

Les coudes β dans les protéines

par Marie-Henriette Loucheux-Lefebvre

(Institut de Recherches sur le Cancer de Lille, BP 3567, 59020 Lille Cedex)

I. Introduction



Il est bien connu que les structures qui, dans les protéines, jouent un rôle essentiel dans leur stabilisation, sont l'hélice α et la structure en chaîne β ; pendant longtemps elles ont d'ailleurs été considérées comme les deux seules structures organisées pouvant exister dans une protéine. En d'autres termes, les parties de la macromolécule qui n'adoptaient aucune de ces deux conformations étaient considérées comme ne présentant aucune structure régulière, comme « inordonnées ».

Mais cette conception a dû être modifiée à la suite des résultats obtenus par les études aux rayons X, qui, ces vingt dernières années, ont permis de connaître la structure tridimensionnelle d'un nombre important de protéines. A ce jour, on compte environ 40 à 50 protéines dont les structures primaire et tridimensionnelle sont parfaitement élucidées, et ce nombre ne cesse d'augmenter. Dès 1971, ces études par rayons X avaient permis à Lewis et ses collaborateurs (1) de mettre en évidence, au sein de ces zones dites « désordonnées », des séquences de quatre résidus d'acides aminés successifs adoptant une conformation telle qu'ils permettaient à la chaîne macromoléculaire de se replier sur elle-même, de « tourner » d'environ 180° . Cette structure appelée coude β fut ensuite confirmée par de nombreux auteurs étudiant un nombre de plus en plus important de protéines (2-5) qui ont montré que ces térapeptides en coude β représentaient une fraction aussi importante que l'hélice α ou que la structure β puisque, à partir de 29 protéines (5), il a pu être montré que les proportions de résidus d'acide aminé adoptant une structure en hélice α , structure β et coudes β étaient respectivement de 38 %, 20 % et 32 % (5).

On peut donc envisager que ce sont ces coudes β qui assurent à la protéine sa forme globulaire et pendant longtemps on a pensé que tel était leur rôle essentiel; les différents termes utilisés alors dans la littérature anglo-saxonne sont, de ce point de vue, significatifs : « β -bends » pour Lewis et ses collaborateurs (1, 2), « hairpin-loops » pour Kuntz (3), « reverse-turns » pour Crawford et ses collaborateurs (4), « 3_{10} bends » pour Dickerson et ses collaborateurs (6).

Mais on s'est aperçu au cours du temps que leur rôle ne se bornait pas à cela et le but de cette mise au point est justement d'essayer de démontrer leur importance dans la structuration fonctionnelle des protéines.

La mise en évidence de l'importance des coudes β s'est faite parallèlement à l'utilisation des méthodes de prédiction de conformation des protéines. Parmi les nombreuses méthodes qui ont été publiées sur ce sujet au cours des dernières années, et dont on trouvera une revue dans la référence 7, nous ne citerons que celle de

Chou et Fasman (8) qui reste incontestablement une des plus utilisées à la fois à cause de sa simplicité de mise en œuvre que de la bonne précision qu'elle permet d'obtenir. Sans entrer dans le détail des règles énoncées par ces auteurs, rappelons-en simplement les bases. Elles ont été établies à partir d'un certain nombre de protéines dont la structure primaire est connue ainsi que la conformation tridimensionnelle par l'étude aux rayons X. Une étude statistique permet alors de déterminer l'aptitude que présente chacun des 20 acides aminés naturels à adopter telle ou telle conformation ; des paramètres structuraux peuvent ainsi être calculés : on définira donc pour un acide aminé donné les paramètres structuraux P_α et P_β correspondant aux structures α et β (avec $P_\alpha = f_\alpha / \langle f_\alpha \rangle$ et $P_\beta = f_\beta / \langle f_\beta \rangle$ dans lesquels f_α et f_β représentent le nombre de fois que le résidu considéré est présent dans les séquences en hélice α ou en structure β et où $\langle f_\alpha \rangle$ et $\langle f_\beta \rangle$ représentent le nombre moyen d'acides aminés constituant ces deux sortes de séquences). Dans le cas de prédiction des coudes β , les choses sont un peu plus délicates puisque la position, sur le coude, du résidu de l'acide aminé considéré, intervient elle aussi. Pour cette raison, à côté d'un paramètre conformationnel P_i , défini comme les P_α et P_β précédents, il faut tenir compte d'une probabilité

$$P_i = (f_i \times f_{i+1} \times f_{i+2} \times f_{i+3})$$

dans laquelle $f_i, f_{i+1}, f_{i+2}, f_{i+3}$ représentent les fréquences d'existence des différents résidus soit en position 1 du coude (i), ou 2 ($i+1$), ou 3 ($i+2$), ou 4 ($i+3$). Si on a $P_i > 1 \times 10^{-4}$, un coude β en position i pourra être prédit.

II. Définition d'un coude β

Pour définir un coude β , il faut d'abord considérer que, pratiquement, la liaison peptidique est plane. La chaîne polypeptidique n'a que deux degrés de liberté par résidu : la rotation autour de la liaison $C_\alpha-N$, définie par l'angle Φ , et celle autour de la liaison $C_\alpha-C'$, définie par l'angle Ψ (Figure 1). Dans cette nomenclature, C_α définit le carbone asymétrique, C' le carbone sur lequel l'oxygène est attaché. Pour définir ces angles diédraux Φ et Ψ , la convention proposée par la commission de nomenclature en biochimie de l'IUPAC-IUB en 1970 a été adoptée (9) : dans ce cas, la chaîne totalement étendue correspond à $\Phi = \Psi = 180^\circ$. Pour des structures comme l'hélice α ou la structure β , les valeurs des angles Φ et Ψ sont identiques pour tous les résidus situés dans la séquence qui adopte cette conformation : pour l'hélice α , $(\Phi, \Psi) \simeq -60^\circ, -50^\circ$; pour la structure β parallèle, $(\Phi, \Psi) \simeq -120^\circ, 110^\circ$; pour la structure β anti-parallèle, $(\Phi, \Psi) \simeq -140^\circ, 135^\circ$. Par contre, pour les coudes β , il n'existe pas une structure unique, mais plusieurs qui toutes permettent à la chaîne macromoléculaire, par l'intermédiaire du térapeptide qui adopte cette structure en coude β , de se replier sur elle-même. Deux d'entre eux sont représentés sur la figure 2. Depuis les travaux de Venkatachalam (10) qui, le premier, a donné les caractéristiques de trois coudes β , il a été possible, à partir de données cristallographiques d'en identifier onze (5) qui sont rapportés dans le tableau 1. Comme le montre la figure 2, ce sont les angles diédraux Φ, Ψ correspondant aux résidus en position 2 ($i+1$) et 3 ($i+2$) qui

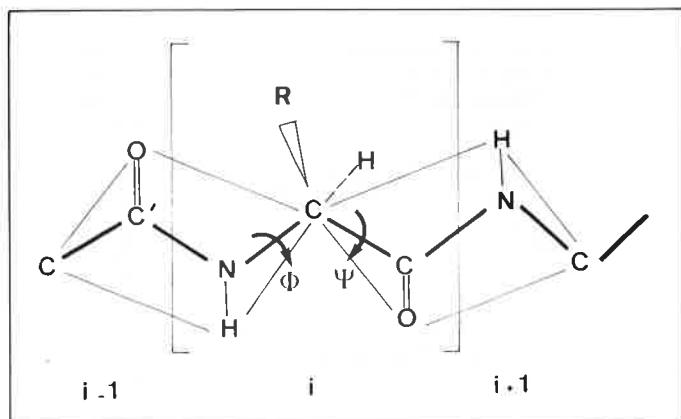


Figure 1. Définition des angles Φ et Ψ .

Il a ainsi été possible de montrer que certains résidus présentaient une préférence plus grande d'occuper telle position dans le coude plutôt que telle autre. C'est le cas, par exemple, de la proline dont la probabilité élevée de se trouver en position 2 du coude (30 %) tombe à 3 % quand il s'agit de la position 3. D'un point de vue plus général, ce sont les résidus d'asparagine, de glycine, de sérine et de proline qui présentent, quelle que soit leur position, la probabilité la plus élevée de se trouver dans un coude β . Par contre, les résidus hydrophobes y seront rarement observés. Enfin, il est intéressant également de considérer la nature des résidus d'acides aminés qui se trouvent dans le voisinage des coudes β . Cette étude a été réalisée par Chou et Fasman (5) qui constatèrent que dans les régions adjacentes aux coudes β on trouve un grand nombre de résidus hydrophobes, c'est-à-dire des résidus favorisant la formation soit d'hélice α , soit de structure β qui sont alors disposées de part et d'autre du coude β . Cette disposition privilégiée favorise des interactions soit hélice α -hélice α , soit structure β -hélice α , aussi bien que la formation de structure β anti-parallèle, qui stabilisent le coude β .

Le rôle essentiel que jouent les coudes β dans les protéines sera illustré par les exemples suivants :

- la conservation des coudes β au cours de l'évolution des protéines,
- la phosphorylation des protéines,
- les glycoprotéines,
- les caséines κ , qui présentent à la fois des sites de phosphorylation et des sites de glycosylation et qui sont connues pour différentes espèces, seront plus spécialement étudiées.

Tableau 1. Classification des différents types de coude suivant la valeur des angles diédraux $(\Phi, \Psi)_2$ et $(\Phi, \Psi)_3$ des deux résidus centraux $i+1$ et $i+2$ (d'après la référence 11).

