carried out on carboxamidomethylated H1 or native H2. Each protein (150 nmol) was dissolved in 4 ml of 70% formic acid and treated with a 100-fold molar excess

(with respect to methionine content) of CNBr in 70% formic acid. The reaction mixtures were flushed with argon and kept under magnetic stirring, for 24 h in the dark at room temperature. The reaction mixtures were then diluted with 20 ml water and lyophilized. Peptides generated from CNBr cleavage were fractionated by gel filtration on Sephadex G100 (Pharmacia) equilibrated with 0.5% formic acid, at a flow rate of 0.2 ml/min. Protein elution was monitored at 280 nm and fractions were pooled, according to the pattern of absorbance.

CNBr-generated fraction B of H1 (20 nmol) was dissolved in 200 µl of a 100 mM ammonium bicarbonate, pH 8.3. 2 µg of endoproteinase Lys-C (in 20 µl ultrapure water) were added. The digestion, carried out at 37°C for 16 h, was stopped by freezing at -20°C. The released fragments were isolated by reverse-phase (RP)-HPLC on a C4 Vydac 214 TP 54 column (250 mm × 4.6 mm internal diameter, 5-µm particle size) equilibrated with 0.05% trifluoroacetic acid in water (solvent A). Isocratic conditions were applied with 4% acetonitrile in solvent A for 10 min, followed by a linear gradient from 4% to 80% acetonitrile in 0.05% trifluoroacetic acid for 70 min, at a flow rate of 0.7 ml/min, with detection by absorbance at 215 nm. Fractions were manually collected.

Automated Edman degradation of peptides and detection of phenylthiohydantoin derivatives were performed on a gas-phase sequencer (Procise 492, Perkin-Elmer) using the pulsed liquid program. Phenylthiohydantoin derivatives of amino acids were identified on-line on a 120 A amino acid analyzer (Perkin-Elmer).

For the peptide Leu586-Hse626 (peptide II; Hse, homoserine) of H2, the adjacent cysteine residues were positively characterized after reduction and alkylation. To this end, 1 nmol peptide dissolved in 100 µl 100 mM ammonium bicarbonate, pH 8.3 containing 6 M guanidine hydrochloride was reduced by adding dithiothreitol at a 100 mM final concentration. After incubation for 30 min at 70°C, the reaction mixture was cooled in ice, treated with acrylamide at a 2 M final concentration and kept chilled for 45 min. The reduced and alkylated peptide was desalted and immobilized on a ProSorb cartridge (Perkin-Elmer) before sequencing. The phenylthiohydantoin derivative of carboxamidoethylcysteine residues was clearly characterized as it eluted between the phenylthiohydantoin derivative of glutamic acid and that of histidine, respectively.

Matrix-assisted laser desorption ionization mass spectrometry (MALDI-TOF-MS) determinations were performed on a Vision 2000 time-of-flight instrument (Finnigan Mat.) equipped with a 337-nm ultraviolet laser. The mass spectra were acquired in positive and reflectron mode with 5 kV accelerating voltage. 0.5 µl of the analyte solutions [15 pmol dissolved in acetonitrile/water (30:70) containing 0.1% trifluoroacetic acid] was mixed with 1 µl of the matrix solution: 2,5-dihydroxybenzoic acid [10 mg/ml dissolved in acetonitrile/water (30:70)]. External calibration was performed using angiotensin I ($m = 1296.7$ Da), bovine insulin ($m = 5733.6$ Da), horse apomyoglobin ($m = 16951.5$ Da) and trypsinogen ($m = 23981.0$ Da).

Limited proteolysis of IαI by *Staphylococcus* V-8 proteinase. IαI and the H1 and H2 heavy chains were equilibrated by dialysis in 100 mM ammonium bicarbonate, pH 7.8. After adjustment of the protein concentration to 1 mg/ml (IαI) and 0.5 mg/ml (heavy chains) in the same buffer, the proteinase was added in an enzyme-to-substrate ratio of 1:50 (by mass) and the mixture incubated at 37°C. Aliquots of 250 µl were withdrawn at 6 h and 24 h and immediately boiled in an equal volume of 2% SDS sample buffer with or without reducing agent before analysis by SDS/PAGE. For immunoblot analysis, IαI (100 µg/ml) was digested by V-8 proteinase for various lengths of time up to 6 h.

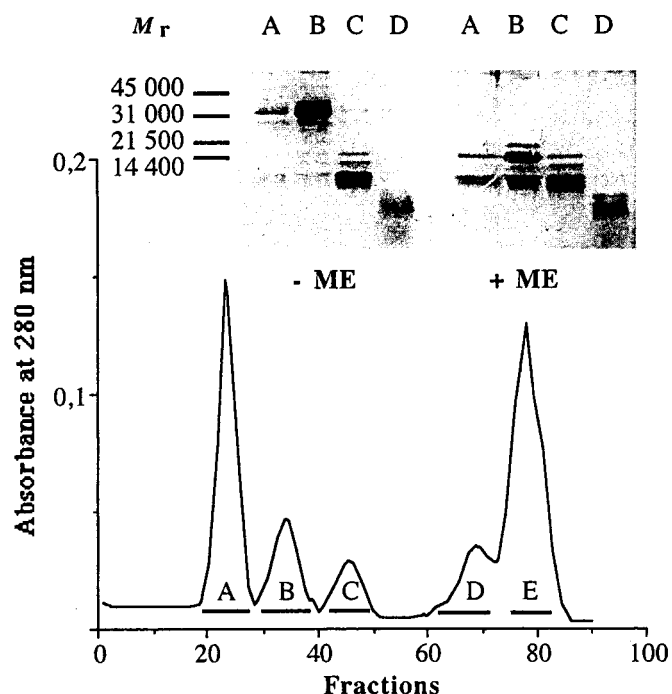


Fig. 1. Separation of the cyanogen bromide cleavage fragments of carboxamidomethylated H1. The fragments generated from the cyanogen bromide cleavage of the carboxamidomethylated H1 (approximately 150 nmol) were loaded onto a Sephadex G-100 column (1.3 cm \times 100 cm) at a flow-rate of 12 ml/h equilibrated with 0.5% formic acid. Fractions of 2 ml were collected. In inset, electrophoretic pattern of the pooled fractions, analysed by 12% SDS/PAGE prior to or after reduction. Detection by Coomassie staining.

Electrophoreses were carried out on 7.5 or 12% (mass/vol.) polyacrylamide slab gels, under reducing or non-reducing conditions, as described by Laemmli [13] with Coomassie staining according to Stephano et al. [14]. Immunodetection after transfer was performed as already described with specific anti-I α I immunoglobulins and anti-H1 and anti-H2 immunosera [9].

Circular dichroism measurements. CD spectra were recorded on a Jobin-Yvon CD6 dichrograph interfaced to a PC microcomputer. Protein solutions were scanned in 0.2-cm or 1-cm quartz cuvettes, depending on the spectral range and the protein concentration. In the 200–250-nm region, the protein concentration was adjusted to 0.3 mg/ml (0.2 cm cells). Higher protein concentrations (0.7–1.28 mg/ml) were used to monitor the CD in the 250–300-nm region, using 1-cm cells. Measurements were carried out in 20 mM sodium phosphate, pH 7.4, containing 150 mM NaCl. Five scans were accumulated and automatically averaged. In each case the base-line spectrum (of the buffer) was subtracted prior to the smoothing procedure. For the determination of the helicity, we have resorted to the method described by Clark et al. [15] based on the measured circular dichroism at 220 nm and using the empirical relationship: % helix = $(-\Delta\epsilon - 0.25)/0.105$. The molar circular dichroism $\Delta\epsilon = \Delta A/l \times c$; ΔA is the circular dichroism amplitude, l the cell path-length and c the total protein concentration (expressed in amino acid residues).

RESULTS

Disulphide bonds assignment in the I α I heavy chains. The amino acid sequence deduced from the cDNA of the H1 heavy chain reveals five cysteine residues. Therefore it is likely that at least one Cys residue has a free sulphhydryl group. Indeed, using

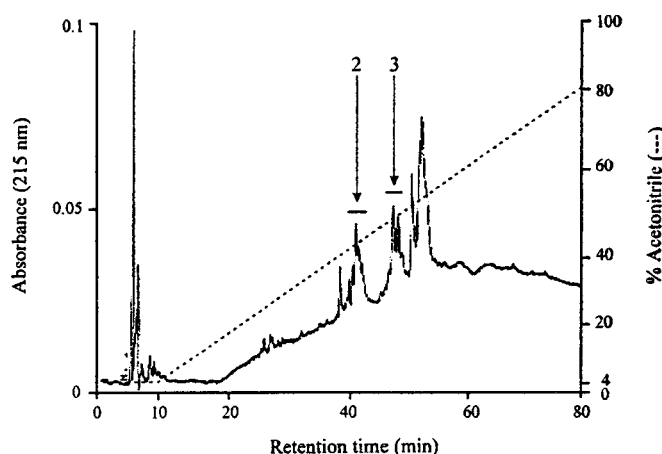


Fig. 2. HPLC of peptides generated from endoproteinase Lys-C digestion. Fraction B from Fig. 1 (about 20 nmol) was hydrolyzed with endoproteinase Lys-C and fractionated by RP-HPLC as indicated in Materials and Methods.

4,4'-dipyridyl disulphide, we obtained a value of 1.2 free sulphhydryl group/mol H1. This result suggests that there are two intramolecular disulphide bonds in the H1 heavy chain. To determine their location, H1 was first alkylated by iodoacetamide, in order to prevent thiol-disulphide exchange reactions which are enhanced when a thiol group and disulphide bonds coexist into a protein. The carboxamidomethylated H1 was subsequently cleaved by CNBr and the cleavage products were fractionated by gel filtration on Sephadex G-100. Five fractions (A–E) were pooled according to the pattern of absorbance at 280 nm (Fig. 1). Aliquots of each peptide fraction were analysed by SDS/PAGE with or without prior reduction and by N-terminal sequencing.

Fraction C contained a peptide with an apparent molecular mass of about 12 000 by SDS/PAGE, with or without prior reduction. It was identified by Edman degradation as the N-terminal peptide of the H1 heavy chain: Ser1–Hse114 (Table 1). It includes only one cysteine residue at position 31, thus carrying the only free sulphhydryl group in the native protein.

SDS/PAGE analysis of fraction B under non-reducing conditions showed a large peptide with apparent molecular mass close to 30 000. By reduction, it generated two bands of similar staining intensity, with apparent molecular masses 18 000 and 12 000, respectively. The major N-terminal sequence found in fraction B corresponded to that of the peptide Glu421–Hse518 containing Cys511. The minor sequence was identified as the N-terminal sequence of peptide Gln146–Hse260. The low yield obtained during the sequencing of this peptide may be attributable to a partial cyclisation of the N-terminal glutamine residue, giving rise to pyroglutamic acid. This peptide Gln146–Hse260 includes the three cysteine residues at positions 215, 218 and 239.

An unequivocal disulphide bond assignment could be obtained by endoproteinase Lys-C digestion of fraction B. The proteolytic fragments were separated by RP-HPLC (Fig. 2). The peptide Gly201–Lys228 (peptide 2a) was identified in fraction 2 by N-terminal sequencing (Table 1). However, Cys215 and Cys218 could not be detected, suggesting that they are linked by a disulphide bond. MALDI-TOF-MS analysis showed a signal at m/z 3074.1, in good agreement with the theoretical mass (3071.5 Da) calculated for this peptide under its oxidized form. Therefore a disulphide linkage between Cys215 and Cys218 was established because carboxamidomethylcysteine was not identified either by sequencing or MALDI-TOF-MS analysis.

The Lys200–Lys228 peptide (2b) coeluted with Gly201–Lys228 peptide by RP-HPLC in the same fraction 2. Its presence

Table 1. Cysteine-containing peptides obtained from the H1 heavy chain. (○) Peptide letters refer to fractions in Fig. 1. (□) Peptide numbers refer to peaks in Fig. 2. (#) The Asn256 residue carries an N-linked desialylated biantennary-type glycan. (*) yields in picomoles; (**) the theoretical masses were calculated for the oxidized form of peptides; n.d., not determined.

After treatment with	Peptides	Molecular mass		N-terminal sequences	
		**calculated	measured		
CNBr	○ C S1-C31-Hse114			S A T G R S S S S S E	
	B Q146-C215-C218-C239-Hse260			Q Y E I V I K V K P	
	B E421-C511-Hse518			E N N G R A Q R I Y	
Endo-Lys C	□ 2a G201-C215-C218-K228	3071.5	3074.1	G H V L F R R P T V S	Q Q Q S C P T C S
	2b K200-C215-C218-K228	3199.6	3201.3	K G H V L F R P T V S	Q Q Q S C P T C S
	3a V229-C239-N256-Hse260			*53.3 76.5 24.6 29.6 32.3 27.4 26.0 21.8 17.2 24.5 26.0 28.0 20.1 8.7 n.d. 6.8 n.d.	
	A496-C511-Hse518	*7792.9	7793	V T Y D V	
	3b I238-C239-N256-Hse260	*6728.7	6732	A D V Q A I C D L L	

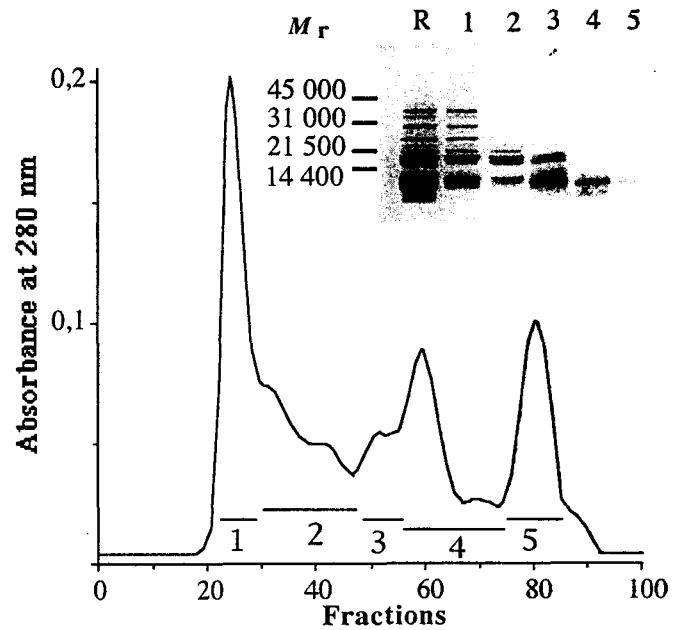


Fig. 3. Separation of the cyanogen bromide cleavage fragments of H2. The fragments generated from cyanogen bromide cleavage of H2 (approximately 150 nmol) were separated by molecular-size-exclusion chromatography on Sephadex G-100 as described in the legend of Fig. 1. In the inset, SDS/PAGE analysis of the pooled fractions under non-reducing conditions. The electrophoretic pattern was not modified after prior reduction of the analysed samples.

results from partial cleavage by endoproteinase Lys-C of the sequence Lys199-Lys200-Gly201. Its characterization by N-terminal sequencing and MALDI-TOF-MS analysis also indicates that there is a disulphide bond between Cys215 and Cys218 (Table 1).

Fraction 3 from RP-HPLC contained two large peptides (3a and 3b). A major N-terminal sequence (ADVQA) identified the expected peptide Ala496-Hse518 that contains Cys511. This one is linked by a disulphide bridge to Cys239. We found, but in a lower yield, two other sequences. The first one begins at Val229, the second at Ile238 (Table 1). Thus peptide 3b results from a partial cleavage by endoproteinase Lys-C of the Lys237-Ile238 bond; the efficiency of the enzyme might have been hindered by the acidic environment induced by the two Asp residues at positions 236 and 240, respectively.

The structure proposed for peptide 3a was also determined by MALDI-TOF-MS analysis. This peptide carries one biantennary-type N-glycan at Asn256, which was probably desialylated during CNBr treatment. Thus, its theoretical mass is 7792.9 Da (6170.9 for the naked peptide and 1622.0 for the N-linked desialylated biantennary-type glycan). The experimental measurement ($m/z = 7793.0$) for this peptide agrees with these data. In the same manner, we have established the structure of peptide 3b, which confirms that there is a disulphide bond between Cys239 and Cys511. For peptides 3a and 3b, the masses corresponding to the disulphide-linked fragments with one or both homoserines in the lactone form (lacking 18 Da and 36 Da, respectively) were not detected.

The H2 heavy chain contains four cysteine residues: two closely located (Cys207 and Cys210) and two adjacent (Cys596 and Cys597). Using 4,4'-dipyridyl disulphide, we did not detect any free sulphhydryl group. Therefore, the H2 heavy chain was cleaved with CNBr without prior alkylation and the fragments were separated by gel filtration on Sephadex G-100 (Fig. 3). Five fractions were pooled as indicated and analysed by SDS/

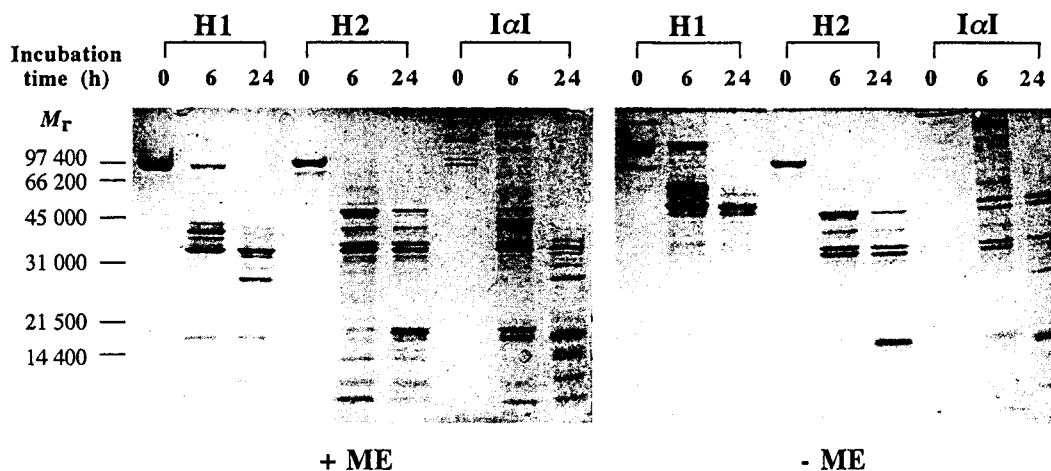


Fig. 4. SDS/PAGE analysis of the digestion products of H1, H2 and I α I by V-8 proteinase. Proteins were simultaneously incubated with V-8 proteinase in an enzyme-to-substrate ratio of 1:50, for 0 min, 6 h and 24 h. The digests were analysed by 12% SDS/PAGE with detection by Coomassie staining, with or without prior reduction as indicated.

Table 2. Cysteine-containing peptides obtained from the H2 heavy chain by CNBr treatment. Peptide I was isolated by RP-HPLC on a C4 Vydac column, from fraction 2 isolated by gel filtration. The separation was performed by using a discontinuous gradient. Solvent A was 0.05% trifluoroacetic acid in water. Solvent B contained 0.05% trifluoroacetic acid, 75% acetonitrile and 25% water. The conditions used were 20% acetonitrile for 10 min, 20 to 37% acetonitrile for 17 min, 37% acetonitrile for 5 min, 37 to 45% for 32 min and 45 to 75% for 30 min. For reduction and alkylation of peptide I, peptide (300 pmol) dissolved in 50 μ l of 100 mM ammonium bicarbonate, pH 8.0, containing 6 M guanidine hydrochloride, was treated at 37°C for 1 h by adding dithiothreitol (5 mM final concentration) and then iodoacetamide (25 mM final concentration) for 35 min at room temperature, under a nitrogen atmosphere. Prior to MS-analysis, the reaction mixture was desalted by RP-HPLC. (O) Cysteine residues were positively identified as carboxamidoethylcysteine after reduction followed by acrylamide treatment; (*) yields in picomoles; (**) the theoretical masses were calculated for both the oxidized and reduced-alkylated forms of peptides; n.d., not determined.

Peptides	N-terminal sequences												Reduction-alkylation	Molecular mass					
														**calculated	measured				
I E109-C207-C210-Hse267	E	N	F	R	T	E	V	N	V	L			-	18204	18197				
	* 33.1	30.8	32.6	19.8	17.2	21.3	26.4	24.6	24.8	30.0			+	18320	18316				
II L586-C596-C597-Hse626	L	A	D	A	P	P	Q	D	P	S	C	C	S	G	A	L	Y		
	* 62.7	56.2	29.2	22.1	24.1	20.7	15.3	8.3	20.6	15.6	n.d.	n.d.	6.7	4.7	6.5	9.5	5.2		

PAGE. The major fragments eluted in a volume which was not correlated to their apparent molecular mass estimated by SDS/PAGE analysis. No change in the migration pattern of the peptides was observed whether the samples were reduced or not, suggesting that closely located cysteine residues were linked together. In order to verify this hypothesis, we looked for the two expected cysteine-containing CNBr fragments by N-terminal sequencing. The first one Glu109-Hse267 (peptide I) was clearly identified in fractions 1 and 2. It was purified by RP-HPLC. Its experimental molecular mass ($m/z = 18197$) determined by MALDI-TOF-MS agreed with the calculated value (18204 Da) and shifted to 18316 upon reduction followed by alkylation with iodoacetamide. This difference in masses clearly demonstrates that Cys207 and Cys210 are linked by a disulphide bridge in the native protein (Table 2).

The second peptide Leu586-Hse626 (peptide II) was the major fragment found in fraction 4, as proved by N-terminal sequencing. All the expected residues were clearly identified, except cysteines at positions 596 and 597. To establish that these two residues are really linked by a disulphide bond, the peptide was reduced by dithiothreitol and then alkylated by acrylamide before N-terminal sequencing. Phenylthiohydantoin derivative of carboxamidoethylcysteine was positively identified at cycles 11 and 12.

Limited proteolysis of I α I and its heavy chains by V-8 proteinase. With respect to disulphide bonds, the H1 and H2 heavy chains are structurally distinct. The large loop due to the disulphide bridge between Cys239 and Cys511 probably stabilizes the tertiary structure of the H1 heavy chain and influences its overall folding. For that reason we compared the susceptibility of the two heavy chains and that of I α I to a limited proteolysis by V-8 proteinase. SDS/PAGE analysis of the digests (Fig. 4) shows that the mobility of the major fragments released from H1 obviously increased upon reduction, with the appearance of a peptide with a molecular mass near 18000. These data suggest that in the region located between Cys239 and Cys511, an exposed site is particularly sensitive to proteolysis by V-8 proteinase. Conversely and in agreement with its disulphide bond topology, the fragmentation pattern of H2 was not different for both the non-reduced and the reduced digests.

H1 is only slightly less susceptible than H2 to proteolysis, as assessed by (a) the persistence of residual H1 after a 6-h incubation; (b) the higher number and the lower size of the cleavage products from H2 digested in the same conditions. The major cleavage products observed after proteolysis of the isolated heavy chains were also visualized after incubation of I α I in the presence of V-8 proteinase, suggesting that the heavy chains were not protected towards proteolysis when they are to-

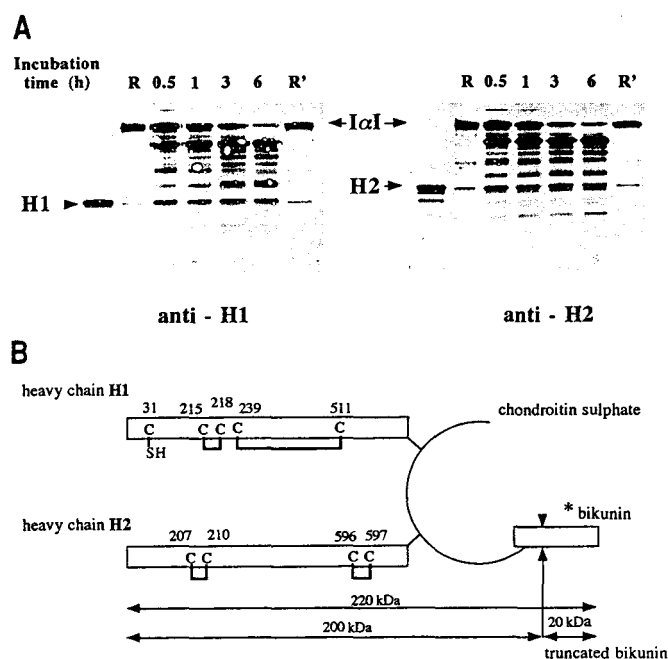


Fig. 5. Immunoprint analysis of the digestion products of I α I by V-8 proteinase after various incubation times. (A) The digestion was carried out as described in Materials and Methods and the digests analysed by 7.5% SDS/PAGE under reducing conditions; after transfer onto nitrocellulose sheets, the detection was achieved by specific anti-H1 and anti-H2 antibodies. R = I α I incubated with the enzyme for 0 min. R' = I α I incubated at 37°C for 6 h without enzyme. An equimolar mixture of the H1 and H2 heavy chains was loaded in the left lane on each blot. (B) Schematic representation (not to scale) of the I α I molecule showing disulphide bridges location and first cleavage products released by V-8 proteinase digestion. The V-8 proteinase digestion shows first the excision of truncated bikunin (the two Kunitz-type domains) and subsequently the gradual digestion of the two heavy chains from their N-terminal extremity, thus releasing compounds of apparent molecular masses ranging from 100 000 to 180 000. (*), except disulphide bridges inside bikunin.

gether covalently linked to the glycosaminoglycan. Thus we conclude that the two heavy chains of I α I are simultaneously degraded by V-8 proteinase at a nearly identical rate.

To establish a possible sequence for the proteolytic events leading to its fragmentation, the cleavage products released early from I α I by incubation with V-8 proteinase were analysed by SDS/PAGE followed by immunoblot detection with specific antibodies raised against each of the heavy chains. As shown in Fig. 5, I α I was progressively hydrolyzed. After only a 30-min incubation, the most intensively stained band migrated with an apparent molecular mass near 200 000. This component was revealed by both anti-H1 and anti-H2 antibodies and results from the early cleavage of bikunin, as previously reported [8]. Truncated bikunin (apparent molecular mass near 20 000) was detected by anti-I α I immunoglobulins (data not shown). The other I α I cleavage products appearing during the course of the digestion with V-8 proteinase have high molecular masses (apparent molecular mass between 100 000 and 180 000). All these fragments were revealed by both anti-H1 and anti-H2 antibodies. However the anti-H1 antibody revealed a component quickly released whose electrophoretic mobility is close to that of the H1 heavy chain. This component was not detected by the anti-H2 antibody and thus results from the cleavage of H1 near its C-terminal extremity. The H2 heavy chain is also partially cleaved in the same region as revealed by the presence of a band only visible with the anti-H2 antibody and exhibiting quite the

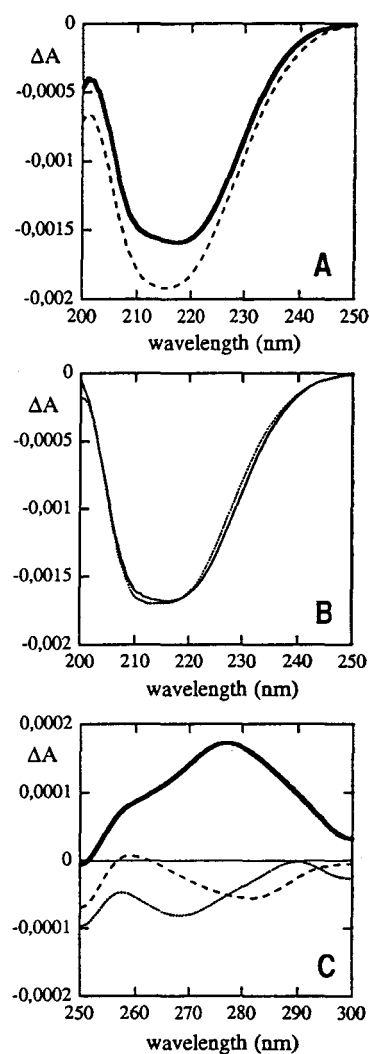


Fig. 6. Analysis of protein conformation by circular dichroism. (A) Far-ultraviolet CD spectra of H1 (thick solid line) and of H2 (dashed line). (B) Far-ultraviolet CD spectra of I α I (dotted line) and of an equimolar mixture of H1+H2 (thin solid line). In (A) and (B), the CD spectra were recorded using a 0.2-cm quartz cuvette and a protein concentration of 0.3 mg/ml. In (C) for near-ultraviolet spectra, we used a 1-cm cell and protein solution at 1.2 mg/ml for H1, 0.95 mg/ml for H2 and 1.28 mg/ml for I α I. Spectra were recorded in 20 mM sodium phosphate, pH 7.4, containing 150 mM NaCl.

same electrophoretic mobility as that of the H2 heavy chain. Furthermore the anti-H2 serum detected in noticeable amounts, several fragments of lower molecular masses, supporting the notion that the susceptibility to proteolysis by V-8 proteinase is slightly more important for H2 than for H1.

Circular dichroism spectra analysis. The conformation of the polypeptide backbone of the H1 and H2 heavy chains was determined by CD measurement in the 200–250-nm range and compared with that of the full-length protein I α I or of the equimolar mixture of H1+H2. As shown in Fig. 6, the CD spectrum of H1 differs slightly from that of H2 (Fig. 6A) whereas the spectrum of I α I is undistinguishable from that of H1+H2 (Fig. 6B). The identical shape of the CD spectra of H1 and H2 suggests that the two proteins are more or less equally folded. These CD spectra can be used to estimate the respective α -helicities of the proteins using an empirical relationship based on the molar circular dichroism at 220 nm (see Materials and Methods). Helix contents of 23.6% and 28.0% were determined for H1 and H2,

respectively. The global helicity of I α I (24.6%) is intermediate between that of H1 and H2, as one would expect from an equal contribution of the two heavy chains. H1 and H2 are equally susceptible to the addition of a reducing agent. Dithiothreitol was added at 1 mM (final concentration) to the protein samples and after 1.5-h incubation at room temperature, in both cases the helix content was reduced significantly by 8–11%.

