

50376
1981
70

50376
1981
70

UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE LILLE

THÈSE

présentée à l'Université de Lille I
pour l'obtention du grade de
DOCTEUR ÈS SCIENCES PHYSIQUES

par

Jean-Pierre AUBERT

Le coude β dans les glycoprotéines

Organisation spatiale et N-glycosylation *in vitro*
de peptides synthétiques



soutenue le 5 mai 1981 devant la Commission d'Examen

Président : M. C. LOUCHEUX

Rapporteurs : M. G. BISERTE
M^{me} M.H. LOUCHEUX
M. P. JOLLÈS

Examineurs : M. P. DEGAND
M. A. VERBERT
M. N. SHARON

A Hélène

A Jérôme

A mes parents

*En témoignage de ma profonde reconnaissance et en remerciement
de leurs sacrifices consentis*

A Madame M.H. LOUCHEUX, Maître de Recherche au CNRS, qui m'a proposé le sujet de ce travail. Je lui dois ma formation de chercheur et lui suis très reconnaissant pour la confiance qu'elle m'a accordée lors de mon arrivée à l'IRCL.

Qu'elle veuille trouver ici l'expression de ma respectueuse gratitude.

Ce travail a été réalisé à l'institut de recherches sur le cancer de Lille (Directeur: Professeur G. BISERTE)

Monsieur C. LOUCHEUX, Professeur à l'Université des Sciences et Techniques de Lille, m'a fait l'honneur de présider le jury de cette thèse.

Qu'il veuille accepter mes respectueux remerciements.

Monsieur le Professeur G. BISERTE nous a accueilli dans son institut et a favorisé notre intégration dans un organisme de recherche. Sa haute autorité scientifique nous a permis d'avoir de fructueuses conversations sur de nombreux sujets abordés par le service de biophysique.

Nous lui exprimons aujourd'hui notre respectueux et sincère attachement.

Nous adressons nos plus vifs remerciements à Messieurs les Professeurs P. DEGAND, P. JOLLES, A. VERBERT et N. SHARON qui nous ont fait l'honneur d'accepter de juger cette thèse.

Je ne voudrais pas terminer sans remercier aussi le personnel du laboratoire du Professeur DEGAND et du laboratoire POF de l'IRCL avec qui les rapports ont toujours été faciles et les collaborations fructueuses.

Je tiens également à remercier tous mes camarades du service de biophysique qui, à titres divers, m'ont aidé dans ce travail. Au sein de ce service, la collaboration avec Mlle N. HELBECQUE a été très efficace et je tiens à lui exprimer ici ma sincère amitié. Mme M.P. HILDEBRAND a été durant de nombreuses années notre collaboratrice technique. Sa haute compétence et sa grande rigueur nous ont été d'une aide précieuse pour la réalisation de nombreux travaux dans le cadre du service commun de biophysique.

Un grand merci également à ma mère pour la patience dont elle a fait preuve lors de la dactylographie de ce mémoire

DOYENS HONORAIRES De l'Ancienne Faculté des Sciences

MM. R.DEFRETIN, H.LEFEBVRE, M.PARREAU.

PROFESSEURS HONORAIRES des Anciennes Facultés de Droit
et Sciences Economiques, des Sciences et des Lettres

MM. ARNOULT, Mme BEAUJEU, BONTE, BROCHARD, CHAPPELON, CHAUDRON, CORDONNIER,
CORSIN, DECUYPER, DEHEUVELS, DEHORS, DION, FAUVEL, FLEURY, P.GERMAIN, GLACET,
GONTIER, HEIM DE BALSAC, HOCQUETTE, KAMPE DE FERIET, KOURGANOFF, LAMOTTE,
LASSERRE, LELONG, Mme LELONG, LHOMME, LIEBAERT, MARTINOT-LAGARDE, MAZET,
MICHEL, PEREZ, ROIG, ROSEAU, ROUELLE, SAVARD, SCHILTZ, WATERLOT, WIEMAN,
ZAMANSKI.

PROFESSEUR EMERITE

M. A.LEBRUN.

ANCIENS PRESIDENTS DE L'UNIVERSITE
DES SCIENCES ET TECHNIQUES DE LILLE

MM. R.DEFRETIN, M.PARREAU, J.LOMBARD.

PRESIDENT DE L'UNIVERSITE
DES SCIENCES ET TECHNIQUES DE LILLE

M. M.MIGEON.

PROFESSEURS - CLASSE EXCEPTIONNELLE

M. DURCHON Maurice	Biologie Expérimentale
M. GABILLARD Robert	Electronique
M. HEUBEL Joseph	Chimie Minérale
M. MONTREUIL Jean	Biochimie
M. PARREAU Michel	Analyse
Mme SCHWARTZ Marie-Hélène	Géométrie
M. TRIDOT Gabriel	Chimie Appliquée
M. VIVIER Emile	Biologie Cellulaire
M. WERTHEIMER Raymond	Physique Atomique et Moléculaire

PROFESSEURS - 1ère Classe

M. BACCHUS Pierre	Astronomie
M. BEAUFILS Jean-Pierre	Chimie Physique
M. BECART Maurice	Physique Atomique et Moléculaire
M. BIAYS Pierre	Géographie
M. BILLARD Jean	Physique du Solide
M. BONNOT Ernest	Biologie Végétale

M. BOUGHON Pierre	Algèbre
M. BOURIQUET Robert	Biologie Végétale
M. CELET Paul	Géologie Générale
M. COEURE Gérard	Analyse
M. CONSTANT Eugène	Electronique
M. CORDONNIER Vincent	Informatique
M. DEBOURSE Jean-Pierre	Gestion des Entreprises
M. DELATTRE Charles	Géologie Générale
M. ESCAIG Bertrand	Physique du Solide
M. FAURE Robert	Mécanique
M. FOOT Jacques	Génie Mécanique
M. FOURET René	Physique du Solide
M. GRANELLE Jean-Jacques	Sciences Economiques
M. GRUSON Laurent	Algèbre
M. GUILLAUME Jean	Microbiologie
M. HECTOR Joseph	Géométrie
M. LABLACHE-COMBIER Alain	Chimie Organique
M. LACOSTE Louis	Biologie Végétale
M. LANSRAUX Guy	Physique Atomique et Moléculaire
M. LAVEINE Jean-Pierre	Paléontologie
M. LEHMANN Daniel	Géométrie
Mme LENOBLE Jacqueline	Physique Atomique et Moléculaire
M. LHOMME Jean	Chimie Organique Biologique
M. LOMBARD Jacques	Sociologie
M. LOUCHEUX Claude	Chimie Physique
M. LUCQUIN Michel	Chimie Physique
M. MAILLET Pierre	Sciences Economiques
M. PAQUET Jacques	Géologie Générale
M. POUZET Pierre	Analyse Numérique
M. PROUVOST Jean	Minéralogie
M. SALMER Georges	Electronique
M. SEGUIER Guy	Electrotechnique
M. STANKIEWICZ François	Sciences Economiques
M. TILLIEU Jacques	Physique Théorique
M. VIDAL Pierre	Automatique
M. ZEYTOUNIAN Radyadour	Mécanique

PROFESSEURS - 2ème Classe

M. AL FAKIR Sabah	Algèbre
M. ANTOINE Philippe	Analyse
M. BART André	Biologie Animale
Mme BATTIAU Yvonne	Géographie
M. BEGUIN Paul	Mécanique
M. BELLET Jean	Physique Atomique et Moléculaire
M. BKOUCHE Rudolphe	Algèbre
M. BOBE Bernard	Sciences Economiques
M. BODARD Marcel	Biologie Végétale
M. BOILLY Bénoni	Biologie Animale
M. BOIVIN Jean-Claude	Chimie Minérale
M. BONNELLE Jean-Pierre	Catalyse
M. BOSCOQ Denis	Probabilités
M. BREZINSKI Claude	Analyse Numérique
M. BRIDOUX Michel	Chimie Physique
M. BRUYELLE Pierre	Géographie
M. CAPURON Alfred	Biologie Animale
M. CARREZ Christian	Informatique
M. CHAMLEY Hervé	Géotechnique
M. CHAPOTON Alain	Electronique