Type de coude	Angles diédraux en degrés			
	Φ_2	Ψ_2	Φ_3	Ψ_3
I	-60	-30	-90	0
I'	60	30	90	0
II	-60	120	80	0
II'	60	-120	-80	0
III	-60	-30	-60	-30
III'	60	30	60	30
IV	Au moins deux des quatre angles $\Phi\Psi$ considérés diffèrent de plus de 40° des valeurs données ci-dessus.			
V	-80	80	80	-80
V'	80	-80	-80	80
VI	Coude caractérisé par la présence d'un résidu Pro de configuration <i>cis</i> en position 3.			
VII	Un coude est créé le long de la chaîne protéique pour $\Psi_2 \simeq 180^\circ$ et $ \Phi_3 < 60^\circ$ ou $ \Psi_2 < 60^\circ$ et $\Phi_3 \simeq 180^\circ$.			

permettront au térapeptide d'adopter une structure telle que la chaîne se repliera sur elle-même. Ce sont donc ces deux séries de valeurs d'angles qui sont données dans le tableau 1. Dans la nomenclature utilisée, le signe « prime », comme dans I', II', III' et V', signifie que les coudes correspondants sont respectivement l'image dans un miroir des coudes I, II, III, V. Cela correspond à des valeurs des angles $(\Phi, \Psi)_2$ et $(\Phi, \Psi)_3$ égales, au signe près. Les coudes I, II et III, donc aussi I', II' et III', avaient été identifiés par Venkatachalam (10). Ceux de type V et V' ont été par Lewis et ses collaborateurs (2). Les résultats cristallographiques sur un certain nombre de protéines montrent que ce sont les coudes de type I qui sont rencontrés le plus fréquemment dans les protéines, puis ceux de type III, et enfin ceux de type II. Ceux de type I', II' et III' ne se rencontrent pas souvent et

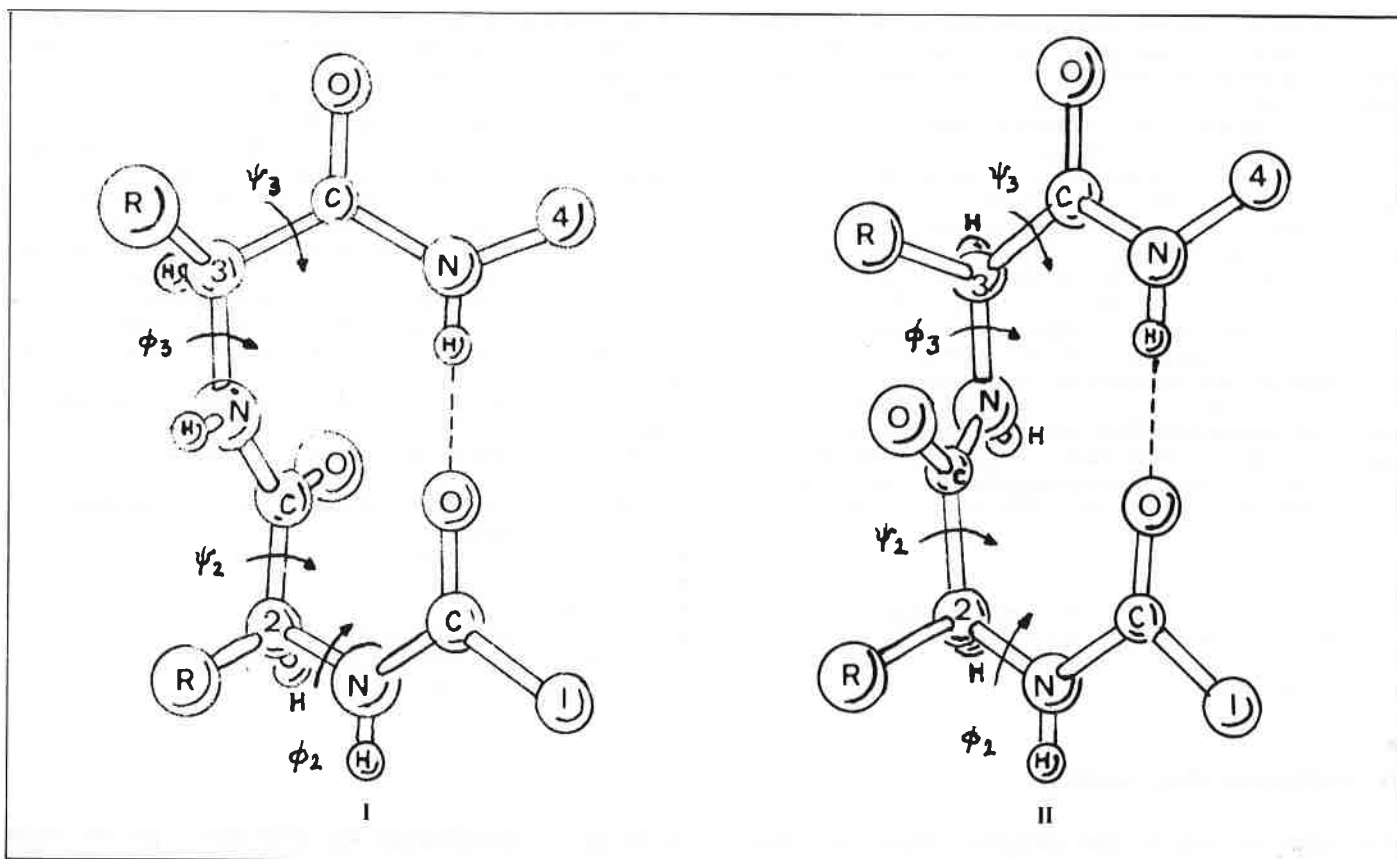


Figure 2. Représentation schématique de deux coudes β .

D'après Dickerson *et al.* (6).

(Les atomes de carbone α sont numérotés 1, 2, 3, 4). Celui de gauche correspond à un coude β de type I avec $(\Phi, \Psi)_2 = (-60^\circ, -30^\circ)$ et $(\Phi, \Psi)_3 = (-90^\circ, 0^\circ)$ et celui de droite à un coude β de type II avec $(\Phi, \Psi)_2 = (-60^\circ, 120^\circ)$ et $(\Phi, \Psi)_3 = (80^\circ, 0^\circ)$.

Tableau 2. Sites de phosphorylation.

La sérine ou la thréonine phosphorylée est indiquée par un astérisque. Les quatre résidus d'acide aminé qui adoptent une structure en coude β sont soulignés. Le résidu basique au voisinage du résidu phosphorylé est écrit en italique. Certains sites de phosphorylation ont été identifiés soit *in vivo* soit *in vitro* et sont indiqués dans le tableau. Quand aucune mention spéciale n'est donnée, il s'agit de sites « probables ».

Phosphoprotéine	Localisation	Séquence	Espèce	Référence
Histone H1	37 ou 38	<i>Lys-Ala-Ser*-Gly-Pro</i>	Foie de Rat <i>in vivo</i> Foie de Veau <i>in vitro</i> Cerveille de Porc <i>in vitro</i> Lapin	34, 35 34, 35 34, 35 36
	105	<i>Ala-Ser-Gly-Ser*-Phe-Lys</i>	Foie de Veau <i>in vitro</i> Foie de Rat <i>in vitro</i> Thymus de Veau <i>in vitro</i> Truite	37, 38 37, 38 37, 38 39
Histone H2A (numérotation : Veau; alignement : cf. réf. 9)	1	<i>Ser*-Gly-Arg-Gly</i>	Testicules de Truite <i>in vitro</i> Veau Oursin Rat Truite Poulet	40 41 42 43 44
	19	<i>Thr-Arg-Ser-Ser*</i>	Cerveille de Porc <i>in vitro</i> Veau Rat Truite	34 41 43 44

Phosphoprotéine	Localisation	Séquence	Espèce	Référence
Histone H2B (numérotation : Veau ; alignements : cf. réf. 17 et 18)	19	<u>Ser-Arg-Ser-Ser*</u>	Oursin Poulet	42 45
	14	<u>Lys-Lys-Gly-Ser*</u>	Cervelle de Porc <i>in vitro</i> Veau Truite	34, 35 46 47
	36	<u>Lys-Glu-Ser-Tyr-Ser</u>	Veau	46
	36	<u>Lys-Glu-Ser-Thr-Ser</u>	Cervelle de Porc <i>in vitro</i> Lymphocytes humains <i>in vitro</i>	34 48, 49
		<u>Lys-Glu-Ser-Tyr-Ala</u> <u>Lys-Glu-Ser-Tyr-Gly</u>	Truite Oursin	47 50, 51
	6	<u>Pro-Ala-Lys-Ser-Ala-Pro-Lys</u>	Testicules de Truite <i>in vitro</i> Veau Truite	52 46 47
	32	<u>Arg-Ser-Arg-Lys</u>	Veau Truite	46 47
	Histone H4	1	<u>Ser-Gly-Arg-Gly*</u>	Testicules de Truites <i>in vivo</i>
Histone H5	45	<u>Ser-Ser-Arg-Gln</u>	Pigeon <i>in vitro</i> Poulet Oie	53 54 55
Phosphorylase b	91	<u>Gly-Ala-Gly-Ser-Ser-Phe-Arg</u>	Pigeon <i>in vitro</i>	53
	14	<u>Ser-Val-Arg-Gly</u>	Muscle de Lapin <i>in vitro</i> Muscle humain <i>in vitro</i>	56, 57, 58 56, 57, 58
Protéine basique de la myéline	12	<u>His-Gly-Ser-Lys</u>	Muscle cardiaque de Bœuf <i>in vitro</i>	59
	110	<u>Gly-Arg-Gly-Leu-Ser-Leu-Ser-Arg</u>	Muscle de Lapin <i>in vitro</i>	60, 61
	55	<u>Arg-Gly-Ser-Gly-Lys</u>	Muscle de Lapin <i>in vitro</i>	62
Phosphorylase ki- nase sous-unité α sous-unité β		<u>Arg-Leu-Ser-Ile-Ser</u>	Muscle de Lapin <i>in vitro</i>	63
		<u>Arg-Ser-Gly-Ser*</u>		64
Troponine 1		<u>Ala-Ile-Thr-Ala-Arg-Arg</u>	Muscle de Lapin <i>in vitro</i>	65, 66
		<u>Val-Lys-Ser-Ser*</u>	Muscle de Lapin <i>in vitro</i>	65, 66
		<u>Val-Arg-Met-Ser*</u>	Muscle de Lapin <i>in vitro</i>	65, 66
Protamine I	6, 7, 8	<u>Ser-Ser-Ser-Arg</u>	Sperme de Truite <i>in vitro</i>	67
	21	<u>Ser-Arg-Arg-Arg</u>	Sperme de Truite <i>in vitro</i>	67
Inhibiteur de la phosphatase		<u>Arg-Pro-Thr-Pro</u>		68
Caséine κ _A	149	<u>Ser-Pro-Glu-Val</u>	Vache	69, 70
	151	<u>Ser-Ser-Glu-Ser</u>	Chèvre	71

ceux de type V et V' encore plus rarement. Lewis et ses collaborateurs (2) ont également mis en évidence d'autres tétrapeptides permettant à la chaîne polypeptidique de se replier sur elle-même, désignés par IV,

VI et VII. Parmi eux, les coudes de type VI et VII ont une probabilité quasi nulle de se rencontrer sur les protéines, alors que celui de type IV l'est assez souvent.