Next we measured the circular dichroism in the 250–300-nm region in order to provide information on the environment of the aromatic amino acid residues (especially tryptophan). Although disulphide bonds may also contribute to the near-ultraviolet CD spectrum of proteins [16], in the present case the CD spectra recorded between 250 nm and 300 nm probably account primarily for the aromatic amino acid side chains because the proteins only contained two disulphide bonds. The spectra in Fig. 6C reveal marked differences between H1 and H2. The CD at 280 nm is positive in sign for H1 whereas it is negative for H2. Therefore we can conclude that the aromatic residues of the two heavy chains are in a different asymmetrical environment. It is important to note also that the CD spectra of H1 and H2 in the 250–300-nm region markedly differ from that of I α I (Fig. 6C). This suggests that the accessibility of the side chains of the aromatic amino acids changes when the two heavy chains are interacting with one another and/or with the bikunin-glycosaminoglycan moiety.

DISCUSSION

Little is known concerning the physiological role and the metabolism of I α I. However, Pratt et al. [17] demonstrated that after intravenous injection in mice, I α I was mainly taken up by liver and then converted into two smaller components (apparent molecular masses near 140 000 and 85 000 in SDS/PAGE). More recently, Kobayashi et al. [18] reported that I α I was bound by hyaluronan to the tumor cell surface in cell cultures and then cleaved by elastase-like proteinases to release bikunin, which thus becomes able to exert its physiological functions.

During inflammatory diseases, I α I might be hydrolyzed by proteinases released from stimulated neutrophils. We have demonstrated that, in this case, leucocyte elastase preferentially cleaves the H2 heavy chain and releases from I α I, high molecular mass compounds (apparent molecular masses near 120–150 000 and 85 000) and then truncated bikunin [5]. Thus the fragmentation pattern of I α I appears rather conserved.

Susceptibility to proteolysis is predominantly associated with sites within unstructured flexible regions of target proteins. Disulphide linkages may influence the tertiary structure of proteins and decrease their susceptibility to proteolysis. Our aim in this work was primarily to determine the topology of disulphide bonds in each of the I α I heavy chains. Each includes two cysteine residues in a sequence CPXC. In this case the distance between the two sulphur atoms seems to promote the formation of a disulphide bridge. Similar bonds have been previously described and the sequence CXXC is found in the active site of thiol-disulphide oxydoreductases [19, 20]. Using amino acid sequencing and MS-analysis, we demonstrate that Cys215 and Cys218 in H1 and Cys at positions 207 and 210 in H2 are disulphide linked.

Disulphide cross-links between adjacent cysteine residues are rarely encountered because the residues involved have a highly constrained conformation. Such a structure is found in regions implicated in local conformational changes [19, 21]. Concerning the H2 heavy chain, the vicinal cysteine residues are linked by a disulphide bond. It is of interest that those residues are found near the C-terminal extremity of the polypeptide

chain, in a region containing a high content of proline residues, which is, therefore, unstructured.

Usually proteins containing disulphide bonds migrate in SDS/PAGE under non-reducing conditions with a lower apparent molecular mass than under reducing conditions. In contrast, the H1 heavy chain exhibits an apparent molecular mass of 96 000 under non-reducing conditions but only 78 000 after previous reduction [5]. The latter estimate is compatible with the molecular mass of H1 ($m/z = 76258$) determined by MALDI-TOF-MS (unpublished results). Thus the unusual behaviour of this polypeptide chain does not appear to be due to other post-translational modifications. As previously reported [22], a large spacing of cysteines involved in an intramolecular disulphide bridge may be responsible for this unusual behaviour. This spacing is 271 residues for the H1 heavy chain. We have also determined that the thiol group of Cys31 is free. Located in the N-terminal region of H1, it would be particularly exposed. Its potential role remains to be elucidated.

The cysteine residues are completely conserved in the different mammalian species (human, pig, mouse, hamster) for which amino acid sequences of the H1 and H2 heavy chains are now known [23]. Therefore, the topology of disulphide bridges is also probably similar in those species. This topology is very different between H1 and H2 and, as recently suggested [23], would be related to the existence of structural constraints belonging to the two heavy chains which might each have a unique function. That hypothesis is supported by near-ultraviolet CD spectrometric analysis that indicates that H1 and H2 are not identical in terms of tertiary structural organization. Although different factors tend to influence the intensities of the aromatic CD bands, the lower CD amplitude measured with H2 may reflect more highly mobile aromatic side chains, when compared with those of H1. The difference observed between the CD spectra of H1, H2 and I α I in the 250–300-nm region indicate that the asymmetric environment of the aromatic amino acid side chain varies significantly when the two heavy chains are associated. It is plausible that the aromatic side chains of H1 are masked by those of H2 and vice versa (and/or by the bikunin moiety). Alternatively, the conformation of the heavy chains may be modified when the two units are interacting but this explanation is relatively unlikely given the identical CD spectra observed in the 200–250-nm region.

The H2 heavy chain lacks the large disulphide loop that characterizes H1. As assessed by the far-ultraviolet CD spectrum, this disulphide bridge is not important to determine the overall structure of the H1 heavy chain. The CD measurements suggest that H1 and H2 are more or less identically folded even though their helix contents differ slightly. Overall the two heavy chains exhibit the same secondary structural organization, whatever they are alone or in an equimolar mixture or covalently linked to bikunin by the glycosaminoglycan chain to make up I α I. Therefore we can conclude from the CD measurements that the H1 and H2 heavy chains differ at the tertiary structure level but present a roughly identical secondary structure.

To explain the proteolytic susceptibility of H2 which is preferentially cleaved near its C-terminal extremity when I α I was *in vitro* incubated in the presence of neutrophil proteinases [5, 6], each heavy chain and I α I were separately digested by V-8 proteinase. This enzyme was chosen for several reasons: (a) its specificity is very different from that of neutrophil proteinases since it hydrolyzes highly specifically peptide bonds at the carboxylic side of the glutamate residues in the conditions used; (b) it is less cationic than human leucocyte elastase and thus its possible affinity for the glycosaminoglycan could be less important; (c) according to Potempa et al. [8] the fragmentation patterns of I α I by V-8 proteinase and neutrophil proteinases are

different. Our results show that the susceptibilities of both heavy chains to proteolysis by V-8 proteinase are quite similar. Thus we have characterized a region particularly sensitive to proteolysis inside the amino acid sequence spacing the two disulphide-linked cysteines (Cys239 and Cys511) of the H1 heavy chain. Likewise Huang et al. [24] reported that H1, complexed with hyaluronan, was cleaved by V-8 proteinase in the same region, between Glu357 and Ser358. Moreover we demonstrate that the proteolytic susceptibility of both heavy chains is not decreased when they are covalently linked to make up IaI, thus confirming that the heavy chains, free or associated inside the IaI molecule, present a similar secondary structure.

The IaI cleavage by V-8 proteinase quickly occurs in the vicinity of the glycosaminoglycan chain, either in the C-terminal region of the heavy chains or in the N-terminal part of bikunin between Glu18 and Val19 [8]. In this case, beside truncated bikunin, a high-molecular-mass fragment near 200000 was released from IaI. It was detected by immunoblot analysis with specific antibodies directed against each of the two heavy chains (H1 and H2). Those cleavages probably occur inside a very hydrophilic surface-exposed region of the IaI molecule. During the course of the incubation, other high-molecular-mass fragments appeared. They were also immunologically revealed by the two anti-H1 and anti-H2 antibodies, thus establishing that V-8 proteinase progressively cleaves the two heavy chains of IaI from their N-terminal extremities. This fragmentation pattern is therefore very different from what we have previously described for IaI digested with leucocyte elastase [5]. Thus we conclude that the mechanism of the IaI degradation is dependent on the proteinase involved. We consider that the preferential cleavage of IaI in the C-terminal part of the heavy chains is due to the affinity of leucocyte elastase for the anionic glycosaminoglycan that is linked to the C-terminal Asp of both H1 and H2. In the same way, the particular susceptibility of the H2 heavy chain to elastase might be explained by the presence in its C-terminal part, of a cluster of sialylated O-glycans (unpublished results).

For many proteins, disulphide bonds are important for stabilization of tertiary structures which strongly influence their biological activity. A von Willebrand type A domain (vWA) has been detected in each heavy chain of IaI [25]. Those domains are involved in the recognition of various ligands, among which heparin. The vWA domains are also characterized by the absence of cysteine residues. However, the cysteines at the boundaries of those domains can form an intrachain loop protruding towards the outside, enhancing the accessibility for ligand interaction [26]. We note that for H1, vWA is included in the large loop between the disulphide cross-linked cysteines (Cys239 and Cys511).

The two heavy chains are responsible for the binding of IaI to heparin-Sepharose [10]. The site responsible for this interaction was not located; however, when the heavy chains were released from IaI by alkaline treatment, H1 was retained slightly better than H2 by heparin-Sepharose. In contrast, when IaI was dissociated by non-denaturing chondroitinase digestion, both the heavy chains developed a similar affinity for heparin. These results suggest that alkaline treatment may modify the three-dimensional structure and thus the biological activity of the heavy chains. The effects of the alkaline treatment will probably be more important for H2 which is more flexible, than for H1 which is stabilized by its disulphide bridges.

This work was supported in part by the *Ministère de l'Enseignement Supérieur et de la Recherche* (contract DRED E. A. 1052), the *convention INSERM-CFB* and by the *Université de Lille II*. We are grateful to Anne Dickès and Karine Gauthier for their skillful assistance. We are indebted to Guy Ricart (*Laboratoire de Spectrométrie de Masse, Centre*

Commun de Mesures, Université des Sciences et Technologies de Lille I, F-59655, Villeneuve d'Ascq, France) for his helpful assistance in the mass spectrometry interpretation.

REFERENCES

1. Salier, J. P., Rouet, P., Raguenez, G. & Daveau, M. (1996) The inter- α -inhibitor family: from structure to regulation, *Biochem. J.* **315**, 1–9.
2. Witte, J., Jochum, M., Scherer, R., Schramm, W., Hochstrasser, K. & Fritz, H. (1982) Disturbances of selected plasma proteins in hyperdynamic septic shock, *Int. Care Med.* **8**, 215–222.
3. McKeehan, W. L., Sakagami, Y., Hoshi, H. & McKeehan, K. A. (1986) Two apparent human endothelial cell growth factors from human hepatoma cells are tumor-associated proteinase inhibitor, *J. Biol. Chem.* **261**, 5378–5383.
4. Chen, L., Mao, J. T., McLean, L. R., Powers, R. W. & Larsen, W. J. (1994) Proteins of the inter- α -trypsin inhibitor family stabilize the cumulus extracellular matrix through their direct binding with hyaluronic acid, *J. Biol. Chem.* **269**, 28282–28287.
5. Balduyck, M., Piva, F., Mizon, C., Maes, P., Malki, N., Gressier, B., Michalski, C. & Mizon, J. (1993) Human leucocyte elastase (HLE) preferentially cleaves the heavy chain H2 of inter- α -trypsin inhibitor (ITI), *Biol. Chem. Hoppe-Seyler* **374**, 895–901.
6. Albani, D., Balduyck, M., Mizon, C. & Mizon, J. (1997) Inter- α -inhibitor as marker for neutrophil proteinase activity: an in vitro investigation, *J. Lab. Clin. Med.* **130**, 339–347.
7. Pratt, C. W. & Pizzo, S. V. (1987) Mechanism of action of inter- α -trypsin inhibitor, *Biochemistry* **26**, 2855–2863.
8. Potempa J., Kwon, K., Chawla, R. & Travis, J. (1988) Inter- α -trypsin inhibitor: inhibition spectrum of native and derived forms, *J. Biol. Chem.* **264**, 15109–15114.
9. Mizon, C., Balduyck, M., Albani, D., Michalski, C., Burnouf, T. & Mizon, J. (1996) Development of an enzyme-linked immunosorbent assay for human plasma inter- α -trypsin inhibitor (ITI) using specific antibodies against each of the H1 and H2 heavy chains, *J. Immunol. Methods* **190**, 61–70.
10. Michalski, C., Piva, F., Balduyck, M., Burnouf, T., Huart, J. J. & Mizon, J. (1994) Preparation and properties of a therapeutic inter- α -trypsin inhibitor concentrate from human plasma, *Vox Sang.* **67**, 329–336.
11. Carrette, O., Mizon, C., Sautière, P., Sesboué, R. & Mizon, J. (1996) Purification and characterization of pig inter- α -inhibitor and its constitutive heavy chains, *Biochim. Biophys. Acta* **1338**, 21–30.
12. Mantle, M., Stewart, G., Zayast, G. & King, M. (1990) The disulphide-bond content and rheological properties of intestinal mucins from normal subjects and patients with cystic fibrosis, *Biochem. J.* **266**, 597–604.
13. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* **227**, 680–685.
14. Stephano, J. L., Gould, M. & Rojas-Galicia, L. (1986) Advantages of picrate fixation for staining polypeptides in polyacrylamide gels, *Anal. Biochem.* **152**, 308–313.
15. Clark, D. J., Hill, C. S., Martin, S. R. & Thomas, J. O. (1988) α -Helix in the carboxy-terminal domains of histones H1 and H5, *EMBO J.* **7**, 69–75.
16. Kelly, S. M. & Price, N. C. (1997) The application of circular dichroism to studies of protein folding and unfolding, *Biochim. Biophys. Acta* **1338**, 161–185.
17. Pratt, C. W. & Pizzo, S. V. (1986) In vivo metabolism of inter- α -trypsin inhibitor and its proteinase complexes: evidence for proteinase transfer to α 2-macroglobulin and α 1-proteinase inhibitor, *Arch. Biochem. Biophys.* **248**, 587–596.
18. Kobayashi, H., Gotoh, J., Hirashima, Y. & Terao, T. (1996) Inter- α -trypsin inhibitor bound to tumor cells is cleaved into the heavy chains and the light chain on the cell surface, *J. Biol. Chem.* **271**, 11362–11367.
19. Lengweiler, S., Schaller, J. & Rickli, E. E. (1996) Identification of disulfide bonds in the ninth component (C9) of human complement, *FEBS Lett.* **380**, 8–12.
20. Holmgren, A. (1995) Thioredoxin structure and mechanism: conformational changes on oxidation of the active-site sulfhydryls to a disulfide, *Structure* **3**, 239–243.

21. Kao, P. N. & Karlin, A. (1986) Acetylcholine receptor binding site contains a disulfide cross-link between adjacent half-cystinyl residues, *J. Biol. Chem.* 261, 8085–8088.
22. Spiess, C., Happersberger, P., Glocker, M. O., Spiess, E., Rippe, K. & Ehrmann, M. (1997) Biochemical characterization and mass spectrometric disulfide bond mapping of periplasmic α -amylase MalS of *Escherichia coli*, *J. Biol. Chem.* 272, 22125–22133.
23. Nakatani, T., Suzuki, Y., Yamamoto, T. & Sinohara, H. (1997) Molecular cloning and sequencing of cDNAs encoding three heavy-chain precursors of the inter- α -trypsin inhibitor in syrian hamster: implications for the evolution of the inter- α -trypsin inhibitor heavy chain family, *J. Biochem. (Tokyo)* 122, 71–82.
24. Huang, L., Yoneda, M. & Kimata, K. (1993) A serum-derived hyaluronan-associated protein (SHAP) is the heavy chain of the inter- α -trypsin inhibitor, *J. Biol. Chem.* 268, 26725–26730.
25. Bork, P. & Rohde, K. (1991) More von Willebrand factor type A domains? Sequence similarities with malaria thrombospondin-related anonymous protein, dihydropyridine-sensitive calcium channel and inter- α -trypsin inhibitor, *Biochem. J.* 279, 908–910.
26. Colombatti, A. & Bonaldo, P. (1991) The superfamily of proteins with von Willebrand factor type A-like domains: one theme common to components of extracellular matrix, hemostasis, cellular adhesion, and defense mechanisms, *Blood* 77, 2305–2315.

2) Discussion et résultats complémentaires

Travaillant parallèlement, Olsen et coll [136] ont déterminé un schéma de pontage des cystéines des chaînes lourdes de l' α I identique au nôtre. En fait, ces deux études montrent que la topologie des ponts disulfure des chaînes lourdes H1 et H2 de l' α I humain est différente. En effet, contrairement à la chaîne H2 dont un pont disulfure implique deux cystéines voisines, la chaîne H1 renferme un pont disulfure formé entre deux cystéines distantes de 271 acides aminés. Comme le démontre la différence de mobilité électrophorétique de la chaîne H1 évaluée en SDS/PAGE avec ou sans réduction préalable, celui-ci influence probablement sa conformation. Cependant, les chaînes lourdes H1 et H2 renferment, dans la même région, un motif (CPXC) pratiquement identique comportant une liaison disulfure. Par conséquent, les deux uniques cystéines de la chaîne H3 qui forment également le motif (CPXC) sont très probablement liées de manière covalente. Un motif sensiblement similaire (CXXC) est présent dans le site actif de protéines présentant des fonctions d'oxydoréduction (thioredoxine). Selon Olsen et coll [136], celui-ci est également trouvé dans des protéines telles que la triglycéride lipase ou la protéinase à acide aspartique de *Rhizopus*. Cependant, la fonction biologique de ce motif (CPXC) retrouvé dans la séquence de toutes les chaînes lourdes exceptées celle de la chaîne H4 reste à démontrer.

Dans cette étude, nous démontrons également la présence d'un groupement thiol libre dans la partie N-terminale de la chaîne H1. Au contraire, Olsen et coll [136] identifient cette cystéine 31 de la chaîne H1 comme glycosylée par deux résidus de glucose ou de galactose. En effet, l'analyse structurale du peptide purifié Ser²⁶-Arg³⁶ montre (i) une différence de masse de 323,3 Da entre les masses théoriques et expérimentales de ce peptide, (ii) l'absence de résidu de phénylthiohydantoïne de la cystéine lors de la dégradation d'Edman et (iii) une réduction-alkylation impossible de ce peptide Ser²⁶-Arg³⁶. Pour en décider, nous avons purifié, par RP-HPLC (reverse phase-high performance liquid chromatography) le peptide Ser¹-Hse¹¹⁴ (Hse, homosérine) obtenu par clivage au BrCN de la chaîne H1 carboxamidométhylée (CAM). Par dégradation d'Edman, une séquence unique (S¹ATGR) était caractérisée. La masse moléculaire de ce peptide ($[M+Na]^+$ = 12362 Da), mesurée par spectrométrie de masse de type

MALDI, correspond à la masse théorique ($[M+Na]^+ = 12366$ Da) du peptide carboxamidométhylé Ser¹-Cys³¹(CAM)-Hse¹¹⁴. Nous avons ensuite clivé ce peptide par l'endoprotéinase arginine-C. L'analyse des masses de peptide purifié par RP-HPLC, permettait d'identifier 90 % de la séquence prévue. Les masses expérimentales 2685,8 et 3281,7 Da correspondent aux masses théoriques des peptides carboxamidométhylés : Gln¹²-Cys³¹(CAM)-Arg³⁶ (2687,1 Da) et Ser²⁶-Cys³¹(CAM)-Arg⁵⁴ (3279,7 Da). Par conséquent, la fonction thiol de la cystéine³¹ est réellement libre. La discordance entre les résultats de Olsen et coll [136] et les nôtres pourrait provenir des différentes techniques utilisées pour dissocier l'IαI. La dissociation enzymatique (chondroitinase ABC) des chaînes lourdes de l'IαI pourrait permettre sans doute de lever cette ambiguïté. La signification fonctionnelle de cette cystéine localisée dans une région particulièrement exposée de la protéine reste de toute manière indéfinie.

Compte tenu de la stricte conservation du nombre et de la localisation des résidus de cystéine des chaînes lourdes H1, H2 et H3 dans les différentes espèces animales [27, 127], **la topologie des ponts disulfure des chaînes lourdes de ces espèces est probablement similaire à celle de l'homme.**

Les régions flexibles des protéines sont généralement plus sensibles à l'action des protéinases que celles plus structurées. La protéolyse ménagée peut donc permettre d'évaluer la structure secondaire et tertiaire des protéines, de même que l'influence des ponts disulfure sur la stabilité moléculaire. Sous l'action de la protéinase V8, les chaînes lourdes H1 et H2 présentent une sensibilité similaire à la protéolyse ménagée. La région de la chaîne H1 située entre les cystéines pontées 239 et 511 présente un site exposé, particulièrement sensible à la dégradation. Balduyck et coll [131], utilisant l'élastase leucocytaire humaine, ont été les premiers à rapporter une dégradation protéolytique C-terminale des chaînes lourdes ainsi qu'une susceptibilité de protéolyse plus importante pour la chaîne H2 que pour la chaîne H1. Récemment, Blom et coll [137] ont complété ces résultats. Ils ont montré que, sous l'action de diverses protéinases (thermolysine, chymotrypsine, élastase et trypsine), les régions C-terminales des chaînes lourdes H1 et H2 et en particulier la région Cys²³¹-Cys⁵³⁹ de la chaîne H1 sont majoritairement protéolysées. Par conséquent, ces régions présentent très

certainement une conformation spatiale non structurée. D'ailleurs, ce sont des régions riches en résidus de proline [126]. Au contraire, les régions N-terminales des chaînes lourdes semblent protégées de la protéolyse ménagée. Elles présentent donc une conformation plus structurée.

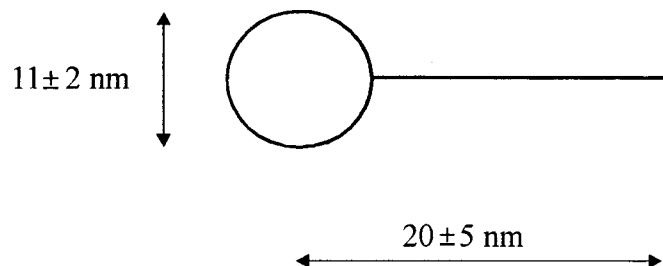
Associées en complexe au sein de l'I α I, les chaînes lourdes ne sont pas mieux protégées de la protéolyse par la protéinase V8 que les chaînes lourdes isolées. Ainsi, les chaînes H1 et H2, libres ou associées en complexe, présentent une conformation spatiale similaire qui serait faite d'une moitié N-terminale structurée et au contraire d'une région C-terminale désordonnée. Les chaînes polypeptidiques de l'I α I ne semblent donc pas se stabiliser entre elles.

Par dichroïsme circulaire, nous démontrons que les chaînes lourdes H1 et H2 présentent une structure secondaire identique faite respectivement de 24 et 28% d'hélicité de type α . Blom et coll [137] aboutissent aux mêmes conclusions mais rapportent des taux d'hélicité de l'ordre de 34% pour les deux chaînes lourdes. D'autre part, les courbes de dichroïsme circulaire dans l'ultraviolet lointain sont similaires que les chaînes lourdes H1 et H2 soient séparées ou en mélange équimolaire ou covalentiellement associées au sein de l'I α I. Ce fait témoigne de structures secondaires similaires. Par contre, prises isolément, les chaînes lourdes H1 et H2 diffèrent en terme de structures tertiaires comme le montrent les spectres dichroïques dans l'ultraviolet proche. De plus, dans cette région du spectre ultraviolet, les différences observées entre les spectres de H1, H2 et de l'I α I témoignent de l'existence d'interactions inter chaînes au sein de la molécule d'I α I.

En microscopie électronique [137], les chaînes lourdes purifiées apparaissent constituées d'un domaine globulaire et d'une région étendue de 15 nm de long. L'I α I humain, quant à lui, apparaît composé de deux domaines globulaires reliés par une fine structure de 31 nm présentant, en son centre, une petite structure globulaire. Ces structures, observées en microscopie électronique, sont présentées sous forme de schéma dans la figure 14.

L'ensemble de ces résultats suggère que la moitié N-terminale des chaînes lourdes est **structurée** et qu'elle forment un **domaine globulaire**. Par contre, l'autre moitié (C-terminale) apparaît comme une **structure étendue et désordonnée** qui correspond, en microscopie électronique, aux structures fines. Ainsi, dans la mesure où les structures observées en microscopie électronique seraient réellement représentatives de celles existant *in vivo*, les domaines globulaires des chaînes lourdes H1, H2 ou de la bikunine seraient distants les uns des autres et par conséquent, ne pourraient s'influencer mutuellement pour l'établissement de leur structure secondaire. Par contre, quand elles sont, soit libres, soit associées à la chaîne de GAG, les régions C-terminales des chaînes lourdes pourraient différer en terme de structure tertiaire. Par conséquent, la formation de la liaison PGP influencerait la structure désordonnée des régions C-terminales des chaînes lourdes de l'I α I.

Chaînes lourdes



Inter- α -inhibiteur

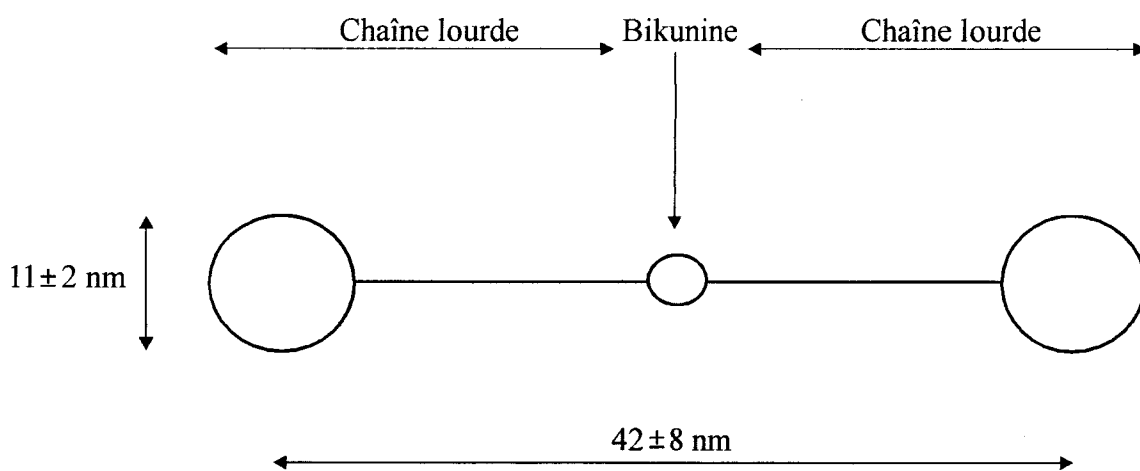


Figure 14 : Représentation schématique des structures observées en microscopie électronique selon Blom et coll [137].

II Structures glycanniques des chaînes lourdes de l'I α I humain

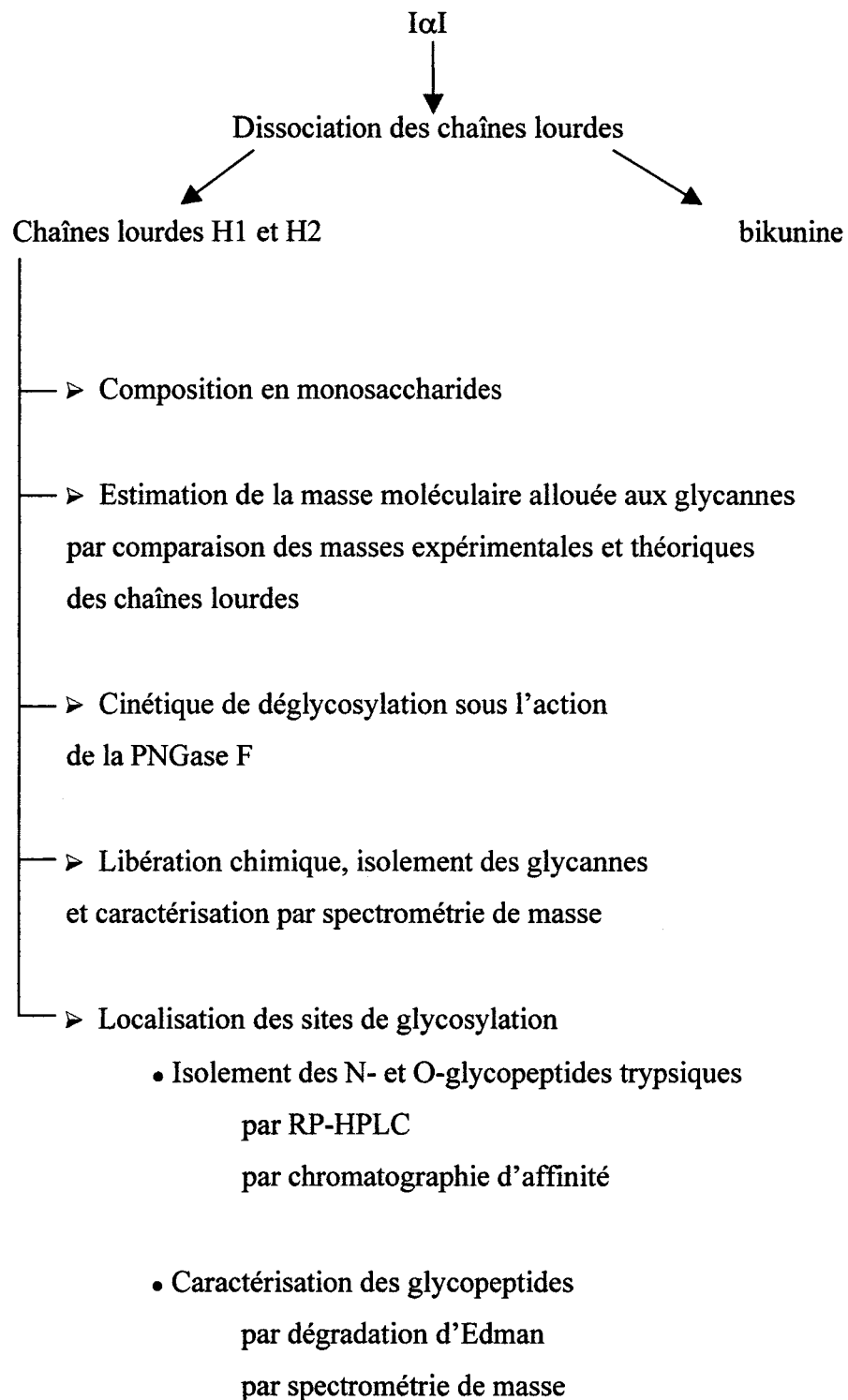
La glycosylation des protéines [138] est une modification post-traductionnelle ubiquitaire qui survient dans le réticulum endoplasmique et dans l'appareil de Golgi des cellules. Elle se caractérise par l'addition, sur certains acides aminés, de séquences glucidiques variées. Elle dépend de la protéine elle-même, de l'équipement enzymatique des organites précédemment cités mais aussi de l'état physiologique de la cellule.