M.	COQUERY Jean-Marie	Psychophysiologie
Mme	CORSIN Paule	Paléontologie
M.	CORTOIS Jean	Physique Nucléaire et Corpusculaire
M.	COUTURIER Daniel	Chimie Organique
M.	CRAMPON Norbert	Hydrogéologie et Environnement
M.	CROSNIER Yves	Electronique
Mle	DACHARRY Monique	Géographie
M.	DEBRABANT Pierre	Géologie Appliquée
M.	DEGAUQUE Pierre	Electronique
M.	DELORME Pierre	Physiologie Animale
M.	DEMUNTER Paul	Sociologie
M.	DE PARIS Jean-Claude	Analyse
M.	DEPREZ Gilbert	Physique du Solide et Cristallographie
M.	DERIEUX Jean-Claude	Microbiologie
Mle	DESSAUX Odile	Spectroscopie de la Réactivité Chimique
M.	DEVRAINNE Pierre	Chimie Minérale
M.	DHAINAUT André	Biologie Animale
Mme	DHAINAUT Nicole	Biologie Animale
M.	DORMARD Serge	Sciences Economiques
M.	DOUKHAN Jean-Claude	Physique du Solide
M.	DUBOIS Henri	Spectroscopie Hertzienne
M.	DUBRULLE Alain	Spectroscopie Hertzienne
M.	DUEE Gérard	Géologie
M.	DYMENT Arthur	Mécanique
Mme	EVRAUD Micheline	Chimie Appliquée
M.	FLAMME Jean-Marie	Technologie de Construction
M.	FONTAINE Hubert	Dynamique des Cristaux
M.	FONTAINE Jacques	Electronique, Electrotechnique, Automatique
M.	FOURNET Bernard	Biochimie Structurale
M.	GERVAIS Michel	Gestion
M.	GLORIEUX Pierre	Physique Moléculaire et Rayonnements Atmosphériques
M.	GOBLOT Rémi	Algèbre
M.	GOSELIN Gabriel	Sociologie
M.	GOUDMAND Pierre	Chimie Physique
M.	GREVET Patrick	Sciences Economiques
M.	GUILBAULT Pierre	Physiologie Animale
M.	HENRY Jean-Pierre	Génie Mécanique
M.	HERMAN Maurice	Physique Spatiale
M.	HOUDART René	Physique Atomique et Moléculaire
M.	JACOB Gérard	Informatique
M.	JACOB Pierre	Probabilités et Statistiques
M.	JACQUILLAT Bertrand	Gestion
M.	JOURNEL Gérard	Spectroscopie Hertzienne
M.	KREMBEL Jean	Biochimie
M.	LAURENT François	Automatique
Mme	LECLERCQ Ginette	Catalyse
Mle	LEGRAND Denise	Algèbre
Mle	LEGRAND Solange	Algèbre
Mme	LEHMANN Josiane	Analyse
M.	LEMAIRE Jean	Spectroscopie Hertzienne
M.	LENTACKER Firmin	Géographie
M.	LEROY Jean-Marie	Méthodologie
M.	LEROY Yves	Electronique, Electrotechnique, Automatique
M.	LEVASSEUR Michel	Sciences Economiques
M.	LHENAFF René	Géographie
M.	LOCQUENEUX Robert	Physique Théorique
M.	LOSFELD Joseph	Informatique
M.	LOUAGE Francis	Electronique
M.	MACKE Bruno	Physique Moléculaire et Rayonnements Atmosphériques

M. MAHIEU Jean-Marie	Physique Atomique et Moléculaire
M. MAIZIERES Christian	Automatique
Mlle MARQUET Simone	Probabilités
M. NESSELYN Jean	Physique Atomique et Moléculaire
M. MIGEON Michel	Chimie Physique
M. MIGNOT Fulbert	Analyse Numérique
M. MONTEL Marc	Physique du Solide
M. MONTUELLE Bernard	Biologie et Biochimie Appliquées
Mme N'GUYEN VAN CHI Régine	Géographie
M. NICOLE Jacques	Chimie Analytique
M. NOTELET Francis	Electronique, Electrotechnique, Automatique
M. PARSY Fernand	Mécanique
Mlle PAUPARDIN Colette	Biologie Physiologie Végétales
M. PECQUE Marcel	Chimie Organique
M. PERROT Pierre	Chimie Appliquée
M. PERTUZON Emile	Physiologie Animale
M. PETIT Francis	Chimie Organique, Minérale et Analytique
M. PONSOLLE Louis	Chimie Physique
M. PORCHET Maurice	Biologie Animale
M. POVY Lucien	Automatique
M. RACZY Ladislas	Electronique
M. RAOULT Jean-François	Géologie Structurale
M. RICHARD Alain	Biologie Animale
M. RIETSCH François	Physique des Polymères
M. ROGALSKI Marc	Analyse
M. ROUSSEAU Jean-Paul	Physiologie Animale
M. ROY Jean-Claude	Psychophysiologie
M. SALAMA Pierre	Sciences Economiques
Mme SCHWARZBACH Yvette	Géométrie
M. SCHAMPS Joël	Spectroscopie Moléculaire
M. SIMON Michel	Sociologie
M. SLIWA Henri	Chimie Organique
M. SOMME Jean	Géographie
Mlle SPIK Geneviève	Biochimie
M. STERBOUL François	Informatique
M. TAILLIEZ Roger	Génie Alimentaire
M. THERY Pierre	Electronique, Electrotechnique, Automatique
M. TOULOTTE Jean-Marc	Automatique
M. VANDORPE Bernard	Chimie Minérale
M. VERBERT André	Biochimie
M. VILETTE Michel	Résistance des Matériaux
M. WALLART Francis	Spectrochimie Infrarouge et Raman
M. WATERLOT Michel	Géologie Générale
M. WERNER Georges	Informatique Fondamentale Appliquée
Mme ZINN-JUSTIN Nicole	Algèbre

GLOSSAIRE

GPRP:	Glycoprotéine riche en proline de la salive parotidienne humaine.
IgM:	Immunoglobuline. M.
DNA:	Acide désoxyribonucléique.
DMSO:	Diméthylsulfoxyde.
CaCl ₂ :	Chlorure de calcium.
SDS:	Dodécyl sulfonate de sodium.
θ :	Ellipticité molaire.
λ :	Longueur d'onde.
Å :	Angström = 10 ⁻¹⁰ mètre.
βT:	Coude β ou βturn.
DC:	Dichroïsme circulaire.
RMN:	Résonance magnétique nucléaire.
Dol:	Dolichol.
Man:	Mannose.
UDP Glc Nac:	N acétylglucosamine.
Glc:	Glucose.

TABLE DES MATIERES

<u>AVANT PROPOS</u>	p. 1
A) INTRODUCTION	p. 2
B) ETUDE DE L'INTERACTION CHAINE GLYCANNIQUE - CHAINE PROTEIQUE	p. 6
<u>I) TECHNIQUES UTILISEES</u>	p. 7
<u>A) Calcul des prédictions de conforma- tion des protéines</u>	p. 7
<u>B) Le Dichroïsme Circulaire</u>	p. 11
- Dichroïsme circulaire des protéines	p. 11
- Dichroïsme circulaire des sucres	p. 15
- Applications du dichroïsme circulaire	p. 15
<u>II) RÉSULTATS</u>	p. 17
<u>A) L'orosomucoïde ou alpha₁-glycopro- téine acide</u>	p. 18
1) <u>Introduction</u>	p. 18
2) <u>Résultats et discussion</u>	p. 19
3) <u>Conclusions</u>	p. 20
Article n°1	p. 20
<u>B) Etude de la glycoprotéine riche en proline de la salive parotidienne humaine (GPRP)</u>	p. 21
1) <u>Introduction</u>	p. 21
2) <u>Obtention des produits</u>	p. 21
3) <u>Résultats</u>	p. 22
4) <u>Conclusions</u>	p. 23

<u>III) CONCLUSIONS GÉNÉRALES</u>	p. 24
C) MISE EN EVIDENCE DU ROLE DU COUDE β DANS LA N-GLYCOSYLATION DES PROTEINES	p. 27
<u>I) TECHNIQUES UTILISÉES</u>	p. 28
<u>II) ÉVALUATION DE LA STRUCTURE SECONDAIRE DE LA PARTIE PEPTIDIQUE AU NIVEAU DE LA LIAISON CHAÎNE GLYCANNIQUE-CHAÎNE PROTÉIQUE</u>	p. 28
<u>A) Introduction</u>	p. 28
<u>B) Résultats</u>	p. 29
<u>C) Conclusions</u>	p. 30
Article n°2	p. 30
<u>III) RELATION STRUCTURE-N-GLYCOSYLATION <i>in vitro</i> DE PEPTIDES SYNTHÉTIQUES</u>	p. 33
<u>A) Introduction</u>	p. 33
<u>B) Méthodes utilisées</u>	p. 35
<u>C) Résultats</u>	p. 36
<u>1) Glycosylation</u>	p. 36
<u>2) Dichroïsme circulaire</u>	p. 38
<u>D) Conclusions</u>	p. 40
Article n°3	p. 40