III. Conservation des coudes β dans les protéines

Nous avons vu précédemment que l'identification des coudes β dans les protéines avait été possible grâce aux études cristallographiques qui ont permis la détermination de la structure tridimensionnelle

d'un certain nombre d'entre elles. Il était donc intéressant de voir ce que devenaient ces coudes β pour une même protéine au cours de l'évolution. Une telle étude peut être facilement envisagée en utilisant

Inhibiteur de la protéase du pancréas de Boeuf	Arg Pro Asp Phe	Cys	Leu Glu Pro Pro Tyr Thr	Gly Pro	Cys	Lys Ala Arg Ile Ile Arg Tyr
Inhibiteur de la protéase du colostrum de Vache	Phe Gln Thr Pro Pro Asp Leu	Cys	Gln Leu Pro Gln Ala Arg	Gly Pro	Cys	Lys Ala Ala Leu Leu Arg Tyr
Inhibiteur de la protéase du blanc d'oeuf de Tortue	(Lys, Glx, Asx, Gly, Arg) Asp Ile	Cys	Arg Leu Pro Pro Glu Gln	Gly Pro	Cys	Lys Gly Arg Ile Pro Arg Tyr
Inhibiteur K de la protéase d'Escargot	Glu Gly Arg Pro Ser Phe	Cys	Asn Leu Pro Ala Glu Thr	Gly Pro	Cys	Lys Ala Ser Phe Arg Gln Tyr
Toxine de la Vipère de Russell	His Asp Arg Pro Thr Phe	Cys	Asn Leu Ala Pro Glu Ser	Gly Arg	Cys	Arg Gly His Leu Arg Arg Ile
Toxine K du Mamba noir	Ala Ala Lys Tyr	Cys	Lys Leu Pro Leu Arg Ile	Gly Pro	Cys	Lys Arg Lys Ile Pro Ser Phe
Toxine I du Mamba noir	(Gln, Pro) Leu Arg Lys Leu	Cys	Ile Leu His Arg Asn Pro	Gly Arg	Cys	Tyr Gln Lys Ile Pro Ala Phe
Inhibiteur de la protéase du pancréas de Boeuf	Phe Tyr Asn Ala Lys Ala Gly Leu	Cys	Gln Thr	Phe Val Tyr Tyr Gly	Gly Cys	Arg Ala Lys Arg Asn Asn Phe
Inhibiteur de la protéase du colostrum de Vache	Phe Tyr Asx Ser Thr Ser Asn Ala	Cys	Glu Pro	Phe Thr Tyr Tyr Gly	Gly Cys	Gln Gly Asn Asn Asx Asn Phe
Inhibiteur de la protéase du blanc d'oeuf de Tortue	Phe Tyr Asn Pro Ala Ser Arg Met	Cys	Glu Ser	Phe Ile Tyr Tyr Gly	Gly Cys	Lys Gly Asn Lys Asn Asn Phe
Inhibiteur K de la protéase d'Escargot	Tyr Tyr Asn Ser Lys Ser Gly	Cys	Gln Gln	Phe Ile Tyr Tyr Gly	Gly Cys	Arg Gly Asn Gln Asn Arg Phe
Toxine de la Vipère de Russell	Tyr Tyr Asn Leu Glu Ser Asn Lys	Cys	Lys Val	Phe Phe Tyr Tyr Gly	Gly Cys	Gly Gly Asn Ala Asn Asn Phe
Toxine K du Mamba noir	Tyr Tyr Lys Trp Lys Ala Lys Gln	Cys	Leu Pro	Phe Asp Tyr Ser	Gly Cys	Gly Gly Asn Ala Asn Arg Phe
Toxine I du Mamba noir	Tyr Tyr Asn Gln Lys Lys Lys Gln	Cys	(Glx, Gly) Phe	Thr Trp Ser	Gly Cys	Gly Gly Asn Ser Asn Arg Phe
Inhibiteur de la protéase du pancréas de Boeuf	Lys Ser Ala Glu Asp	Cys	Met Arg Thr	Cys	Gly Gly Ala	
Inhibiteur de la protéase du colostrum de Vache	Glu Thr Thr Glu Met	Cys	Leu Arg Ile	Cys	Glu Pro Pro Gln Thr Asp Lys Ser	
Inhibiteur de la protéase du blanc d'oeuf de Tortue	Lys Thr Lys Ala Glu	Cys	Val Arg Thr	Cys	Gly Pro Gly Ile Cys Leu	et 60 résidus supplémentaires
Inhibiteur K de la protéase d'Escargot	Asp Thr Thr Gln Gln	Cys	Gln Gly Val	Cys	Val	
Toxine de la Vipère de Russell	Glu Thr Arg Asp Glu	Cys	Arg Gln Thr	Cys	Gly Gly Lys	
Toxine K du Mamba noir	Lys Thr Ile Glu Glu	Cys	Arg Arg Thr	Cys	Val Gly	
Toxine I du Mamba noir	Lys Thr Ile Glu Glu	Cys	Arg Arg Thr	Cys	Ile Arg Lys	

Les onze résidus invariants sont encadrés. La structure tridimensionnelle de l'inhibiteur de la trypsine pancréatique de Boeuf (1^{re} ligne) est connue (12) : les tétrapeptides correspondant à des coudes β sont soulignés, ainsi que ceux prédits (11) dans les autres inhibiteurs.

Figure 3. Séquences primaires de l'inhibiteur de la protéase (d'après Creighton, réf. 13).

	BC1	BC2	C1	3	6	19	21	22	24	25	26	C31	CA1	CA2				
Homme	Arg	Arg	Glu	Ala	Glu	Val	Gly	Gln	Val	Gly	Gln	Leu	Gly	Leu	Gly	Gln	Arg	Arg
Singe			Glu	Ala	Glu	Val	Gly	Gln	Val	Gly	Gln	Leu	Gly	Leu	Gly	Gln		
Cochon d'Inde			Glu	Leu	Glu	Val	Glu	Gln	Val	Gly	Met	Gly	Leu	-	Gln	Gly	Ala	Leu
Chinchilla			Glu	Leu	Glu	Val	Gly	Gln	Val	Gly	Val	Pro	Glu	Ala	Gly	Arg	Leu	Gln
Rat I			Glu	Val	Glu	Val	Pro	Gln	Val	Gly	Gly	Pro	Glu	Ala	Gly	Arg	Leu	Gln
Rat II			Glu	Val	Glu	Val	Pro	Gln	Val	Gly	Gly	Pro	Glu	Ala	Gly	Arg	Leu	Gln
Cheval			Glu	Ala	Glu	Val	Gly	Gln	Val	Gly	Gly	Pro	Gly	Leu	Gly	Gly	Pro	Gln
Cochon			Glu	Ala	Glu	Ala	Gly	Ala	Val	Gly	Gly	Leu	-	-	Gly	Gly	Leu	Gln
Boeuf, Mouton			Glu	Val	Glu	Val	Gly	Ala	Leu	Glu	Leu	Ala	Gly	Gly	Pro	Gly	Ala	Gly

Figure 4. Séquences primaires de neuf homologues du peptide C de la pro-insuline (d'après Snell et Smyth, réf. 21).

Les dix résidus invariants sont encadrés. Les tétrapeptides prédits par Chou et Fasman comme adoptant une structure en coudes β sont soulignés (11). Les résidus basiques reliant entre eux le résidu C-terminal de la chaîne B au résidu N-terminal de peptide C sont appelés BC 1 et BC 2. Les résidus basiques reliant l'extrémité C-terminale du peptide C à l'extrémité N-terminale de la chaîne A sont numérotés CA 1 et CA 2. Les résidus du peptide C sont numérotés de 1 à 31 et l'alignement des différentes séquences, quand il y a des délétions, met en évidence le maximum d'homologies.

Mouton	1	5	10	15	20	25	30
	Gln	Gln	Asn	Gln	Glu	Arg	Phe
Vache							
Mouton	35	40	45	50	55	60	
	Val	Leu	Ser	Arg	Tyr	Gln	Arg
Vache							
Mouton	65	70	75	80	85	90	
	Tyr	Ala	Lys	Pro	Val	Arg	Pro
Vache							
Mouton	95	100	105	110	115	120	
	Gln	Pro	Thr	Ala	Met	Ala	Arg
Vache							
Mouton	125	130	135	140	145	150	
	Ala	Ile	Asn	Thr	Ile	Asn	Thr
Vache							
Mouton	155	160	165	170			
	Ser	Ser	Glu	Pro	Glu	Thr	Asn
Vache							
	Pro	Val	Glu	Pro	Ile	Val	

Figure 6. Séquences primaires des caséines kappa de Mouton et de Vache (33).

1 5 10 15 20 25

Lys	Glu	Thr	Ala	Ala	Lys	Phe	Glu	Arg	Gln	His	Met	Asp	Ser	Ser	Thr	Ser	Ala	Ala	Ser	Ser	Ser	Asn	Tyr
Lys	Glu	Ser	Pro	Ala	Lys	Phe	Glu	Arg	Gln	His	Met	Asp	Ser	Ser	Gly	Ser	Ser	Ala	Ser	Pro	Thr	Asn	Tyr
Lys	Glu	Ser	Ala	Ala	Lys	Phe	Glu	Arg	Gln	His	Met	Asp	Ser	Ser	Thr	Ser	Ser	Ala	Ser	Ser	Ser	Asn	Tyr
Lys	Glu	Ser	Ala	Ala	Lys	Phe	Glu	Arg	Gln	His	Met	Asp	Ser	Ser	Thr	Ser	Ser	Ala	Ser	Ser	Pro	Asx	Tyr
Lys	Glu	Ser	Ala	Ala	Lys	Phe	Glu	Arg	Gln	His	Met	Asp	Ser	Ser	Thr	Ser	Ser	Ala	Ser	Ser	Pro	Asx	Tyr
Lys	Glu	Ser	Pro	Ala	Lys	Phe	Gln	Arg	Gln	His	Met	Asp	Pro	Asp	Ser	Ser	Ser	Ser	Ser	Ser	Ser	Asn	Tyr
Arg	Glu	Ser	Ser	Ala	Asp	Lys	Lys	Arg	Gln	His	Met	Asp	Thr	Glu	Gly	Pro	Ser	Lys	Ser	Ser	Pro	Thr	Tyr
Lys	Glu	Ser	Ala	Ala	Lys	Phe	Glu	Arg	Gln	His	Ile	Asp	Ser	Ser	Thr	Ser	Ser	Val	Ser	Ser	Ser	Asn	Tyr
Ser	Glu	Ser	Ser	Ala	Lys	Phe	Glu	Arg	Gln	His	Met	Asp	Ser	Ser	Arg	Gly	Ser	Pro	Ser	Thr	Asn	Pro	Tyr
Lys	Glu	Ser	Ser	Ala	Met	Lys	Phe	Gln	Arg	Gln	His	Met	Asp	Ser	Ser	Gly	Ser	Pro	Ser	Thr	Asn	Pro	Tyr
Ala	Glu	Ser	Ser	Ala	Met	Lys	Phe	Glu	Arg	Gln	His	Val	Asp	Ser	Gly	Gly	Ser	Ser	Ser	Thr	Asn	Ala	Tyr
Ala	Glu	Ser	Ser	Ala	Met	Lys	Phe	Gln	Arg	Gln	His	Met	Asp	Pro	Glu	Gly	Ser	Pro	Ser	Thr	Asn	Ala	Tyr