On distingue principalement deux types de glycosylation : la N-glycosylation qui n'affecte que les asparagines appartenant à une séquence consensus Asn-Xaa-Ser/Thr et la O-glycosylation qui intervient sur des résidus de sérine ou de thréonine placés dans des régions riches en résidus de sérine, thréonine et proline. Ces deux types de glycosylation ne s'excluent pas l'un l'autre et peuvent, par conséquent, être retrouvés ensemble au sein d'une glycoprotéine. Les structures glycanniques formées diffèrent non seulement en terme de séquence ou de longueur de chaîne mais également en terme de position de liaison et d'anoméricité. Ainsi, la glycosylation génère un ensemble de molécules dénommées glycoformes qui partagent une ossature polypeptidique identique mais se distinguent par le nombre, la localisation et/ou la nature de leurs glycannes

Ces hétérogénéités glycanniques sont responsables de différentes propriétés physiques et biochimiques des glycoprotéines et peuvent aboutir à une diversité de fonction [132, 134]. Par exemple, la N-glycosylation influence la mise en conformation des protéines. De même, la glycosylation, d'un point de vue général, joue un rôle de protection vis-à-vis de l'action de protéinases. Plus important encore, il est désormais clairement établi que la glycosylation des protéines agit en tant que signal de reconnaissance dans de nombreux processus physiologiques ou pathologiques. Ceux-ci incluent, par exemple, la clairance plasmatique des glycoprotéines [139], le trafic intracellulaire des enzymes [140] mais également les interactions cellule-cellule ou molécule-cellule comme l'adressage leucocytaire vers les foyers inflammatoires ou l'adhésion cellulaire de virus ou de micro-organismes [132, 134].

Mise à part la chaîne de GAG, la glycosylation des chaînes polypeptidiques de l'I α I est restée pendant longtemps méconnue. Ainsi, alors que la présence d'un N-glycane sur la bikunine est connue depuis 1981 [118], sa structure n'a été établie que 11 ans plus tard [119]. Il en est de même pour les chaînes lourdes de l'I α I. En effet, les séquences en acides aminés, établies entre 1988 et 1992, permettaient d'identifier respectivement 2 et 3 sites potentiels de N-glycosylation sur les chaînes lourdes H1 et H2. Dès 1988, Gebhard et coll [125] annonçaient la présence de 2 N-glycannes sur la chaîne H2. 4 ans plus tard, Malki et coll [122], par la composition en monosaccharides et la réactivité des chaînes lourdes vis-à-vis de la Concanavaleine A, apportaient la preuve définitive de l'existence de N-glycannes de type complexe sur les chaînes lourdes H1 et H2 mais également celle de O-glycannes sur la chaîne H2. Par la suite Enghild et coll [17], lors d'une étude non destinée à la caractérisation de ces structures, identifiaient un O-glycane trisaccharidique NeuAc-Gal-GalNAc situé sur la thréonine 637 de la chaîne H2.

Par conséquent, en raison de l'importance biologique des structures glycaniques, il nous est apparu nécessaire de préciser le nombre, la structure ainsi que la localisation des glycannes présents sur chaque chaîne lourde de l'I α I. Pour atteindre ce but, nous avons adopté la démarche de recherche présentée dans la page suivante.



1) Publication

Les résultats obtenus ont fait l'objet de la publication suivante :
 « Glycosylation pattern of human inter- α -inhibitor heavy chains ».

Glycosylation pattern of human inter- α -inhibitor heavy chains

Christophe FLAHAUT*†, Calliope CAPON*†, Malika BALDUYCK*, Guy RICART‡, Pierre SAUTIERE§ and Jacques MIZON*¹

*Laboratoire de Biochimie, Faculté de Pharmacie, Université de Lille II, Avenue du Professeur Laguesse, B.P. 83, F-59006 Lille, France, †Laboratoire de Chimie Biologique, Université des Sciences et Technologies de Lille I, F-59655, Villeneuve d'Ascq, France, ‡Laboratoire de Spectrométrie de Masse, Centre Commun de Mesures, Université des Sciences et Technologies de Lille I, F-59655, Villeneuve d'Ascq, France, and §Département V, URA 1309 CNRS, Institut de Biologie et Institut Pasteur de Lille, 1 rue du Professeur Calmette, B.P. 447, F-59021 Lille, France

Human inter- α -inhibitor ($I\alpha I$) is a plasma serine-proteinase inhibitor. It consists of three polypeptide chains covalently linked by a glycosaminoglycan chain: a light chain named bikunin carrying the anti-proteinase activity and two heavy chains, H1 and H2, which exhibit specific properties, e.g. they interact with hyaluronan thus stabilizing the extracellular matrix. In this study, using matrix-assisted laser desorption ionization–time-of-flight MS and amino acid sequencing of tryptic peptides, we provide a detailed analysis of the glycosylation pattern of both heavy chains. H1 carries two complex-type N-glycans of pre-

dominantly biantennary structure linked to asparagine residues at positions 256 and 559 respectively. In contrast, the oligosaccharides attached to H2 are a complex-type N-glycan in the N-terminal region of the protein (Asn⁶⁴) and three to four type-1 core-structure O-glycans mono- or di-sialylated, clustered in the C-terminal region. We propose that these O-glycans might function as a recognition signal for the H2 heavy chain. The biological implications of this hypothesis, notably for the biosynthetic pathway of $I\alpha I$, are discussed.

INTRODUCTION

Inter- α -inhibitor ($I\alpha I$) is the leading member of a family of human plasma serine-proteinase inhibitors. $I\alpha I$ consists of three polypeptide chains, covalently linked by a glycosaminoglycan chain: two heavy chains H1 and H2 (M_r approx. 80000) and a light chain (M_r 26000) named bikunin, because it is structurally related to Kunitz-type proteinase inhibitors (recently reviewed in [1]). For many years, $I\alpha I$ has been considered to be a precursor for smaller anti-proteinases carrying bikunin, able to diffuse more easily and thus participate in tissue protection [2]. Furthermore, besides its proteinase-inhibiting activity, bikunin may play additional roles, e.g. modulation of cellular proliferation, anti-inflammatory activity, etc.

It is now evident that the $I\alpha I$ heavy chains also exert specific functions, e.g. they are responsible for the hyaluronan (HA)-binding capacity of $I\alpha I$ reported previously [3–5]. In this respect, $I\alpha I$ may help to stabilize the extracellular matrix. On the other hand, it has been proposed that the heavy chains mediate the binding of $I\alpha I$ to the cell surface [6]; thus they would serve as specific carriers for bikunin delivery in various tissues.

However, the possible mechanisms for HA binding to $I\alpha I$ remain open to discussion and therefore a thorough structural characterization of the heavy chains might assist in a better understanding of their physiological role. The amino acid sequences of the two heavy chains H1 and H2 are highly similar, with 40% identity [7]. With regard to glycosylation, two asparagine residues were identified as carbohydrate-binding sites in each of the heavy chains [8]. However, we have reported the presence of *N*-acetylgalactosamine (GalNAc) in the H2 heavy chain, thus suggesting the probable occurrence of O-glycosylation [9], and more recently [10] it has been demonstrated that Thr⁶³⁷

in H2 is effectively O-glycosylated. Here, we provide a detailed analysis of the glycosylation pattern of both H1 and H2 of $I\alpha I$.

MATERIALS AND METHODS

Materials

Human $I\alpha I$ was isolated as described recently [11] and supplied by the Laboratoire Français du Fractionnement et des Biotechnologies (Lille, France). Tosylphenylalanylchloromethane (Tos-Phe-CH₂Cl)-treated trypsin from bovine pancreas and endoproteinase Glu-C from *Staphylococcus aureus* were from Sigma and Boehringer respectively. Recombinant peptide-N-glycosidase F (PNGase F) from *Flavobacterium meningosepticum* was from Oxford GlycoSystems (Abingdon, Oxon., U.K.). All the solvents and reagents used for gas-phase sequencing were from Perkin–Elmer (Norwalk, CT, U.S.A.).

Isolation and characterization of the $I\alpha I$ heavy chains

The H1 and H2 heavy chains were isolated from $I\alpha I$ by anion-exchange chromatography following dissociation by treatment with NH₂OH. Briefly, $I\alpha I$ (80 mg) in 100 ml of water was mixed with an equal vol. of 2 M NH₂OH adjusted to pH 9.2 and stirred for 1 h at 37 °C before dialysis overnight at 4 °C against a 20 mM Tris/HCl buffer, pH 7.6, containing 4 mM EDTA (buffer A). The efficacy of the dissociation was checked by SDS/PAGE (7.5% gels) with Coomassie Brilliant Blue staining. The dissociated $I\alpha I$ was then loaded on to a Q-Sepharose fast-flow column (16 cm × 1.5 cm) equilibrated in buffer A at a flow rate of 100 ml/h. After washing the column for 30 min with the same

Abbreviations used: $I\alpha I$, inter- α -inhibitor; GalNAc, *N*-acetylgalactosamine; Tos-Phe-CH₂Cl, tosylphenylalanylchloromethane; PNGase F, peptide-N-glycosidase F; i.d., internal diameter; RP, reverse-phase; TFA, trifluoroacetic acid; MALDI–TOF–MS, matrix-assisted laser desorption ionization–time-of-flight MS; PTH, phenylthiohydantoin; 2-AB, 2-aminobenzamide; HA, hyaluronan; NeuAc, *N*-acetylneuraminic acid.

¹ To whom correspondence should be addressed (e-mail jmizon@phare.univ-lille2.fr).

buffer, the elution was performed by using a linear gradient from 0 to 0.5 M NaCl in buffer A for 4 h and then 1 M NaCl in buffer A was applied to the column. The A_{280} was monitored and the eluted fractions further analysed by SDS/PAGE (10% gels). The fractions containing each heavy chain were collected and desalted if necessary by HPLC using a Synchropak RP1 column [100 mm \times 4.6 mm internal diameter (i.d.); Alltech, Deerfield, IL, U.S.A.].

For PNGase F digestion, each pure heavy chain (60 μ g) in 60 μ l of a 10 mM phosphate buffer, pH 7.5, containing 25 mM EDTA, 0.5% (w/v) SDS and 5% (v/v) 2-mercaptoethanol was heated for 2 min at 100 °C. After cooling the mixture, 20 μ l of 10% (v/v) Triton X-100 and 10 μ l of PNGase F (5 units) were successively added. For H1, the reaction mixture was incubated for 3 h at 25 °C in a water bath.

In order to enhance the deglycosylation of H2, an additional 5 units of PNGase F was added after 65 h and the incubation was prolonged up to 72 h. Samples were taken at different incubation times for SDS/PAGE analysis.

Reduction, alkylation and tryptic hydrolysis of the heavy chains

5 mg of each heavy chain in 5 ml of 100 mM Tris/HCl, pH 7.25, were denatured by 6 M guanidine hydrochloride. The pH was then adjusted to 8.0. The samples were flushed with nitrogen and incubated for 30 min at 50 °C. After addition of 12 mM dithiothreitol, the samples were again flushed with nitrogen and incubated at 50 °C for 4 h. 300 mM iodoacetamide was then added. The reaction was carried out in the dark at room temperature for 30 min under a N₂ atmosphere. After extensive dialysis of the reaction mixtures against a 100 mM Tris/HCl buffer, pH 8.0, Tos-Phe-CH₂Cl-treated trypsin was added with an enzyme-to-substrate ratio of 1:50 (by mass) and the mixture incubated at 37 °C for 24 h. The hydrolysis was stopped by addition of 4-(2-aminoethyl)-benzenesulphonyl fluoride (AEBSF) at 1 mM final concentration. The fragments obtained were separated either by reverse-phase (RP)-HPLC or affinity chromatography using lectin-coupled agarose.

RP-HPLC separation of trypsin-digested peptides of the H2 heavy chain

The tryptic digest of H2 was applied to a C18 RP 5 μ m Zorbax column (250 mm \times 4.6 mm i.d.) equipped with a guard column (50 mm \times 4.6 mm i.d.) of the same phase (Interchim, Montluçon, France) and equilibrated with 0.1% (v/v) trifluoroacetic acid (TFA) in water. For elution, isocratic conditions with the same solvent were applied for 5 min, followed by a linear gradient of acetonitrile from 0 to 80% in 0.1% (v/v) TFA over 120 min at a flow rate of 0.7 ml/min. Eluates were monitored by A_{215} . Fractions 21–23 containing glycopeptides were detected by molar carbohydrate composition and matrix-assisted laser desorption ionization–time-of-flight MS (MALDI–TOF–MS) and further purified on the same column by using a linear gradient of acetonitrile from 20–70% over 100 min.

Separation of glycopeptides by affinity chromatography on immobilized lectins

The peptide digests corresponding to 5 mg (in 5 ml buffer) of H1 or H2, after reduction, alkylation and tryptic hydrolysis, were loaded on to a Concanavalin-A–Sepharose column (5 ml) equilibrated in 50 mM Tris/HCl/100 mM NaCl buffer, pH 7.4, containing 1 mM CaCl₂, 1 mM MnCl₂, 1 mM MgCl₂ and 0.02% (w/v) NaN₃ (buffer B). After washing the column with 10 ml of

buffer B, the elution was performed by 10 ml of 0.5 M α -methylglucoside in buffer B at a flow rate of 20 ml/h, with detection by A_{280} . For the H2 digest, the washed fractions were pooled and loaded at the same flow rate on to a Jacalin–agarose column (2 ml) equilibrated in buffer B. After washing the column (6 ml buffer B), glycopeptides were eluted by 0.5 M α -methylgalactoside in buffer B (6 ml). The different separated fractions were concentrated into 2 ml aliquots in a vacuum centrifuge (Jouan, Saint-Nazaire, France) before desalting on a Sep-Pak C18 cartridge (Waters U.K. Ltd., Watford, Herts., U.K.) according to the manufacturer's instructions. They were then freeze-dried before analysis by MALDI–TOF–MS. In preparation for amino acid sequencing, the glycopeptide fractions were purified using an ABI 130A MicroBore HPLC system (Perkin–Elmer). The freeze-dried fractions were dissolved in 0.05% (v/v) TFA and separated on a C18 Vydac 218 TP column (250 mm \times 1 mm i.d., 10 μ m particle size, 300 nm pore size), using a linear gradient from 0 to 65% acetonitrile in 0.05% (v/v) TFA over 120 min at a flow rate of 50 μ l/min. The A_{215} was monitored.

Analysis of oligosaccharides

Glycans were released by hydrazinolysis from 2 mg of heavy chains, using the GlycoPrep 1000 (Oxford GlycoSystems) with N-mode. The resulting oligosaccharides were desalted by gel-filtration chromatography on a Bio-Gel P2 column (90 cm \times 1.6 cm) with deionized H₂O. Desalted glycans were analysed with respect to molar carbohydrate composition and labelled with 2-aminobenzamide (2-AB) [12]. AB-labelled glycans were then separated by HPLC on a GlycoSep H column (100 mm \times 3 mm i.d., Oxford GlycoSystems), using a gradient of acetonitrile from 0 to 40% in 0.05% (v/v) TFA over 30 min, followed by 5 min at 40% and 15 min from 40–60% at a flow rate of 0.5 ml/min. The A_{254} was monitored. All fractions were further analysed by MALDI–TOF–MS.

Analytical methods

PAGE was performed using 7.5% and 10% polyacrylamide slab gels under reducing or non-reducing conditions [13] with Coomassie Blue staining [14].

Carbohydrate analysis was performed using gas chromatography with a silicone OV 101 capillary column (0.32 mm \times 25 m). Samples were analysed after methanolysis (0.5 M HCl in methanol for 24 h at 80 °C), followed by N-reacetylation and trimethylsilylation, as described in [15] with slight modifications [16].

N-terminal sequencing was carried out on a gas-phase sequencer (Procise 492, Perkin–Elmer) using the pulsed-liquid programme. Phenylthiohydantoin (PTH) derivatives of amino acids were identified on-line on a 120 A amino acid analyser (Perkin–Elmer). The yields expressed in picomol of PTH-amino acids correspond to raw data.

MALDI–TOF–MS was performed on a 'Vision 2000' time-of-flight instrument (Finnigan MAT, Bremen, Germany) equipped with a 337 nm UV Laser. The mass spectra were acquired in linear or reflection modes with positive or negative detection and with or without 10 kV accelerating voltage. Aliquots (1 μ l) of the analyte solution (15 pmol) were mixed with an equal vol. of the matrix solution, 2,5-dihydroxybenzoic acid [10 mg/ml dissolved in acetonitrile–H₂O (30:70, by vol.)]. External calibration was performed using either BSA (M_r 66431; Sigma) for mass determination of the heavy chains or angiotensin I (M_r 1296.7; Sigma) as standards, and bovine insulin (M_r 5733.6; Sigma) for

mass determination of the glycopeptides and the released oligosaccharides respectively.

For endoproteinase Glu-C hydrolysis of O-glycopeptides of H2, pure freeze-dried peptides (6 nmol) were dissolved in 40 μ l of 50 mM ammonium acetate buffer, pH 4.0, and digested at 37 °C for 15 h using an enzyme-to-peptide ratio of 1:50. Digestion was then stopped by the addition of 5 μ l of 10% (v/v) formic acid before freeze-drying and MALDI-TOF-MS analysis.

Chemical desialylation was carried out by heating peptides at 80 °C in a closed reaction vial in the presence of 5% (v/v) formic acid at pH 2.0.

RESULTS

Isolation and characterization of the I α I heavy chains

I α I was dissociated by NH₂OH treatment, which selectively cleaves the ester bonds between an internal GalNAc molecule in the glycosaminoglycan chain and the C-terminal aspartate of each heavy chain, namely H1 and H2. They were then isolated by anion-exchange chromatography and analysed by SDS/PAGE with prior reduction. H1 and H2 moved as a unique protein band; their relative molecular masses were 78000 and 85000 respectively, as described previously [17].

The purity of the preparations was also assessed by MALDI-TOF-MS (Figure 1). The mass of H1, determined from the singly

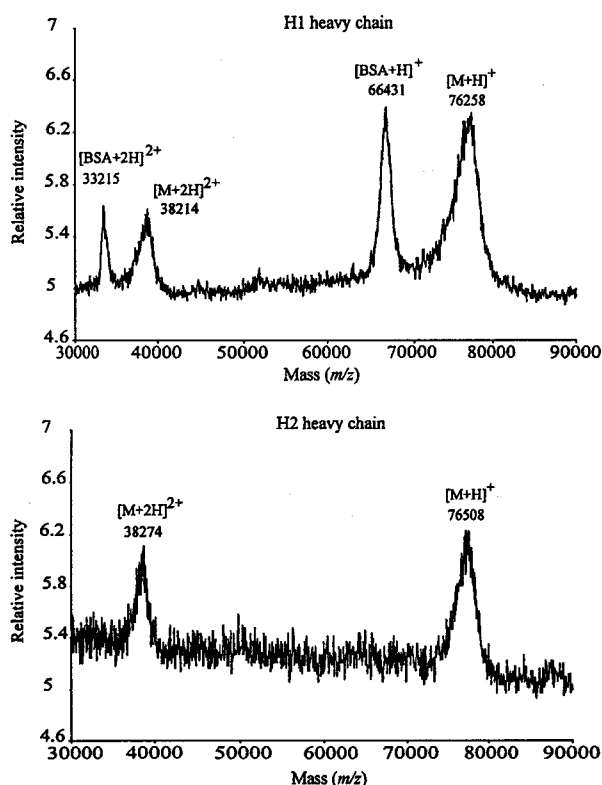


Figure 1 Linear time-of-flight mass spectrum for the H1 and H2 heavy chains

The singly charged ion of H1 (top panel) appeared at m/z 76258; the same signal for H2 (bottom panel) appeared at m/z 76508. The double-charged ions of the two heavy chains were also detected (m/z 38214 for H1; m/z 38274 for H2). BSA was used as an external calibration standard in both cases.

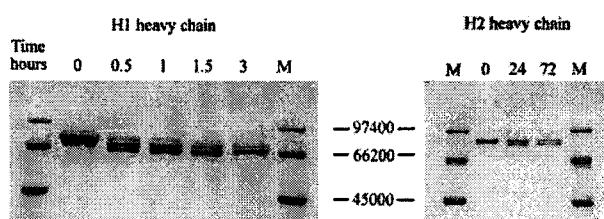


Figure 2 PNGase F digestion patterns of the H1 and H2 heavy chains

Samples (1 μ g), after different incubation times as indicated, were analysed by SDS/PAGE (10% gels) under reducing conditions with Coomassie Blue staining. Left panel: H1; right panel: H2. M, low-molecular-mass markers from Bio-Rad: phosphorylase B, 97400 Da; BSA, 66200 Da; egg white albumin, 45000 Da.

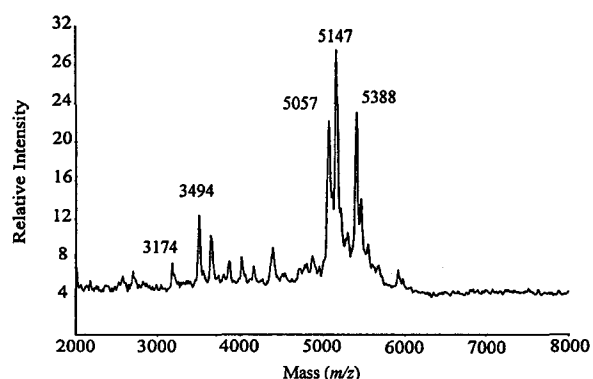
charged ion, was 76258. Subtraction of the protein mass of 71892, calculated from the amino acid sequence, resulted in a mass of 4366 for the carbohydrate moiety. By employing the same method, the molecular mass on the basis of the amino acid sequence of H2 (72432) was compared with the mass 76508 determined by MALDI-TOF-MS. The difference (4076) indicates an increase in mass due to a post-translational modification, presumably glycosylation. For each I α I heavy chain, MS analysis showed a broader peak than for BSA used as external calibrator; the broadening of the peak was related to the heterogeneity of glycosylation. Thus the glycosylation patterns of H1 and H2 were successively studied.

Glycosylation pattern of the H1 heavy chain

Carbohydrate composition of H1 (Table 1) was in agreement with the presence of N-linked glycans, mainly of biantennary type. The presence of two N-linked glycans was established by PNGase F digestion of H1 and subsequent SDS/PAGE analysis. Indeed, two bands (M_r , approx. 76000 and 74000) successively appeared during the incubation time (Figure 2).

To identify the glycosylation sites, glycopeptides obtained by tryptic digestion of reduced and carboxamidomethylated H1 heavy chain were extracted by affinity chromatography on Concanavalin-A-Sepharose. The glycopeptides eluted by α -methylglucoside were desalted on Sep-Pak cartridges and analysed by MALDI-TOF-MS. The spectrum (Figure 3) showed three strong ion signals at m/z 5388, 5147 and 5057. To determine the structure of these components, the glycopeptides were further purified by RP-HPLC and subjected to N-terminal sequencing. The intense signals 5388 and 5147 were assigned to peptides of Asp²³⁶-Lys²⁶² and Ile²³⁸-Lys²⁶² respectively, which both comprise the potential N-glycosylation acceptor site Asn²⁵⁶-Leu-Thr and a di-sialylated biantennary glycan. The signal observed at m/z 5057 was consistent with an identical di-sialylated biantennary structure linked to the second potential N-glycosylation site Asn⁵⁵⁹-Leu-Ser located in the peptide Ala⁵⁵⁸-Arg⁵⁸³. As indicated in Figure 3, we also assigned a mass composition to other signals (3494 and 3174). Interestingly, the signal at m/z 3494 corresponds to the peptide Phe²⁵²-Lys²⁶², which results from cleavage at the C-terminal side of Phe²⁵¹. Such a cleavage by trypsin has been previously observed [18], even when chymotryptic activity had been chemically inhibited. Therefore our results confirm that the two potential N-glycosylation sites of H1 (namely Asn²⁵⁶ and Asn⁵⁵⁹) are effectively occupied.

Furthermore, total glycans were released from H1 by hydrazinolysis, labelled with 2-AB and analysed on a GlycoSep H column. The chromatographic elution profile was followed by



Glycopeptides			
Observed Mass	N-terminal Sequence	Identification or position in sequence	Theoretical Mass
5388	DKICD	D ³³⁶ -N ²⁵⁶ -K ²⁶²	5378
5147	ICDLL	I ²³⁸ -N ²⁵⁶ -K ²⁶²	5135
5057	A-LSS	A ⁵⁵⁸ -N ⁵⁵⁹ -R ⁵⁸³	5061
3494	FAPQ-	F ²⁵² -N ²⁵⁶ -K ²⁶²	3481
3174	A-LSS	A ⁵⁵⁸ -N ⁵⁵⁹ -R ⁵⁶⁶	3163

Figure 3 Linear time-of-flight MS and assignment of signals to the N-glycopeptides of H1 obtained by affinity chromatography on Concanavalin-A-Sepharose

The conditions used are as described in the Materials and methods section. Asn at positions 256 and 559 correspond to potential sites of N-glycosylation.

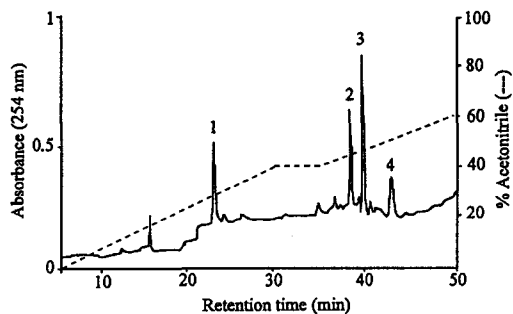


Figure 4 Fractionation of 2-AB-labelled oligosaccharides from the H1 heavy chain

The labelled oligosaccharides were fractionated on a GlycoSep H column (100 mm × 3 mm, i.d.) developed with an acetonitrile gradient, as described in the Materials and methods section. According to MALDI-TOF-MS analysis, peak 1 corresponds to 2-AB in excess; peaks 2 and 3 contain mono- and di-sialylated biantennary structures respectively; peak 4 corresponds to a triantennary structure.

monitoring A_{254} (Figure 4). The major fractions corresponding to peaks 2 and 3 were identified by MALDI-TOF-MS as mono- and di-sialylated biantennary structures respectively. They represent about 28% and 47% of the released glycans respectively. The minor fraction corresponding to peak 4 represents

Table 1 Carbohydrate composition of the glycans linked to the $\alpha 1$ heavy chains

The molar ratios were calculated on the basis of three mannose units. Since the GlcNAc-Asn linkage is only partially cleaved by methanolysis [15,16], the number of GlcNAc is underestimated (about one residue for three mannose units).

Heavy chain	Molar ratios						
	Fuc	Gal	Man	Glc	GalNAc	GlcNAc	NeuAc
H1	—	3	3	—	—	2.5	1.7
H2	—	5.6	3	—	3.1	3	7

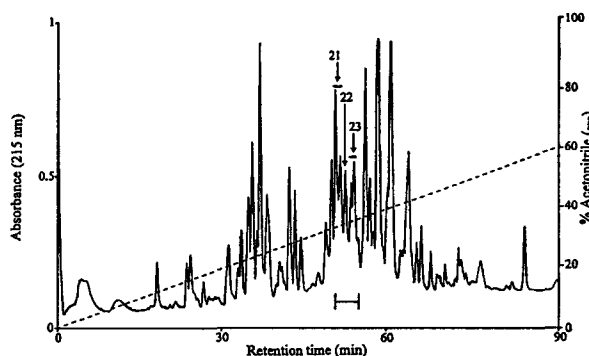


Figure 5 RP-HPLC elution profile of the tryptic digest from H2

The peptides generated by tryptic hydrolysis were injected on to a C18 RP 5 μ m Zorbax column (250 mm × 4.6 mm i.d.) equipped with a guard column (50 mm × 4.6 mm i.d.) of the same phase and equilibrated with 0.1% (v/v) TFA in H₂O. The elution was performed as described in the Materials and methods section.

23% of these glycans and was characterized as a triantennary structure.

Glycosylation pattern of the H2 heavy chain

Carbohydrate composition of H2 (Table 1) appeared more complex compared with that of the H1 heavy chain. The increased contents of sialic acid and galactose and the presence of GalNAc suggest the co-existence of mucin-type O-linked glycans and complex-type N-glycans.

Characterization of the N-glycopeptide of H2

H2 contains three N-glycosylation sites that are candidates for modification by addition of carbohydrates: Asn⁴²-Asn-Ser, Asn⁶⁴-Phe-Ser and Asn³⁹¹-Ile-Ser. SDS/PAGE analysis showed that when H2 was incubated in the presence of PNGase F, its relative molecular mass shifted slowly from 85000 (native form) to 83000 (deglycosylated form) (Figure 2). Thus only one N-linked carbohydrate chain per mol of H2 would be present.

To identify attachment sites and structures of glycans on H2, the glycoprotein was reduced, carboxamidomethylated and then digested by trypsin. The resulting peptide fragments were separated by RP-HPLC (Figure 5). The purified fractions were analysed by MALDI-TOF-MS.