D)	CONCLUSIONS ET PERSPECTIVES	p. 43
E)	APPENDICE TECHNIQUE	p. 46
I)	<u>ENONCÉ DES RÈGLES PERMETTANT DE RECHER-</u> <u>CHER LES STRUCTURES ORGANISÉES SELON CHOU ET</u> <u>FASMAN</u>	p. 47
A)	<u>Recherche des régions α hélicoïdales</u>	p. 47
B)	<u>Recherche des feuillets plissés β</u>	p. 47
C)	<u>Recherche des coudes β</u>	p. 48
II)	<u>DICHROÏSME CIRCULAIRE</u>	p. 48
A)	Origine phénoménologique	p. 48
B)	Conformation des chaînes peptidiques	p. 49
	BIBLIOGRAPHIE	p. 52

AVANT-PROPOS.

Les glycoprotéines forment avec les glycolipides l'importante classe des glycoconjugués. Les glycoprotéines sont des macromolécules formées de deux espèces macromoléculaires totalement différentes et liées de façon covalentes.

La partie protéique est formée d'un enchaînement linéaire d'acides aminés de configuration L et la partie glycannique d'un enchaînement linéaire ou ramifié de glucides.

De nombreuses équipes de chercheurs ont compris depuis quelques années l'intérêt des glycoprotéines tant au point de vue biologique que pathologique. De très gros efforts ont été faits en particulier pour déterminer les séquences primaires aussi bien de la partie protéique que de l'enchaînement glucidique, pour connaître la biosynthèse ainsi que le rôle des glycoprotéines. Ces travaux n'ont jamais été faciles à réaliser étant donné la diversité des composants des deux macromolécules constituant une glycoprotéine ainsi que la très grande microhétérogénéité due le plus souvent à la partie glycannique.

Lorsque nous avons entrepris notre travail, l'étude des glycoprotéines avait pris un nouvel essor depuis quelques années mais pas ou peu de travaux avaient été entrepris dans le domaine de l'organisation spatiale des glycoprotéines. C'est donc dans cette voie que nous nous sommes efforcé de poursuivre nos recherches.

A) INTRODUCTION.

Les glycoprotéines étant formées de deux constituants, leur rôle dépend de la nature de chacun d'eux et de leur organisation réciproque.

Le rôle de la partie peptidique est directement lié à la nature, à l'enchaînement des acides aminés constitutifs et par conséquent à la conformation, comme dans toutes les protéines. On peut citer comme exemple les glycoprotéines membranaires formées de régions d'acides aminés hydrophobes permettant l'interaction avec les lipides et de régions d'acides aminés hydrophiles permettant la mise en contact des chaînes glycaniques avec le milieu extracellulaire ou intracellulaire.

Les rôles joués par la partie glycanique sont peut-être un peu mieux définis:

- La coupure des résidus d'acides sialiques terminaux des glycoprotéines sériques signe leur " arrêt de mort " par un transfert immédiat de ces molécules du plasma aux hépatocytes ou elles sont dégradées (Théorie d' Ashwell) (Ashwell et Morell 1974). Des récepteurs de surface des cellules hépatiques nommées H B P (Hepatic binding protein) spécifiques de résidus galactose terminaux, ont été mis en évidence. Ces récepteurs sont eux-mêmes des glycoprotéines (Kawasaki et Ashwell 1976) .

- Les glycoprotéines présentes à la surface des cellules peuvent être modifiées lors de phénomènes pathologiques et en particulier dans le cas de tumeurs. On peut observer soit l'augmentation (Codington et al 1978) soit la diminution (Vaheiri et Mosher 1978) de certaines glycoprotéines.

- Des perturbations de propriétés physiques de la surface cellulaire telles l'adhésion et la reconnaissance cellulaire (Pena et Hughes 1978) ont leur origine dans les modifications glucidiques.

- Les glycoprotéines peuvent jouer le rôle d'antigènes de surface et de récepteur de virus (Marchesi et al 1976).

- Dans certains cas, les glycoprotéines ne peuvent être excrétées de la cellule qu'à la condition d'être glycosylées. On a longtemps pu croire d'ailleurs que les chaînes glycaniques ne servaient qu'à l'excrétion (Eylar 1965).

- Les chaînes glycaniques ont parfois un rôle protecteur vis-à-vis de l'action des protéases (Gilbert et al 1973). En conclusion, nous pouvons dire que les glycoprotéines se retrouvent dans pratiquement tous les systèmes biologiques.

Parallèlement à ces études sur le rôle des glycoprotéines s'est développée la connaissance de la biosynthèse de celles-ci. Nous avons résumé, dans un schéma emprunté à Lodish et Rothman, la biosynthèse de la glycoprotéine du V.S.V. (Virus de la Stomatite Vésiculaire) (Lodish et Rothman 1979) (Figure I).

La protéine est synthétisée sur les polysomes attachés aux membranes du réticulum endoplasmique. Pour les protéines N-glycosylées, c'est à ce niveau qu'a lieu la glycosylation; elle se fait par l'intermédiaire de lipides transportant une chaîne oligomannosidique et c'est au niveau du Golgi que la synthèse des chaînes glycaniques est terminée. La glycoprotéine achevée apparaît ensuite au niveau de la membrane plasmique. Pour la O-glycosylation, le mécanisme n'est pas encore établi avec certitude. Il semble cependant que la mannosylation des résidus sérine ou thréonine se fasse par l'intermédiaire de lipides (Bretthawer et Wu 1975 ; Gold et Hahn 1976 ; Soliday et Kolattukudy 1979) alors que la fixation d'un résidu galactosamine (Babczinshi 1980) ou galactose (Iannino et al 1979) se fasse par transfert du nucléotide-ose correspondant.

D'après ce que nous venons de dire, nous voyons que la N-glycosylation des protéines n'est pas un phénomène post-