Gly Glu Ser

26 30 35 40 45 50

Cys	Asn	Gln	Met	Met	Lys	Ser	Arg	Asn	Leu	Thr	Lys	Asp	Arg	Cys	Lys	Pro	Val	Asn	Thr	Phe	Val	His	Glu	Ser
Cys	Asn	Gln	Met	Met	Lys	Arg	Arg	Asn	Met	Thr	Gln	Gly	Arg	Cys	Lys	Pro	Val	Asn	Thr	Phe	Val	His	Glu	Pro
Cys	Asn	Gln	Met	Met	Lys	Ser	Arg	Asn	Leu	Thr	Gln	Asp	Arg	Cys	Lys	Pro	Val	Asn	Thr	Phe	Val	His	Glu	Ser
Cys	Asn	Gln	Met	Met	Gln	Ser	Arg	Asn	Leu	Thr	Gln	Asp	Arg	Cys	Lys	Pro	Val	Asn	Thr	Phe	Val	His	Glu	Ser
Cys	Asn	Gln	Met	Met	Gln	Ser	Arg	Asn	Met	Thr	Gln	Gly	Arg	Cys	Lys	Pro	Val	Asn	Thr	Phe	Val	His	Glu	Ser
Cys	Asn	Leu	Met	Met	Ser	Arg	Arg	Gln	Met	Thr	Gln	Gly	Ser	Cys	Lys	Pro	Val	Asn	Thr	Phe	Val	His	Glu	Pro
Cys	Asn	Gln	Met	Met	Thr	Ser	Arg	Asn	Leu	Thr	Gln	Asp	Arg	Cys	Lys	Pro	Val	Asn	Thr	Phe	Val	His	Glu	Ser
Cys	Asn	Glu	Met	Met	Lys	Ser	Arg	Asn	Met	Thr	Gln	Gly	Arg	Cys	Lys	Pro	Val	Asn	Thr	Phe	Val	His	Glu	Pro
Cys	Asn	Glu	Met	Met	Lys	Gly	Arg	Asn	Met	Thr	Gln	Gly	Tyr	Cys	Lys	Pro	Val	Asn	Thr	Phe	Val	His	Glu	Pro
Cys	Asn	Glu	Met	Met	Lys	Lys	Arg	Arg	Met	Thr	Lys	Asp	Arg	Cys	Lys	Pro	Val	Asn	Thr	Phe	Val	His	Glu	Pro
Cys	Asn	Val	Met	Met	Ile	Arg	Arg	Asn	Met	Thr	Gln	Gly	Arg	Cys	Lys	Pro	Val	Asn	Thr	Phe	Val	His	Glu	Ser

51 55 60 65 70 75

Leu	Ala	Asp	Val	Gln	Ala	Val	Cys	Gln	Lys	Asn	Val	Ala	Gln	Thr	Asn	Cys	Thr	Gln	Ser
Leu	Ala	Asp	Val	Gln	Ala	Ile	Cys	Gln	Lys	Asn	Ile	Thr	Gln	Ser	Asn	Cys	Tyr	Gln	Ser
Leu	Ala	Asp	Val	Gln	Ala	Val	Cys	Gln	Lys	Asn	Val	Ala	Gln	Thr	Asn	Cys	Tyr	Gln	Ser
Leu	Ala	Asp	Val	Gln	Ala	Val	Cys	Gln	Lys	Asn	Val	Ile	Gln	Ser	Asn	Cys	Tyr	Gln	Ser
Leu	Ala	Asp	Val	Gln	Ala	Val	Cys	Gln	Lys	Asn	Val	Ala	Gln	Ser	Asn	Cys	Tyr	Gln	Ser
Leu	Glu	Asp	Val	Gln	Ala	Val	Cys	Gln	Ile	Asn	Val	Asn	Gln	Thr	Asn	Cys	Tyr	Gln	Ser
Leu	Ala	Asp	Val	Gln	Ala	Val	Cys	Gln	Lys	Asn	Val	Ala	Arg	Asp	Asn	Cys	His	Lys	Ser
Leu	Ala	Asp	Val	Gln	Ala	Val	Cys	Gln	Lys	Asn	Val	Ala	Gln	Thr	Asn	Cys	Tyr	Gln	Ser
Leu	Ala	Asp	Val	Gln	Ala	Val	Cys	Gln	Lys	Asn	Val	Leu	Gln	Thr	Asn	Cys	Tyr	Gln	Ser
Leu	Ala	Asp	Val	Gln	Ala	Val	Cys	Gln	Lys	Asn	Val	Pro	Gln	Ser	Asn	Cys	Tyr	Gln	Ser
Leu	Ala	Glu	Val	Gln	Ala	Val	Cys	Gln	Arg	Asn	Val	Ser	Gln	Thr	Asn	Cys	Tyr	Gln	Ser
Leu	Ala	Asp	Val	Gln	Ala	Val	Cys	Gln	Lys	Asn	Val	Pro	Cys	Thr	Asn	Cys	Tyr	Gln	Ser

Boeuf	76	Tyr	Thr	Met	Ser	Ile	Thr	Asp	Cys	Arg	Glu	Thr	Gly	Ser	Ser	Lys	Tyr	Pro	Asn	Cys	Ala	Tyr	Lys	Thr	Thr	100
		Ser	Ser	Met	His	Ile	Thr	Asp	Cys	Arg	Leu	Thr	Ser	Gly	Ser	Lys	Tyr	Pro	Asn	Cys	Ala	Tyr	Gln	Thr	Ser	
Cheval	80	Tyr	Thr	Met	Ser	Ile	Thr	Asp	Cys	Arg	Glu	Thr	Gly	Ser	Ser	Lys	Tyr	Pro	Asn	Cys	Ala	Tyr	Lys	Thr	Thr	
		Ser	Ser	Met	His	Ile	Thr	Asp	Cys	Arg	Glu	Thr	Ser	Gly	Ser	Lys	Tyr	Pro	Asn	Cys	Ala	Tyr	Gln	Thr	Ser	
Chèvre	85	Asn	Ser	Ala	Met	His	Ile	Thr	Asp	Cys	Arg	Glx	Ser	Gly	Asx	Ser	Lys	Tyr	Pro	Asn	Cys	Val	Lys	Thr	Thr	
		Ser	Ser	Ala	Met	His	Ile	Thr	Asp	Cys	Arg	Glx	Ser	Gly	Asx	Ser	Lys	Tyr	Pro	Asn	Cys	Val	Lys	Ala	Thr	
Chevreuil	90	Asn	Ser	Thr	Met	His	Ile	Thr	Asp	Cys	Arg	Gln	Thr	Gly	Ser	Ser	Lys	Tyr	Pro	Asn	Cys	Ala	Tyr	Lys	Ala	Thr
		Ser	Ser	Thr	Leu	Arg	Ile	Thr	Asp	Cys	Arg	Leu	Lys	Gly	Ser	Ser	Lys	Tyr	Pro	Asn	Cys	Ala	Tyr	Lys	Ala	Thr
Cerf	95	Asn	Ser	Ala	Met	His	Ile	Thr	Asp	Cys	Arg	Glu	Thr	Gly	Ser	Ser	Lys	Tyr	Pro	Asn	Cys	Ala	Tyr	Lys	Ala	Thr
		Ser	Ser	Ala	Met	His	Ile	Thr	Asp	Cys	Arg	Glu	Thr	Gly	Ser	Ser	Lys	Tyr	Pro	Asn	Cys	Ala	Tyr	Lys	Ala	Thr
Cochon	100	Asn	Ser	Thr	Met	His	Ile	Thr	Asp	Cys	Arg	Gln	Thr	Gly	Ser	Ser	Lys	Tyr	Pro	Asn	Cys	Ala	Tyr	Lys	Ala	Thr
		Ser	Ser	Thr	Leu	Arg	Ile	Thr	Asp	Cys	Arg	Leu	Lys	Gly	Ser	Ser	Lys	Tyr	Pro	Asn	Cys	Ala	Tyr	Lys	Ala	Thr
Rat	105	Tyr	Ser	Ala	Met	His	Ile	Thr	Asp	Cys	Arg	Glu	Thr	Gly	Ser	Ser	Lys	Tyr	Pro	Asn	Cys	Ala	Tyr	Lys	Ala	Thr
		Ser	Ser	Ala	Met	His	Ile	Thr	Asp	Cys	Arg	Leu	Thr	Ser	Ser	Lys	Tyr	Pro	Asn	Cys	Ala	Tyr	Gln	Thr	Thr	
Girafe	110	Asn	Ser	Asn	Met	His	Ile	Thr	Asp	Cys	Arg	Val	Thr	Ser	Asn	Ser	Asp	Tyr	Pro	Asn	Cys	Ser	Tyr	Arg	Thr	Ser
		Ser	Ser	Asn	Met	His	Ile	Thr	Asp	Cys	Arg	Leu	Thr	Ser	Asn	Ser	Lys	Tyr	Pro	Asn	Cys	Ser	Tyr	Arg	Thr	Ser
Coypu	115	Tyr	Ser	Ser	Met	His	Ile	Thr	Asp	Cys	Arg	Leu	Thr	Ser	Gly	Ser	Lys	Tyr	Phe	Pro	Asn	Cys	Ser	Tyr	Arg	Thr
		Ser	Ser	Arg	Met	Arg	Ile	Thr	Asp	Cys	Arg	Val	Thr	Ser	Ser	Ser	Lys	Tyr	Phe	Pro	Asn	Cys	Ser	Tyr	Arg	Thr
Chinchilla	120	Tyr	Ser	Arg	Met	Arg	Ile	Thr	Asp	Cys	Arg	Val	Thr	Ser	Ser	Lys	Tyr	Phe	Pro	Asn	Cys	Ser	Tyr	Arg	Thr	Ser
		Ser	Ser	Arg	Met	Arg	Ile	Thr	Asp	Cys	Arg	Val	Thr	Ser	Ser	Lys	Tyr	Phe	Pro	Asn	Cys	Ser	Tyr	Arg	Thr	Ser
Cochon d'Inde A	125	Gln	Ala	Asn	Lys	His	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Lys	Glu	Arg	His	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Cochon d'Inde B	130	Gln	Ala	Asn	Lys	His	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Lys	Glu	Arg	His	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Boeuf	135	Gln	Ala	Asn	Lys	His	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Lys	Glu	Arg	His	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Cheval	140	Gln	Ala	Glx	Arg	His	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Ala	Glx	Arg	His	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Chèvre	145	Glx	Ala	Glx	Arg	His	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Ala	Glx	Arg	His	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Chevreuil	150	Glx	Ala	Glx	Arg	His	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Ala	Glx	Arg	His	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Cerf	155	Gln	Glu	Gln	Arg	His	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Ala	Gln	Arg	His	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Cochon	160	Asn	Ser	Glu	Arg	His	Ile	Val	Ala	Cys	Asp	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Glu	Arg	Lys	His	Ile	Val	Val	Ala	Cys	Asp	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Rat	165	Gln	Ala	Glu	Lys	His	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Ala	Glu	Lys	His	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Girafe	170	Gln	Glu	Glu	Lys	Ser	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Ala	Glu	Lys	Ser	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Coypu	175	Arg	Glu	Asn	Lys	Gly	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Ala	Gln	Lys	Ser	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Chinchilla	180	Gln	Ala	Gln	Lys	Ser	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Ala	Gln	Lys	Ser	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Cochon d'Inde A	185	Gln	Ala	Gln	Lys	Ser	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Ala	Gln	Lys	Ser	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Cochon d'Inde B	190	Gln	Ala	Gln	Lys	Ser	Ile	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
		Ala	Gln	Lys	Ser	Ile	Val	Val	Ala	Cys	Glu	Gly	Asn	Pro	Tyr	Val	Pro	Val	His	Phe	Asp	Ala	Ser	Val	Gln	
Boeuf	195	Pro	Ser	Thr																						

Figure 5. Séquences primaires de douze homologues de la ribonucléase (d'après les références 25, 26, 27, 28).
 Les résidus invariants sont encadrés. La structure tridimensionnelle de la ribonucléase de Boeuf (1^{re} ligne) est connue (24) : les tétrapeptides correspondant à des coudes β y sont soulignés, ainsi que ceux prédicts dans les autres ribonucléases (11).