Sequence $G^{59}-A-F-I-S-N^{64}-F-S-M-T-V-D-G-K^{72}$
 Yields in picomoles 95-155-172-182-37 - 110-nd-nd-nd-nd-nd-nd-nd

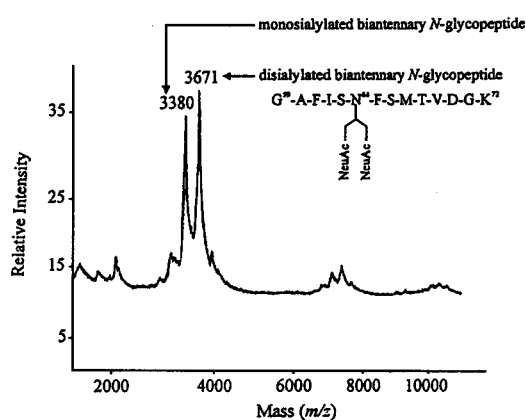


Figure 6 MALDI-TOF-MS analysis and N-terminal sequencing of the N-glycopeptide Gly⁵⁹-Lys⁷² from the H2 heavy chain

The glycopeptide isolated by RP-HPLC was analysed by MALDI-TOF-MS using a linear negative mode. Ions at m/z 3671 and 3380 correspond to the theoretical mass of the peptide Gly⁵⁹-Lys⁷² (1473 Da) with an additional N-glycan biantennary structure, either di-sialylated (2204 Da) or monosialylated (1913 Da). nd, not determined.

Two peptides with masses calculated theoretically from the known sequences around the potential N-glycosylation site were found: Val⁴⁰-Asn⁴²-Lys⁵⁸ and Glu³⁸⁶-Asn³⁹¹-Lys⁴⁰⁹. However, the naked peptide Gly⁵⁹-Asn⁶⁴-Lys⁷² was not detected: the predicted mass of this peptide, including the biantennary complex-type N-glycan, would be approx. 3500 Da. Glyco-

peptides with masses of 3500 Da were indeed characterized in fractions 21, 22 and 23 from HPLC (Figure 5). These fractions were pooled and rechromatographed as indicated in the Materials and methods section. The expected glycopeptides were clearly identified by Edman degradation (Figure 6); Asn⁶⁴ at the glycosylation site was, of course, not detected.

The structure of the N-linked carbohydrate chain was mainly of the biantennary type, as shown by MS data (Figure 6); the major signals at m/z 3671 and 3380 account for the presence of biantennary structure, di- and mono-sialylated respectively, attached to the peptide. Secondly, the structure was also shown by chromatographic analysis on a GlycoSep H column of oligosaccharides released from H2 by hydrazinolysis and then labelled by reductive amination with 2-AB; biantennary structure without, or with, one or two sialic acid residues represented 15%, 52% and 32% of N-glycans respectively. Thirdly, the glycopeptide Gly⁵⁹-Lys⁷² was easily purified from a tryptic digest of H2 by affinity chromatography on Concanavalin-A-Sepharose.

Characterization of the O-glycopeptides of H2

The carbohydrate composition (presence of GalNAc) and MS analysis (family of peaks with m/z approx. 6000) of the fractions 21, 22 and 23, obtained by RP chromatography of the tryptic digest of H2 (Figure 5), indicate the presence of O-glycopeptides in addition to the N-glycopeptide characterized previously. The pooled fractions were again fractionated by RP-HPLC. Several fractions in low amounts were obtained that, as indicated by amino acid sequencing, correspond to the C-terminal region of the H2 heavy chain, beginning at Val⁶⁰⁷. It was previously demonstrated that Thr⁶³⁷ is O-glycosylated and MS data suggested that the O-linked carbohydrate chain has a type-I core structure Gal β 1-3-GalNAc-Thr recognized by Jacalin [10]. Indeed, by affinity chromatography on Jacalin-agarose followed by RP-HPLC on a microbore column, we separated a glycopeptide

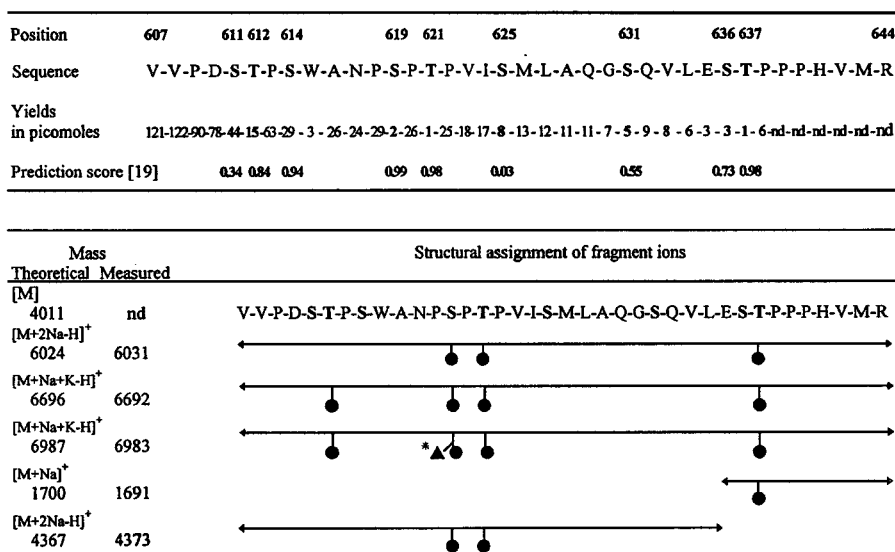


Figure 7 Structural characterization and assignment of the O-glycopeptide Val⁶⁰⁷-Arg⁶⁴⁴ from the H2 heavy chain

Schematic representation showing the location of O-linked glycans on the O-glycopeptide Val⁶⁰⁷-Arg⁶⁴⁴ from H2. The N-terminal sequencing data expressed in picomol of PTH-amino acids correspond to raw data. The O-linked oligosaccharides are indicated with symbols: (●), NeuAc-Gal-GalNAc; (▲), additional NeuAc residue carried by NeuAc-Gal-GalNAc; *, undefined residue position; nd, not determined.

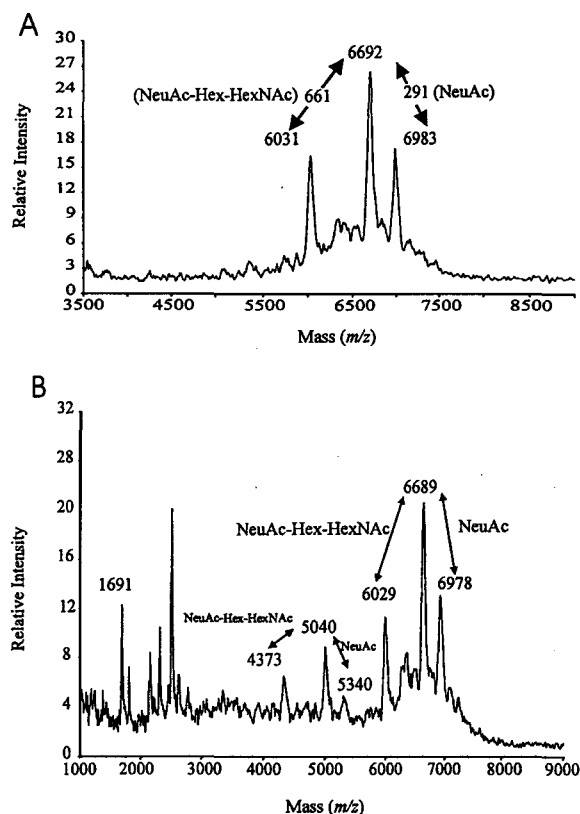


Figure 8 Linear time-of-flight spectrum of the O-glycopeptidic fraction isolated by Jacalin-agarose chromatography of the tryptic digest of H2

(A) Mass spectrum of the native O-glycopeptides; (B) mass spectrum of the same fraction after partial endoproteinase Glu-C digestion.

identified by N-terminal sequencing as the C-terminal part of H2, starting at Val⁶⁰⁷ (Figure 7). The predicted molecular mass of the peptide Val⁶⁰⁷-Arg⁶⁴⁴ is 4011 Da. MALDI-TOF-MS analysis of the fraction eluted from Jacalin revealed three products with masses of 6983, 6692 and 6031 Da respectively (Figure 8A). Subtraction of the theoretical mass of the peptide (4011 Da) from the ion at *m/z* 6031 suggested the presence of three mucin-type O-glycans: *N*-acetylneuraminic acid (NeuAc)-Gal-GalNAc (656 Da). In the same manner, the ion at *m/z* 6692 was identified as the same peptide Val⁶⁰⁷-Arg⁶⁴⁴ carrying four trisaccharides, whereas the ion at *m/z* 6983 was the same peptide carrying three trisaccharides and one di-sialylated tetrasaccharide. Indeed, by mass analysis after acid-catalysed desialylation of the glycopeptide, two components were characterized simultaneously at *m/z* 5152 and 5518 (results not shown). They correspond to the same peptide (4011 Da) carrying three or four O-glycans Gal-GalNAc (365 Da) respectively. This model is in good agreement with amino acid sequence analysis. The first 32 residues were positively identified by Edman degradation except for Ser⁶¹⁹, Thr⁶²¹ and Thr⁶³⁷ thus suggesting that these three residues are fully glycosylated. Moreover, the low yield of PTH-Thr⁶¹² indicates a partial glycosylation of this residue (Figure 7).

The sequence of the glycopeptide was submitted to the Net O-Glyc Server [19] for prediction of O-glycosylation sites. All the

O-glycosylated residues were correctly identified. However, Ser⁶¹⁴ and Ser⁶³⁶ also exhibited a high potential for being glycosylated.

To confirm the structure of the O-glycopeptide Val⁶⁰⁷-Arg⁶⁴⁴ eluted from Jacalin, it was digested with endoproteinase Glu-C and the digest analysed by MALDI-TOF-MS (Figure 8B). Fragment ions at *m/z* 1691 were assigned to be the peptide Ser⁶³⁶-Arg⁶⁴⁴ with Thr⁶³⁷ fully glycosylated by the trisaccharide NeuAc-Gal-GalNAc. The three signals with *m/z* 4373, 5040 and 5340 corresponded to the mass of the peptide Val⁶⁰⁷-Glu⁶³⁵ carrying two or three O-linked carbohydrate chains.

Taken together, our results demonstrate that the H2 heavy chain of IαI contains three or four mucin-type O-glycans clustered in the C-terminal region of the protein and one N-glycan, in a biantennary structure, linked to the Asn⁶⁴ residue, whereas the H1 heavy chain carries two N-glycans possessing mainly the same biantennary structure. The different post-translational modifications concerning the human IαI molecule are depicted in Figure 9.

DISCUSSION

On the basis of protein sequence data, the human IαI heavy chains H1 and H2, which contain 644 and 648 amino acid residues respectively, are recognized as being highly homologous [7]. Indeed, their masses estimated in this work by MALDI-TOF-MS are very similar: 76258 for H1 and 76508 for H2. In the same way, with regard to their glycosylation patterns, two N-glycans were considered to be present on each chain [8].

However, as early as 1979, IαI was classified among the human serum proteins carrying alkali-labile carbohydrate chains able to be recognized by Jacalin [22]. Because carbohydrate moieties may play important physiological roles, we undertook a detailed analysis of the glycans bound to each heavy chain. We confirm that the two potential N-glycosylation sites of H1 are effectively fully occupied by complex-type N-glycans. They are predominantly of biantennary type, as demonstrated by MALDI-TOF-MS analysis of N-glycans released by hydrazinolysis.

However, we have clearly established that the H2 heavy chain carries only one complex-type N-glycan attached to Asn⁶⁴. Indeed, in the tryptic digest of H2, we detected by MALDI-TOF-MS analysis two naked peptides containing the two other potent glycosylation sites, Asn⁴²-Asn-Ser and Asn³⁹¹-Ile-Ser. Likewise, we demonstrated by amino acid sequencing that Asn⁶¹⁷, which belongs to a rarely glycosylated site (Asn-Pro-Ser), is effectively not modified (Figure 7).

The main result of our study is to show that, in addition to the O-glycan carried by Thr⁶³⁷ [10], two or three additional O-linked carbohydrate chains are present on H2. All these O-glycans consist of a type-1 core structure with one or two NeuAc moieties and are clustered in the C-terminal part of the peptide chain. Our results are on the basis of MS analysis and amino acid sequencing.

As suggested, the high frequency of proline residues in the C-terminal region of the different heavy chains (namely H1, H2 and H3), all belonging to the IαI family, may promote their maturation processing and their covalent linkage to the glycosaminoglycan carried by bikunin [23]. However, mucin-type O-linked glycosylation also occurs in regions with high contents of serine, threonine and proline. Thus to estimate the relative propensity of hydroxy amino acids to be glycosylated, we submitted the C-terminal sequences of the three heavy chains to the predictive method described in [19]. Our results show that the O-glycosylated sites on H2 were correctly identified (Figure 7). However, other predicted O-glycosylation sites were identified in

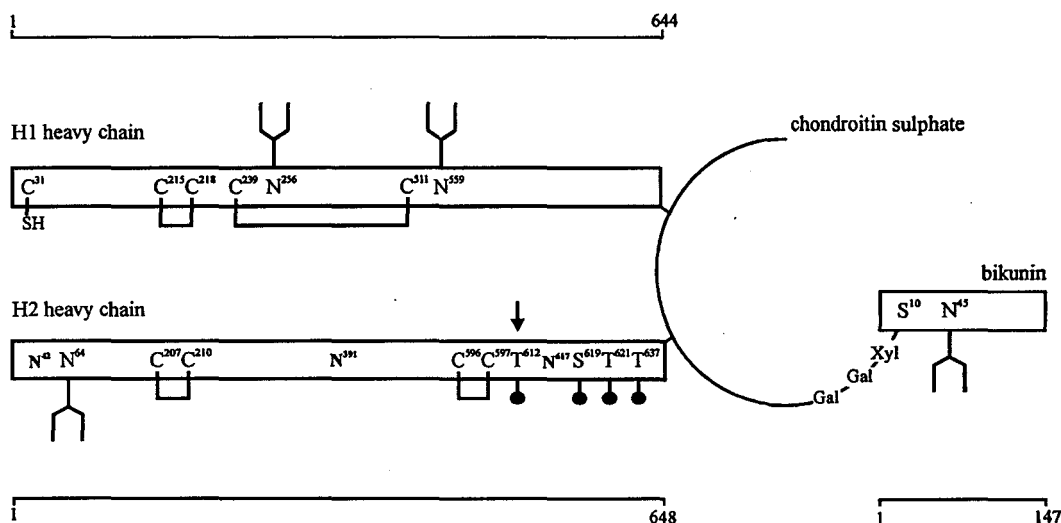


Figure 9 Schematic representation (not to scale) of the human I α I molecule showing post-translational modifications

The I α I molecule consists of three polypeptide chains covalently linked by a chondroitin sulphate chain. H1 and H2 are linked to the chondroitin sulphate chain via ester bonds between their C-terminal Asp residues and an internal *N*-acetylgalactosamine of the glycosaminoglycan chain [20]. The oligosaccharide structures are depicted with following symbols: (Y), N-glycan carrying either zero, one or two NeuAc residue(s) at the non-reducing terminal position. The biantennary N-glycan on bikunin has been described previously [21]. (●), O-linked glycan: NeuAc-Gal-GalNAc. The black arrow indicates the only site partially O-glycosylated. The disulphide bridges of human I α I [36] are depicted with solid lines. The disulphide bridges in bikunin are omitted for simplicity.

each heavy chain, whereas we demonstrated that, as assessed by their GalNAc content, H1 (this study) as well as H3 [24] did not contain O-glycans. Therefore we conclude that the glycosylation pattern of H2 (one N-glycan and a cluster of O-glycans) is clearly different from that of H1 or H3 (two N-glycans). These differences might be conserved between species; indeed, we note that the three fully O-glycosylated residues and the N-glycosylated Asn of human H2 are strictly conserved in the amino acid sequence of mouse H2 [25], as well as Syrian hamster H2 [26]. Furthermore, recent findings indicate that the segment including the cluster of O-glycans on H2 corresponds to the region in which the amino acid sequences of the three heavy chains (H1, H2 and H3) diverged during evolution, before the emergence of mammals [26]. To investigate this further, the structural study of the I α I heavy chains in different animal species might be rewarding.

Concerning the physiological implications of the presence of the O-glycan domain, several functions may be assigned. As originally described [27,28], the two heavy chains H1 and H2 are associated within the I α I structure and we have previously demonstrated by selective immunoadsorption [17] that each I α I molecule contains one H1 and one H2 heavy chain, both linked to bikunin. Obviously, O-glycosylation does not interfere in the formation of the glycosaminoglycan-protein cross-link. However, the mechanism that permits the control in hepatocytes of I α I biosynthesis by preferential assembly of these peptide chains remains unknown. We suggest that O-glycans located in the C-terminal part of H2 in an area exposed on the glycoprotein surface might carry the requisite information for a specific reconnaissance of the partners concerned. Indeed, two types of I α I have been described: a human type designated as I α I_{H1,H2}, B and a bovine type that consists of H2 and H3 linked to bikunin [1]. Thus in each case, a unique highly O-glycosylated chain (H2) is associated with another peptide chain (H1 or H3). Interestingly, it has been reported that both types of I α I (human and bovine) co-exist in Syrian hamster [29]. However, the I α I_{H1,H3}, B I α I type is still unknown.

Our hypothesis is supported by the fact that the chain assembly occurs in the latter part of the biosynthetic pathway, most likely in the *trans*-Golgi network or in secretory vesicles [30,31], whereas O-linked glycosylation is initiated in the *cis*-Golgi compartment, where polypeptide *N*-acetylgalactosaminyltransferases are located.

With regard to HA-binding capacity of the heavy chains, it has been suggested that the C-terminal half of the heavy chains participates in this binding [32]. Because O-linked glycosylation of CD44 decreases its ability to adhere to HA [33], it would be worthwhile to study further the influence of O-linked glycosylation of H2 on its affinity for HA.

At last, the metabolism of I α I might also be influenced by the presence of O-glycans linked on H2. Therefore during inflammatory processes such as septic shock, where the harmful role of leucocyte elastase has been suggested, a large consumption of I α I has been reported [34]. We have previously demonstrated that this proteinase preferentially cleaves the H2 heavy chain in its C-terminal part [35] and we hypothesized that charge-mediated interactions between the glycosaminoglycan chain of I α I and elastase, which is cationic, might explain this particular sensitivity to proteolysis. This one would be obviously enhanced by the O-glycan chains located in the C-terminal part of H2.

Multivalent O-glycosylation is largely implicated in glycoprotein functions. The O-glycosylation of H2 we describe here has to be taken into account in further studies concerning I α I.

This work was supported in part by the Ministère de l'Enseignement Supérieur et de la Recherche (DRED E.A. 1052) and by the Université de Lille II.

REFERENCES

- 1 Salier, J. P., Rouet, P., Raguenez, G. and Daveau, M. (1996) *Biochem. J.* **315**, 1-9
- 2 Gebhard, W. and Hochstrasser, K. (1986) in *Proteinase Inhibitors* (Barrett, A. J. and Salvesen, G., eds.), pp. 389-401. Elsevier Science Publishers BV
- 3 Sandson, J., Hamerman, D. and Schwick, G. (1965) *Trans. Ass. Am. Physicians* **28**, 304-313

- 4 Chen, L., Simon, J. T. and Larsen, W. J. (1992) *J. Biol. Chem.* **267**, 12380–12386
- 5 Blom, A., Pertoff, H. and Fries, E. (1995) *J. Biol. Chem.* **270**, 9698–9701
- 6 Kobayashi, H., Gotoh, J., Hirashima, Y. and Terao, T. (1996) *J. Biol. Chem.* **271**, 11362–11367
- 7 Diarra-Mehrpour, M., Bourguignon, J., Bost, F., Sesbotié, R., Muschio, F., Sarafan, N. and Martin, J. P. (1992) *Biochim. Biophys. Acta* **1132**, 114–118
- 8 Gebhard, W., Schreitmüller, T., Hochstrasser, K. and Wachter, E. (1989) *Eur. J. Biochem.* **181**, 571–576
- 9 Malki, N., Balduyck, M., Maes, P., Capon, C., Mizon, C., Han, K. K., Tartar, A., Fournet, B. and Mizon, J. (1992) *Biol. Chem. Hoppe-Seyler* **373**, 1009–1018
- 10 Enghild, J. J., Salvesen, G., Thøgersen, I. B., Valnickova, Z., Pizzo, S. and Hefta, S. A. (1993) *J. Biol. Chem.* **268**, 8711–8716
- 11 Michalski, C., Piva, F., Balduyck, M., Mizon, C., Burnouf, T., Huart, J. J. and Mizon, J. (1994) *Vox Sang.* **67**, 329–336
- 12 Bigge, J. C., Patel, T., Bruce, J. A., Goulding, P. N., Charles, S. M. and Parekh, R. B. (1995) *Anal. Biochem.* **230**, 229–238
- 13 Laemmli, U.K. (1970) *Nature (London)* **227**, 680–685
- 14 Stephano, J. L., Gould, M. and Rojas-Galicia, L. (1986) *Anal. Biochem.* **152**, 308–313
- 15 Kamerling, J. P., Gerwig, G. J., Vliegenthart, J. F. G. and Clamp, J. R. (1975) *Biochem. J.* **151**, 491–495
- 16 Montreuil, J., Bouquelet, S., Debray, H., Fournet, B., Spik, G. and Strecker, G. (1986) in *Carbohydrate Analysis: a Practical Approach* (Chaplin, M. F. and Kennedy, J. F., eds.), pp. 143–204. IRL Press, Oxford
- 17 Mizon, C., Balduyck, M., Albani, D., Michalski, C., Burnouf, T. and Mizon, J. (1996) *J. Immunol. Methods* **190**, 61–70
- 18 Gohlke, M., Baude, G., Nuck, R., Grunow, D., Kannick, C., Bringmann, P., Donne, P. and Reutter, W. (1996) *J. Biol. Chem.* **271**, 7381–7386
- 19 Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L. and Brunak, S. (1998) *Glycoconjugate J.* **15**, 115–130
- 20 Morelle, W., Capon, C., Balduyck, M., Sautière, P., Kouach, M., Michalski, C., Fournet, B. and Mizon, J. (1994) *Eur. J. Biochem.* **221**, 881–888
- 21 Toyada, H., Ikey, T., Demachi, Y., Toida, T. and Imanari, T. (1992) *Chem. Pharm. Bull.* **40**, 2882–2884
- 22 Chatterjee, B., Vaith, P., Chatterjee, S., Karduck, D. and Uhlenbruck, G. (1979) *Int. J. Biochem.* **10**, 321–327
- 23 Bourguignon, J., Diarra-Mehrpour, M., Thiberville, L., Bost, F., Sesbotié, R. and Martin, J. P. (1993) *Eur. J. Biochem.* **212**, 771–776
- 24 Mizon, C., Héron, A., Capon, C., Sautière, P., Michalski, C., Sesbotié, R. and Mizon, J. (1997) *J. Chromatogr. B* **692**, 281–291
- 25 Chan, P., Risler, J. L., Raguenez, G. and Salier, J. P. (1995) *Biochem. J.* **306**, 505–512
- 26 Nakatani, T., Suzuki, Y., Yamamoto, T. and Sinohara, H. (1997) *J. Biochem.* **122**, 71–82
- 27 Balduyck, M., Laroui, S., Mizon, C. and Mizon, J. (1989) *Biol. Chem. Hoppe-Seyler* **370**, 329–336
- 28 Enghild, J. J., Thøgersen, I. B., Pizzo, S. V. and Salvesen, G. (1989) *J. Biol. Chem.* **264**, 15975–15981
- 29 Yamamoto, T., Yamamoto, K. and Sinohara, H. (1996) *J. Biochem.* **120**, 145–152
- 30 Héron, A., Bourguignon, J., Callé, A., Borghi, H., Sesbotié, R., Diarra-Mehrpour, M. and Martin, J. P. (1994) *Biochem. J.* **302**, 573–580
- 31 Thøgersen, I. B. and Enghild, J. J. (1995) *J. Biol. Chem.* **270**, 18700–18709
- 32 Huang, L., Yoneda, M. and Kimata, K. (1993) *J. Biol. Chem.* **268**, 26725–26730
- 33 Dasgupta, A., Takahashi, K., Cutler, M. and Tanabé, K. (1996) *Biochem. Biophys. Res. Commun.* **227**, 110–117
- 34 Witte, J., Jochum, M., Scherrer, R., Schramm, W., Hochstrasser, K. and Fritz, H. (1982) *Int. Care Med.* **8**, 215–222
- 35 Balduyck, M., Piva, F., Mizon, C., Maes, P., Malki, N., Gressier, B., Michalski, C. and Mizon, J. (1993) *Biol. Chem. Hoppe-Seyler* **374**, 895–901
- 36 Flahaut, C., Mizon, C., Aumercier-Maes, P., Colson, P., Bailly, C., Sautière, P. and Mizon, J. (1998) *Eur. J. Biochem.*, in the press

Received 5 February 1998/24 April 1998; accepted 14 May 1998

2) Discussion et résultats complémentaires

Nous avons, pour la première fois, estimé par spectrométrie de masse de type MALDI les masses moléculaires des chaînes lourdes H1 et H2 (76258 et 76508 Da, respectivement). Ces masses sont en accord avec la valeur de 74100 Da caractérisée pour la chaîne lourde H3 [16]. Dès lors, la masse moléculaire allouée aux glycanes peut être évaluée par la différence existant entre les masses moléculaires théoriques et expérimentales des chaînes lourdes. D'autre part, sous l'action de la PNGase F, nous montrons la présence de 2 et de 1 N-glycanes sur les chaînes H1 et H2 respectivement. Par conséquent, les 2 sites potentiels de N-glycosylation de la chaîne H1 sont tous les 2 occupés. Par contre, parmi les 3 sites potentiels de H2, un seul site semble effectivement occupé.

Les compositions en monosaccharides, réalisés par chromatographie en phase gazeuse, ont également révélé la présence de N-glycanes sur les 2 chaînes lourdes. Par contre, comme précédemment rapporté [122], seule la chaîne H2 semblait posséder des O-glycanes si on considère les taux de GalNAc mesurés. Utilisant une stratégie légèrement différente, Olsen et coll [136] ont indépendamment caractérisé la structure des glycanes présents sur les chaînes lourdes de l'IcI. Ainsi, pour déterminer la composition en monosaccharides, ils ont utilisé la méthode FACE (fluorophore assisted carbohydrate electrophoresis) et démontrent la présence de O-glycanes également sur la chaîne lourde H1. Le rapport plus élevé existant entre les taux de N- et de O-glycanes présents sur cette chaîne lourde peut expliquer la sous estimation de la GalNAc dosée par chromatographie en phase gazeuse.

Ultérieurement, nous avons confirmé le résultat rapporté par Olsen et coll [136], en isolant par RP-HPLC, à partir des fractions générées par clivage chimique de H1 au BrCN, le peptide C-terminal (Leu⁶⁰⁹-Asp⁶⁴⁴). Les masses moléculaires expérimentales de ce peptide (4276,4 et 4565,8 Da), analysées par spectrométrie de masse de type MALDI, diffèrent par rapport à sa masse théorique (3908,3 Da) de 368,1 et 657,5 Da, respectivement. Ces différences de masse correspondent aux masses d'un O-glycane de type Gal-GalNAc sialylé (656 Da) ou non sialylé (365 Da). Par dégradation d'Edman, les 18 premiers acides aminés de ce peptide étaient

positivement identifiés excepté le résidu de thréonine en position 624. Par conséquent, la chaîne lourde H1 possède effectivement un O-glycane di- ou trisaccharidique localisé sur la Thr⁶²⁴.

La chaîne lourde H2, quant à elle, possède une région fortement O-glycosylée composée de 3 ou 4 O-glycannes principalement monosialylés à proximité de son extrémité C-terminale. Cette région anionique pourrait donc augmenter les interactions de charge pouvant exister entre la chaîne de CS et l'élastase (cationique) expliquant ainsi la sensibilité plus marquée de la chaîne H2 à l'action de cette protéinase. Les chaînes lourdes H1 et H2 ne sont donc pas homologues en terme de O-glycosylation. Ainsi, compte tenu de la sélectivité d'assemblage observée au sein de l'I α I qui comporte toujours une chaîne H1 et une chaîne H2, nous suggérons que les O-glycannes, localisés dans une région non structurée et très proche de l'acide aspartique C-terminal, pourraient porter l'information nécessaire à la reconnaissance spécifique des différentes chaînes lourdes au cours de la biosynthèse de l'I α I. Cette hypothèse pourrait être testée par mutagenèse dirigée.

En général, un certain degré d'hétérogénéité de la structure des N-glycannes portés par une glycoprotéine est observé. D'une manière surprenante, Olsen et coll [136], travaillant à partir de fractions glycopeptidiques isolées par l'intermédiaire d'une colonne de Concanavaleine A immobilisée mais ne tenant pas compte de la spécificité d'action de cette lectine, s'étonnent de la nature exclusivement biantennée des N-glycannes caractérisés. Si effectivement, aussi bien sur la bikunine [119] que sur les chaînes lourdes, les N-glycannes sont principalement de type biantennés di-, mono- ou non sialylés, la chaîne H1 renferme néanmoins un faible pourcentage de structures triantennées. De leur côté, Olsen et coll rapportent également un taux de fucosylation de l'ordre de 10 %.

Finalement, les chaînes lourdes H1 et H2 diffèrent également en terme de N-glycosylation. La chaîne H2 ne contient qu'un seul N-glycane situé au niveau de son extrémité N-terminale (Asn⁶⁴). Au contraire, la chaîne H1 possède 2 N-glycannes situés d'une part dans sa région globulaire et d'autre part au début de son extrémité C-terminale non structurée. Dès lors, il est intéressant de remarquer

(Figure 13, p 68) que la chaîne H3 qui porte 2 N-glycannes [16] présente la moitié des caractéristiques de chaque chaîne. En effet, le premier site potentiel de N-glycosylation se situe au niveau de l'extrémité N-terminale, quant au second, il se situe dans la région C-terminale.

D'un point de vue fonctionnel, comme de nombreuses glycoprotéines plasmatiques, la sialylation des glycannes protège sans doute l'I α I d'une élimination hépatique rapide. L'incubation d'hépatocytes primaires en présence de tunicamycine, un inhibiteur spécifique de la N-glycosylation, pourrait permettre d'étudier l'influence de la N-glycosylation sur l'établissement de la conformation comme sur les mécanismes de maturation, d'assemblage des chaînes lourdes ou de sécrétion de l'I α I.