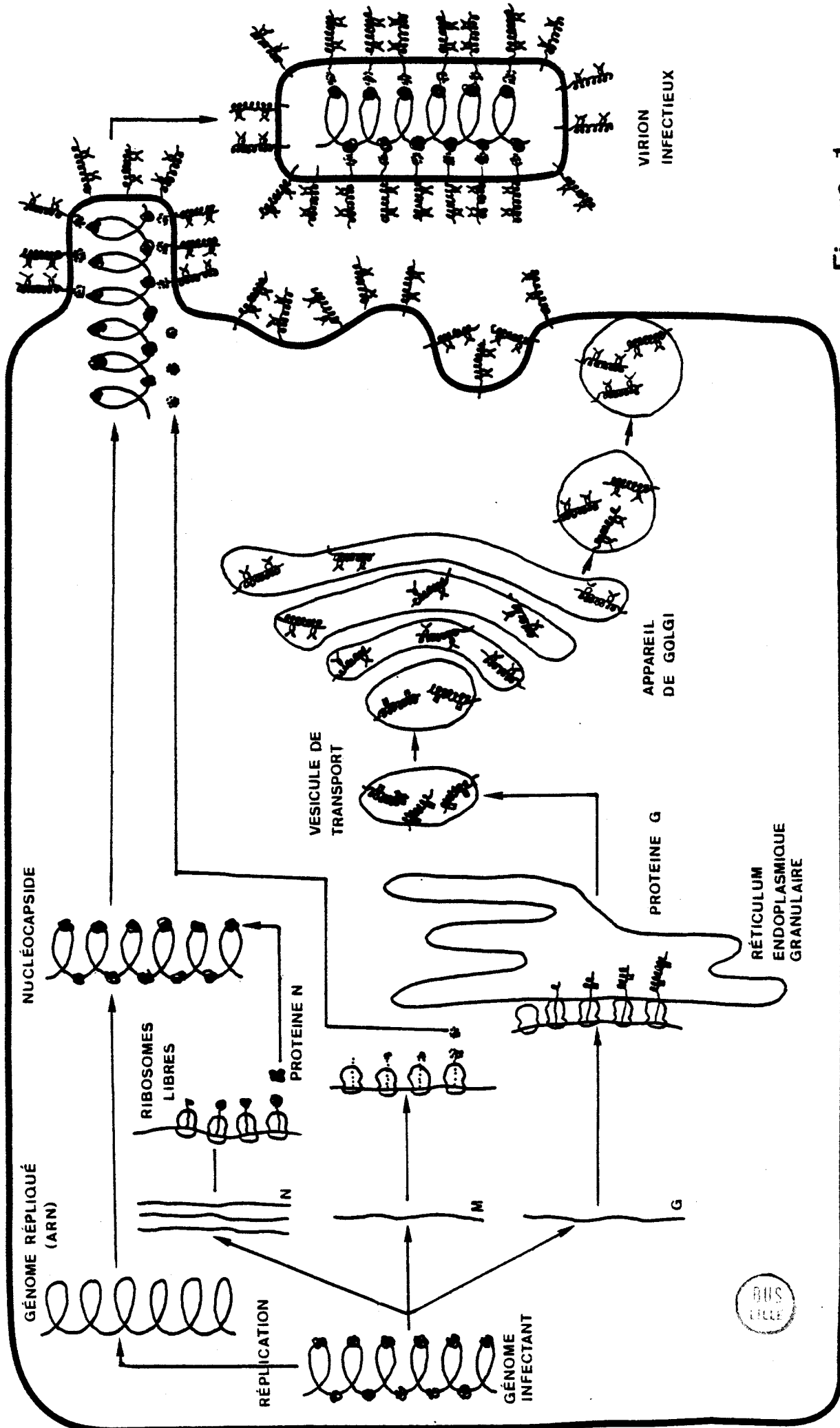


Figure 1

TRANSCRIPTION DE M, N TRANSCRIPTION DE G FIN DE LA GLYCOSYLATION
 TRANSCRIPTION DE G DEBUT DE LA GLYCOSYLATION



transcriptionnel comme l'acétylation ou la phosphorylation (Turkington et Topper 1966; Singh et al 1967) mais une réaction ayant lieu au cours de l'élongation de la chaîne peptidique.

D'autre part il semble, d'après des travaux très récents, que lorsqu'une protéine possède plusieurs sites de N-glycosylation, les chaînes glycaniques d'un même variant soient identiques. Ceci est vrai pour l' α_1 -foetoprotéine (Bayard et Kerckaert 1981) et pour un variant de l' α_1 -glycoprotéine acide (Bayard et Kerckaert 1980). Ce résultat n'était pas évident: en effet lorsqu'on isole une glycoprotéine, elle possède toujours une hétérogénéité glycanique.

Ceci amène à se poser une question: une cellule fabrique-t-elle un type unique de chaîne glycanique ou existe-t-il plusieurs systèmes de glycosyl transférases dans une même cellule? Le problème reste entier et ceci nous montre que nous sommes encore loin de connaître parfaitement la biosynthèse des glycoprotéines.

Lorsque nous avons entrepris ce travail, très peu de choses avaient été publiées concernant la structure et l'organisation spatiale d'une glycoprotéine et ceci était essentiellement dû à la dualité des composants et à la difficulté de les séparer l'un de l'autre. De plus, la liaison qui unit ces deux macromolécules peut être de nature différente.

Rappelons en effet que les glycoprotéines sont classées en deux grands types:

- Les protéines O-glycosylées impliquant un résidu sérine ou thréonine d'une part et un résidu soit de xylose, de N-acétylgalactosamine, de mannose, de galactose ou fucose d'autre part. Dans ce type de glycoprotéines on peut également rencontrer mais plus rarement des liaisons impliquant un résidu d'hydroxyllysine et un résidu de galactose.

- Les protéines N-glycosylées impliquant toujours un résidu asparagine et un résidu N-acétylglucosamine.

Généralement les glycoprotéines sont soit N- soit O-glycosylées. On peut cependant rencontrer dans une même glycoprotéine les deux types de liaison, comme par exemple la chaîne β de l'hormone H C G (Hormone Chorion Gonadotrope) et la glycophorine qui est une glycoprotéine membranaire.

Avec le développement des techniques spectroscopiques et physiques en général, on voit paraître de plus en plus d'études physicochimiques des glycoprotéines. Un des exemples le plus significatif est celui de l'étude des chaînes glycaniques par résonance magnétique nucléaire à haut champ. Ce travail réalisé par les laboratoires de Vliegthart et Montreuil a mis en évidence le fait qu'il était possible de connaître la structure complète d'une chaîne glycanique par R.M.N. du proton à 360 MHz (Dorland et al 1977 a) et b) ; Dorland et al 1978 a) et b); Strecker et al 1978).

La diffusion de neutrons commence également à faire son apparition dans l'étude des glycoprotéines (Martel et al 1980; Li et al 1980). Etant donné qu'il est très difficile de cristalliser une glycoprotéine, les chercheurs utilisent donc d'autres techniques pour approcher l'organisation spatiale de ces dernières.

Dans ce travail nous ne considérerons que les protéines N-glycosylées et en particulier la structure secondaire de la partie peptidique au niveau du point d'attache de la chaîne glycanique.

On sait que pour qu'une protéine puisse être N-glycosylée, il faut que dans sa structure primaire existe une séquence code Asn-X-Ser /Thr (Marshall et Neuberger 1970) dans laquelle X est n'importe quel résidu d'acide aminé sauf celui de la proline. Nous avons donc entrepris de vérifier si ce codage constituait la condition nécessaire et suffisante à la N-glycosylation ou si elle n'était que nécessaire. Pour cela nous avons étudié la structure secondaire au niveau de cette

séquence code. Nous avons utilisé le dichroïsme circulaire ainsi que des calculs d'évaluation de structure secondaire pour démontrer l'existence d'une structure privilégiée autour de la liaison sucre-protéine. Ces études ont été réalisées sur des glycoprotéines natives et sur des peptides synthétiques modèles; ces derniers constituent la partie peptidique de glycopeptides isolés de glycoprotéines natives. Nous avons ainsi pu étudier la N-glycosylation *in vitro* de ces peptides synthétiques et établir une corrélation entre structure et glycosylation. Nous avons également cherché à voir si une interaction forte existait entre les chaînes glycaniques et la partie peptidique. Nous avons surtout cherché à établir si cette éventuelle interaction pouvait interdire l'étude de la structure secondaire de la partie peptidique par des méthodes classiques.

Nous présenterons donc tout d'abord les travaux qui nous ont permis d'affirmer qu'il n'existait pas d'interactions chaînes glycaniques-chaîne protéique; ensuite nous mettrons en évidence le rôle d'une conformation privilégiée et nous montrerons qu'il s'agit d'un coude β -lors de la N-glycosylation des protéines.

B) ETUDE DE L'INTERACTION CHAINE GLYCANNIQUE -CHAINE
PROTEIQUE.

I) TECHNIQUES UTILISÉES.

Deux techniques principales ont été utilisées. Il s'agit du calcul de la prédiction de conformation des protéines d'une part et du dichroïsme circulaire d'autre part. Nous développerons donc ces deux techniques en précisant que les détails seront donnés dans l'appendice technique.

A) Calcul des prédictions de conformation des protéines.

Au fur et à mesure que paraissaient les structures primaires de protéines ainsi que leur structures tertiaires déterminées par cristallographie, les chercheurs ont tenté d'établir une corrélation entre structure primaire et structure tertiaire puis structure primaire et structure secondaire. De ce fait au cours de ces dernières années, de nombreuses méthodes de prédiction de conformations ont été publiées (pour une revue voir Stenberg et Thornton 1978).

Une des premières séries de méthodes de visualisation de la structure secondaire des protéines fut illustrée par les travaux de Schiffer et Edmundson (Schiffer et Edmundson 1967) et de Prothero (Prothero 1966). Ces auteurs utilisent le concept de zones hydrophobes qui, dans une protéine, auront tendance à former des structures organisées; à cette époque, il n'était d'ailleurs question que d'hélice α . Les différentes méthodes de représentation utilisées permettent de visualiser les interactions entre les différentes hélices. De plus, les interactions intrahélicoïdales, qui peuvent être très importantes pour la stabilité thermodynamique (Némethy et Scheraga 1962; Guzzo 1965), ne doivent pas être négligées.