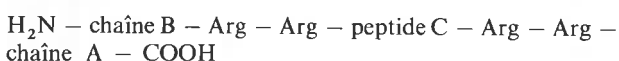
les méthodes de prédiction de conformation à partir de la structure primaire : en effet, les variations de séquence pouvant exister d'une protéine à une autre sont plus facilement mises en évidence que n'est réalisée une étude cristallographique sur chacune d'elle. Ce sont Chou et Fasman eux-mêmes qui ont réalisé une telle étude (11). Nous donnons ici les résultats qu'ils ont obtenus à partir de trois classes de protéines : plusieurs inhibiteurs de protéases, le peptide C de différentes proinsulines et certaines ribonucléases.

III.1. Inhibiteurs de protéases

Laskowski et ses collaborateurs (12) ainsi que Creighton (13) ont mis en évidence des homologies de séquence de sept inhibiteurs de protéases, malgré leur grande différence d'ordre phylogénétique (inhibiteur de la trypsine pancréatique de Bœuf, de protéases du colostrum de Vache, du blanc d'œuf de Tortue, de l'Escargot, ainsi que la toxine de la Vipère de Russel (*Vipera russelli*) et les toxines K et I du Mamba noir) (Figure 3). Dans tous les cas, les six résidus de cystéine peuvent toujours être appariés pour donner entre les résidus 5-55, 30-51 et 14-38 les ponts disulfurés et, pour ces résidus, il n'y a ni délétion, ni insertion. Si l'on considère les sept séquences de la figure 3, on constate que seuls cinq autres résidus d'acides aminés sont invariants, autrement dit, qu'il n'y a finalement que 20 % de résidus qui sont conservés pour ces sept protéines. Les probabilités d'existence des coudes β ont été calculées et les profils correspondants en ont été tracés (11). Ils montrent que les coudes β sont conservés. De plus, puisque la structure tridimensionnelle de la protéine d'origine bovine a été déterminée aux rayons X par Deisenhofer et Steigemann (14), il est intéressant de constater que les prédictions de conformation donnent des résultats tout à fait comparables. Les tétrapeptides 12-15, 25-28, 37-40 et 41-44 adoptent une structure en coude β . De plus, sur les onze résidus invariants, quatre (Gly₁₂, Cys₁₄, Gly₃₇, Cys₃₈) sont situés dans les coudes 12-15 et 37-40. Trois autres (Tyr₂₃, Cys₃₀ et Phe₄₃) sont situés tout près des coudes 25-28 et 41-44. Cette étude met, non seulement, en évidence l'homologie qui existe entre les différents inhibiteurs, mais elle a permis aussi d'infirmes l'hypothèse de Reddy et de ses collaborateurs (15) selon laquelle ces différentes protéines étaient aussi homologues de l'inhibiteur de la bromélaïne d'ananas comme auraient pu le faire penser les résultats préliminaires concernant la séquence de cette protéine. Cependant, lorsque la séquence primaire fut entièrement connue (13), les prédictions de conformation ont permis de montrer qu'il n'existait aucune analogie entre sa structure secondaire et celles des sept autres inhibiteurs.

III.2. Peptide C de la pro-insuline

Le précurseur de l'insuline, la pro-insuline, est synthétisé sous forme d'une chaîne polypeptidique unique que l'on peut caractériser par le schéma suivant :



dans lequel les deux chaînes polypeptidiques A et B sont réunies par l'intermédiaire d'un peptide possédant deux résidus Arg à chacune de ses extrémités N- et C-terminales. La transformation de la pro-insuline en insuline s'effectue grâce à une protéolyse limitée de la pro-insuline (16, 17) dans laquelle le peptide C est libéré (c'est-à-dire le peptide de jonction entre les chaînes A et B, moins les quatre résidus Arg). La longueur du peptide C varie beaucoup d'une espèce à l'autre (Figure 4) et, en particulier, il est constitué de vingt-six résidus d'acides aminés chez le Bœuf et le Mouton, vingt-neuf chez le Porc et le Cochon d'Inde, et trente-et-un chez l'Homme, le Singe, le Chinchilla, le Cheval et le Rat chez qui deux variants ont pu être caractérisés (18). Bien que le rôle du peptide C ne soit pas encore

connu, il est intéressant de noter que la fraction de la chaîne protéique de la proinsuline, qui deviendra ensuite l'insuline elle-même, possède la même structure dans la proinsuline que dans l'insuline (19, 20). Snell et Smyth (21) proposèrent une structure secondaire pour les neuf peptides C de pro-insuline indiqués figure 4 prêtes à partir des règles empiriques de Chou et Fasman (8). Une telle détermination fut également réalisée par Vogt et ses collaborateurs (22) et enfin par Chou et Fasman eux-mêmes (11) qui utilisent alors des paramètres conformationnels obtenus à partir d'un nombre plus important de protéines. Les trois séries de résultats (11, 21, 22) sont tout à fait comparables : on observe que, bien qu'il n'y ait que le tiers environ des résidus qui soient invariants pour les neuf peptides considérés, une remarquable conservation de structure secondaire est observée : un tétrapeptide adoptant une structure en coude β est toujours prévu au centre du peptide C, dans la région 12-20 alors que cette probabilité est quasi nulle dans le reste de la molécule, où deux séquences hélicoïdales sont prévues, de part et d'autre du coude β (11). Chou et Fasman pensent qu'une telle configuration qui, parce qu'elle permettrait un repliement approprié des parties hélicoïdales, servirait à masquer la région de l'insuline qui se lie au récepteur (23) justifiant ainsi le fait que la pro-insuline soit inactive.

III.3. Ribonucléases

La structure tridimensionnelle de la ribonucléase S pancréatique de Bœuf a été déterminée aux rayons X (24). A partir de cette étude, il a été possible de mettre en évidence onze coudes β correspondant aux tétrapeptides 16-19, 23-26, 34-37, 36-39, 65-68, 66-69, 75-78, 87-90, 91-94, 92-95 et 112-115 (Figure 5). Puisque la séquence d'un certain nombre de ribonucléases A de différentes espèces est connue, il était intéressant de rechercher si les coudes β étaient conservés pour ces différentes protéines. C'est ce que Chou et Fasman (11) ont réalisé à partir des ribonucléases de Bœuf (25), de Cheval, de Chèvre, de Chevreuil, de Cerf, de Cochon, de Rat (26), de Girafe (27) de Coypu, de Chinchilla et de Cochon d'Inde (28). Huit des douze espèces considérées ont des longueurs de chaînes de 124 résidus. Les séquences de la ribonucléase varient d'environ 25 % dans les espèces étudiées. L'invariabilité de la distribution des résidus hydrophobes aurait pour effet de stabiliser la structure secondaire et ce rôle a été discuté récemment par Leustra et ses collaborateurs (29). Les probabilités de formation de coudes β ont été calculés (11). Les tétrapeptides correspondants ont été soulignés dans la figure 5. Pour la ribonucléase S, ce sont les coudes β identifiés par rayons X qui ont été soulignés. Comme on peut le remarquer, seule la séquence 65-68 est constituée de résidus d'acide aminé identiques pour les douze espèces. Elle correspond au tétrapeptide Cys-Lys-Asn-Gly qui présente une probabilité très élevée d'adopter une structure en coude β .

Pour les douze espèces étudiées, aucun coude β n'est prévu entre les résidus 40 et 65. Il est intéressant de constater, à partir de l'étude tridimensionnelle de la ribonucléase S, que cette région contient deux séquences en structure β , 42-48 et 60-65, entourant une zone hélicoïdale, 50-59. On constate de plus que l'accessibilité au solvant de cette zone est quasi nulle (24). Par contre, les résidus pour qui l'accessibilité du solvant est élevée sont trouvés dans des coudes β , indiquant que les régions de repliement de la chaîne se trouvent surtout à la surface de la protéine. Enfin, la position des ponts disulfurés (Cys₄₀-Cys₇₂, Cys₈₄-Cys₁₁₀) montre que ce sont les coudes β qui doivent induire l'organisation spatiale favorable à leur formation. Il semble donc que les coudes β jouent un rôle particulièrement important puisqu'ils permettent la conservation au cours de l'évolution de caractéristiques structurales importantes qui permettraient de préserver certaines fonctions biologiques.

IV. Phosphorylation des protéines

Le rôle de la phosphorylation des protéines dans les phénomènes de régulation métabolique est maintenant bien connu (30). Cette phosphorylation, bien que n'impliquant que les résidus de sérine et de thréonine, se fait en des sites très spécifiques. Il n'a cependant pas été possible de mettre en évidence des homologies de séquences pour expliquer cette spécificité (31). Par contre, un examen systématique

des séquences entourant les résidus phosphorylés dans un certain nombre de phosphoprotéines intracellulaires montre que l'on trouve généralement au voisinage de la sérine ou de la thréonine phosphorylée un résidu d'acide aminé dont la chaîne latérale porte une charge positive. Plus précisément, cet acide aminé basique (lysine ou arginine) qui peut être situé aussi bien du côté N- que C-terminal du

Tableau 3. Sites de glycosylation.

Cas des chaînes N-glycosylées. Le groupement asparagine portant la chaîne glycanique est indiqué par un astérisque. Le coude β est souligné.

Glycoprotéine	Localisation	Séquence	Référence
Lactoferrine humaine	6	<u>Gly-Ser-Asn-Val-Thr</u> ^{**}	77
Protéine de la salive parotidienne humaine	15	<u>Gly-Pro-Asn-Gln-Ser</u> ^{**}	78
		Thr-Asn-Ala-Thr ^{**}	79
Orosomucoïde humain	38	<u>Tyr-Asn-Lys-Ser</u> ^{**}	79
		<u>Thr-Pro-Asn-Lys-Thr</u> ^{**}	79
Avidine de Poule	75	<u>Tyr-Asn-Ser-Ser</u> ^{**}	79
		Asn-Gly-Thr-Val-Ser	79
Bromelaïne d'ananas	17	<u>Gly-Ser-Asn-Met-Thr</u> ^{**}	80
Chaîne γ du fibrinogène humain	20	<u>Pro-Arg-Asn-Asn-Glu-Ser</u> ^{**}	81
		<u>Asn-Lys-Thr-Ser</u> ^{**}	82
Chaîne β de l'haptoglobine humaine	32	<u>Asx-Asn-Ser-Thr</u> ^{**}	83
Hormone lutéinisante de Mouton. Chaîne α	56	<u>Pro-Lys-Asn-Ile-Thr</u> ^{**}	84
		<u>Glu-Asn-His-Thr</u> ^{**}	84
Transferrine de Poulet	6	<u>Asn-Arg-Thr-Gly</u> ^{**}	85
		<u>Met-Asn-Asn-Ser</u> ^{**}	85
	336		

Tableau 3 (suite).