III Schéma de glycosylation des chaînes lourdes de l'I α I porcine

La détermination des séquences primaires des précurseurs des chaînes lourdes chez d'autres espèces que chez l'homme a permis de mettre en évidence des taux d'homologie de séquence en acides aminés très élevés [27, 127]. Ainsi, l'utilisation d'un programme d'alignement de séquences [141] indique que les chaînes lourdes H1 et H2 de l'homme et du porc présentent 81 % d'homologie. D'autre part, comme pour les précurseurs des autres espèces, le nombre et la position des résidus de cystéine sont strictement conservés (Figure 13, p 68) entre les chaînes lourdes de l'I α I humain et porcine. Ainsi, quelle que soit l'espèce, la topologie des ponts disulfure est très probablement identique. Pour preuve, la mobilité électrophorétique en gel SDS de la chaîne H1 porcine est, comme celle de la chaîne humaine, plus importante après réduction [24]. Par conséquent, la conformation spatiale des chaînes lourdes porcines, comme celle des autres espèces, doit très probablement être similaire à celle existant chez l'homme.

La glycosylation des protéines est une donnée structurale en étroite relation avec la notion de conformation spatiale des protéines. Dès lors, la détermination des structures glycaniques et surtout de leur distribution le long de la chaîne polypeptidique peut apporter une preuve supplémentaire de la conservation de la conformation de chaque chaîne lourde entre les espèces. Bien évidemment, cette hypothèse suppose que les chaînes lourdes des différentes espèces présentent d'abord une conservation de leurs sites potentiels de N- et de O-glycosylation.

Les sites potentiels de N-glycosylation (Asn-Xaa-Ser/Thr, avec Xaa différent de la proline) sont, pour une chaîne lourde donnée, très conservés au point de vue de leur localisation ou de leur séquence [127]. Par exemple, les chaînes H2 matures de l'homme et du porc renferment toutes deux 3 sites potentiels de N-glycosylation strictement identiques en terme de localisation et très similaires, voire identiques en terme de séquence. Par contre, sur les 2 sites de la chaîne lourde H1 humaine, un seul site est conservé (localisation et séquence identique) dans la chaîne porcine.

L'analyse des nombreuses séquences O-glycosylées déjà caractérisées [142, 143] montre qu' aucune séquence consensus ne permet de prédire la présence de O-

glycannes. Cependant, il est clairement établi que la O-glycosylation intervient majoritairement dans des régions peptidiques présentant un pourcentage élevé de résidus de sérine, thréonine et proline, ce qui est le cas des régions C-terminales des chaînes lourdes humaines [126] où nous avons démontré la présence de O-glycannes. Plusieurs auteurs [143, 144] ont développé des algorithmes de prédiction des sites de O-glycosylation. Ainsi, si on soumet les séquences primaires C-terminales des chaînes lourdes des différentes espèces à ce type de prédiction [144], plusieurs résidus de sérine et de thréonine présentent une probabilité de O-glycosylation non négligeable. Par conséquent, comme chez l'homme, ces régions pourraient être O-glycosylées.

Dès lors, profitant des techniques de purification de l'I α I porcin récemment développées au laboratoire [24], nous avons étudié le schéma de glycosylation de ces chaînes lourdes en profitant de l'expérience acquise par l'étude des structures glycanniques caractérisées sur l'I α I humain. Il nous est enfin apparu intéressant de comparer les séquences primaires porteuses de glycannes des chaînes lourdes de l'I α I de l'homme et du porc à celles des autres espèces.

1) Résultats

Les résultats obtenus feront l'objet d'une publication dont l'ébauche est présentée dans les pages suivantes sous le titre : « Pig inter- α -inhibitor heavy chains are N- and O-glycosylated as their human counterparts ».

Pig inter- α -inhibitor heavy chains are N- and O-glycosylated as their human counterparts.

Christophe FLAHAUT^{1,2}, Calliope CAPON^{1,2}, Hervé DROBECQ³, Jean-Claude MICHALSKI² and Jacques MIZON^{1*}

From ¹ Laboratoire de Biochimie, UPRES E.A. 2681, Faculté des Sciences Pharmaceutiques et Biologiques, Université de Lille II, Avenue du Professeur Laguesse, B.P. 83, 59006 Lille, France, ² Laboratoire de Chimie Biologique, UMR 8576 CNRS, Université des Sciences et Technologies de Lille I, 59655 Villeneuve d'Ascq, France and ³ Département V, URA 1309 CNRS, Institut de Biologie et Institut Pasteur de Lille, 1 rue de Professeur Calmette, B P 447, 59021 Lille, France.

* To whom correspondence should be addressed :

Laboratoire de Biochimie, Faculté des Sciences Pharmaceutiques et Biologiques, Université de Lille II, Avenue du Professeur Laguesse, B.P. 83, 59006 Lille, France. Tel.:333-20-96-40-45 ; Fax: 333-20-95-90-09 ; e-mail : jmizon@phare.univ-lille2.fr

Running title : glycosylation pattern of porcine I α I heavy chains

Abstract

Inter- α -inhibitor ($I\alpha I$) is a plasma serine-proteinase inhibitor. It consists of three polypeptide chains covalently linked by a glycosaminoglycan chain : a light chain named bikunin and two heavy chains, H1 and H2. Recent findings indicate that the human heavy chains are N- and O-glycosylated. In this study, using a combined mass spectrometry and Edman degradation approach, we have characterized the glycosylation pattern of H1 and H2 heavy chains from pig $I\alpha I$. Analysis reveals two N-linked oligosaccharides located on Asn²⁴⁹ of pig H1 heavy chains and on Asn⁶⁴ of pig H2 heavy chains. In addition, the pig $I\alpha I$ heavy chains carry several O-linked glycans located (i) on a partially glycosylated site Ser⁶¹¹ and on a fully glycosylated site Thr⁶¹⁹ of pig H1 and (ii) on two fully glycosylated sites (Thr⁶²¹ and Thr⁶³⁷) of pig H2. All these O-glycans consist of a T antigen monosialylated on pig H1 and mono- and disialylated on pig H2.

INTRODUCTION

As recently reported (Flahaut *et al.*, 1998a; Olsen *et al.*, 1998), the heavy chains H1 and H2 (hH1 and hH2) of human I α I are N- and O-glycosylated. The heavy chain hH1 carries two complex-type N-glycans of predominantly biantennary structure and one O-glycan, located in its C-terminal extremity. In contrast, the oligosaccharides attached to the hH2 are (i) a biantennary N-glycan in the N-terminal region and (ii) three to four O-glycans, clustered in the C-terminal region.

The amino acid sequences of the two heavy chains H1 and H2 from pig (pH1 and pH2) are known. In the peptide sequence, all consensus sequences Asn-Xaa-Ser/Thr (Kornfeld *et al.*, 1985) are potential N-glycosylation sites, except if such amino acid residues (e.g. Pro) are present in the position Xaa. According to the Kornfeld's rules, pH1 contains two potential N-glycosylation sites Asn⁴¹-Asn-Thr and Asn²⁴⁹-Leu-Thr against three putative sites Asn⁴²-His-Ser, Asn⁶⁴-Phe-Ser and Asn³⁹¹-Val-Ser for pH2. These putative N-glycosylation sites are more or less conserved between species. For example, the three potential sites of heavy chains H2 from human and pig are strictly conserved in term of location, whereas the heavy chains H1 from these same species only possess one common potential site. In addition, a prediction of Ser or Thr residues able to bind glycans is possible with different prediction methods (Hansen *et al.*, 1998; Elhammer *et al.*, 1993). The NetOGlyc 2.0 prediction results indicate the presence of cluster of putative O-glycosylation sites in the C-terminal part of both pig heavy chains. Indeed, this computer program predicted that pH1 chain residues Ser⁶¹⁴, Thr⁶¹⁶, Ser⁶¹⁹, Ser⁶²⁰, Ser⁶²¹ and pH2 chain residues Thr⁶²¹, Ser⁶³², Thr⁶³⁷ were likely to be O-glycosylated. Therefore, as their human counterparts, pH1 and pH2 are probably N- and O-glycosylated. Thus, the glycosylation patterns of pH1 and pH2 were successively studied.

MATERIALS AND METHODS

Materials

Porcine I α I (pI α I) was purified from blood of healthy pigs. pI α I was dissociated by hydroxylamine treatment and its constitutive heavy chains were isolated according to Carrette *et al.*, 1997. Human I α I (hI α I) was isolated according to Michalski *et al.*, 1994 and supplied by the Laboratoire Français du Fractionnement et des Biotechnologies (Lille, France). Human H1 and H2 heavy chains were prepared according to Flahaut *et al.*, 1998a and H3 heavy chain from human pre- α -inhibitor (P α I) was isolated as described by Mizon *et al.*, 1997. Trypsin from bovine pancreas and DIG glycan differentiation kit were from Boehringer Mannheim (Meylan, France). Peptide-N-glycosidase F (PNGase F) from *Flavobacterium meningosepticum* was from Bio Labs (Beverly, MA, USA). 1,2-diamino-4,5-methylene dioxybenzene (DMB) were from Molecular Probes, Inc (Eugene, OR, USA). Concanavalin-A Sepharose and Jacalin-agarose was from Amersham Pharmacia Biotech (Saclay, France). All other chemicals were of reagent grade and were purchased from commercial sources.

Carbohydrate analysis of porcine H1 and H2 heavy chains

Carbohydrate composition Monosaccharide composition was performed using gas chromatography with a silicone OV 101 capillary column (0.32 mm x 25 m). Samples were analyzed after methanolysis (0.5 M HCl in methanol for 24 h at 80°C), followed by N-reacetylation and trimethylsilylation, as described by Kamerling *et al.*, 1975 with slight modifications according to Montreuil *et al.*, 1986.

Qualitative sialic acid analysis To release sialic acids, desalted samples of human and porcine I α I and human H1 and H2 heavy chains were dissolved in 200 μ l of formic acid solution pH 2.0 and heated twice for 1 hour at 80°C. The sialic acids were purified by ultrafiltration onto MICROSEP centrifugal concentrator (Pall Filtron Corporation, Northborough, MA, USA ; cutoff value, 3kDa) and then freeze-dried. The derivatization with DMB was carried out by heating at 50°C for 2.5 h in the dark in 7 mM DMB, 0.75 M β -mercaptoethanol and 18 mM sodium hydrosulfite in 1.4 M acetic acid (Hara *et al.*, 1989). DMB-derivatized sialic acids were separated by RP-HPLC using a C18 Zorbax column (250 mm x 4.6 mm, particule size 5 μ m) equipped with a guard column (50 mm x 4.6 mm) of the same phase (Interchim, Montluçon,

France). Elution was achieved using a linear gradient from 4/6/90 to 11/7/82 of acetonitrile/methanol/water over 60 min at a flow rate of 0.7 ml/min. On-line fluorescence detection was performed with a FL 2000 fluorescence detector (ThermoQuest, Paris) operating at an excitation wavelength of 373 nm and an emission wavelength of 448 nm. Sialic acids from bovine submaxillary mucin were used as standard (Klein *et al.*, 1997).

Electrophoresis and blotting with digoxigenin-labelled lectins An aliquot (0.3 µg) of each heavy chain and control glycoproteins (1 µg) was resolved on 10 % SDS/PAGE under reducing conditions. After migration, proteins were electrophoretically transferred onto nitrocellulose membrane (Schleicher and Schüll, Dassel, Germany). Blots were saturated in 3 % bovine serum albumin (BSA) in TBS-Tween [15 mM Tris, 140 mM NaCl, 0.05% (v/v) Tween] for 45 min and washed three times for 15 min with TBS-Tween. The four blots were incubated with each digoxigenin-labelled lectins diluted in TBS-Tween: *Sambucus nigra* agglutinin (SNA-dig, 1 µg/ml), *Maackia amurensis* agglutinin (MAA-dig, 5 µg/ml), *Concanavalia ensiformis* agglutinin (Con A-dig, 10 µg/ml) and *Arachis hypogaea* agglutinin (PNA-dig, 10 µg/ml). Prior the PNA-revelation, the blot was chemically desialylated by incubation in a formic acid solution pH 2.0 during 1 h at 80°C. After three washing with TBS-Tween for 15 min, blots were incubated for 1 h in TBS-Tween in presence of anti-digoxigenin alkaline phosphatase-labelled Fab fragments (1 µg/ml). Labelled glycoproteins were revealed by 4-nitro blue tetrazolium chloride 5-bromo-4-chloro-3-indolyl phosphate staining.

PNGase F digestion Glycoproteins were denaturated at 100 °C for 10 min in 100 µl of a 50 mM sodium phosphate buffer pH 7.5 containing 0.5 % SDS and 1 % β-mercaptoethanol. Nonidet-P40 was added to a final concentration of 1 % before addition of 7 milliunits (IUB) of PNGase F. Digestion was carried out up to 72 h at 37°C and was checked by SDS/PAGE 10% under reducing conditions.

Glycopeptide purification

Reduction, alkylation and tryptic hydrolysis of the heavy chains 4 mg of pH1 and 2 mg of pH2 heavy chains, at a final concentration of 1 mg/ml in a 100 mM Tris/HCl buffer pH 7.25, were denaturated, reduced and carboxamidomethylated as previously described (Flahaut *et al.*, 1998a). After extensive dialysis of the reaction mixtures against a 100mM Tris/HCl buffer, pH 8.0, trypsin was added with a final

enzyme-to-substrate ratio of 1/50 (w/w) and incubated at 37°C for 24 hours. The hydrolysis was stopped by addition of 4-(2-aminoethyl)-benzenesulfonylfluoride (AEBSF) at a final concentration of 1mM.

Separation of glycopeptides by affinity chromatography on immobilized lectins

The peptide digests of porcine H1 and H2 were loaded onto a Concanavalin-A-Sepharose column (5 ml) equilibrated in 50 mM Tris/HCl/100 mM NaCl buffer, pH 7.4, containing 1 mM CaCl₂, 1 mM MnCl₂, 1 mM MgCl₂ and 0.02 % (w/v) NaN₃ (buffer A). After washing the column with 10 ml of buffer A, the elution was performed by 10 ml of 0.5 M α -methylglucoside in buffer A at a flow rate of 20 ml/h, with detection by absorbance at 280 nm. The unbound fractions were pooled and loaded at the same flow rate onto a Jacalin-agarose column (1 ml) equilibrated in buffer A. After washing the column (6 ml buffer A), O-glycopeptides were eluted by 0.5 M α -methylgalactoside in buffer A (6 ml). The different fractions were concentrated and desalted on a Sep-Pak C18 cartridge (Waters U.K. Ltd., Watford, Herts., U.K.) according to the manufacturer's instructions. They were then freeze-dried and analyzed by MALDI-TOF-MS before and after chemical desialylation. The chemical desialylation was carried out by heating peptides at 80 °C in a closed reaction vial in the presence of a formic solution at pH 2.0 for 1 hour.

Purification process of peptides and N-glycans from N-glycopeptides

N-glycopeptides isolated by Con-A Sepharose column were, after chemical desialylation, digested for 24 h by PNGase F as described above without addition of detergents or reductor agent. Two mini columns were packed, the first with 10 mg of C18 phase (Sep-Pak, Waters U.K. Ltd.), the other one with 10 mg of porous graphite carbon (PGC, hypercarb[®], Shandon Scientific, Cheshire, UK), in order to purify, in one step, peptides and N-glycans, respectively. PNGase F-digested N-glycopeptides were first loaded onto C18 column and the C18-unretained materials was then loaded onto the PGC mini column. Salts were removed by washing mini columns with 0.5 ml of water acidified by 0.1 % TFA. The C18-retained peptides were eluted with 0.5 ml of acetonitrile/water (80/20, v/v) containing 0.1 % TFA and analyzed by MALDI-TOF-MS and Edman degradation. N-glycans were recovered from PGC mini column by elution with 0.5 ml of acetonitrile/water (50/50, v/v) containing 0.1 % TFA (Packer *et al.*, 1998) and analyzed by MALDI-TOF-MS.

RP-HPLC separation of Jacalin-isolated glycopeptides The jacalin-isolated glycopeptides of each heavy chain were applied onto the above-described Zorbax column equilibrated in 0.1 % trifluoroacetic acid (TFA) in water. For elution, isocratic conditions were applied for 5 min, followed by a linear gradient of acetonitrile from 0 to 80 % in 0.1 % TFA over 80 min at a flow rate of 1 ml/min. Eluates were monitored by UV absorbance at 215 nm. The O-glycopeptides were further submitted to Edman degradation in order to identify the O-glycosylation sites.

Analytical methods

SDS/Polyacrylamide Gel Electrophoresis SDS/PAGE was performed using 7.5 % and 10 % polyacrylamide slab gels (Laemmli., 1970) under reducing or non-reducing conditions with Coomassie Blue staining (Stephano et al., 1986).

Sequence analysis Automated Edman degradation of peptides or glycopeptides was performed on a gas-phase sequencer (Procise 492, Perkin-Elmer, Courtaboeuf, France) using the pulsed-liquid program. Phenylthiohydantoin derivatives of amino acids were identified on-line on a 120 A amino acid analyzer (Perkin-Elmer).

Matrix-assisted Laser Desorption Ionization – time of flight – Mass Spectrometry MALDI-TOF-MS analysis was performed on a 'Vision 2000' time-of-flight instrument (Finnigan MAT, Bremen, Germany) equipped with a 337 nm ultraviolet laser. The mass spectra were acquired in linear or reflectron mode with positive or negative detection. Aliquots of glycopeptides, peptides and N-glycans were mixed with an equal volume of the matrix solution, 2,5-dihydroxybenzoic acid [10mg/ml dissolved in acetonitrile/water (30/70, v/v)]. External calibration was performed using angiotensin I (1296.7 Da, Sigma, France) and bovine insulin (5733.6 Da, Sigma, France).

RESULTS

Porcine I α I was isolated from plasma by a double-step procedure and further dissociated by hydroxylamine treatment. The heavy chains thus obtained were purified by anion-exchange chromatography (Carrette *et al.*, 1997) and their purity was assessed by 10 % SDS-PAGE with or without reducing conditions and Coomassie Brilliant blue staining.

Carbohydrate analysis from native heavy chains

In order to confirm the presence of both N- and O-glycans, the different peptide chains were stained with a panel of four digoxigenin-labelled lectins (SNA-dig, MAA-dig, Con A-dig and PNA-dig). As shown in Figure 1, the use of Con A-dig lectin (panel A), which mainly reacts against α -D-glucosyl and α -D-mannosyl (Kornfeld *et al.*, 1975), confirms the occurrence of N-linked glycans in human (hH1, hH2 and hH3) and pig heavy chains (pH1 and pH2). In addition, it has been previously demonstrated (Olsen *et al.*, 1998) that human H1 and H2 contain in their C-terminal part sialylated O-glycans. Thus, to establish if pH1, pH2 or hH3 carry the same glycan structures, a blot was, after chemical desialylation, revealed by PNA-dig lectin (Lotan *et al.*, 1975), which recognizes the T antigen (Gal β 1-3GalNAc α 1-0-Ser/Thr). As shown in panel B, the desialylated pH1 and pH2 are positively stained indicating the presence of the Gal β 1-3GalNAc α 1-0-Ser/Thr structure. By opposition, the desialylated hH3 was not recognized by PNA-dig lectin suggesting that this desialylated heavy chain doesn't possess the T antigen. This result is in good agreement with the previously reported studies (Mizon *et al.*, 1997) in which the absence of GalNAc residues were demonstrated. Moreover, the heavy chains of pig as well as human were strongly recognized by SNA-dig lectin (panel C) whereas MAA-dig lectin (panel D) weakly revealed them. The specificity of these two lectins differs with the NeuAc linkage. SNA lectin recognizes the NeuAc α 2-6Gal/GalNAc structure (Shibuya *et al.*, 1987) whereas MAA lectin recognizes the NeuAc α 2-3Gal structure (Wang *et al.*, 1988). Compared to the control glycoproteins, these experiments indicated that sialic acid residues are mainly attached by an α 2-6 linkage.

In order to study sialic acid diversity, human and porcine I α I were submitted to formic acid hydrolysis. After purification, DMB derivatization and RP-HPLC chromatography (Figure 2), the DMB-derivatized sialic acids were identified by

comparison of their retention time to those of reference sialic acids from BSM (panel A). These results indicate that human α I molecule (panel B) as well as porcine α I (panel C) mainly possess N-acetylneuraminic acid (Neu5Ac). Only 3% of N-glycolylneuraminic acid (Neu5Gc) was detected on the porcine α I molecule.

PNGase F digestions of heavy chains and subsequent SDS/PAGE analysis after 24h, 48h and 72h incubation times (data not shown) indicated that pH1 and pH2 carry only one N-linked glycan. Indeed, in both cases, only one new band, corresponding to the relative molecular masses of heavy chains approximatively shifted of 2000 Da, appeared whatever the incubation times.

Molar carbohydrate compositions (Table I), performed by gas chromatography, show that the molar ratios of N-acetylglucosamine (GlcNAc) and mannose (Man) are consistent with the presence of N-glycans, mainly of biantennary type. A fucose residue is detected, only in the case of pH1, with a fucose-to-N-glycan ratio of 1/1. The presence of N-acetylgalactosamine (GalNAc) is usually related to the presence of O-linked glycans. The mannose and N-acetylgalactosamine ratio is in favour of one N-glycan to one O-glycan for pH1 against one N-glycan to two O-glycans for pH2.

Glycosylation pattern of porcine heavy chains

To identify what N- and O-glycosylation sites were effectively occupied and to obtain confirmation by molecular masses of attached-glycan structures, porcine H1 and H2 heavy chains were submitted to a reduction, carboxamidomethylation and tryptic hydrolysis. The heavy chains H1 and H2 of the pig α I are revealed by Con A-dig and PNA-dig lectins. Therefore, the pig heavy chain peptides containing N- and O-linked oligosaccharides were purified by Concanavalin-A and Jacalin affinity chromatography, respectively.

Structural characterization of N-glycosylated peptides

After desalting by Sep-Pak cartridge, the Con-A isolated N-glycopeptides from pH1 and pH2 were analyzed by MALDI-TOF-MS. The sialic acid number was assessed by the spectra comparison of the native and chemically desialylated N-glycopeptides. Then, in order to rapidly identify by their molecular masses the peptides and the desialylated N-glycans from the desialylated compounds, these one

were submitted to PNGase F digestion followed by one purification step. This latter allows to segregate the peptide and glycan pools onto C18 phase and porous graphitized carbon mini columns, respectively. These purified and desalted fractions were then analyzed by MALDI-TOF-MS and the peptide fractions were also studied by Edman degradation.

The porcine H2 heavy chain contains one N-glycan for three potential N-glycosylation sites : Asn⁴²-His-Ser, Asn⁶⁴-Phe-Ser and Asn³⁹¹-Val-Ser. MALDI-TOF-MS analysis of the Con A-isolated N-glycopeptides (Figure 3A) reveals two experimental molecular masses : 3781.8 and 3494.5 Da. After chemical desialylation, the mass differences (580.9 Da and 293.6 Da) between the measured masses of the native and desialylated N-glycopeptides (panel B) indicate that the N-glycan structures are terminated by one or two N-acetylneuraminic acid (291.3 and 582.6 Da). After PNGase F treatment of the desialylated N-glycopeptides, mass spectrometry analysis of the peptide fraction (panel C) shows one ion signal at m/z 1576.6. This experimental mass corresponds to the calculated mass of the peptide Gly⁵⁹-Lys⁷³ (1576.7 Da), which contains the N-glycosylation site Asn⁶⁴-Phe-Ser. The Edman degradation confirm the peptide sequence and identify an aspartic acid residue instead of an asparagine acid residue (Table II). This one results from the PNGase F action, which transforms the Asn accommodating the N-glycan in an Asp residue (Plummer *et al.*, 1984). In addition, with respect to the purification process, the measured mass ($[M+Na]^+$) for the desialylated carbohydrate structure (1663.5 Da) correlates with the calculated mass of the (HexNAc)₅-(Hex)₄ structure. Obviously, the sum ($[M+H]^+ = 3197.9$ Da) of peptide and desialylated N-glycan masses is equivalent to the experimental mass (3200.9 Da) obtained in case of the desialylated fraction (panel B). Therefore, among the three potential N-glycosylation sites, only the Asn⁶⁴ of pH2 carries a N-glycan.

The porcine H1 heavy chain contains one N-glycan for two potential N-glycosylation sites : Asn⁴¹-Asn-Thr and Asn²⁴⁹-Phe-Thr. After Con A purification, the mass spectra of the native and desialylated N-glycopeptides of pH1 were complex and reflected a probable peptide heterogeneity. Therefore, to rapidly identify with accuracy the peptide and desialylated N-glycans from the desialylated N-glycopeptides, this fraction was treated by PNGase F. After purification, MALDI-TOF-MS analysis of peptides from PNGase F-treated N-glycopeptides of pH1 (Figure 4D) shows five peptide sequences with respective molecular masses of 2857.4,

2599.5, 1700.8, 1422.7 and 1067.4 Da. As summarized in Table II, all these experimental masses correspond to calculated masses of peptides containing the same glycosylation site Asn²⁴⁹-Leu-Ser. The Edman degradation (Table II) performed on these PNGase F-treated peptides confirms the peptide identification and allows to characterize, in one peptide, the asparagine residue transformed in an aspartic acid residue by the PNGase F.

The mass spectrum of desialylated N-glycans from PNGase F-treated N-glycopeptides of pH1 is presented in Figure 4C. The ion signal detected at m/z 1810.8 ($[M+Na]^+$) correlates with the calculated mass (1809.5 Da) of a (HexNAc)₄-(Hex)₅ structure possessing an additional fucose residue (structure A). The other signals at m/z 1664.7 and 1461.7 correspond to the same structure A without fucose residue (structure B) and without fucose and N-acetylglucosamine residues (structure C), respectively. Therefore, the peptide masses incremented by the desialylated N-glycan masses must correspond to the masses measured for the desialylated N-glycopeptides. However, the MALDI-TOF-MS analysis of chemically desialylated N-glycopeptides (Figure 4B) shows mainly four major signals at m/z 2858.9, 3492.0, 4391.6 and 4649.5 ($[M+Na]^+$), which match with the calculated masses ($[M+Na]^+$) of the main detected peptides bearing a biantennary N-glycan always fucosylated. Therefore, the desialylated N-glycan structures B and C must likely be due to the PNGase F digestion. Indeed, this enzyme is not a recombinant PNGase F. Thus, weak fucosidase and endoglycosidase H activities could be responsible of the N-glycan degradation. Treatment of the desialylated N-glycopeptides with a recombinant PNGase F and/or a fucosidase should confirm this hypothesis. Finally, the mass differences between the desialylated and untreated glycopeptides (Figure 4B and 4A) indicate that the carbohydrate structure A is terminated by one or two NeuAc residues, demonstrating that, in regard to the purification process, the oligosaccharide linked to pH1 Asn²⁴⁹ is a di-sialylated biantennary structure mainly fucosylated.

Structural characterization of O-glycosylated peptides

Jacalin-purified glycopeptides were also analyzed before and after chemical desialylation by MALDI-TOF-MS. Sequencing of O-glycopeptides from pH1 indicates the presence of a peptide with N-terminal end located in position 606 (Table III). The MALDI-TOF-MS analysis of the desialylated O-glycopeptides reveals a major molecular mass at m/z 2365.2 and a minor one at m/z 2731.7, which correspond

to masses of the naked peptide Phe⁶⁰⁶-Lys⁶²⁴ (1999.2 Da) with one or two disaccharides Gal β 1-3GalNAc α 1, respectively. The experimental molecular masses (3309.8 and 2659.7 Da) characterized from untreated glycopeptides are in agreement with the calculated masses (3312.4 and 2655.8 Da) of this same peptide added with the mass of one or two trisaccharides NeuAc-(Gal β 1-3GalNAc α 1), respectively. Further RP-HPLC purification followed by Edman degradation demonstrated that the fully O-glycosylated site is the Thr⁶¹⁶ residue whereas the partially O-glycosylated site correspond to the Ser⁶¹¹ residue.

Results of MALDI-TOF-MS analysis and Edman degradation (Table III) were more complex for pH2 than pH1. Indeed, the Edman degradation (5 cycles) of desialylated O-glycopeptides indicates the presence of two N-terminal sequences S⁶¹¹LPSW⁶¹⁵ and V⁶²⁴PLPA⁶²⁸. These sequences were recovered in the pH2 C-terminal. MALDI-TOF-MS analysis of desialylated O-glycopeptides showed three molecular masses of 4342.5, 2531.0 and 1830.2 Da. By comparison with the amino acid sequence, the mass of 4342.5 Da might be attributed to the calculated mass (4341.9 Da) of the peptide Ser⁶¹¹-Arg⁶⁴⁴ with two disaccharides Gal β 1-3GalNAc α 1. The experimental masses of 2531.0 and 1830.2 Da were in a good agreement with the theoretical masses (2530.9 and 1830.0 Da) of peptides Ser⁶¹¹-Arg⁶²³ and Val⁶²⁴-Arg⁶⁴⁴, each carrying a disaccharide Gal β 1-3GalNAc α 1. Therefore, the peptide Ser⁶¹¹-Arg⁶²³, where Ser⁶¹¹ and Ser⁶¹⁴ residues were positively identified by Edman degradation, carry its O-glycan on the alone remaining O-glycosylation site, which is the Thr⁶²¹ residue. To characterize the exact location of the second site of O-glycosylation, the peptide Val⁶²⁴-Arg⁶⁴⁴ was purified by RP-HPLC and submitted to 15 cycles of Edman degradation. Only the Thr⁶³⁷ was not characterized indicating the presence of a O-glycan on this residue. Finally, the number of sialic acid residues beared by O-glycans was determined by mass differences between the masses of desialylated and untreated glycopeptides. Thus, the disaccharide carried by the peptides Ser⁶¹¹-Arg⁶²³ and Val⁶²⁴-Arg⁶⁴⁴ possess only one NeuAc, whereas the O-glycopeptide Ser⁶¹¹-Arg⁶⁴⁴ carries one to three NeuAc residues (Table III). Since no disialylated O-glycans is observed on peptides Ser⁶¹¹-Arg⁶²³ and Val⁶²⁴-Arg⁶⁴⁴, it may be assumed that the additional NeuAc carried by the peptide Ser⁶¹¹-Arg⁶⁴⁴ likely corresponds to a tetrasaccharide (NeuAc)₂-Gal β 1-3GalNAc α 1 located on the Thr⁶²¹.