Ces vingt dernières années, grâce à l'étude par rayons X permettant de déterminer la structure tridimensionnelle d'un nombre important de protéines, les techniques de

prévision de structure secondaire se sont beaucoup développées.

Les premières tentatives ne concernaient que la détermination et la localisation des fragments hélicoïdaux. Ces méthodes, basées sur des déterminations statistiques, ont permis de diviser les différents acides aminés en "hydrophobes" ou "hydrophiles" (Ptitsyn 1969; Finkelstein et Ptitsyn 1971; Palau et Puigdomenech 1974). A chaque acide aminé était ainsi attribué un pouvoir stabilisateur ou déstabilisateur d'hélices.

Kabbat et Wu (Kabbat et Wu 1973; Kabbat et Wu 1974) ont été les premiers à s'intéresser non seulement à la prévision des structures hélicoïdales mais aussi à celles en feuillet plissés β . Ces auteurs ont considéré les voisins les plus proches d'un résidu d'acide aminé donné. Ils ont donc calculé le nombre d'apparitions d'un résidu d'acide aminé (n) dans une conformation donnée en se basant sur la nature des résidus (n-1) et (n+1) . Ce calcul a été réalisé à partir de douze protéines. Les renseignements que donne cette approche peuvent être très intéressants pour la détermination du début ou de la fin d'un segment organisé.

Ptitsyn (Ptitsyn 1974) a proposé une analyse statistique basée sur la compétition entre les liaisons hydrogène des chaînes latérales polaires courtes et les liaisons hydrogène de la chaîne peptidique. Cette compétition serait essentielle à l'organisation d'une protéine.

En 1974 sont apparues deux nouvelles méthodes: celle de Lim (Lim 1974) et celle de Chou et Fasman (Chou et Fasman 1974).

- Lim a utilisé un algorithme pour la prédiction des régions en hélice α ou en structure β dans les protéines globulaires. Pour utiliser cet algorithme, Lim tient compte tout d'abord des interactions à courte distance des acides aminés hydrophobes (Cys, Cys-Cys, Val, Ile, Leu, Met, Phe, Tyr, Trp, Ala, Pro,) ou hydrophiles (Lys, Arg, Glu, Gln, His, Ser, Thr, Asp, Asn,). Il obtient ainsi des domaines auxquels il suf-

fit d'appliquer certaines règles empiriques d'interactions à longue distance pour trouver les régions hélicoïdales. Ces règles empiriques ont été déterminées à partir de 25 protéines. On applique ensuite le même traitement aux régions restantes pour y déceler les régions en feuillets plissés β . Il semble que seul Lim sache utiliser facilement son algorithme. Nous avons pour notre part eu beaucoup de difficultés à l'exploiter correctement.

A la même époque, Chou et Fasman publièrent leur méthode de prédiction de conformation qui est, elle, extrêmement facile à utiliser. Ils déterminent d'abord l'aptitude que présente chacun des vingt acides aminés à adopter telle ou telle structure selon leur fréquence d'apparition dans une conformation donnée. Cette fréquence est tirée de compilations faites sur un certain nombre de protéines dont la structure tertiaire est connue par cristallographie. Les fréquences de chaque acide aminé dans une structure donnée ont d'ailleurs été affinées depuis (Chou et Fasman 1977; Chou et Fasman 1978). La méthode de Chou et Fasman permet la prédiction de trois structures organisées: l'hélice α , la structure β et le coude β .

D'autres méthodes permettent également de telles déterminations comme par exemple celles de Burgess et al (Burgess et al 1974), Nagano (Nagano 1977) et Robson et Suzuki (Robson et Suzuki 1976). Ces trois dernières méthodes sont avec celle de Chou et Fasman les plus intéressantes car elles permettent la prédiction des coudes β ; mais elles nécessitent l'utilisation de l'ordinateur.

La méthode de Chou et Fasman est donc la plus facile à utiliser et elle s'avère être parmi les plus précises. En effet la comparaison de ces différentes méthodes montre qu'aucune ne semble être plus précise qu'une autre (Schulz et al 1974; Lenstra 1977).

Une autre méthode peut parfois apporter des renseignements complémentaires pour la détermination des régions hélicoïdales. C'est celle de Lewis et Bradbury (Lewis et Bradbury 1974). Cette méthode tient compte des interactions électrostatiques (attraction ou répulsion) le long de l'hélice. Si on considère un résidu i chargé, on tient compte de ses interactions possibles avec les résidus $i\pm 1$, $i\pm 2$, $i\pm 3$, $i\pm 4$ et $i\pm 7$. Si le nombre de charges répulsives est important, on admet que l'hélice est déstabilisée.

Nous avons donc choisi de développer le modèle de Chou et Fasman. Ayant cet outil en mains nous l'avons utilisé non seulement pour résoudre nos problèmes mais aussi pour réaliser un certain nombre de collaborations (Loucheux-Lefebvre et Aubert 1976; Aubert et al 1976 a) et b); Aubert et al 1977b; Loucheux et al 1978; Fiat et al 1980).

Nous ne rappellerons ci-après que les principes qui ont servi à la mise au point des méthodes de prédiction; les détails techniques qui y correspondent seront développés dans l'appendice.

La fréquence d'apparition d'un acide aminé dans une structure donnée est donc obtenue à partir de compilations faites sur des protéines dont la structure secondaire est connue par cristallographie.

La fréquence des vingt résidus d'acides aminés dans les états α , β et βT est pour une protéine:

$$F_{j,k} = \frac{n_{j,k}}{n_j}$$

α , β et βT représentent respectivement la structure α hélicoïdale, les feuillets plissés β et les coudes β ou β turns. $n_{j,k}$ est le nombre de fois où le résidu j apparaît dans la structure k .

n_j est le nombre de fois où le résidu j apparaît dans la protéine.

Les paramètres conformationnels d'un résidu j pour les trois structures k dans une protéine sont obtenus par:

$$P_{j, k} = \frac{Pr_{j,k}}{Pr_j}$$

$Pr_{j,k}$ est la probabilité de trouver le résidu j dans la structure k .

Pr_j est la probabilité de trouver le résidu j dans la protéine.

Pour la recherche des coudes β , il faut en plus tenir compte de la position de chaque acide aminé dans le coude β . On détermine ainsi une autre probabilité :

$$p(\beta T) = f_i \times f_{i+1} \times f_{i+2} \times f_{i+3}$$

C'est une probabilité dans laquelle $f_i, f_{i+1}, f_{i+2}, f_{i+3}$ représentent les fréquences d'existence des différents résidus en position $i, i+1, i+2, i+3$.

Au cours du temps, les paramètres ont été affinés en fonction des nouvelles structures primaires et tertiaires qui étaient connues. Ce sont les valeurs les plus récentes qui sont données dans les tableaux I et II (Chou et Fasman 1977; Chou et Fasman 1978).

B) Le dichroïsme circulaire.

- Dichroïsme circulaire des protéines

Nous ne développerons ici que l'utilisation du dichroïsme circulaire à la détermination de la structure secondaire des protéines. Les détails concernant cette technique seront donnés dans l'appendice.

Rappelons tout d'abord qu'une protéine peut adopter

Tableau I

Acides aminés	Code 3 lettres	Code 1 lettre	Hélice α P< α >	Structure β P< β >	β -turn P< β T>
Alanine	Ala	A	<u>1.42</u> H α	0.83 i β	0.66
Arginine	Arg	R	0.98 i α	0.93 i β	0.95
Asparagine	Asn	N	0.67 b α	0.89 i β	<u>1.56</u>
Acide aspartique	Asp	D	1.01 I α	0.54 B β	<u>1.46</u>
Cystéine	Cys	C	0.70 i α	1.19 h β	1.19
Acide glutamique	Glu	E	<u>1.51</u> H α	0.37 B β	0.74
Glutamine	Gln	Q	1.11 h α	1.10 h β	0.98
Glycocolle	Gly	G	0.57 B α	0.75 b β	<u>1.56</u>
Histidine	His	H	1.00 I α	0.87 i β	0.95
Isoleucine	Ile	I	1.08 h α	<u>1.60</u> H β	0.47
Leucine	Leu	L	<u>1.21</u> H α	1.30 h β	0.59
Lysine	Lys	K	1.16 h α	0.74 b β	1.01
Méthionine	Met	M	<u>1.45</u> H α	1.05 h β	0.60
Phénylalanine	Phe	F	1.13 h α	1.38 h β	0.60
Proline	Pro	P	0.57 B α	0.55 B β	<u>1.52</u>
Serine	Ser	S	0.77 i α	0.75 b β	<u>1.43</u>
Thréonine	Thr	T	0.83 i α	1.19 h β	0.96
Tryptophane	Trp	W	1.08 h α	1.37 h β	0.96
Tyrosine	Tyr	Y	0.69 b α	<u>1.47</u> H β	1.14
Valine	Val	V	1.06 h α	<u>1.70</u> H β	0.50

Paramètres conformationnels concernant les trois structures organisées :
hélice α , structure β et coude β .