Glycoprotéine	Localisation	Séquence	Référence
Macroglobuline Ou de Waldenstrom. Chaîne μ	170	<u>Asn-Asn-Asn-Lys-Ser</u> **	86
	332	<u>Gln-Asn-Ala-Ser</u> **	86
	395	<u>His-Thr-Asn-Ile-Ser</u> **	86
	402	<u>His-Pro-Asn-Ala-Thr</u> **	86
	563	<u>Tyr-Asn-Val-Ser</u> **	86
Ribonucléase pancréatique de Porc	21	<u>Ser-Ser-Asn-Ser-Ser</u> **	87
	34	<u>Arg-Arg-Asn-Met-Thr</u> **	
	76	<u>Gln-Ser-Asn-Ser-Thr</u> **	
Ribonucléase pancréatique de Boeuf	34	<u>Ser-Arg-Asn-Leu-Thr</u> **	88
Hormone cortico-gonadotropine humaine (HCG) sous-unité	52	<u>Lys-Asn-Val-Thr-Ser</u> **	89
	78	<u>Glu-Asn-His-Thr</u> **	89
Prothrombine de Boeuf	77	<u>Tyr-Arg-Glu-Asn-Val-Ser</u> **	90
	101	<u>Glu-Ile-Asn-Ser-Thr-Thr</u> **	90
	376	<u>Trp-Asp-Lys-Asn-Phe-Thr</u> **	90
Fibrinopeptide Lamproie	9	<u>Thr-Asn-Asn-Ser-Asp-Pro</u> **	90
Lipases pancréatiques L _A et L _B de Porc		<u>Thr-Asn-Gly-Thr-Ile</u> **	91

Tableau 3 (suite).

Glycoprotéine	Localisation	Séquence	Référence
Céruleoplasmine (Homme et Porc)	(Porc)	** Tyr-Asn-Asp-Thr	92
		** Asn-Val-Ser-Asn	92
		** Thr-Lys-Glu-Asn	92
Ovalbumine de Poule		** Glu-Lys-Tyr-Asn-Leu-Thr	93
Phosvitine de Poule		** Gly-His-Gly-Asn-Lys-Ser-Ser	94
Fétuine de Boeuf		** Ser-Asn-Gly-Ser	95
		** Asn-Asp-Ser-Arg	95

résidu phosphorylé n'en est séparé que par un seul acide aminé qui peut être de n'importe quelle nature. Il était donc tentant de rechercher l'existence d'un « signal de reconnaissance » au niveau de la structure secondaire. A partir des vingt-quatre protéines phosphorylées qui ont été étudiées jusqu'ici (32, 33), il a été possible de montrer que pratiquement 80 % des sérine et thréonine impliquées étaient situées dans des coude β. Parmi les 20 % restants, la grande majorité est localisée dans des régions adjacentes à des coude β. Seuls deux résidus de sérine phosphorylés n'ont été trouvés ni dans ni près d'un coude β (Tableau 2) : la sérine 110 de la protéine basique de la myéline et une sérine de la phosphorylase kinase.

V. Étude des glycoprotéines

Les glycoprotéines font partie d'un groupe très important de biopolymères que l'on trouve dans les tissus conjonctifs, les cellules sanguines, le sérum, l'urine, la salive etc... et qui présentent des fonctions biologiques très diversifiées. Elles sont essentiellement constituées d'un squelette protéique au niveau duquel différentes chaînes glycaniques sont liées de façon covalente. Non seulement les proportions et l'agencement des différents sucres de ces glycanes peuvent varier, mais le nombre des chaînes glycaniques et leur localisation peuvent être très différents d'une glycoprotéine à l'autre. Le type de liaison glycanne-peptide présente dans les glycoprotéines est caractérisé à la fois par le résidu monosaccharidique et l'acide aminé impliqués dans la liaison (voir Neuberger *et al.*, référence 72). Deux types de liaison entre la chaîne glycanique et la protéine sont caractérisés : l'un, de type N-glycosidique, lorsque l'asparagine est impliquée dans la liaison ; l'autre, de type O-glycosidique, lorsque la liaison se fait par l'intermédiaire d'un résidu de sérine ou de thréonine. Certaines glycoprotéines contenant plusieurs chaînes glycaniques peuvent présenter les deux types de liaison. Théoriquement, la caractérisation totale d'une glycoprotéine implique la connaissance de la structure de sa chaîne glycanique et de sa séquence primaire en acides aminés. Mais une telle détermination n'est pas toujours aisée et seules quelques glycoprotéines sont actuellement entièrement connues. Par contre, à l'aide de coupures enzymatiques ou chimiques de la chaîne protéique, il a été possible d'obtenir un grand nombre de glycopeptides dans lesquels la chaîne glycanique est pratiquement intacte et la partie peptidique constituée d'un nombre relativement petit d'acides aminés, mais suffisant pour envisager des déterminations de configuration.

A partir de l'étude d'un certain nombre de ces glycopeptides, Marshall et Neuberger (73, 74) ont pu démontrer que, dans le cas où la chaîne glycanique est liée à un résidu d'asparagine, celui-ci est inclus dans une séquence Asn-X-Ser (Thr), dans laquelle le résidu X peut être n'importe quel résidu d'acide aminé. Marshall (75) a émis l'hypothèse que le groupement hydroxyle de la sérine peut créer une liaison hydrogène avec le groupement carbonyle de la chaîne latérale de l'asparagine, diminuant ainsi la constante de dissociation du groupement amide, ce qui aurait pour effet de faciliter la liaison de la N-acétylglucosamine au résidu d'asparagine. Il a été suggéré par plusieurs auteurs (73, 74, 76) que cette séquence Asn-X-Ser (Thr) pouvait constituer une condition nécessaire — mais non suffisante — pour que la glycosylation enzymatique puisse se faire. Il était donc tentant de regarder si d'autres conditions, d'ordre conformationnel cette fois, ne devaient pas aussi être remplies pour qu'une telle glycosylation enzymatique puisse se faire. La recherche d'une conformation préférentielle que pourrait adopter la partie protéique au voisinage du site de fixation de la chaîne glycanique a donc été systématiquement entreprise. Elle l'a été non seulement sur les glycopeptides N-glycosylés (Tableau 3) dont nous venons de parler mais aussi sur les glycopeptides O-glycosylés (Tableau 4) bien que, dans ce cas, aucune séquence « code » n'ait pu être mise en évidence. L'utilisation des règles de prédiction de conformation a permis de prévoir que le résidu d'acide aminé sur lequel était attaché la chaîne glycanique présentait une probabilité élevée d'être situé dans un coude β (107). Ces résultats, repris par la suite sur un plus grand nombre de glycoprotéines (108, 109) ont montré que cinquante-neuf des soixante et une chaînes N-glycosidiques considérées (soit plus de 95 %) se liaient à un résidu d'asparagine situé sur un coude β et que cette proportion atteignait 100 % pour les chaînes O-glycosylées : la totalité des liaisons considérées — soit 21 — se font sur un résidu de sérine (ou de thréonine) appartenant à un coude β.

Tableau 4. Sites de glycosylation.

Cas des chaînes O-glycosylées. La sérine ou la thréonine impliquée dans la liaison glycanique est indiquée par un astérisque. Le coude β est souligné.

Glycoprotéine	Localisation	Séquence	Référence
Protéine de la maladie ZUC		<u>Pro-Gly-Gly-Ser</u> ^{**}	96
Chondroïtine sulfate du cartilage de Porc		<u>Glu-Gly-Ser-Gly</u> ^{**}	97
Mucine sous-maxillaire de Boeuf		<u>Ser-Thr-Gly-Ser</u> ^{**}	98
		<u>Ser-Thr-Gly-Ser</u> ^{**}	98
Immunoglobuline γ G de Lapin		<u>Pro-Thr-Cys-Pro</u> ^{**}	99
Hormone cortico-gonadotropine (HCG) sous-unité β	121	<u>Ser-Ser-Ser-Ser</u> ^{**}	89
	127	<u>Pro-Pro-Ser-Leu</u> ^{**}	89
	132	<u>Ser-Pro-Ser-Arg</u> ^{**}	89
	138	<u>Gly-Pro-Ser-Thr</u> ^{**}	89
α -Amylase		<u>Ser-Glu-Asp-Gly</u> ^{**}	100
Fétuine de Boeuf		<u>Gly-Pro-Ser-Pro</u> ^{**}	95
		<u>Ser-Thr-Pro-Gly</u> ^{**}	95
Immunoglobuline IgA ₁		<u>Thr-Pro-Ser-Pro</u> ^{**}	101
		<u>Pro-Thr-Cys-Pro</u> ^{**}	102

Tableau 4 (suite).

Glycoprotéine	Localisation	Séquence	Référence
Caséine κ_A de Vache		** Thr-Pro-Thr-Thr	103
Caséine λ_A de Mouton		** Thr-Pro-Thr-Thr	104
Apolipoprotéine (Apo C-III)		** Arg-Pro-Thr-Ser	105
Protéine encéphalitogénique		** Thr-Pro-Arg-Thr	106