The presence of this glycan might prevent the trypsin cleavage between the Arg⁶²³ and Val⁶²⁴ residues.

These results demonstrate the presence of O-glycans in the C-terminal region of both porcine heavy chains (Figure 5). The pH1 contains O-linked trisaccharides NeuAc-Gal β 1-3GalNAc α 1 located on a fully glycosylated site Thr⁶¹⁶ and on a partially glycosylated site Ser⁶¹¹, whereas pH2 possess the same O-linked trisaccharide on the Thr⁶³⁷ residue and either a trisaccharide or a tetrasaccharide (NeuAc)₂-Gal β 1-3GalNAc α 1 linked to the Thr⁶²¹.

Discussion

Although some degree of glycan heterogeneity is usually observed among the serum glycoproteins, it was previously demonstrated that human I α I possesses a weak heterogeneity of its constitutive N-linked oligosaccharide chains, independantly from the N-glycosylation site (Olsen *et al.*, 1998 ; Flahaut *et al.*, 1998a). Indeed, the oligosaccharide structures previously determined for the three N-glycans on the human I α I heavy chains are mainly of disialylated biantennary structures : two attached to Asn²⁵⁶ and Asn⁵⁵⁹ of hH1 and one linked to hH2 Asn⁶⁴. Moreover, the bikunin, the third peptide chain constituting the human I α I Also possesses one N-glycosylation site (Asn⁴⁵) bearing this same oligosaccharide structure (Toyada *et al.*, 1992). In addition to the N-glycans, the human I α I heavy chains also carry five O-linked glycans : one trisaccharide NeuAc-Gal β 1-3GalNAc attached to the hH1 Thr⁶²⁴ residue and a cluster of three to four O-glycan structures (Gal β 1-3GalNAc α 1) mono- or disialylated, all located near the C-terminal extremity of hH2. In this present study, by using the combined mass spectrometry and Edman approaches of glycopeptides isolated by affinity chromatography, we clearly demonstrate the N- and O-glycosylation patterns of porcine H1 and H2 heavy chains.

Using of an original purification procedure based on the selective retention of peptides and N-glycans from a PNGase F digestion of Con A-purified glycopeptides, we have characterized, in one step, the different peptides containing the N-glycosylation sites and the N-glycan structures attached to these sites. Indeed, among the potential N-glycosylation sites of pH1 and pH2, the analyses by mass spectrometry and Edman degradation of the peptide and N-glycan fractions show that, firstly the pH1 Asn²⁴⁹ residue carries a disialylated biantennary N-glycan with a fucose residue. Secondly, only the Asn⁶⁴ residue of pH2 is effectively occupied by a disialylated biantennary N-glycan never fucosylated.

The main result of our study is to show that, as human I α I heavy chains, porcine H1 and H2 heavy chains carry O-linked glycans. The NetOglyc2.0 computer program predicted that pH1 chain residues Ser⁶¹⁴, Thr⁶¹⁶, Ser⁶¹⁹, Ser⁶²¹ and pH2 chain residues Thr⁶²¹, Ser⁶³², Thr⁶³⁷ were likely to be O-glycosylated. Since discrepancies exist between experimental mapping and predictive algorithms, using Edman degradation, we have experimentally located these sites. The pH1 heavy chain

possesses one partially and one fully occupied O-glycosylation site on the Ser⁶¹¹ and Thr⁶¹⁶ residues, respectively, whereas the pH2 heavy chain contains two fully O-glycosylated sites on the Thr⁶²¹ and Thr⁶³⁷ residues. All these O-glycans consist of a type-1 core-structure Gal β 1-3GalNAc α 1 with one or two NeuAc residues, as demonstrated by (i) the fact that each heavy chain is recognized by the PNA lectin after chemical desialylation, (ii) the O-glycopeptide retention on Jacalin immobilized and finally (iii) the molecular mass increments measured by MALDI-TOF-MS analysis of the Jacalin-isolated glycopeptides. Therefore, as their human counterparts, the N- and O-glycosylation patterns (Figure 5) reveal a weak heterogeneity of the oligosaccharide structures carried by the porcine I α I heavy chains.

In recent years, it has become apparent that N-glycans, which are attached in the endoplasmic reticulum before the folding of the peptide chain, play an important role in the conformational maturation of most glycoproteins. Disulphide linkage also influence the tertiary structure of proteins and are often crucial for their biological function. The BLASTP 2.0.8 computer program (Altschul *et al.*, 1997) reveals amino acid sequence homologies of 81 % between the heavy chain precursors of human and pig. Moreover, the cysteine residues are strictly conserved in the different mammalian species (human, pig, mouse and hamster) for which amino acid sequences of H1 and H2 heavy chains are now known. Therefore, the topology of disulphide bridges described for the human I α I heavy chains (Olsen *et al.*, 1998 ; Flahaut *et al.*, 1998b) is also probably similar in those species. However, although various oligosaccharide structures depending upon tissue source, physiological status, pathological state and species may be found associated to each N-glycosylation site, it is interesting to note that, firstly, the only common N-glycosylation site between the H1 heavy chains of human and pig is always recovered glycosylated. In human, this asparagine residue belongs to the globular domain of the H1 heavy chain, described by Blom *et al.*, 1999. Secondly, the three putative N-glycosylation sites of human and porcine H2 heavy chains are identical. Nevertheless, only the asparagine residue in position 64 is effectively glycosylated and the oligosaccharide structures recovered are always the same, whatever the species. Therefore, taking into account the very weak oligosaccharide diversity, we postulate that the presence of these N-glycans recovered both in human and porcine heavy chains, must play a role in the correct folding of

these highly conserved regions but also in biological activity as cell addressing or plasmatic clearance.

The assembly of mucin-type oligosaccharides occurs by sequential addition of monosaccharide units to the growing oligosaccharide chain, a process invariably initiated by the transfer of an N-acetylgalactosamine unit to a Ser or Thr residue on the acceptor glycoprotein (Roseman., 1970). As protein O-glycosylation occurs after the folding process, the O-glycosylation more likely depends of the accessibility of surface residues rather than any specific sequence or structure, as demonstrated by the various sequences of O-glycosylation already characterized (Elhammer *et al.*, 1993). The presence of Pro, Ser or Thr tends to enhance glycosylation and then random coiled acceptor peptides are better substrates. On both heavy chains, the attachment O-glycosylation sites correspond to proline rich sequences (Bourguignon *et al.*, 1993), which have been described as random coiled structures (Blom *et al.*, 1999). The sequence alignments of C-terminal extremities, which allow the assembly by PGP cross-link between the different polypeptide chains, reveal conserved regions, specially around O-glycosylation sites (Figure 6). If we compare the amino acid sequences established for diverse species, it appeared that, in hH1, the threonine residue, always recovered glycosylated as well as in human than in porcine, also exists in other species. Likewise, in pH1, we have characterized a partial O-glycosylation site on the Ser⁶¹¹. This site is respectively mutated by Leu and Ile in human and mouse but exists as a Thr residue in hamster. For the H2 heavy chains, the partially O-glycosylated residue Thr⁶¹² in human is replaced, in the other species, by Leu or Ile residues of course not glycosylatable. The hH2 chain residue Ser⁶¹⁹ is also mutated in pig but exists in mouse and hamster. The two last hH2 chain residues Thr⁶²¹ and Thr⁶³⁷ are O-glycosylation sites strictly conserved between the compared species. These two sites are effectively found glycosylated in this study. Recent findings indicate that the segment including the clusters of O-glycans corresponds to the region in which amino acid sequences of the three heavy chains (H1, H2 and H3) diverged during evolution, before the emergence of mammals (Nakatani *et al.*, 1997). Therefore, it appear that the O-glycosylation of IαI heavy chains seems to be a conservative process.

The assembly mechanism of heavy chains could depend of this conserved O-glycosylation process. The presence of O-glycans in the neighbouring of PGP cross-links could participate to the assembly selectivity of heavy chains. Indeed, we confirm

the absence of O-glycans on the human H3 heavy chain constituting the pre- α -inhibitor (P α I) molecule. Therefore, the establishment of the ester linkage between the heavy chains and the chondroitin sulfate chain is not dependent of the occurrence of O-glycans. Among the bikunin-linked proteins, P α I is the only positive acute-phase protein (Daveau et al., 1993) and is never found associated with an other heavy chain in the human serum. Therefore, the O-glycosylated C-terminal extremities of H1 and H2 heavy chains may serve as a selective signal for the establishment of the I α I molecule. In contrast, the H3 heavy chain, in which the C-terminal part is unglycosylated, only form the P α I molecule. Therefore, the structural study of oligosaccharides and the identification of glycosylation sites of heavy chains from different species and specially from the bovine I α I (Castillo *et al.*, 1993), which encompass the bovine H2 and H3 heavy chains linked to bikunin might be rewarding.

REFERENCES

- Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. *Nucleic. Acids. Res.*, **25**, 3389-3402.
- Blom,A.M., Mörgelin,M., Öyen,M., Jarvet,J. and Fries,E. (1999) Structural characterization of inter- α -inhibitor : evidence for an extended shape. *J. Biol. Chem.*, **274**, 298-304.
- Bourguignon,J., Diarra-Mehrpour,M., Thiberville,L., Bost,F., Sesboué,R. and Martin,J.P. (1993) Human pre- α -trypsin inhibitor-precursor heavy chain cDNA and deduced amino acid sequence. *Eur. J. Biochem.*, **212**, 771-776.
- Carrette,O., Mizon,C., Sautière,P., Sesboué,R. and Mizon,J. (1997) Purification and characterization of pig inter- α -inhibitor and its constitutive heavy chains. *Biochem. Biophys. Acta.*, **1338**, 21-30.
- Castillo,G.M. and Templeton,D.M. (1993) Subunit structure of bovine ESF (extracellular-matrix stabilizing factor(s)). A chondroitin sulfate proteoglycan with homology to human I alpha i (inter-alpha-trypsin inhibitor). *FEBS Letter.*, **318**, 292-296.
- Daveau,M., Rouet,P., Scotte,M., Faye,L., Hiron,M., Lebreton,J.P. and Salier,J.P. (1993) Human inter- α -inhibitor family in inflammation : simultaneous synthesis of positive and negative acute-phase proteins. *Biochem. J.*, **292**, 485-492.
- Elhammer,A.P., Poorman,R.A., Brown,E., Maggiora,L.L., Hoogerheide,J.G. and Kézdy,F.J. (1993) The specificity of UDP-GalNAc : polypeptide N-acetylgalactosaminyltransferase as inferred from a database of *in vivo* substrates and from the *in vitro* glycosylation of proteins and peptides. *J. Biol. Chem.*, **268**, 10029-10038.

Flahaut,C., Capon,C., Balduyck,M., Ricart,G., Sautière,P. and Mizon,J. (1998a) Glycosylation pattern of human inter- α -inhibitor heavy chains. *Biochem. J.*, **333**, 749-756.

Flahaut,C., Mizon,C., Aumercier-Maes,P., Colson,P., Bailly,C., Sautière,P. and Mizon,J. (1998b) Disulphide bonds assignment in the inter- α -inhibitor heavy chains : structural and functional implications. *Eur. J. Biochem.*, **225**, 107-115.

Hansen,J.E., Lund,O., Tolstrup,N., Gooley,A.A., Williams,K.L. and Brunak,S. (1998) NetOGlyc : prediction of mucin-type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate J.*, **15**, 115-130.

Hara,S., Yamaguchi,M., Takemori,Y., Furuhashi,K., Ogura,H. and Nakamura,M. (1989) Determination of mono-O-acetylated N-acetylneuraminic acids in human and rat sera by fluorometric high-performance liquid chromatography. *Anal. Biochem.*, **179**, 162-166.

Kamerling,J.P., Gerwig,G.J., Vliegthart,J.F.G. and Clamp,J.R. (1975) Characterization by gas-liquid chromatography-mass spectrometry and proton-magnetic-resonance spectroscopy of pertrimethylsilyl methyl glucosides obtained in the methanolysis of glycoproteins and glycopeptides. *Biochem. J.*, **151**, 491-495.

Klein,A., Diaz,S., Ferreira,I., Lamblin,G., Roussel,P. and Manzi,A.E. (1997) New sialic acids from biological sources identified by a comprehensive and sensitive approach : liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) of SIA quinoxalinones. *Glycobiology.*, **7**, 421-432.

Kornfeld,R. and Ferris,C. (1975) Interactions of immunoglobulin glycopeptides with Concanavalin A. *J. Biol. Chem.*, **250**, 2614-2619.

Kornfeld,R. and Kornfeld,S. (1985) Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.*, **54**, 631-664.

Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly on the head of bacteriophage T4. *Nature.*, **227**, 680-685.

Lotan,R., Skutelsky,E., Danon,D. and Sharon,N. (1975) The purification, composition and specificity of the anti-T lectin from peanut (*Arachis hypogaea*). *J. Biol. Chem.*, **250**, 8518-8523.

Michalski,C., Piva,F., Balduyck,M., Mizon,C., Burnouf,T., Huart,J.J. and Mizon,J. (1994) Preparation and properties of a therapeutic inter-alpha-trypsin inhibitor concentrate from human plasma. *Vox. Sang.*, **67**, 329-336.

Mizon,C., Héron,A., Capon,C., Sautière,P., Michalski,C., Sesboüé,R. and Mizon,J. (1997) Human pre- α -inhibitor : isolation from a by-product of industrial scale plasma fractionation and structural analysis of its H3 heavy chain. *J. Chromatogr. B.*, **692**, 281-291.

Montreuil,J., Bouquelet,S., Debray,H., Fournet,B., Spik,G. and Strecker,G. (1986) in Carbohydrate analysis : a Practical Approach (Chaplin, M. F. and Kennedy, J. F., eds.), pp. 143-204, IRL Press, Oxford.

Nakatani,T., Suzuki,Y., Yamamoto,T. and Sinohara,H. (1997) Molecular cloning and sequencing of cDNAs encoding three heavy-chain precursors of the inter- α -trypsin inhibitor in Syrian hamster : implications for the evolution of the inter- α -trypsin inhibitor heavy chain family., *J. Biochem. (Tokyo)*, **122**, 71-82.

Olsen,E.H.N., Rahbek-Nielsen,H., Thogersen,I.B., Roepstorff,P. and Enghild,J.J. (1998) Posttranslational modifications of human inter- α -inhibitor : identification of glycans and disulfide bridges in heavy chains 1 and 2. *Biochemistry.*, **37**, 408-416.

Packer,N.H., Lawson,M.A., Jardine,D.R. and RedmondJ.W. (1998) A general approach to desalting oligosaccharides released from glycoproteins. *Glycoconjugate. J.*, **15**, 737-747.

Plummer, T.H. Jr., Elder, J.H., Alexander, S., Phelan, A.W. and Tarentino, A.L. (1984) Demonstration of peptide: N-glycosidase F activity in endo-beta-N-acetylglucosaminidase F preparations. *J. Biol. Chem.*, **259**, 10700-10704.

Roseman, S. (1970) The synthesis of complex carbohydrates by multiglycosyltransferase systems and their potential function in intracellular adhesion. *Chem. Phys. Lipids.*, **5**, 270-297.

Salier, J.P., Rouet, P., Raguenez, G. and Daveau, M. (1996) The inter- α -inhibitor family : from structure to regulation. *Biochem. J.*, **315**, 1-9.

Shibuya, N., Golstein, I.J., Broekaert, W.F., Nsimba-Lubaki, M., Peters, B. and Peumans, W.J. (1987) The elderberry (*Sambucus nigra* L.) bark lectin recognizes the NeuAc(α 2-6)Gal/GalNAc sequence. *J. Biol. Chem.*, **262**, 1596-1601.

Stephano, J.L., Gould, M. and Rojas-Galicia, L. (1986) Advantages of picrate fixation for staining polypeptides in polyacrylamide gels. *Anal. Biochem.*, **152**, 308-313.

Toyada, H., Ikei, T., Demachi, Y., Toida, T. and Imanari, T. (1992) Structural analysis of the N-linked oligosaccharides from human urinary trypsin inhibitor. *Chem. Pharm. Bull.*, **40**, 2882-2884.

Wang, W.C. and Cumming, R.D. (1988) The immobilized leucoagglutinin from the seeds of *Maackia amurensis* binds with high affinity to complex-type Asn-linked oligosaccharides containing terminal sialic acid-linked α -2,3 to penultimate galactose residues. *J. Biol. Chem.*, **263**, 4576-4585.

LEGEND TO THE FIGURES

Fig. 1. Electrophoresis and blotting with digoxigenin-labelled lectins.

Panels correspond to blots incubated in presence of digoxigenin-labelled lectins: (A), *Concanavalia ensiformis* agglutinin (Con A-dig); (B), *Arachis hypogaea* agglutinin (PNA-dig); (C), *Sambucus nigra* agglutinin (SNA-dig); (D), *Maackia amurensis* agglutinin (MAA-dig). The abbreviations used to identify glycoproteins are : pH1 and pH2, porcine H1 and H2 heavy chains; hH1, hH2 and hH3, heavy chains from human. Numbers correspond to the control glycoproteins: (1), carboxypeptidase Y; (2), asialo fetuin; (3), fetuin; (4), human transferrin; (5), bovine serum albumin; (6), porcine transferrin. The migration of standard masses are depicted on the right.

Fig. 2. RP-HPLC of sialic acid quinoxalinones.

Fluorescence detection was performed with an excitation and an emission wavelength of 373 nm and 448 nm, respectively. (A), typical RP-HPLC profile of DMB-derivatized sialic acids from bovine submaxillary mucin used as standard; (B), profile of DMB-derivatized sialic acids from human I α I; (C), profile obtained from porcine I α I. Neu5Gc, *N*-glycolylneuraminic acid; Neu5Ac, *N*-acetylneuraminic acid; Neu5,7Ac₂, 7-*O*-acetyl-*N*-acetylneuraminic acid; Neu5Gc9Ac, 9-*O*-acetyl-*N*-glycolylneuraminic acid; Neu5,8Ac₂, 8-*O*-acetyl-*N*-acetylneuraminic acid; Neu5,9Ac₂, 9-*O*-acetyl-*N*-acetylneuraminic acid; Neu5,7,9Ac₃, 7,9-Di-*O*-acetyl-*N*-acetylneuraminic acid.

Fig. 3. MALDI time-of-flight spectra from N-glycopeptides of pH2.

Con A-purified N-glycopeptides of pH2 were analyzed by MALDI before (panel A) and after (panel B) chemical desialylation. The desialylated N-glycopeptides were then digested by PNGase F. The desialylated N-glycan (panel D) and peptide (panel C) pools were purified and desalted onto PGC and C18 column, respectively and further analyzed by mass spectrometry.

Fig. 4. MALDI time-of-flight spectra from N-glycopeptides of pH1.

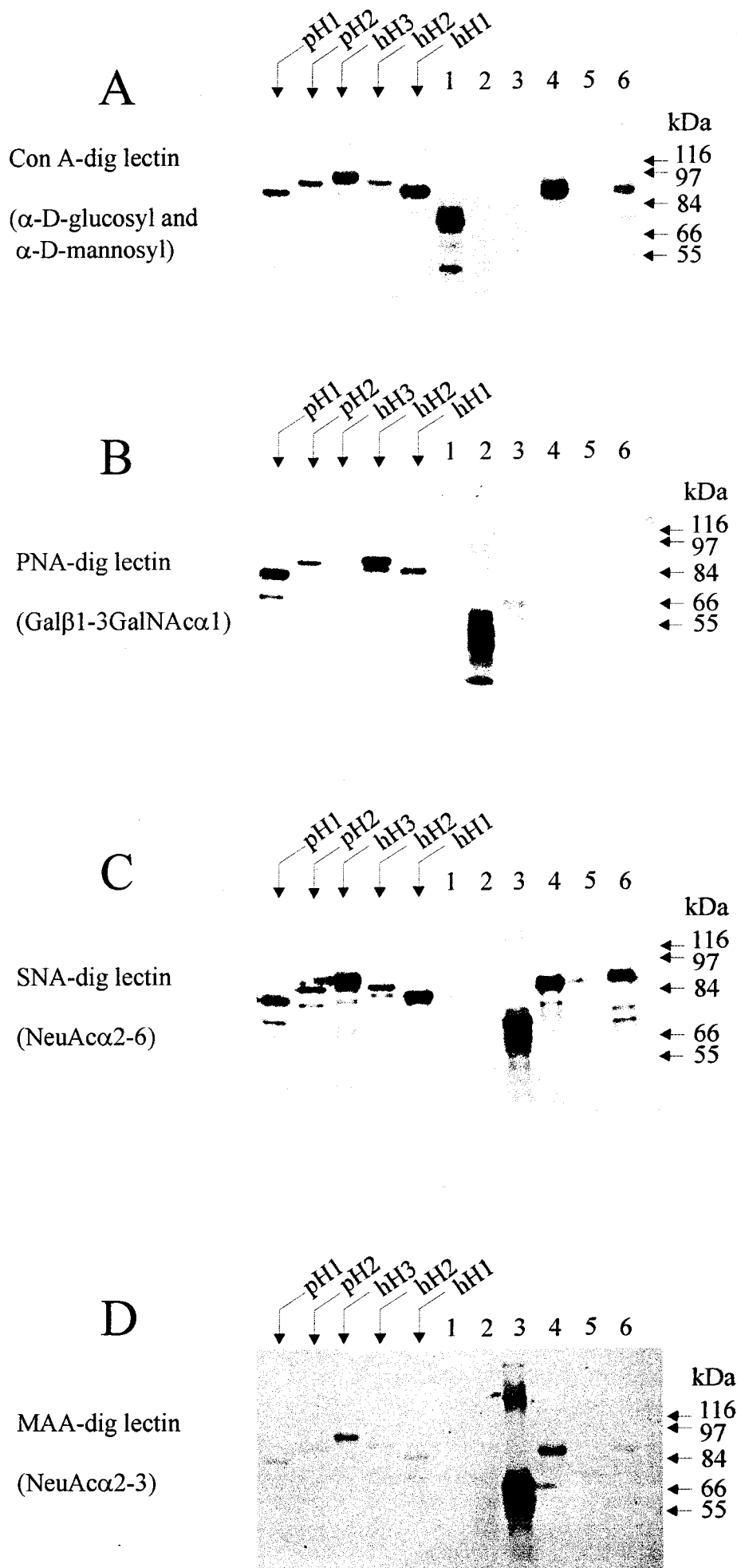
After tryptic hydrolysis and isolation onto Con A-Sepharose, N-glycopeptides of pH1 were analyzed by MALDI in linear mode before (panel A) and after (panel B) chemical desialylation. The desialylated fraction was then digested by PNGase F. After purification, N-glycans (panel C) and peptides (panel D) were analyzed by MALDI in reflectron mode.

Fig. 5. Schematic representation of glycosylation pattern of porcin heavy chains.

The heavy chains are assembled via ester bonds between their C-terminal Asp residues and an internal N-acetylgalactosamine residue of the chondroitin sulfate chain that originates from Ser¹⁰ of bikunin. Monosaccharides are depicted with symbols: (●), N-acetylglucosamine; (◇), N-acetylgalactosamine; (O), mannose; (■), galactose; (Δ), sialic acid; (□), fucose. Monosaccharides enclosed in brackets are only present in some of the glycans. (*), The pH1 Ser⁶¹¹ chain residue is a partially occupied O-glycosylation site.

Fig. 6. Sequence comparison of O-linked sites of IαI heavy chains from several species.

A, sequence alignments of H1 heavy chains; B, sequence alignments of H2 heavy chains. The O-glycosylation sites are indicated in bold. The black box indicates the amino acids conserved between the different species.



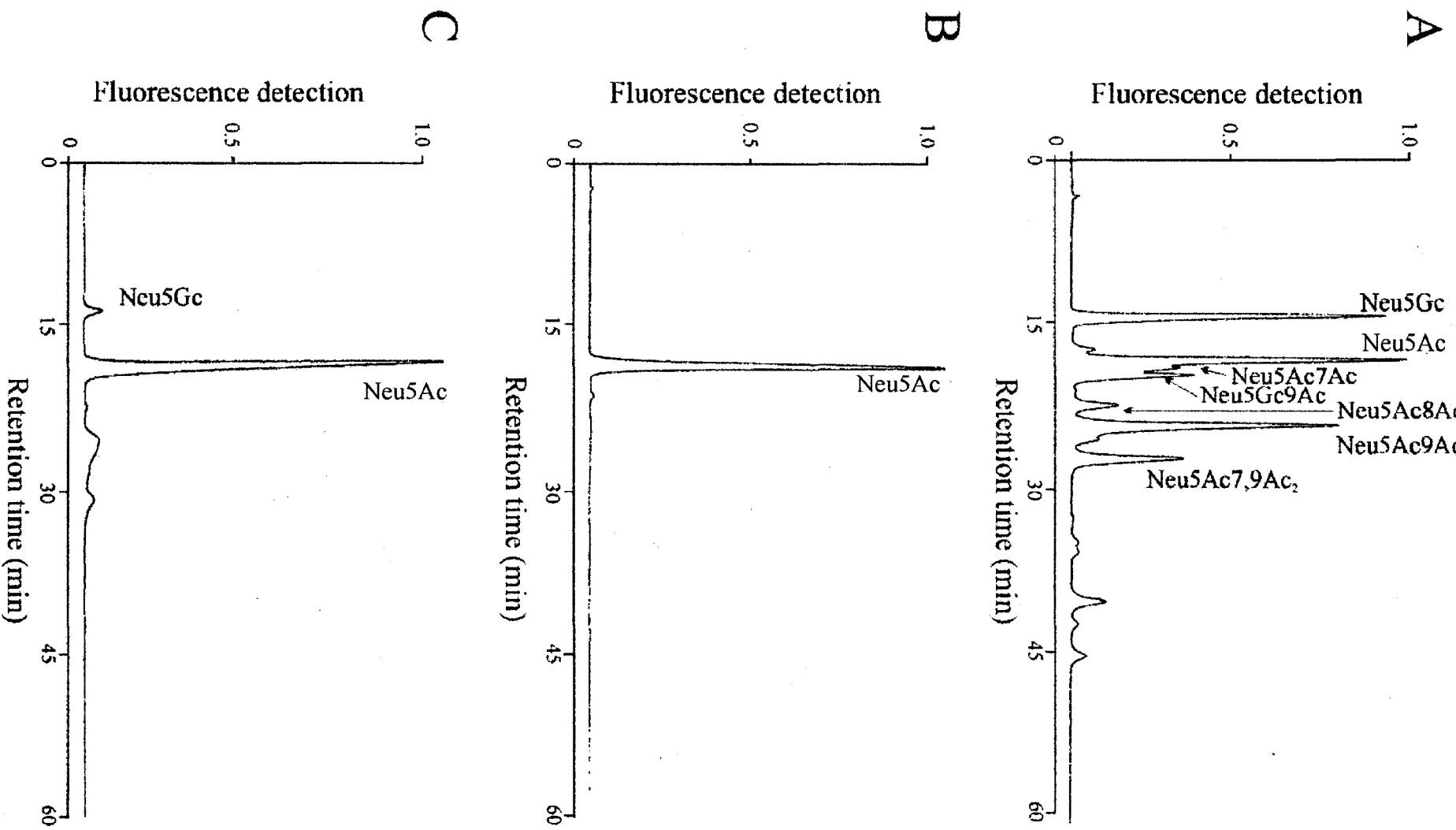
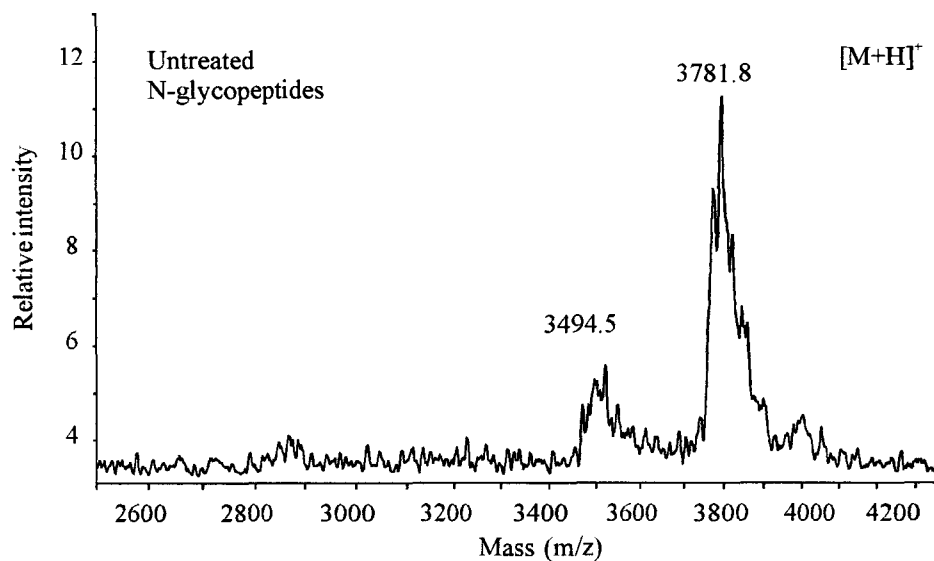
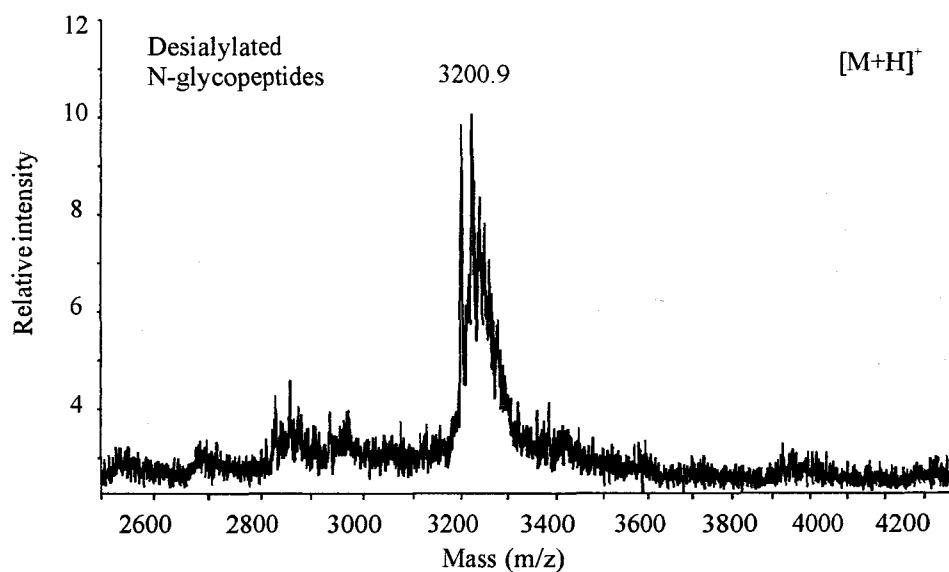