(Les valeurs soulignées représentent les meilleurs stabilisateurs de la structure concernée. Les lettres H, h, I, i, B et b représentent respectivement la faculté de formation (H, h), l'"indifférence" (I, i) et la rupture (B, b) d'une structure donnée).



Tableau II

Acides aminés		Position i	Position i+1	Position i+2	Position i+3
Ala	A	0,086	0,089	0,052	0,060
Arg	R	0,066	<u>0,132</u>	0,076	<u>0,104</u>
Asn	N	<u>0,159</u>	<u>0,100</u>	<u>0,224</u>	0,088
Asp	D	<u>0,190</u>	<u>0,115</u>	<u>0,172</u>	0,081
Cys	C	<u>0,122</u>	0,027	<u>0,122</u>	0,095
Gln	Q	0,073	<u>0,106</u>	0,033	<u>0,114</u>
Glu	E	0,057	0,057	0,098	0,049
Gly	G	<u>0,120</u>	<u>0,120</u>	<u>0,194</u>	<u>0,162</u>
His	H	<u>0,175</u>	0,032	<u>0,111</u>	0,016
Ile	I	0,058	0,045	0,032	0,083
Leu	L	0,032	0,037	0,041	0,083
Lys	K	0,057	<u>0,109</u>	0,083	<u>0,109</u>
Met	M	0,087	0,087	0,022	0,087
Phe	F	0,032	0,021	0,074	0,042
Pro	P	<u>0,134</u>	<u>0,277</u>	0,036	0,080
Ser	S	<u>0,150</u>	<u>0,143</u>	<u>0,135</u>	<u>0,135</u>
Thr	T	<u>0,105</u>	<u>0,140</u>	0,085	<u>0,115</u>
Trp	W	0,073	0,018	0,073	<u>0,200</u>
Tyr	Y	<u>0,114</u>	0,071	<u>0,114</u>	<u>0,129</u>
Val	V	0,033	0,061	0,025	0,045

Fréquences d'apparition des acides aminés en fonction de leur place dans le coude β .

(Les valeurs soulignées contribuent à stabiliser le coude β).

quatre structures principales : l'hélice α , la structure en feuillets plissés β , le coude β (ou virage β ou β turn) et la structure désordonnée.

Les deux premières structures sont parfaitement définies. Il s'agit de l'hélice α de Pauling caractérisée par des angles Φ et Ψ ayant une même valeur le long de l'axe polypeptidique et de la structure en feuillets plissés β pouvant exister sous forme parallèle ou antiparallèle. Cette dernière est d'ailleurs favorisée par l'existence des coudes β .

Le coude β (figure II) est formé d'un enchaînement de quatre résidus d'acides aminés. Venkatachalam (Venkatachalam 1968) a défini de façon théorique ces coudes β qui expliquent le repliement obligatoire d'une chaîne peptidique. Leur existence a ensuite été mise en évidence par les cristallographes grâce à l'amélioration du pouvoir de résolution des mesures de RX. Jusque là, les coudes β étaient assimilés à de la structure inorganisée.

Deux conditions sont nécessaires à la formation d'un coude:

+ Il faut que la distance $C^{\alpha}_i - C^{\alpha}_{i+3}$ soit inférieure ou égale à $5,7 \text{ \AA}$ ce qui permet la formation d'une liaison hydrogène entre le groupement amide du résidu i et le groupement carbonyle du résidu $i+3$. Cette liaison hydrogène n'est pas nécessaire, mais elle stabilise le coude.

+ Il faut également que les angles Φ_{i+1}, Ψ_{i+1} et Φ_{i+2}, Ψ_{i+2} adoptent certaines valeurs particulières.

On a ainsi pu déterminer 11 types de turns différents.
(Tableau III)

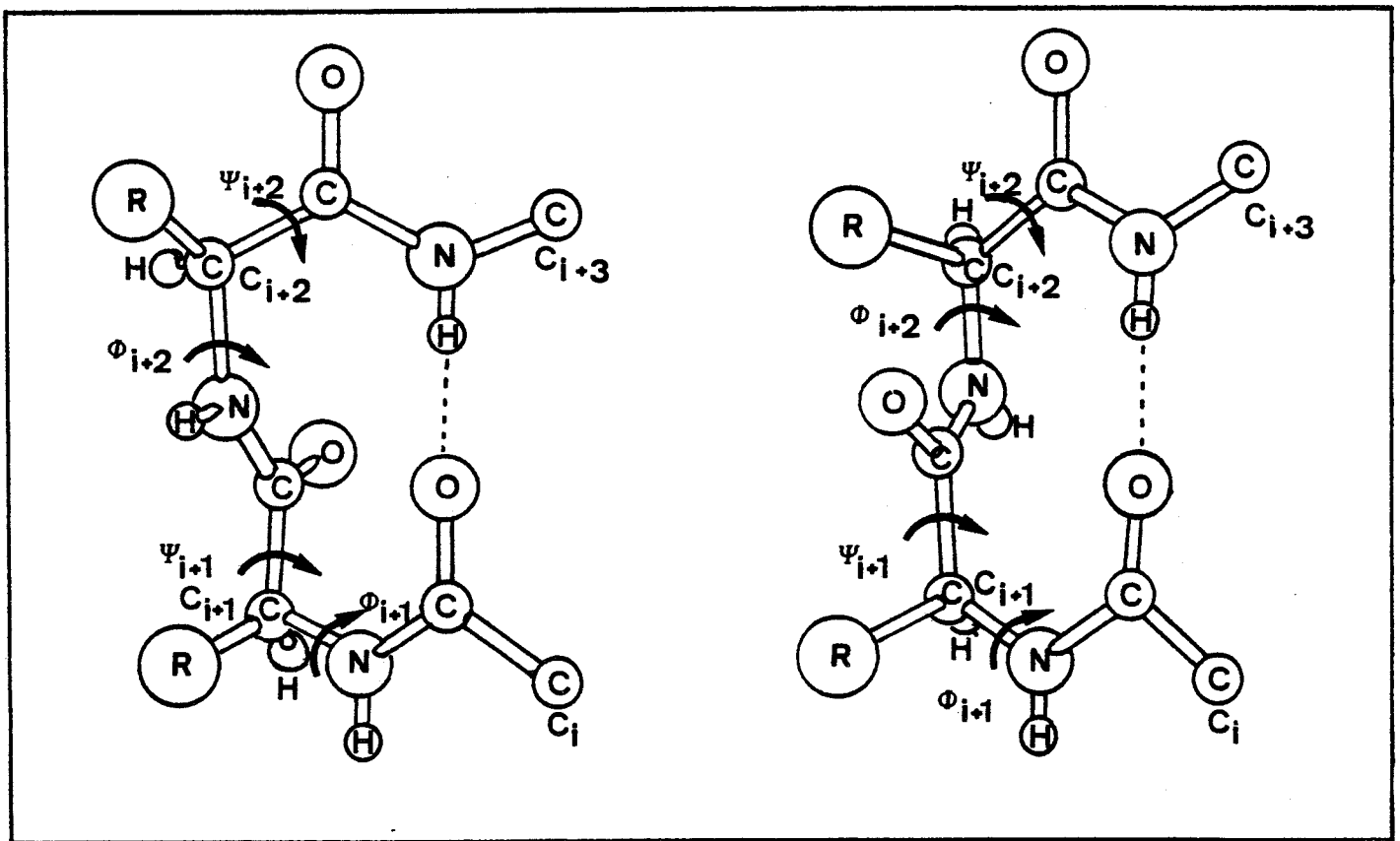


Figure 2

Tableau III

Type de coude	Angles diédraux en degrés			
	Φ_{i+1}	ψ_{i+1}	Φ_{i+2}	ψ_{i+2}
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 Φ et ψ considérés différent 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 $\psi_{i+1} \approx 180^\circ$ et $ \Phi_{i+1} < 60^\circ$ ou $ \psi_{i+1} < 60^\circ$ et $\Phi_{i+1} \approx 180^\circ$.			