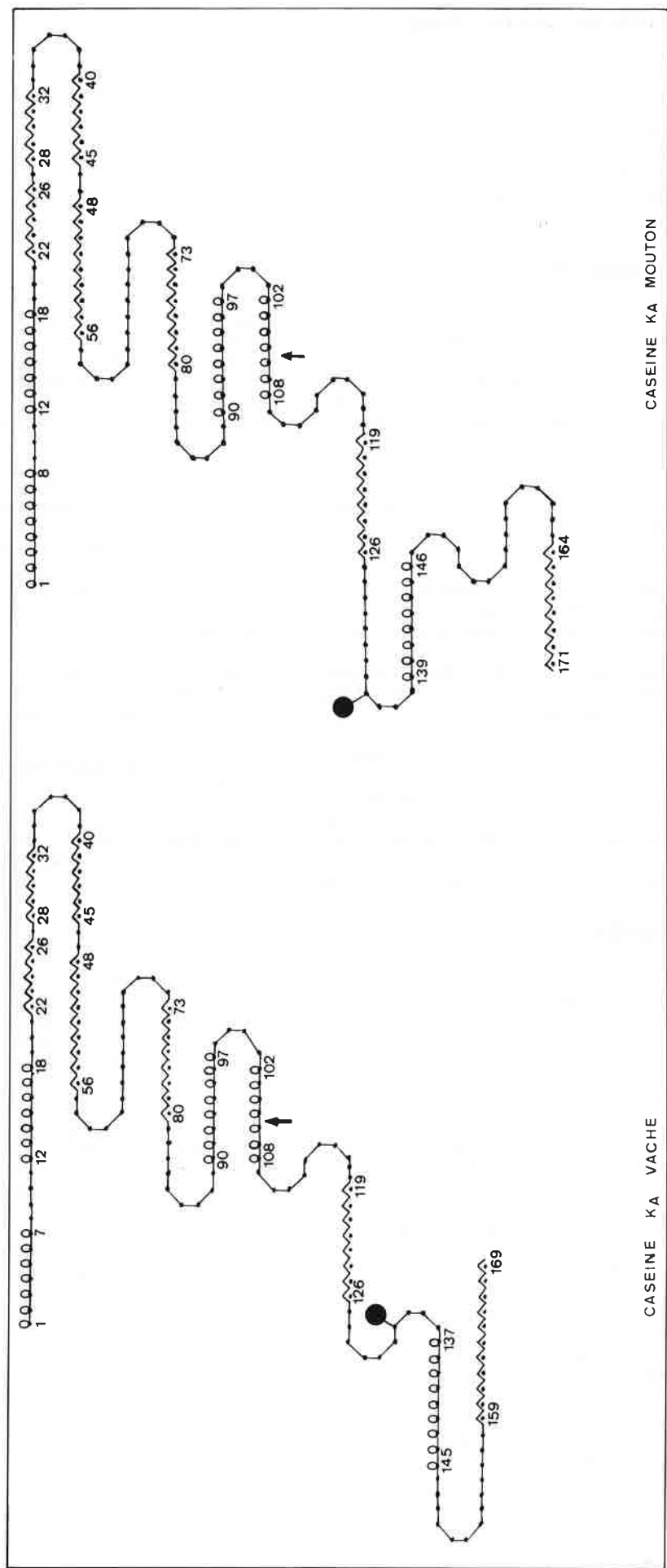


Figure 7. Représentation schématique de la structure secondaire prédite pour les caséines kappa de Vache et de Mouton. Les résidus sont représentés selon la conformation qu'ils adoptent : en hélice α (O), en structure β (∇), en structure inordonnée (-) ou en coude β (ζ) (33).

VI. Étude des caséines kappa

L'étude des caséines kappa illustre bien ce qui vient d'être dit : ce sont des glycoprotéines qui possèdent un site de phosphorylation et dont la conformation a été prédite pour deux espèces différentes, la Vache et le Mouton (33). La chaîne peptidique (Figure 6) peut être coupée par la chymosine entre les résidus 107 et 108 (Phe-Met) donnant ainsi une partie N-terminale, insoluble et une partie C-terminale, soluble, le caséino-glycopeptide. C'est dans cette partie que l'on trouve le plus de variations de la structure primaire quand on passe de la Vache au

Mouton. Mais c'est aussi dans cette partie que l'on trouve les sites d'intérêt biologique : les résidus sérine 149 ou 151, respectivement pour la Vache et le Mouton, sont phosphorylés et les résidus thréonine 133 ou 131 portent une chaîne glycanique. On observe que ces résidus sont toujours situés dans des coudes β (Figure 7) : les modifications intervenant dans la structure primaire se font d'une manière telle qu'elles permettent aux coudes β d'être conservés.

VII. Conclusion

Nous avons voulu démontrer dans les exemples ci-dessus le rôle privilégié que joue le coude β dans le problème de la transformation chimique des protéines. Il est intéressant de discuter ce fait à la lumière de deux constatations.

La première est que, comme cela a été précédemment mis en évidence (3), les coudes β présentent une probabilité élevée d'être localisés à la surface de la protéine. Ils correspondent en effet à des zones hydrophiles ; les acides aminés qui les constituent auront donc tendance à être en contact avec l'eau environnante ; les résidus hydrophobes se trouveront alors de préférence au centre de la protéine, c'est-à-dire qu'ils chercheront à éviter au maximum le contact avec le solvant (110).

La seconde constatation concerne les modifications chimiques post-transcriptionnelles que subissent les protéines. Une récente revue (111) dans ce domaine montre en effet que si le nombre d'acides aminés qui constituent les protéines n'est que de vingt, il a été cependant possible d'identifier, à partir d'un grand choix de protéines appartenant à des organismes différents, 120 dérivés d'acides aminés en plus des vingt constituants.

En d'autres termes, une fois la protéine biosynthétisée à partir des vingt acides aminés naturels, elle subit un certain nombre de modifications chimiques, modifications par lesquelles des liaisons covalentes peuvent être formées. Il semble logique de penser que ces transformations ne peuvent se faire qu'en certains sites correspondant à une « présentation privilégiée » de leur conformation spatiale. Si l'on considère ce qui se passe par exemple pour certains enzymes

dont la modulation est assurée par une réaction réversible de phosphorylation-déphosphorylation (comme par exemple la phosphorylase musculaire et hépatique, la pyruvate kinase etc...), on est tenté de penser qu'une telle réaction doit se faire plus facilement à la surface de la protéine et dans une structure qui permettra à la chaîne latérale de l'acide aminé impliqué dans la phosphorylation d'être correctement exposée. On imagine aisément qu'un coude β permette une telle exposition privilégiée et ce d'autant plus que sa stabilité est conditionnée par son environnement immédiat ou proche, donc par la séquence primaire de la protéine.

Dans les exemples que nous avons considérés, nous avons mis en évidence l'importance de l'accessibilité de la chaîne latérale des résidus d'asparagine pour la glycosylation (107), de sérine et de thréonine pour la phosphorylation (107) ou la phosphorylation (32). Cela a aussi été démontré lors de l'hydroxylation des résidus prolyl du collagène (112). Ces considérations d'ordre conformationnel éclairent d'un jour nouveau le problème des modifications post-transcriptionnelles des protéines qui interviennent soit dans leur modulation précise, soit dans les fonctions de structuration, soit dans les régulations.

Remerciements

L'auteur remercie vivement M. le Professeur G. Biserte, Directeur de l'Unité 124 de l'INSERM, pour les discussions très fructueuses qu'elle a souvent eues avec lui sur le problème des coudes β , ainsi que pour la lecture critique de son manuscrit.

Bibliographie

- (1) P.N. Lewis, F.A. Momany et H.A. Scheraga, *Proc. Nat. Acad. Sci. USA*, 1971, **68**, 2293.
- (2) P.N. Lewis, F.A. Momany et H.A. Scheraga, *Biochim. Biophys. Acta*, 1973, **303**, 211.
- (3) I.D. Kuntz, *J. Amer. Chem. Soc.*, 1972, **94**, 4009.
- (4) J.L. Crawford, W.N. Lipscomb et C.G. Schellman, *Proc. Nat. Acad. Sci. USA*, 1973, **70**, 538.
- (5) P.Y. Chou et G.D. Fasman, *J. Mol. Biol.*, 1977, **115**, 135.
- (6) R.E. Dickerson, T. Takano, D. Eisenberg, O.B. Kallai, L. Samson, A. Cooper et E. Margoliash, *J. Biol. Chem.*, 1971, **246**, 1511.
- (7) M.J.E. Sternberg et J.M. Thornton, *Nature*, 1978, **271**, 15.
- (8) P.Y. Chou et G.D. Fasman, *Biochemistry*, 1974, **13**, 211.
- (9) IUPAC-IUB Commission on Biochemical Nomenclature, *Biochemistry*, 1970, **9**, 3471.
- (10) C.M. Venkatachalam, *Biopolymers*, 1968, **6**, 1425.
- (11) P.Y. Chou et G.D. Fasman, *Adv. Enzym.*, 1978, **47**, 45.
- (12) M. Jr. Laskowski, I. Kato, T.R. Leary, J. Schrode et R.W. Sealock, In « *Proteinase Inhibitors* », (Edited by H. Tschesche, L.J. Greene, E. Truscheit) Springer-Verlag, 1974, Berlin.
- (13) T.E. Creighton, *Nature*, 1975, **255**, 743.
- (14) J. Deisenhofer et W. Steigemann, *Acta Cryst.*, 1975, **B-31**, 238.
- (15) M.N. Reddy, P.S. Keim, R.L. Heinrikson et F.J. Kezdy, *J. Biol. Chem.*, 1975, **250**, 1741.
- (16) D.F. Steiner et P. Oyer, *Proc. Nat. Acad. Sci. USA*, 1967, **57**, 473.
- (17) R.E. Chance, R.M. Ellis et W.W. Bromer, *Science*, 1968, **161**, 165.
- (18) J. Markussen et F. Sundby, *Eur. J. Biochem.*, 1972, **25**, 153.
- (19) B.H. Frank et A.J. Veros, *Biochem. Biophys. Res. Commun.*, 1968, **32**, 155.
- (20) A.H. Pekar et B.H. Frank, *Biochemistry*, 1972, **11**, 4013.
- (21) C.R. Snell et D.G. Smyth, *J. Biol. Chem.*, 1975, **250**, 6291.
- (22) H.P. Vogt, A. Vollmen, V.K. Naithani et H. Zahn, *Hoppe-Seyler's Z. Physiol. Chem.*, 1976, **357**, 107.
- (23) R.A. Pullen, D.G. Lindsay, S.P. Wood, I.J. Tickle, T.L. Blundell, A. Wollmer, G. Krail, D. Brandenburg, H. Zahn, J. Gliemann et S. Gammeltoft, *Nature*, 1976, **259**, 369.
- (24) F.M. Richards et H.W. Wyckoff, In « *Atlas of Molecular Structures in Biology. Ribonuclease S* », 1973, Clarendon Press, Oxford.
- (25) D.G. Smyth, W.H. Stein et S. Moore, *J. Biol. Chem.*, 1963, **238**, 227.
- (26) A.J. Scheffer et J.J. Beintema, *Eur. J. Biochem.*, 1974, **46**, 221.
- (27) W. Gaastra, G. Groen, G.W. Welling et J.J. Beintema, *FEBS-Letters*, 1974, **41**, 227.
- (28) A. Van den Berg et J.J. Beintema, *Nature*, 1975, **253**, 207.
- (29) J.A. Lenstra, J. Hofsteenge et J.J. Beintema, *J. Mol. Biol.*, 1977, **109**, 185.
- (30) F. Huijing et E.Y.C. Lee, In « *Protein Phosphorylation in Control Mechanisms* », 1973, Academic Press, New York.
- (31) R.E. Williams, *Science*, 1976, **192**, 473.
- (32) D. Small, P.Y. Chou et G.D. Fasman, *Biochem. Biophys. Res. Commun.*, 1977, **79**, 341.
- (33) M.H. Loucheux-Lefebvre, J.P. Aubert et P. Jollès, *Biophys. J.*, 1978, **23**, 323.
- (34) S.V. Shlyapnikov, A.A. Arutyunyan, S.N. Kurochkin, L.V. Memelova, M.V. Nesterova, L.P. Saschenko et E.S. Severin, *FEBS-Letters*, 1975, **53**, 316.