Figure 2

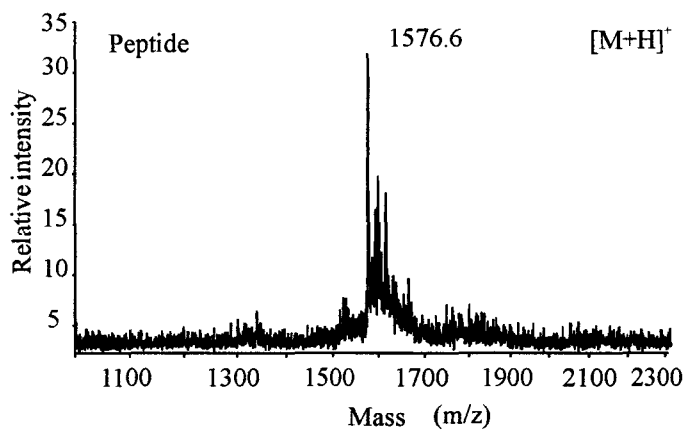
A



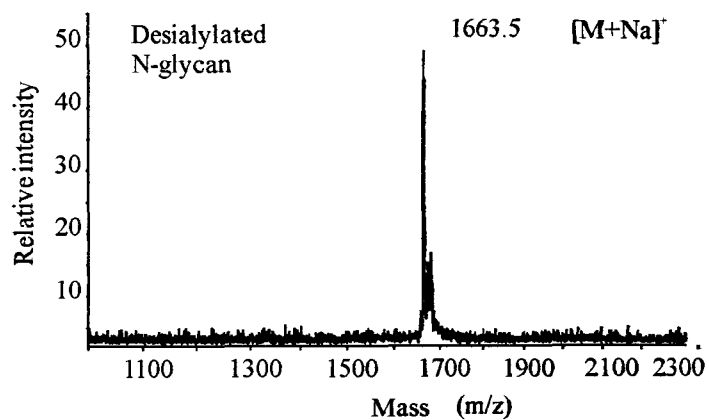
B

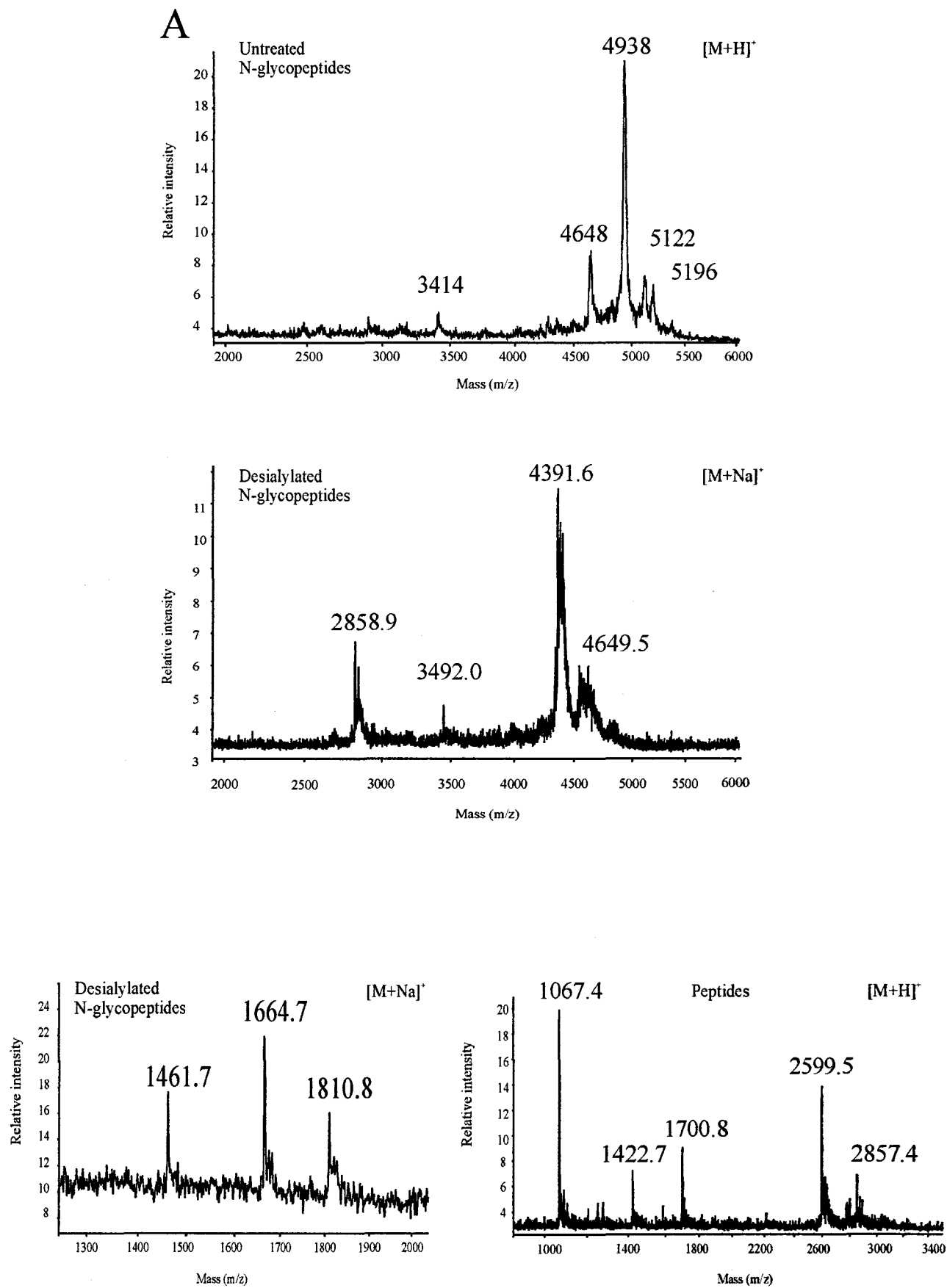


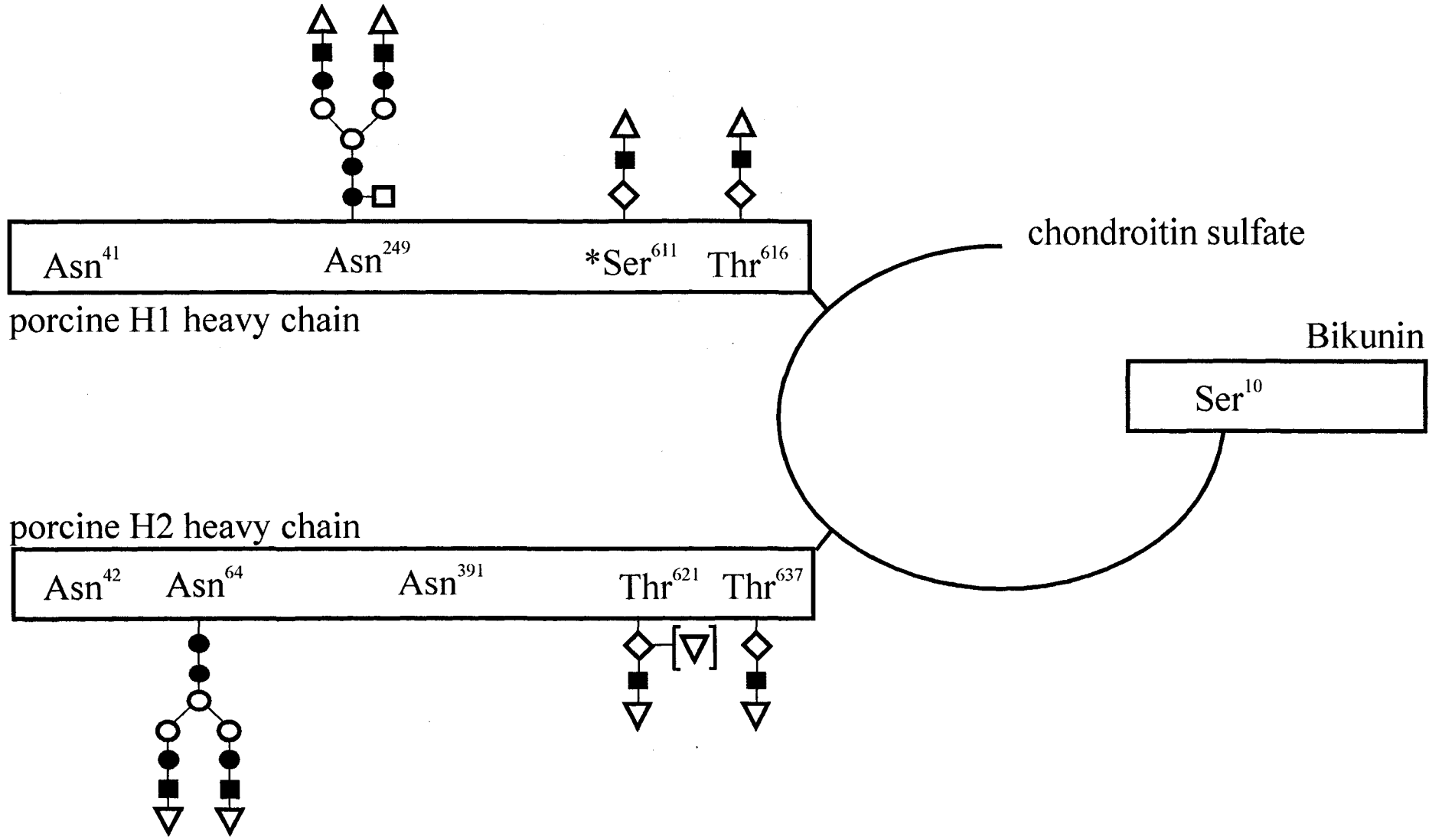
C



D







A

ITH1-human	G P R R T	F V L S A L	Q P S P <u>T</u>	H S S S N T Q	R L P D R	V T G V T D
ITH1-pig	G P R R K	F V L S A S	Q P S P <u>T</u>	H P S S S I Q	K L P D R	V T G V T D
ITH1-mouse	G P R R T	F V L S A I	Q P S P T	A H P T D - S	K L P L R	V T G V T D
ITH1-hamster	G P R R T	F V L S A T	Q P S P T	A R S S V V S	K L P N Q	V T G V T D

B

133

ITH2-human	K V V P D S <u>T</u>	P S W A N	P <u>S</u>	P <u>T</u> P	V I S M L A	Q G S Q V L	E S	<u>T</u> P P P	H V M R	V E N D
ITH2-pig	K V T P N S L	P S W V N	P L	P <u>T</u> P	R V P L P A	V G P S V I	E A	<u>T</u> P P P	H V M R	V E N D
ITH2-mouse	K V A S G P I	P S W A N	P S	P T P	M S A M L A	V G A K P L	E S	T P P T	H L N Q	V E N D
ITH2-hamster	K V A S A S I	P S W A S	P S	P T P	V M A M L A	V G A N R L	E S	T P P P	H V I R	V E N D

Figure 6

Table I. Monosaccharide composition of porcine I α I heavy chains.

Heavy chain	Molar ratios ^a						
	Fuc	Gal	Man	Glc	GalNAc	GlcNAc	NeuAc
pH1	1	3.1	3	-	0.85	2.6	3.0
pH2	-	4.3	3	-	1.7	3.0	3.2

^aThe molar ratios were calculated on the basis of three mannose units.

Table II. Structural identification of polypeptide backbones and N-glycans from porcine H1 and H2 heavy chains^a

Peptides ^{b,c}	Molecular mass [M+H] ⁺ (Da)			Edman degradation	N-glycan structures	Molecular mass [M+Na] ⁺ (Da)	
	observed	calculated				observed	calculated
pH1 Glu ²²⁹ -Lys ²⁵²	2857.4	2856.3	E K L - D L L	(HexNAc) ₄ -(Hex) ₅ -Fuc	1810.8	1809.5	
Lys ²³¹ -Lys ²⁵²	2599.5	2599.0	L - D L L V A	(HexNAc) ₄ -(Hex) ₅	1664.7	1663.5	
Asn ²³⁹ -Lys ²⁵²	1700.8	1699.9	N Y F A H F F	(HexNAc) ₃ -(Hex) ₅	1461.7	1460.5	
Phe ²⁴¹ -Lys ²⁵²	1422.7	1422.6	F A H F F A P				
Phe ²⁴⁴ -Lys ²⁵²	1067.4	1067.2	F F A P Q D L				
pH2 Gly ⁵⁹ -Lys ⁷³	1576.6	1576.7	G A F I S D F	(HexNAc) ₄ -(Hex) ₅	1663.5	1663.5	

^a N-glycosylated tryptic peptides of porcine H1 and H2 were isolated by affinity chromatography on Con A-Sepharose column. After chemical desialylation and PNGase F digestion, peptides and N-glycans were separately isolated onto C18 column and PGC column, respectively. They were then analyzed by MALDI-TOF-MS in positive reflectron mode.

^b The mass of peptides containing a cysteine residue are increased of 57 Da (-CH₂-CONH₂). Gaps (-) symbolize a carboxamidomethylated cysteine residue.

^c Peptides are modified by PNGase F action, which transforms the asparagine carrying the N-glycan in a aspartic acid residue.

Table III. Structural identification of peptides and O-linked carbohydrates from porcine H1 and H2 heavy chains^a

	Peptides	Treatment	Molecular mass [M+H] ⁺ (Da)		Edman degradation	Proposed carbohydrate
			observed	calculated		
pH1	Phe ⁶⁰⁶ -Lys ⁶²⁴	Untreated	3309.8	3312.4	nd	(HexNAc-Hex-NeuAc) ₂
			2659.7	2655.8	nd	HexNAc-Hex-NeuAc
		Desialylated	2731.7	2729.8	FVLSA-QPSP-HPSSSI	(HexNAc-Hex) ₂
			2365.2	2364.5	FVLSASQPSP-HPSSSI	HexNAc-Hex
		Naked	nd	1999.2	nd	none
pH2	Ser ⁶¹¹ -Arg ⁶⁴⁴	Untreated	5205.9	5215.8	nd	(HexNAc-Hex) ₂ -NeuAc ₃
			4917.9	4924.5	nd	(HexNAc-Hex) ₂ -NeuAc ₂
			4634.0	4633.2	nd	(HexNAc-Hex) ₂ -NeuAc
		Desialylated	4342.5	4341.9	SLPSW	(HexNAc-Hex) ₂
		Naked	nd	3611.3	nd	none
pH2	Ser ⁶¹¹ -Arg ⁶²³	Untreated	2118.9	2121.3	nd	HexNAc-Hex-NeuAc
		Desialylated	1830.2	1830.0	SLPSW	HexNAc-Hex
		Naked	nd	1464.7	nd	none
pH2	Val ⁶²⁴ -Arg ⁶⁴⁴	Untreated	2821.0	2822.2	nd	HexNAc-Hex-NeuAc
			2529.2	2530.9	nd	HexNAc-Hex
		Desialylated	2531.0	2530.9	VPLPAVGPSIEA-P	HexNAc-Hex
		Naked	nd	2165.6	nd	none

^a O-glycosylated tryptic peptides of porcine H1 and H2 were isolated by Jacalin-agarose chromatography. Subsequently, the carbohydrates were chemically desialylated. MALDI-TOF-MS analysis were performed in linear (untreated glycopeptides) and reflectron (desialylated glycopeptides) mode. The carbohydrate structures were deduced from the mass changes. Gaps (-) symbolize the threonine residue carrying the O-glycans. nd, not determined.

Discussions et Perspectives

DISCUSSIONS ET PERSPECTIVES

La caractérisation dans le sérum et dans l'urine d'un inhibiteur acido stable de la trypsine a été à l'origine de la définition d'une famille d'inhibiteurs de sérine-protéinases circulant dans le plasma de l'homme et de nombreuses espèces animales dont le principal représentant est l' α_1 I. Il était logique que les recherches alors conduites aient été orientées sur une meilleure définition du motif responsable de l'activité antiprotéasique et de son rôle physiologique éventuel. Les résultats obtenus sont importants. La structure tridimensionnelle de la bikunine qui peut être obtenue par génie génétique est connue. Le spectre des protéinases inhibées est étendu puisqu'il comprend des enzymes impliquées dans la réaction inflammatoire ; élastase, plasmine... Ce spectre s'élargit encore : la bikunine vient d'être reconnue comme un inhibiteur très spécifique du granzyme K [145]. Enfin, la bikunine qui est utilisée en thérapeutique comme anti inflammatoire notoire pourrait tirer son activité d'autres propriétés intéressantes : inhibition de la synthèse ou de l'activité de cytokines, de métalloprotéinases...

Les membres de la famille de l' α_1 I comportent le plus souvent la bikunine associée covalentiellement à une ou deux chaînes lourdes par l'intermédiaire d'un glycosaminoglycane de type chondroïtine sulfate. Une telle structure n'a jamais été jusqu'ici retrouvée dans un autre composant de l'organisme.

Les chaînes lourdes ont été initialement caractérisées par l'identification des ARNm qui les codent et paradoxalement, leurs structures tridimensionnelles et leurs rôles physiologiques sont restées totalement énigmatiques jusqu'à la mise en évidence de récepteurs membranaires capables de les reconnaître et à celle de leur affinité pour l'acide hyaluronique qui leur permet d'aider à la cohésion de la matrice extracellulaire.

Il devient donc essentiel de mieux préciser la structure conformationnelle de ces protéines.

Dans une première étape, nous démontrons que les chaînes H1 et H2 qui entrent dans la constitution de l'I α I présentent une topologie des ponts disulfure nettement distincte. Cependant, celle-ci ne semble pas influencer de manière déterminante sur leur structure tridimensionnelle étudiée par leur spectre de dichroïsme circulaire dans l'ultraviolet lointain. Ainsi, la réactivité du domaine von Willebrand qui leur est commun pourrait être conservée et participer à leur capacité à reconnaître des glycosaminoglycannes. Cet aspect mériterait sans doute une étude plus poussée.

De même, la présence de la structure « thioredoxine like » commune à toutes les chaînes lourdes (Cys-Pro-Xaa-Cys) est intrigante. Peut-elle se réduire dans le milieu intra ou extracellulaire et ainsi participer à certains processus d'oxydoréduction ?

L'I α I était connu depuis longtemps comme une des protéines plasmatiques relativement peu nombreuses susceptibles d'être O-glycosylées. Nous démontrons dans la deuxième partie de notre travail qu'effectivement les chaînes lourdes H1 et H2 humaines sont O-glycosylées et que les O-glycannes sialylés qui dérivent de l'antigène T (Gal-GalNAc) sont fixés dans la région C-terminale riche en proline et donc peu structurée de ces chaînes.

A quelle nécessité leur présence répond-t-elle ? Ils semblent importants puisque nous les retrouvons dans les chaînes isolées de l'I α I porcin. Ils ne paraissent pas fixés par hasard en raison des particularités de séquence de l'extrémité C-terminale des chaînes lourdes puisque nous ne les trouvons pas dans la chaîne H3 pourtant très homologue. Ils n'interviennent pas dans les mécanismes de formation de la liaison protéine-glycosaminoglycane-protéine puisque cette liaison permet l'union de la chaîne H3 à la bikunine dans le pré- α -inhibiteur. Participent-ils à la spécificité du choix des chaînes lourdes associées à la bikunine au cours de la biosynthèse des différents membres de la famille de l'I α I ? Le pré- α -inhibiteur comporte une seule chaîne lourde H3 non O-glycosylée qui peut être synthétisé *in vitro* par des cellules COS transfectées [23, 115]. Au contraire, l'I α I associe les chaînes lourdes H1 et H2 qui sont O-glycosylées mais n'a pu jusqu'ici

être obtenu *in vitro* par transfection de cellules COS [115] alors qu'il peut être produit par des hépatocytes primaires en culture mieux capables peut-être de reproduire les modifications post-traductionnelles des chaînes lourdes. Il devient possible de tester expérimentalement ces hypothèses. L'I α I assurément apparaît comme une protéine plasmatique très originale. Ne reste-t-elle pas aussi par bien des aspects très « énigmatique ».

Bibliographie

1. Steinbuch, M., Loeb, J. (1961) Isolation of an alpha-2-globulin from human plasma. *Nature*, **192**, 1196.
2. Mizon, C., Balduyck, M., Albani, D., Michalski, C., Burnouf, T., Mizon, J. (1996) Development of an enzyme-linked immunosorbent assay for human plasma inter- α -trypsin inhibitor (ITI) using specific antibodies against each of the H1 and H2 heavy chains. *J. Imm. Meth.*, **190**, 61-70.
3. Heide, K., Heimburger, N. and Haupt, H. (1965) An inter-alpha-trypsin inhibitor of human serum. *Clin. Chim. Acta*, **11**, 82-85.
4. Gebhard, W. and Hochstrasser, K. (1986) Inter- α -trypsin inhibitor and its close relatives, *In Proteinase inhibitors* (Barrett, A.J. and Salvesen, G., eds.) pp. 309-401, Elsevier Sci. Publ. BV, Amsterdam.
5. Bourguignon, J., Vercaigne, D., Sesboüé, R. and Martin, J.P. (1983) Inter-alpha-trypsin inhibitor (ITI): two distinct mRNAs in baboon liver argue for a discrete synthesis of ITI and ITI derivatives. *FEBS Lett.*, **162**, 379-383.
6. Schreitmüller, T., Hochstrasser, K., Reisinger, P., Wachter, E. and Gebhard, W. (1987) cDNA cloning of human inter- α -trypsin inhibitor discloses three different proteins. *Biol. Chem. Hoppe-Seyler*, **368**, 963-970.
7. Salier, J.P., Diarra-Mehrpour, M., Sesboüé, R., Bourguignon, R., Ohkubo, I., Kurachi, S., Kurachi, K. and Martin, J.P. (1987) Isolation and characterization of cDNAs encoding the heavy chain of human inter- α -trypsin inhibitor ($I\alpha$ TI): unambiguous evidence for multipolypeptide chain structure of $I\alpha$ TI. *Proc. Natl. Acad. Sci. USA*, **84**, 8272-8276.
8. Hochstrasser, K., Bretzel, G., Feuth, H., Hilla, W. and Lempart, K. (1976) The inter- α -trypsin inhibitor as a precursor of the acid-stable proteinase inhibitors in human serum and urine. *Hoppe-Seyler's Z. Physiol. Chem.*, **357**, 153-162.

9. Balduyck, M., Hayem, A., Kerckaert, J.P., Mizon, C. and Mizon, J. (1982) Isolation of a human urinary trypsin inhibitor. *Biochem. Biophys. Res. Commun.*, **109**, 1247-1255.
10. Balduyck, M., Mizon, C., Loutfi, H., Richet, C., Roussel, P. and Mizon, J. (1986) The Major urinary trypsin inhibitor is a proteoglycan. *Eur. J. Biochem.*, **158**, 417-422.
11. Jessen, T.E., Faarvang, K.L. and Ploug, M. (1988) Carbohydrate as covalent crosslinkin human inter- α -trypsin inhibitor : a novel plasma protein structure. *FEBS Lett.*, **230**, 195-200.
12. Balduyck, M., Laroui, S., Mizon, C. and Mizon, J. (1989) A proteoglycan related to the Urinary Trypsin Inhibitor (UTI) links the two heavy chains of inter- α -trypsin inhibitor. *Biol. Chem. Hoppe-Seyler*, **370**, 329-336.
13. Enghild, J.J., Thogersen, I.B., Pizzo, S.V. and Salvesen, G. (1989) Analysis of inter- α -trypsin inhibitor and a novel trypsin inhibitor, pre- α -trypsin inhibitor, from human plasma. Polypeptide chain stoichiometry and assembly by glycan. *J. Biol. Chem.* **264**, 15975-81.
14. Enghild, J.J., Salvesen, G., Hefta, S.A., Thogersen, I.B., Rutherford, S. and Pizzo, S.V. (1991) Chondroitin 4-sulfate covalently cross-links the chains of the human blood protein Pre- α -inhibitor. *J. Biol. Chem.*, **266**, 747-751.
15. Salier, J.P., Rouet, P., Raguenez, G. and Daveau, M. (1996) The inter- α -inhibitor family : from structure to regulation. *Biochem. J.*, **315**, 1-9.
16. Mizon, C., Héron, A., Capon, C., Sautière, P., Michalski, C., Sesboüé, R. and Mizon, J. (1997) Human pre- α -inhibitor : isolation from a by-product of industrial scale plasma fractionation and structural analysis of its H3 heavy chain. *J. Chromatogr.*, **692**, 281-291.

17. Enghild, J.J., Salvesen, G., Thøgersen, I.B., Valnickova, Z., Pizzo, S.V. and Hefta, S.A. (1993) Presence of the protein-glycosaminoglycan-protein covalent cross-link in the inter- α -inhibitor related proteinase inhibitor heavy chain 2/bikunin. *J. Biol. Chem.*, **268**, 8711-8716.
18. Hammer, C.H., Jacobs, R.M., and Franck, M.M. (1989) Isolation and characterization of a novel plasma protein which binds to activated C4 of the classical complement pathway. *J. Biol. Chem.*, **264**, 2283-2291.
19. Choi-Miura, N.H., Sano, Y., Oda, E., Nakano, Y., Tobe, T., Yanagishita, T., Taniyama, M., Katagiri, T. and Tomita, M. (1995) Purification and characterization of a novel glycoprotein which has significant homology to heavy chains of inter- α -trypsin inhibitor family from human plasma. *J. Biochem.*, **117**, 400-407.
20. Nishimura, H., Kakizaki, I., Muta, T., Sasaki, N., Pu, P.X., Yamashita, T. and Nagasawa, S. (1995) cDNA and deduced amino acid sequence of human PK-120, a plasma kallikrein-sensitive glycoprotein. *FEBS Lett.*, **357**, 207-211.
21. Jean, L., Risler, J.L., Nagase, T., Coulouarn, C., Nomura, N. and Salier, J.P. (1999) The nuclear protein PH5P of the inter- α -inhibitor superfamily : a missing link between poly(ADP-ribose)polymerase and the inter- α -inhibitor family and a novel actor as DNA repair ? *FEBS Lett.* **446**, 6-8
22. Héron, A., Bourguignon, J., Diarra-Mehrpour, M., Dautréaux, B., Martin, J.P. and Sesboué, R. (1995) Involvement of the three inter- α -trypsin inhibitor (ITI) heavy chains in each member of the serum. *FEBS Lett.* **374**, 195-198.
23. Blom, A.M., Thuveson, M. and Fries, E. (1997) Intracellular coupling of bikunin and the heavy chain of rat pre- α -inhibitor in Cos-1 cells. *Biochem. J.*, **328**, 185-191.
24. Carrette, O., Mizon, C., Sautière, P., Sesboué, R. and Mizon, J. (1997) Purification and characterization of pig inter- α -inhibitor and its constitutive heavy chains. *Biochim. Biophys. Acta*, **1338**, 21-30.

25. Castillo, G.M. and Templeton, D.M. (1993) Subunit structure of bovine ESF (extracellular-matrix stabilizing factor(s)). A chondroitin sulfate proteoglycan with homology to human I alpha i (inter-alpha-trypsin inhibitor). *FEBS Lett.*, **318**, 292-296.
26. Yamamoto, T., Yamamoto, K. and Sinohara, H. (1996) Inter- α -trypsin inhibitor and its related proteins in syrian hamster urine and plasma. *J. Biochem.*, **120**, 145-152.
27. Nakatani, T., Suzuki, Y., Yamamoto, T. and Sinohara, H. (1997) Molecular cloning and sequencing of cDNAs encoding three heavy-chain precursors of the inter- α -trypsin inhibitor in syrian hamster : implications for the evolution of the inter- α -trypsin inhibitor heavy chain family. *J. Biochem.*, **122**, 71-82.
28. Salier, J.P., Verga, V., Doly, J., Diarra-Mehrpour, M. and Erickson, R.P. (1992) The genes for the inter- α -inhibitor family share a homologous organization in human and mouse. *Mammal. Genome.*, **2**, 233-239.
29. Wachter, E., Hochstrasser, K., Bretzel, G. and Heindl, S. (1979) Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter- α -trypsin inhibitor .II. Characterization of the second inhibitory inactive domaine by amino acid sequence determination. *Hoppe-Seyler's Z. Physiol. Chem.*, **360**, 1297-1303.
30. Aubry, M. and Bieth, J. (1976) A kinetic study of the inhibition of human and bovine trypsins and chymotrypsins by the inter- α -inhibitor from human plasma. *Biochem. Biophys. Acta*, **438**, 221-230.
31. Potempa, J., Kwon, K., Chawla, R. and Travis, J. (1989) Inter- α -trypsin inhibitor : inhibition spectrum of native and derived forms. *J. Biol. Chem.*, **264**, 15109-15114.
32. Pratt, C.W. and Pizzo, S.V. (1987) Mechanism of action of inter- α -trypsin inhibitor. *Biochemistry*, **26**, 2855-2863.