Les résultats cristallographiques montrent que ce sont les coudes de type I qui sont rencontrés le plus souvent dans les protéines, puis ceux de type II, et enfin ceux de type III.

La structure inorganisée quant à elle est caractérisée par le fait qu'un certain nombre d'acides consécutifs possède des angles ϕ et ψ dont les valeurs ne sont caractéristiques d'aucune structure organisée.

En ce qui concerne les signaux dichroïques propres à chaque structure, nous pouvons dire que seul celui de l'hélice α est parfaitement défini quoique variant un peu avec la longueur des segments hélicoïdaux (Rinaudo et Domard 1976). Pour la structure β , les valeurs varient quelque peu mais à notre connaissance, il reste un problème à résoudre quant aux spectres des structures parallèles et antiparallèles.

Des spectres expérimentaux de peptides adoptant la structure en coude β ont été publiés (Howard et al 1975; Brahms et al 1977; Kawai et Fasman 1978; Bush et al 1978; Brahmachari et al 1979; Siemon et al 1980). Là encore, apparaît une grande diversité. Des spectres de coudes β ont été calculés par Woody (Woody 1974) et il semble qu'ils soient actuellement les meilleures références. Une revue concernant les coudes β dans les protéines vient d'être publiée par Smith et Pease (Smith et Pease 1980).

Malgré ces difficultés, il a quand même été possible d'aborder l'étude de la conformation des protéines par le dichroïsme circulaire mais aucune protéine n'adoptant 100% d'une structure organisée, le problème des spectres références a été délicat à résoudre .

Il a fallu attendre 1971 (Saxena et Wetlanfer 1971) puis 1972 (Chen et al 1972) pour que Chen et al mettent au

point une détermination de spectres dichroïques de chaque structure de référence à partir de protéines dont la structure secondaire était parfaitement connue par cristallographie. Dans ce premier calcul, ils ne considèrent que trois structures : l'hélice α , la structure β et le "reste" incluant la structure inordonnée et les coudes β . Ils mettent à profit la propriété d'additivité du dichroïsme circulaire $[\theta]$ pour écrire à une longueur d'onde donnée λ .

$$[\theta]_{i\lambda} = F_{i\alpha} [\theta]_{\alpha\lambda} + F_{i\beta} [\theta]_{\beta\lambda} + F_{iR} [\theta]_{R\lambda}$$

dans laquelle i désigne une protéine connue par cristallographie, F_α , F_β et F_R les pourcentages respectifs de structure en hélice α , en feuillets plissés β et en "reste".

Il suffit donc théoriquement d'étudier au moins trois protéines i connues par cristallographie pour pouvoir résoudre cette équation. A l'inverse, lorsque l'on connaît les structures de base, il suffit d'utiliser cette même équation pour obtenir cette fois le pourcentage de chacune des structures dans une protéine inconnue. Cela se fait à l'aide d'un programme de traitement par ordinateur basé sur une méthode de moindres carrés non linéaires. De plus on impose les deux conditions suivantes:

$$F_\alpha + F_\beta + F_R = 1$$

$$0 \leq F_i \leq 1$$

Au laboratoire, J.M. Freyssinet a essayé de réaliser le même travail en tenant compte des 4 structures. Chang et al (Chang et al 1978) ont également tenté de faire cette étude. Les résultats obtenus par ces auteurs ne sont pas concluants car il existe de nombreux types de coudes β , qui ne sont pas parfaitement différenciés par cristallographie, mais qui donnent des signaux différents en dichroïsme circulaire.

En pratique, on constate que plus il y a d'hélices α dans une protéine, plus il est facile de déterminer sa structure secondaire de façon précise. Parallèlement, s'il y a beaucoup de structure β , il est impossible d'obtenir la structure secondaire car le signal dichroïque est très petit et fortement perturbé par des bandes dues aux chaînes latérales aromatiques. Un article récent (Brahms et Brahms 1980) montre que la détermination est beaucoup plus précise si on travaille à très basse longueur d'onde c'est à dire entre 200 et 150 nm. Cela nécessite un appareillage très particulier et non commercialisé.

- Dichroïsme circulaire des sucres.

Les effets Cotton observables en dichroïsme circulaire dans le domaine de l'U.V. lointain sont ceux dus au chromophore amide. En effet pour les oses non acétylés, les bandes d'absorption se situent à basse longueur d'onde (environ 160 nm) et ne sont donc pas accessibles avec un dichrographe classique.

Les sucres possédant le chromophore acétamido montrent 2 bandes dichroïques à environ 210 et 190 nm. Par analogie avec le chromophore CO-NH peptidique, ces bandes sont décrites comme les transitions $n \rightarrow \pi^*$ et $\pi \rightarrow \pi^*$

- Application du dichroïsme circulaire.

Outre la détermination de la structure secondaire d'un peptide ou d'un polypeptide, d'une protéine ou d'une glycoprotéine, le dichroïsme circulaire permet de réaliser un certain nombre d'études. Il est impossible ici de donner toutes les applications et la bibliographie concernant ce sujet; nous n'en citerons que quelques exemples.

Le dichroïsme circulaire permet d'étudier la stabilité d'une protéine en fonction de différents agents dénaturants, d'étudier la stabilité et la structure d'un DNA en

fonction de la température, d'étudier les interactions protéine-protéine ou protéine-ligand ou protéine - DNA, d'étudier toutes les molécules organiques susceptibles d'être dichroïques, d'étudier les oligosaccharides.

Nous avons, pour notre part, établi un certain nombre de collaborations avec différents groupes qui nous ont permis d'utiliser les nombreuses possibilités d'un dichrographe et ont conduit aux publications suivantes: (Aubert et al 1975; Aubert et Loucheux-Lefebvre 1975; Degand et al 1976; Le Gaillard et al 1976; Aubert et Loucheux-Lefebvre 1976; Aubert et al 1977 a); Aubert et al 1977 b); Loucheux-Lefebvre et al 1978; Aubert et al 1981 a); Mazurier et al 1976).

II) RESULTATS

Nous avons donc utilisé ces techniques à l'étude de deux glycoprotéines : l' α_1 -glycoprotéine acide et la glycoprotéine riche en proline de la salive parotidienne humaine (GPRP).

Cette étude est compliquée par le fait qu'il est impossible d'obtenir la partie protéique indépendamment de la chaîne glycanique. Aussi, allons nous rappeler brièvement les différentes techniques qu'il est possible d'utiliser pour isoler chacune des chaînes.

Nous en citerons trois. La première d'entre elles permet d'isoler la chaîne glycanique, la seconde la chaîne peptidique, la troisième chacune des chaînes constituantes.

- L'hydrazinolyse (Bayard et Roux 1975) est une méthode chimique hydrolysant toutes les liaisons CO-NH. Si on la fait suivre d'une réacétylation (Reading et al 1978), on récupère donc les chaînes glycaniques intactes. Par contre, la chaîne peptidique est dégradée entièrement sous forme d'acides aminés libres.

- Le traitement par l'acide fluorhydrique (Lenard 1969; Mort et Lamport 1977) a pour but de détruire les chaînes glycaniques en ne laissant au point d'attache qu'un seul résidu de N-acétylglucosamine plus ou moins dégradé. L'utilisation de cette technique peut parfois produire des coupures de l'axe peptidique.

- L'action d'une 4-L-aspartyl glucosylamine amidohydrolase (Takahashi 1977) permet de libérer les chaînes glycaniques de la partie protéique en ne modifiant que d'une manière minime cette dernière : le résidu asparagine est transformé

en résidu acide aspartique. Il peut là aussi se produire des coupures de l'axe peptidique. Ajoutons également, qu'en général, les rendements sont meilleurs avec des glycopeptides qu'avec des glycoprotéines natives.

Ces trois techniques ont été utilisées dans les travaux que nous avons réalisés. L'hydrazinolyse a été faite en collaboration avec B. Bayard, le traitement par l'acide fluorhydrique en collaboration avec N. Porchet et l'action de la 4-L-aspartyl glucosylamine amidohydrolase en collaboration avec A. Boersma.