- (35) T.A. Langan, *Proc. Nat. Acad. Sci. USA*, 1969, **64**, 1276.
- (36) S.C.R. Elgin et H. Weintraub, *Ann. Rev. Biochem.*, 1975, 725.
- (37) T.A. Langan, *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, 1971, **30**, 1089.
- (38) T.A. Langan, *Ann. N.Y. Acad. Sci.*, 1971, **185**, 160.
- (39) A.R. MacLeod, N.C.W. Wong et G.H. Dixon, *Eur. J. Biochem.*, 1977, **78**, 281.
- (40) M.T. Sung et G.H. Dixon, *Proc. Nat. Acad. Sci. USA*, 1970, **67**, 1616.
- (41) P. Sautière, D. Tyrou, B. Laine, J. Mizon, P. Ruffin et G. Biserte, *Eur. J. Biochem.*, 1974, **41**, 563.
- (42) D. Wouters, P. Sautière et G. Biserte, *Eur. J. Biochem.*, 1978, **90**, 231.
- (43) B. Laine, P. Sautière et G. Biserte, *Biochemistry*, 1976, **15**, 1640.
- (44) G.S. Bailey et G.H. Dixon, *J. Biol. Chem.*, 1973, **248**, 5463.
- (45) B. Laine, D. Kmićik, P. Sautière et G. Biserte, *Biochimie (Paris)*, 1978, **60**, 147.
- (46) K. Iwai, K. Ishikawa et H. Hayashi, *Nature*, 1970, **226**, 1056.
- (47) A. Kootstra et G.S. Bailey, *FEBS-Letters*, 1976, **68**, 76.
- (48) A. Farago, T. Romhanyi, F. Antoni, A. Takato et F. Fabian, *Nature*, 1975, **254**, 88.
- (49) E. Hashimoto, M. Takeda, Y. Nishizuka, K. Hamana et K. Iwai, *Biochem. Biophys. Res. Commun.*, 1975, **66**, 547.
- (50) M. Strickland, W.N. Strickland, W.F. Brandt et C. Von Holt, *Eur. J. Biochem.*, 1977, **77**, 263.
- (51) W.N. Strickland, M. Strickland, W.F. Brandt et C. Von Holt, *Eur. J. Biochem.*, 1977, **77**, 277.
- (52) A.J. Louie, E.P.M. Candido et G.H. Dixon, *Cold Spring Harbor Symp. Quant. Biol.*, 1974, **38**, 803.
- (53) S.N. Kurochkin, N.B. Andreeva, K.G. Gasaryan, E.S. Severin et I.V. Kurchatov, *FEBS-Letters*, 1977, **76**, 112.
- (54) P. Sautière, G. Briand, D. Kmićik, O. Loy, G. Biserte, A. Garel et M. Champagne, *FEBS-Letters*, 1976, **63**, 164.
- (55) M. Yaguchi, C. Roy, M. Dove et V. Selegy, *Biochem. Biophys. Res. Commun.*, 1977, **76**, 100.
- (56) K. Titani, P. Cohen, K.A. Walsh et H. Neurath, *FEBS-Letters*, 1975, **55**, 120.
- (57) E.H. Fisher, D.G. Graves, E.R. Snyder-Crittenden et E.G. Krebs, *J. Biol. Chem.*, 1959, **234**, 1698.
- (58) D.P. Wolf, E.H. Fisher et E.G. Krebs, *Biochemistry*, 1970, **9**, 1923.
- (59) P. Daile et P.R. Carnegie, *Biochem. Biophys. Res. Commun.*, 1974, **61**, 852.
- (60) P.R. Carnegie, B.E. Kemp, P.R. Dunkley et A.W. Murray, *Biochem. J.*, 1973, **135**, 569.
- (61) P. Daile, P.R. Carnegie et J.D. Young, *Nature*, 1975, **257**, 416.
- (62) P.R. Carnegie, P.R. Dunkley, B.E. Kemp et A.W. Murray, *Nature*, 1974, **249**, 147.
- (63) P. Cohen, D.C. Watson et G.H. Dixon, *Eur. J. Biochem.*, 1975, **51**, 79.
- (64) S.J. Yeman, P. Cohen, D.C. Watson et G.H. Dixon, *Biochem. J.*, 1977, **162**, 411.
- (65) T.S. Huang, D.B. Bylund, J.T. Stall et E.G. Krebs, *FEBS-Letters*, 1974, **42**, 253.
- (66) A.J.G. Moir, J.M. Wilkinson et S.V. Perry, *FEBS-Letters*, 1974, **42**, 253.
- (67) M.M. Sanders et G.H. Dixon, *J. Biol. Chem.*, 1972, **247**, 851.
- (68) P. Cohen, D.B. Rylatt et G.A. Nimmo, *FEBS-Letters*, 1977, **76**, 182.
- (69) J. Jollès, A.M. Fiat, C. Alais et P. Jollès, *FEBS-Letters*, 1973, **30**, 173.
- (70) J.C. Mercier, J. Uro, B. Ribadeau-Dumas et F. Grosclaude, *Eur. J. Biochem.*, 1972, **27**, 535.
- (71) J.C. Mercier, J.M. Chobert et F. Addeo, *FEBS-Letters*, 1976, **72**, 208.
- (72) A. Neuberger, A. Gottschalk, R. Marshall et R.G. Spiro, In « *Glycoproteins: their Composition, Structure and Function* », (A. Gottschalk, ed.), 1972, Vol. 5, Part A, p. 450, Elsevier, Amsterdam.
- (73) A. Neuberger et R.D. Marshall, In « *Carbohydrates and their Roles* », (H.W. Schultz, R.F. Cain et R.W. Wrolstao, eds), 5th Symp. on Foods at Oregon State University, AVI Publishing, Wesport, 1968, p. 115.
- (74) R.D. Marshall et A. Neuberger, *Adv. Carb. Chem. Biochem.*, 1970, **25**, 407.
- (75) R.D. Marshall, *Ann. Rev. Biochem.*, 1972, **41**, 673.
- (76) H.C. Sox et L. Hood, *Proc. Nat. Acad. Sci. USA*, 1970, **66**, 975.
- (77) P. Charet, M. Monsigny, G. Spik et J. Montreuil, *C.R. Acad. Sci. Paris*, 1969, **D269**, 1019.
- (78) P. Degand, Communication personnelle.
- (79) K. Schmid, H. Kaufman, S. Isemura, F. Bauer, J. Emura, T. Motoyama, M. Ishiguro et S. Nanno, *Biochemistry*, 1973, **12**, 2711.
- (80) R.J. Delange et T.S. Huang, *J. Biol. Chem.*, 1971, **246**, 698.
- (81) T. Murachi et N. Takahashi, In « *Structure-Function Relationships of Proteolytic Enzymes* », (P. Desnuelle, H. Neurath et M. Ottesem, eds.), p. 298, Munksgaard, Copenhagen, 1970.
- (82) S. Iwanaga, B. Blomback, N.J. Grondhal, B. Hessel et P. Wallen, *Biochim. Biophys. Acta*, 1968, **160**, 280.
- (83) M.O. Dayhoff, In « *Atlas of Protein Sequence and Structure* », National Biomedical Research Foundation, Washington, 1972.
- (84) W.K. Liu, H.S. Nahm, C.M. Sweeney, H.N. Baker, W.M. Lamkin et D.N. Ward, *Res. Comm. Chem. Path. Pharmacol.*, 1971, **2**, 168.
- (85) T.C. Elleman et J. Williams, *Biochem. J.*, 1970, **116**, 515.
- (86) A. Shimizu, C. Paul, H. Köhler, T. Shimoda et F.W. Putnam, *Science*, 1971, **173**, 629.
- (87) R.L. Jackson et C.H.W. Hirs, *J. Biol. Chem.*, 1970, **245**, 635.
- (88) D.G. Smyth, W.H. Stein et S. Moore, *J. Biol. Chem.*, 1963, **238**, 227.
- (89) F.J. Morgan, S. Birken et R.E. Camfield, *Mol. Cell. Biochem.*, 1973, **2**, 97.
- (90) Handbook of Biochemistry and Molecular Biology, Vol. III, 1975 (G.D. Fasman, ed.) CRC Press, Cleveland.
- (91) T.H. Plummer Jr. et L. Sarda, *J. Biol. Chem.*, 1973, **248**, 7865.
- (92) L. Ridén et D. Eaker, *Eur. J. Biochem.*, 1974, **44**, 171.
- (93) M.O. Dayhoff, in « *Atlas of Protein Sequence and Structure* », Vol. V, National Biomedical Research Foundation, Washington, 1972.
- (94) R. Shainkin et G.E. Perlmann, *J. Biol. Chem.*, 1971, **246**, 2278.
- (95) R. Begbie, *Biochim. Biophys. Acta*, 1974, **371**, 549.
- (96) B. Frangione et C. Milstein, *Nature*, 1969, **224**, 597.
- (97) N. Katsura et E.A. Davidson, *Biochim. Biophys. Acta*, 1966, **121**, 120.
- (98) T. Ozeki et Z. Yosizawa, *Arch. Biochem. Biophys.*, 1971, **142**, 177.
- (99) D.S. Smyth et S. Utsumi, *Nature*, 1967, **216**, 332.
- (100) A. Tsugita et B. Akabori, *J. Biochem. (Tokyo)*, 1959, **46**, 695.
- (101) J. Baenziger et S. Kornfeld, *J. Biol. Chem.*, 1974, **249**, 7270.
- (102) M.W. Fanger et D.G. Smyth, *Biochem. J.*, 1972, **127**, 767.
- (103) J. Jollès, F. Schoentgen, C. Alais, A.M. Fiat et P. Jollès, *Helv. Chim. Acta*, 1972, **55**, 2872.
- (104) J. Jollès, A.M. Fiat, F. Schoentgen, C. Alais et P. Jollès, *Biochem. Biophys. Acta*, 1974, **365**, 335.
- (105) H.B. Brewar Jr., R. Schulman, P. Herbert, R. Roman et K. Wehrly, *J. Biol. Chem.*, 1974, **249**, 4975.
- (106) A. Hagopian, J.S. Whitehead et F.C. Westall, *Fed. Proc.*, 1971, **30**, 1186.
- (107) J.P. Aubert, G. Biserte, M.H. Loucheux-Lefebvre, *Arch. Biochem. Biophys.*, 1976, **175**, 410.
- (108) J.G. Beeley, *Biochem. Biophys. Res. Commun.*, 1977, **76**, 1051.
- (109) M. Nagarajan et V.S.R. Rao, *Curr. Sci.*, 1977, **46**, 395.
- (110) G.D. Rose, *Nature*, 1978, **272**, 586.
- (111) R. Ury et F. Wold, *Science*, 1977, **198**, 890.
- (112) S.K. Brahmachari et V.S. Ananthanarayanan, *Curr. Sci.*, 1978, **47**, 107.