33. Dietl, T., Dobrinski, W. and Hochstrasser, K. (1979) Human inter- α -trypsin inhibitor : limited proteolysis by trypsin, plasmin, kallikrein and granulocytic elastase and inhibitory properties of the cleavage products. *Hoppe-Seyler's Z. Physiol. Chem.*, **360**, 1313-1318.
34. Hirose, J., Ozawa, T., Miura, T., Isaji, M., Nagao, Y., Yamashiro, K., Nii, A., Kato, K. and Uemura, A. (1998) Human neutrophil elastase degrades inter- α -trypsin inhibitor to liberate urinary trypsin inhibitor related proteins. *Biol. Pharm. Bull.*, **21**, 651-656.
35. Potempa, J., Korzus, E. and Travis, J. (1994) The serpin superfamily of proteinase inhibitors : structure, function, and regulation. *J. Biol. Chem.*, **269**, 15957-15960.
36. Pratt, C.W., Roche, P.A. and Pizzo, S.V. (1987) The role of inter- α -trypsin inhibitor and the other proteinase inhibitors in the plasma clearance of the neutrophil elastase and plasmin. *Arch. Biochem. Biophys.*, **258**, 591-599.
37. Jönsson, B. and Ohlsson, K. (1984) Interactions of the complex between human urinary trypsin inhibitor and human leukocyte elastase with α 1-proteinase inhibitor and α 2-macroglobulin. *Hoppe-Seyler's Z. Physiol. Chem.*, **365**, 1403-1408.
38. Ohnishi, H., Kosuzume, H., Ashida, Y., Kato, K. and Honjo, I. (1984) Effects of urinary trypsin inhibitor on pancreatic enzymes and experimental acute pancreatitis. *Dig. Dis. Sci.*, **29**, 26-32.
39. Tani, S., Otsuki, M., Itoh, H., Nakamura, T., Fujii, M., Okabayashi, Y., Fujisawa, T. and Baba, S. (1988) The protective effect of trypsin inhibitor urinastatin on cerulein-induced acute pancreatitis in rats. *Pancreas*, **3**, 471-476.
40. Nakakuki, M., Yamasaki, F., Shinkawa, T., Kudo, M., Watanabe, M. and Mizota, M. (1995) Protective effect of human ulinastatine against gentamicin-induced acute renal failure in rats. *Can. J. Physiol. Pharmacol.*, **74**, 104-111.

41. Kato, Y., Kudo, M., Shinkawa, T., Mochizuki, H., Isaji, M., Shiromizu, I. and Hoshida, K. (1998) Role of O-linked carbohydrate of human urinary trypsin inhibitor on its lysosomal membrane-stabilizing property. *Biochem. Biophys. Res. Com.*, **243**, 377-383.
42. Balduyck, M. and Mizon, J. (1991) L'inter-alpha-trypsine inhibiteur et ses dérivés plasmatiques et urinaires. *Ann. Biol. Clin.*, **49**, 273-281.
43. Tani, T., Aoki, H., Yoshioka, T., Lin, K.J. and Kodama, M. (1993) Treatment of septic shock with a protease inhibitor in a canine model : a prospective, randomized, controlled trial. *Crit. Care. Med.*, **21**, 925-930.
44. Endo, S., Inada, K., Taki, K., Hoshi, S. and Yoshida, M. (1990) Inhibitory effects of ulinastatin on the production of cytokines : implication for the prevention of septicemic shock. *Clin. Therap.*, **12**, 323-326.
45. Endo, S., Inada, K., Yamashita, H., Takakuva, T., Nakae, H., Yamada, Y., Kasai, T., Taki, K. and Yoshida, M. (1993) The inhibitory actions of protease inhibitors on the production of polymorphonuclear elastase and interleukin 8. *Res. Commun. Chem. Pathol. Pharmacol.*, **82**, 27-33.
46. Maehara, K., Kanayama, N., Halim, A., El Maradny, E., Oda, T., Fujita, M. and Terao, T. (1995) Down-regulation of interleukin-8 gene expression in HL60 cell line by human Kunitz-type trypsin inhibitor. *Biochem. Biophys. Res. Comm.*, **206**, 927-934.
47. Kanayama, N., Halim, A., Maehara, K., Kajiwara, Y., Fujie, M. and Terao, T. (1995) Kunitz-type trypsin inhibitor prevents LPS-induced increase of cytosolic free Ca²⁺ in human neutrophils and HUVEC cells. *Biochem. Biophys. Res. Comm.*, **207**, 324-330.
48. Kanayama, N., Maehara, K., Suzuki, M., Fujise, Y. and Terao, T. (1997) The role of chondroitin sulfate chains of urinary trypsin inhibitor in inhibition of LPS-induced

- increase of cytosolic free Ca^{2+} in HL60 cells and HUVEC cells. *Biochem. Biophys. Res. Comm.*, **238**, 560-564.
49. Atmani, F., Mizon, J. and Khan, S.R. (1996) Identification of uronic-acid-rich protein as urinary bikunin, the light chain of inter- α -inhibitor. *Eur. J. Biochem.*, **236**, 984-990.
50. Imada, K., Ito, A., Kanayama, N., Terao, T. and Mori, Y. (1997) Urinary trypsin inhibitor suppresses the production of interstitial procollagenase/proMMP-1 and prostromelysin 1/ proMMP-3 in human uterine cervical fibroblasts and chorionic cells. *FEBS Letters.*, **417**, 337-340.
51. McKeehan, W.L., Sakagami, Y., Hoshi, H. and McKeehan, K.A. (1986) Two apparent human endothelial cell growth factors from human hepatoma cells are tumor-associated proteinase inhibitors. *J. Biol. Chem.*, **261**, 5378-5383.
52. Chawla, R.K., Lawson, D.H. and Travis, J. (1990) Plasma inter- α -trypsin inhibitor related urinary glycoprotein EDC1 inhibits the growth of a Burkitt's lymphoma cell line. *J. Cell. Biochem.*, **42**, 207-217.
53. Perry, J.K., Scott, K. and Tse, C.A. (1994) Modulation of proliferation of cultured human cells by urinary trypsin inhibitor. *Biochem. Biophys. Acta*, **1221**, 145-152.
54. Kawaguchi, T., Qin, L., Shimomura, T., Kondo, J., Matsumoto, K., Denda, K. and Kitamura, N. (1997) Purification and cloning of hepatocyte growth factor activator inhibitor type 2, a Kunitz-type serine protease inhibitor. *J. Biol. Chem.*, **272**, 27558-27564.
55. Ruoslahti, E. (1989) Proteoglycans in cell regulation. *J. Biol. Chem.*, **264**, 13369-13372.

56. Praillet, C., Grimaud, J.A. and Lortat-Jacob, H. (1998) Les protéoglycannes I. Molécules aux multiples fonctions... futures molécules thérapeutiques ? *Médecine/Sciences.*, **14**, 412-420.
57. Rudman, D., Chawla, R.K., Hendrickson, L.J., Vogler, W.R. and Sophianopoulos, A.J. (1976) Isolation of a novel glycoprotein (EDC1) from the urine of a patient with acute myelocytic leukemia. *Cancer. Res.*, **36**, 1837-1846.
58. Chawla, R.K., Wadsworth, A.D. and Rudman, D. (1978) Relation of the urinary cancer-related glycoprotein EDC1 to plasma inter- α -trypsin inhibitor. *J. Immunol.*, **121**, 1636-1639.
59. Kobayashi, H., Shinohara, H., Ohi, H., Sugimura, M., Terao, T. and Fujie, M. (1994) Urinary trypsin inhibitor (UTI) and fragments derived from UTI by limited proteolysis efficiently inhibit tumor cell invasion. *Clin. Exp. Metastasis.*, **12**, 117-128.
60. Kobayashi, H., Shinohara, H., Takeuchi, K., Itoh, M., Fujie, M., Saitoh, M. and Terao, T. (1994) Inhibition of the soluble and tumor cell receptor-bound plasmin by urinary trypsin inhibitor and subsequent effects on tumor cell invasion and metastasis. *Cancer. Res.*, **54**, 844-849.
61. Kobayashi, H., Hirashima, Y., Wei Sun, G., Fujie, M., Shibata, K., Tamotsu, S., Miura, K., Sugino, D., Tanaka, Y., Kondo, S. and Terao, T. (1998) Identification and characterization of the cell-associated binding protein for urinary trypsin inhibitor. *Biochim. Biophys. Acta.*, **1383**, 253-268.
62. Kobayashi, H., Gotoh, J., Fujie, M. and Terao, T. (1994) Characterization of the cellular binding site for the urinary trypsin inhibitor. *J. Biol. Chem.*, **269**, 20642-20647.
63. Kobayashi, H., Gotoh, J., Kanayama, N., Hirashima, Y., Terao, T. and Sugino, D. (1995) Inhibition of tumor cell invasion through matrigel by a peptide derived from domain II region in urinary trypsin inhibitor. *Cancer. Res.*, **55**, 1847-1852.

64. Kobayashi, H., Gotoh, J., Hirashima, Y. and Terao, T. (1996) Inter- α -trypsin inhibitor bound to tumor cells is cleaved into the heavy chains and the light chain on the cell surface. *J. Biol. Chem.*, **271**, 11362-11367.
65. Kobayashi, H., Shinohara, H., Gotoh, J., Fujie, M., Fujishiro, S. and Terao, T. (1995) Anti-metastatic therapy by urinary trypsin inhibitor in combination with an anti-cancer agent. *Br. J. Cancer.*, **72**, 1131-1137.
66. Kobayashi, H., Gotoh, J., Hirashima, M., Fujie, M., Sugino, D. and Terao, T. (1995) Inhibitory effect of a conjugate between human urokinase and urinary trypsin inhibitor on tumor cell invasion *in vitro*. *J. Biol. Chem.*, **270**, 8361-8366.
67. Kobayashi, H., Sugino, D., She, M.Y., Odi, H., Hirashima, Y., Shinihara, H., Fujie, M., Shibata, K. and Terao, T. (1998) A bifunctional hybrid molecule of the amino-terminal fragment of urokinase and domain II of bikunin efficiently inhibits tumor cell invasion and metastasis. *Eur. J. Biochem.*, **253**, 817-826.
68. Sandson, J. Hamerman, D. and Schwick, G. (1965) Altered properties of pathological hyaluronate due to a bound inter-alpha-trypsin inhibitor. *Trans. Assoc. Am. Physicians.*, **78**, 304-313.
69. Jessen, T.E., Odum, L. and Johnsen, A. H. (1994) *In vivo* binding of human inter- α -trypsin inhibitor free heavy chains to hyaluronic acid. *Biol. Chem. Hoppe-Seyler*, **375**, 521-526.
70. Yoneda, M., Suzuki, S. and Kimata, K. (1990) Hyaluronic acid associated with the surfaces of cultured fibroblasts is linked to a serum-derived 85-kDa protein. *J. Biol. Chem.*, **265**, 5247-5257.
71. Chen, L., Mao, S.J.T. and Larsen, W.J. (1992) Identification of a factor in fetal bovine serum that stabilizes the cumulus extracellular matrix : a role for a member of the inter- α -trypsin inhibitor family. *J. Biol. Chem.*, **267**, 12380-12386.

72. Chen, L., Mao, S.J.T., McLean, L.R., Powers, R.W. and Larsen, W.J. (1994) Proteins of the inter- α -trypsin inhibitor family stabilize the cumulus extracellular matrix through their direct binding with hyaluronic acid. *J. Biol. Chem.*, **269**, 28282-28287.
73. Camaioni, A., Hascall, V.C., Yanagishita, M. and Salustri, A. (1993) Effects of exogenous hyaluronic acid and serum on matrix organization and stability in the mouse cumulus cell-oocyte complex. *J. Biol. Chem.*, **268**, 20473-20481.
74. Huang, L., Yoneda, M. and Kimata, K. (1993) A serum-derived hyaluronan-associated protein (SHAP) is the heavy chain of the inter- α -trypsin inhibitor. *J. Biol. Chem.*, **268**, 26725-26730.
75. Zhao, M., Yoneda, M., Ohashi, Y., Kurono, S., Iwata, H., Ohnuki, Y. and Kimata, K. (1995) Evidence for the covalent binding of SHAP, heavy chains of inter- α -trypsin inhibitor, to hyaluronan. *J. Biol. Chem.*, **270**, 26657-26663.
76. Chen, L., Zhang, H., Powers, R.W., Russell, P.T. and Larsen, W.J. (1996) Covalent linkage between proteins of the inter- α -inhibitor family and hyaluronic acid is mediated by a factor produced by granulosa cells. *J. Biol. Chem.*, **271**, 19409-19414.
77. Thomas, G.J., Yung, S. and Davies, M. (1998) Bikunin present in human peritoneal fluid is part derived from the interaction of serum with peritoneal mesothelial cells. *Am. J. Pathol.*, **153**, 1267-1276.
78. Blom, A., Pertoft, H. and Fries, E. (1995) Inter- α -inhibitor is required for the formation of the hyaluronan-containing coat on fibroblasts and mesothelial cells. *J. Biol. Chem.*, **270**, 9698-9701.
79. Bost, F., Diarra-Mehrpour, M. and Martin, J.P. (1998) Inter- α -trypsin inhibitor proteoglycan family : a group of proteins binding and stabilizing the extracellular matrix. *Eur. J. Biochem.*, **252**, 339-346.

80. Wisniewski, H.G., Burgess, W.H., Oppenheim, J.D. and Vilcek, J. (1994) TSG-6, an arthristis-associated hyaluronan binding protein, forms a stable complex with the serum protein inter- α -inhibitor. *Biochemistry*, **33**, 7423-7429.
81. Wisniewski, H.G., Hua, J.C., Poppers, D.M., Naime, D., Vilcek, J. and Cronstein, B.N. (1996) TNF/IL-1-inducible protein TSG-6 potentiates plasmin inhibition by inter- α -inhibitor and exerts a strong anti-inflammatory effect *in vivo*. *J. Immunol.*, **156**, 1609-1615.
82. Eriksen, G.V., Carlstedt, I., Mörgelin, M., Uldbjerg, N. and Malmström, A. (1999) Isolation and characterization of proteoglycans from human follicular fluid. *Biochem. J.*, **340**, 613-620.
83. Hascall, V.C., Calabro, A., Midura, R.J. and Yanagishita, M. (1994) Isolation and characterization of proteoglycans. *Meth. Enzy.*, **230**, 390-417.
84. Kjellén, L. and Lindahl, U. (1991) Proteoglycans : structures and interactions. *Annu. Rev. Biochem.*, **60**, 443-475.
85. Petitou, M., Lormeau, J.C. and Choay, J. (1988) Interaction of heparin and antithrombin III. The role of O-sulfate groups. *Eur. J. Biochem.*, **176**, 637-640.
86. Bourdon, M.A., Krusius, T., Campbell, S., Schwartz, N.B. and Ruoslahti, E. (1987) Identification and synthesis of a recognition signal for the attachment of glycosaminoglycans to proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 3194-3198.
87. Levesque, H., Chassagne, P., Frebourg, T. and Delpech, B. (1990) Acide hyaluronique : intérêts et perspectives de son dosage sérique. *Presse Méd.*, **19**, 1001-1005.
88. Hardingham, T.E. and Fosang, A.J. (1992) Proteoglycans : many forms and many functions. *FASEB J.*, **6**, 861-870.

89. Giry-Loziquez, C., Kleman, J.P. and van der Rest, M. (1994) Modules et interactions moléculaires au sein des matrices extracellulaires. *Médecine/Sciences.*, **10**, 1234-1243.
90. Wight, T.N., Kinsella, M.G. and Qwarnström, E.E. (1992) The role of proteoglycans in cell adhesion, migration and proliferation. *Curr. Opin. Cell Biol.*, **4**, 793-801.
91. Esko, J.D. (1991) Genetic analysis of proteoglycan structure, function and metabolism. *Curr. Opin. Cell Biol.*, **3**, 805-816.
92. Chirat, F., Balduyck, M., Mizon, C., Laroui, S., Sautière, P. and Mizon, J. (1991) A chondroitin-sulfate chain is located on serine-10 of the urinary trypsin inhibitor. *Int. J. Biochem.*, **23**, 1201-1203.
93. Morelle, W., Capon, C., Balduyck, M., Sautière, P., Kouach, M., Michalski, C., Fournet, B. and Mizon, J. (1994) Chondroitin sulphate covalently cross-links the three polypeptide chains of inter- α -trypsin inhibitor. *Eur. J. Biochem.*, **221**, 881-888.
94. Yamada, S., Oyama, M., Yuki, Y., Kato, K. and Sugahara, K. (1995) The uniform galactose 4-sulfate structure in the carbohydrate-protein linkage region of human urinary trypsin inhibitor. *Eur. J. Biochem.*, **233**, 687-693.
95. Yamada, S., Oyama, M., Kinugasa, H., Nakagawa, T., Kawasaki, T., Nagasawa, S., Khoo, K.H., Morris, H.R., Dell, A. and Sugahara, K. (1995) The sulphated carbohydrate-protein linkage region isolated from chondroitin 4-sulphate chains of inter- α -trypsin inhibitor in human plasma. *Glycobiology*, **5**, 335-341.
96. Enghild, J.J., Thogersen, I.B., Cheng, F., Fransson, L.A., Roepstorff, P. and Rahbek-Nielsen, H. (1999) Organization of the inter- α -inhibitor heavy chains on the chondroitin sulfate originating from Ser10 of bikunin : posttranslational modification of I α I-derived bikunin. *Biochemistry*, **38**, 11804-11813.

97. Toyada, H., Kobayashi, S., Sakamoto, S., Toida, T. and Imanari, T. (1993) Structural analysis of low-sulfurated chondroitin sulfate chain in human urinary trypsin inhibitor. *Biol. Pharm. Bull. (Tokyo)*, **16**, 945-947.
98. Sjöberg, E.M., Blom, A., Larsson, B.S., Alston-Smith, J., Sjöquist, M. and Fries, E. (1995) Plasma clearance of rat bikunin : evidence for receptor-mediated uptake. *Biochem. J.*, **308**, 881-887.
99. Diarra-Mehrpour, M., Bourguignon, J., Sesboüé, R., Mattei, M.G., Passage, E., Salier, J.P. and Martin, J.P. (1989) Human plasma inter- α -trypsin inhibitor is encoded by four genes on three chromosomes. *Eur. J. Biochem.*, **179**, 147-154.
100. Bost, F., Bourguignon, J., Martin, J.P., Sesboüé, R., Thiberville, L. and Diarra-Mehrpour, M. (1993) Isolation and characterization of the human inter- α -inhibitor heavy chain H1 gene. *Eur. J. Biochem.*, **218**, 283-291.
101. Diarra-Mehrpour, M., Sarafan, N., Bourguignon, J., Bonnet, F., Bost, F. and Martin, J.P. (1998) Human inter- α -trypsin inhibitor heavy chain H3 gene : genomic organization, promoter analysis and gene linkage. *J. Biol. Chem.*, **273**, 26809-26819.
102. Diarra-Mehrpour, M., Bourguignon, J., Sesboüé, R., Salier, J.P., Lévillard, T. and Martin, J.P. (1990) Structural analysis of the human inter- α -trypsin inhibitor light-chain gene. *Eur. J. Biochem.*, **191**, 131-139.
103. Tavakkol, A. (1991) Molecular cloning of porcine α 1-microglobulin/HI-30 reveals developmental and tissue-specific expression of two variant messenger ribonucleic acids. *Biochim. Biophys. Acta.*, **1088**, 47-56.
104. Lindquist, A. and Akerström, B. (1992) Bovine α 1-microglobulin/HI-30. Isolation and characterization of liver cDNA and urinary α 1-microglobulin. *Biochim. Biophys. Acta.*, **1306**, 98-106.

105. Kaumeyer, J.F., Polazzi, J.O. and Kotick., M.P. (1996) The mRNA for a proteinase inhibitor related to the HI-30 domain of inter-alpha-trypsin inhibitor also encodes alpha-1-microglobulin (Protein HC). *Nucleic Acids Res.*, **14**, 7839-7850.
106. Thuveson, M. and Fries, E. (1999) Intracellular proteolytic processing of the heavy chain of rat pre- α -inhibitor. *J. Biol. Chem.*, **274**, 6741-6746.
107. Bourguignon, J., Sesboué, R., Diarra-Mehrpour, M., Daveau, M. and Martin, J.P. (1989) Human inter- α -trypsin inhibitor : synthesis and maturation in hepatoma HepG2 cells. *Biochem. J.*, **261**, 305-308.
108. Odum, L. (1992) Biosynthesis of inter- α -trypsin inhibitor and α 1-microglobulin in a human hepatoma cell line. *Int. J. Biochem.*, **24**, 215-222.
109. Sjöberg, E.M and Fries, E. (1992) Biosynthesis of bikunin (urinary trypsin inhibitor) in rat hepatocytes. *Arch. Biochem. Biophys.*, **295**, 217-222.
110. Thogersen, I.B. and Enghild, J.J. (1995) Biosynthesis of bikunin proteins in the human carcinoma cell line HepG2 and in primary human hepatocytes : polypeptide assembly by glycosaminoglycan. *J. Biol. Chem.*, **270**, 18700-18709.
111. Héron, A., Bourguignon, J., Callé, A., Borghi, H., Sesboué, R., Diarra-Mehrpour, M. and Martin, J.P. (1994) Post-translational processing of the inter- α -trypsin inhibitor in the human hepatoma HepG2 cell line. *Biochem. J.*, **302**, 573-580.
112. Daveau, M., Jean, L., Soury, E., Masson, S., Lyoumi, S., Chan, P., Hiron, M., Lebreton, J.P., Husson, A., Jegou, S., Vaudry, H. and Salier, J.P. (1998) Hepatic and extra-hepatic transcription of inter- α -inhibitor family genes under normal and acute inflammatory conditions in rat. *Arch. Biochem. Biophys.*, **350**, 315-323.
113. Mizushima, S., Nii, A., Kato, K. and Uemura, A. (1998) Gene expression of the two heavy chains and one light chain forming the inter-alpha-trypsin inhibitor in human tissues. *Biol. Pharm. Bull.*, **21**, 167-169.

114. Daveau, M., Rouet, P., Scotte, M., Faye, L., Hiron, M., Lebreton, J.P. and Salier, J.P. (1993) Human inter- α -inhibitor family in inflammation : simultaneous synthesis of positive and negative acute-phase proteins. *Biochem. J.*, **292**, 485-492.
115. Martin-Vandeleet, N., Paris, S., Bourguignon, J., Sesboüé, R., Martin, J.P. and Diarra-Mehrpour, M. (1999) Assembly and secretion of recombinant chains of human inter- α -trypsin inhibitor in COS-7 cells. *Eur. J. Biochem.*, **259**, 476-484.
116. Hochstrasser, K. and Wachter, E. (1979) Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter- α -trypsin inhibitor, I. Determination of the amino acid sequence of the antitryptic domain by solid-phase Edman degradation. *Hoppe-Seyler's Z. Physiol. Chem.*, **360**, 1285-1296.
117. Xu, Y., Carr, P.D., Guss, J.M. and Ollis, D.L. (1998) The crystal structure of bikunin from the inter- α -trypsin complex : a serine protease inhibitor with two Kunitz domains. *J. Mol. Biol.*, **276**, 955-966.
118. Hochstrasser, K., Schönberger, O.L., Rossmann, I. and Wachter, E. (1981) Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter- α -trypsin inhibitor, V. Attachment of carbohydrates in the human urinary trypsin inhibitor isolated by affinity chromatography. *Hoppe-Seyler's Z. Physiol. Chem.*, **362**, 1357-1362.
119. Toyoda, H., Ikei, T., Demachi, Y., Toida, T. and Imanari, T. (1992) Structural analysis of the N-linked oligosaccharides from human urinary trypsin inhibitor. *Chem. Pharm. Bull. (Tokyo)*, **40**, 2882-2884.
120. Wachter, E. and Hochstrasser, K. (1981) Kunitz-type proteinase inhibitors derived by limited proteolysis of the inter- α -trypsin inhibitor, IV. The amino acid sequence of the human urinary trypsin inhibitor isolated by affinity chromatography. *Hoppe-Seyler's Z. Physiol. Chem.*, **362**, 1351-1355.

121. Selloum, L., Davril, M., Mizon, C., Balduyck, M. and Mizon, J. (1997) The effect of the glycosaminoglycan chain removal on some properties of the human urinary trypsin inhibitor. *Biol. Chem. Hoppe-Seyler*, **368**, 47-55.
122. Malki, N., Balduyck, M., Maes, P., Capon, C., Mizon, C., Han, K.K., Tatar, A., Fournet, B. and Mizon, J. (1992) The heavy chains of human plasma inter- α -trypsin inhibitor: their isolation, their identification by electrophoresis and partial sequencing. Differential reactivity with Concanavalin A. *Biol. Chem. Hoppe-Seyler*, **373**, 1009-1018.
123. Gebhard, W., Schreitmuller, T., Hochstrasser, K. and Wachter, E. (1989) Two out of the three kinds of subunits of inter- α -trypsin inhibitor are structurally related. *Eur. J. Biochem.*, **181**, 571-576.
124. Diarra-Mehrpour, M., Bourguignon, J., Bost, F., Sesboué, R., Muschio, F., Sarafan, N. and Martin, J.P. (1992) Human inter- α -trypsin inhibitor: full-length cDNA sequence of the heavy chain H1. *Biochim. Biophys. Acta.*, **1132**, 114-118.
125. Gebhard, W., Schreitmuller, T., Hochstrasser, K. and Wachter, E. (1988) Complementary cDNA and derived amino acid sequence of the precursor of one of the three protein components of the inter- α -trypsin inhibitor complex. *FEBS Lett.*, **229**, 63-67.
126. Bourguignon, J., Diarra-Mehrpour, M., Thiberville, L., Bost, F., Sesboué, R. and Martin, J.P. (1993) Human pre- α -trypsin inhibitor-precursor heavy chain cDNA and deduced amino acid sequence. *Eur. J. Biochem.*, **212**, 771-776.
127. Chan, P., Risler, J.L., Raguenez, G. and Salier, J.P. (1995) The three heavy chain precursors for the inter- α -inhibitor family in mouse: new members of the multicopper oxidase protein group with differential transcription in liver and brain. *Biochem. J.*, **306**, 505-512.

128. Saguchi, K., Tobe, T., Hashimoto, K., Sano, Y., Nakano, Y, Miura, N. H. and Tomita, M. (1995) Cloning and characterization of cDNA for inter-alpha-trypsin inhibitor family heavy chain-related protein (IHRP), a novel human plasma glycoprotein. *J. Biochem (Tokyo)*, **117**, 14-18.
129. Kornfeld, R. and Kornfeld, S. (1985) Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.*, **54**, 631-664.
130. Chatterjee, B., Vaith, P., Chatterjee, S., Karduck, D. and Uhlenbruck, G. (1979) Comparative studies of new marker lectins for alkali-labile and alkali-stable carbohydrate chains in glycoproteins. *Int. J. Biochem.*, **10**, 321-327.
131. Balduyck, M., Piva, F., Mizon, C., Maes, P., Malki, N., Gressier, B., Michalski, C. and Mizon, J. (1993) Human leucocyte elastase (HLE) preferentially cleaves the heavy chain H2 of inter- α -trypsin inhibitor (ITI). *Biol. Chem. Hoppe-Seyler*, **374**, 895-901.
132. Lis, N. and Sharon, N. (1993) Protein glycosylation : structural and functional aspects. *Eur. J. Biochem.*, **218**, 1-27.
133. Gerken, T.A., Owens, C.L. and Pasumarthy, M. (1997) Effects on O-glycosylation on protein structure : role and characterization of heavily O-glycosylation mucin domains. Townsend, R. R. and Hotchkiss, A. T. (eds), in *Techniques in glycobiology*, Marcel Dekker, inc., 270 Madison avenue, New York, USA, pp. 247-269.
134. Hounsell, E.F., Davies, M.J. and Renouf, D.V. (1996) O-linked protein glycosylation structure and function. *Glycoconjugate J.*, **13**, 19-26.
135. Michalski, C., Piva, F., Balduyck, M., Mizon, C., Burnouf, T., Huart, J.J. and Mizon, J. (1994) Preparation and properties of a therapeutic inter- α -trypsin inhibitor concentrate from human plasma. *Vox. Sang.*, **67**, 329-336.

136. Olsen, E.H.N., Rahbek-Nielsen, H., Thogersen, I.B., Roepstorff, P. and Engild, J.J. (1998) Posttranslational modifications of human inter- α -inhibitor : identification of glycans and disulfide bridges in heavy chains 1 and 2. *Biochemistry*, **37**, 408-416.
137. Blom, A.M., Mörgelin, M., Öyen, M., Jarvet, J. and Fries, E. (1999) Structural characterization of inter- α -inhibitor : evidence of an extended shape. *J. Biol. Chem.*, **274**, 298-304.
138. Opdenakker, G., Rudd, P.M., Ponting, C.P. and Dwek, R.A. (1993) Concepts and principles of glycobiology. *FASEB J.*, **7**, 1330-1337.
139. Ashwell, G. and Harford, J. (1982) Carbohydrate-specific receptors of the liver. *Annu. Rev. Biochem.*, **51**, 531-554.
140. Kornfeld, S. (1992) Structure and function of the mannose 6-phosphate/insulinlike growth factor II receptors. *Annu. Rev. Biochem.*, **61**, 307-330.
141. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST : a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389-3402.
142. Elhammer, A.P., Poorman, R.A., Brown, E., Maggiora, L.L., Hoogerheide, J.G. and Kézdy, F.J. (1993) The specificity of UDP-GalNAc : polypeptide N-acetylgalactosaminyltransferase as inferred from a database of *in vivo* substrates and from *in vitro* glycosylation of proteins and peptides. *J. Biol. Chem.*, **268**, 10029-10038.
143. Elhammer, A.P., Poorman, R.A. and Kézdy, F.J. (1997) Predicting mucin-type O-glycosylation sites. Townsend, R. R. and Hotchkiss, A. T. (eds), in *Techniques in glycobiology*, Marcel Dekker, inc., 270 Madison avenue, New York, USA, pp. 271-297.

144. Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L. and Brunak, S. (1998) NetOGlyc : prediction of mucin-type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate J.*, **15**, 115-130.
145. Wilharm, E., Parry, M.A., Friebel, R., Tschesche, H., Matschiner, G., Sommerhoff, C.P. and Jenne, D.E. (1999) Generation of catalytically active granzyme K from *Escherichia coli* inclusion bodies and identification of efficient granzyme K inhibitors in human plasma. *J. Biol. Chem.*, **274**, 27331-27337.