A) L'OROSOMUCOÏDE OU ALPHA₁-GLYCOPROTÉINE ACIDE.

1) Introduction

Nous nous sommes intéressé à l'orosomucoïde (Schmid et Kamiyama 1963) car c'est la glycoprotéine du sérum humain dont la teneur en sucres est la plus élevée (environ 40%). De plus, la séquence primaire faite de 181 résidus d'acides aminés est parfaitement établie (Schmid et al 1973) ainsi que la position des cinq chaînes glycaniques. La composition, les enchaînements et les anoméries des différents sucres constituant les chaînes glycaniques sont parfaitement connus (Schwarzmann et al 1973; Hatcher et Jeanloz 1973; Bayard et Fournet 1975).

Le dichroïsme circulaire, qui est une très bonne technique de détermination de structure secondaire de protéines, ne peut être utilisé sans précaution dans le cas des glycoprotéines car la contribution dichroïque des chaînes glycaniques dans la région 180-250 nm pourrait ne pas être négligeable.

Cette étude de l'orosomucoïde nous a permis de montrer qu'il existait un bon accord entre le dichroïsme expérimental et celui calculé à partir de l'évaluation de struc-

ture secondaire. De plus, il nous a été possible d'affirmer que, pour cette glycoprotéine, il n'existait pas de forte interaction chaîne glycanique-chaîne protéique.

2) Résultats et discussion

Nous avons joint à cette thèse l'article n° 1 correspondant à l'étude conformationnelle de l'orosomucoïde (Aubert et Loucheux-Lefebvre 1976). Nous nous contenterons donc de résumer les résultats.

+ La structure primaire de l'orosomucoïde établie par Schmid et al (Schmid et al 1973) est donnée sur la figure I (article n°1). Un certain nombre de variants existe et on dénombre 21 substitutions. La plupart de celles-ci sont conservatives. Dans le tableau I (article n°1), nous donnons les résultats obtenus à partir de différentes méthodes d'évaluation ainsi que notre choix définitif. La figure II (article n°1) donne une représentation de cette structure secondaire qui est constituée de 21% d'hélice α , 21% de structure β , 18% de coudes β , et 40% de structure inordonnée. Ces valeurs nous ont fourni un spectre dichroïque calculé de la façon suivante:

$$[\theta]_{\lambda} = F_H [\theta]_{\lambda}^H + F_{\beta} [\theta]_{\lambda}^{\beta} + F_{\beta T} [\theta]_{\lambda}^{\beta T} + F_{RC} [\theta]_{\lambda}^{RC}$$

où F_H , F_{β} , $F_{\beta T}$ et F_{RC} correspondent respectivement aux pourcentages d'hélice α , de structure β , de coude β et de structure inordonnée.

$[\theta]_{\lambda}^h$, $[\theta]_{\lambda}^{\beta}$, $[\theta]_{\lambda}^{\beta T}$ et $[\theta]_{\lambda}^{RC}$ correspondent respectivement aux ellipticités à une longueur d'onde donnée de l'hélice α (Chen et al 1974) de la structure β (Chen et al 1974), les coudes β (Venkatachalam 1968) et de la structure inorganisée (Chen et al 1974).

+ Le spectre dichroïque expérimental de l'orosomucoïde native est caractéristique d'un mélange de structures (figure V article n°1).

+ Parallèlement, le spectre dichroïque des chaînes glycaniques réacétylées a été enregistré entre 180 et 250 nm (figure III article n°1). Il est caractérisé par deux bandes dichroïques centrées à 210 nm (négative) et à 192 nm (positive) et il est comparé à la somme pondérée des spectres dichroïques des composants des chaînes glycaniques (c'est à dire : mannose, galactose, fucose, N-acétylglucosamine et acide N-acétylneuraminique). Comme on peut le voir sur la figure III de l'article n°1, il existe une très bonne superposition entre ces deux spectres. Ce résultat nous permet de dire qu'il n'existe aucune organisation privilégiée des chaînes glycaniques dans l'eau. En effet, le spectre dichroïque reste inchangé lorsque l'on passe d'un ensemble de monomères à un hétéropolymère. Il est cependant impossible de dire si ce manque d'organisation existe lorsque les chaînes glycaniques sont fixées sur la protéine.

+ En additionnant au spectre calculé à partir de l'évaluation de conformation le spectre des chaînes glycaniques et en comparant cette somme avec le spectre dichroïque expérimental de la glycoprotéine intacte, nous remarquons une très bonne superposition (figure V article n°1).

3) Conclusions

-- -----

Nous avons pu montrer que la contribution des chaînes glycaniques au dichroïsme circulaire était négligeable ce qui permet d'envisager l'étude de la structure secondaire des glycoprotéines par cette technique.

De plus, la présence des chaînes glycaniques ne perturbe pas la structure secondaire de l'axe peptidique ce qui nous indique qu'aucune interaction forte n'existe entre la partie protéique et la partie glycanique.

B) ÉTUDE DE LA GLYCOPROTÉINE RICHE EN PROLINE

DE LA SALIVE PAROTIDIENNE HUMAINE (GPRP)

Ce travail a été réalisé en collaboration avec le laboratoire du Professeur P. Degand (U.16 INSERM) et est actuellement soumis pour publication.

1) Introduction

--- -----

La GPRP est une glycoprotéine caractérisée par sa forte teneur en résidus proline, glycine et acide glutamique qui lui confèrent une structure secondaire assez particulière. En effet, nous avons pu montrer qu'elle était constituée d'environ 60% de structure hélicoïdale de type polyproline II (Aubert et al 1975).

La composition, les propriétés et le rôle de la GPRP sont discutés dans un travail de Boersma (Boersma 1977).

Dans ce chapitre, nous considèrerons l'influence des chaînes glycaniques sur la structure secondaire de la glycoprotéine (Aubert et al 1981 b).

2) Obtention des produits

--- -----

La GPRP a été déglycosylée par deux techniques: l'action de la 4-L-aspartyl glucosylamine amidohydrolase, et par action de l'acide fluorhydrique (Porchet et al 1981). Nous avons ainsi deux possibilités de comparer la GPRP native et la GPRP déglycosylée. Dans ce travail, des informations supplémentaires ont pu être obtenues grâce aux études réalisées sur des glycopeptides correspondants à la GPRP. En effet, cette dernière soumise à une hydrolyse trypsique donne des glycopeptides qui ont été caractérisés comme étant formés d'environ 20 acides aminés: ces glycopeptides ont pratiquement tous la même structure primaire (Porchet et al 1981). Ils

ont pu également être déglycosylés par l'acide fluorhydrique. Lors de cette hydrolyse sont aussi libérés des petits peptides appelés "intercalaires" qui relient entre eux les différents glycopeptides, assurant ainsi leur répétitivité.

Nous disposons également du glycopeptide obtenu après hydrolyse de la GPRP par un mélange papaïne-pronase dont la structure primaire est Gly-Asn-Gln-Ser (Degand et al 1975), ainsi que de son homologue déglycosylé par action de la 4-L-aspartyl glucosylamine amidohydrolase.

Nous possédions donc par la même occasion des chaînes glycaniques complètes.

3) Résultats

-- -----

Sur la figure III sont montrés les spectres dichroïques de la GPRP native, des glycopeptides trysiques et du glycopeptide papaïne-pronase ainsi que ceux de leurs homologues déglycosylés. Aucune différence ne peut être décelée que les chaînes glycaniques soient présentes ou non. Il est intéressant de considérer cette étude dichroïque à la lumière de ce que nous avons démontré précédemment sur la GPRP et sa stabilité vis-à-vis des ions calcium (Aubert et al 1975): ces derniers déstabilisent la structure hélicoïdale de la GPRP qui est de type polyproline II (Tiffany et Krimm 1968).

Nous avons donc repris cette étude sur chacun des constituants pour voir si les chaînes glycaniques pouvaient jouer un rôle protecteur ou au contraire déstabilisant vis-à-vis de la chaîne peptidique.

La figure IV montre les transitions de $[\theta]_{203 \text{ nm}}$ caractéristique de la structure en polyproline II en fonction de la concentration en chlorure de calcium. Dans tous les cas, on observe une transition entre 0.5 M et 1.5 M que le pro-

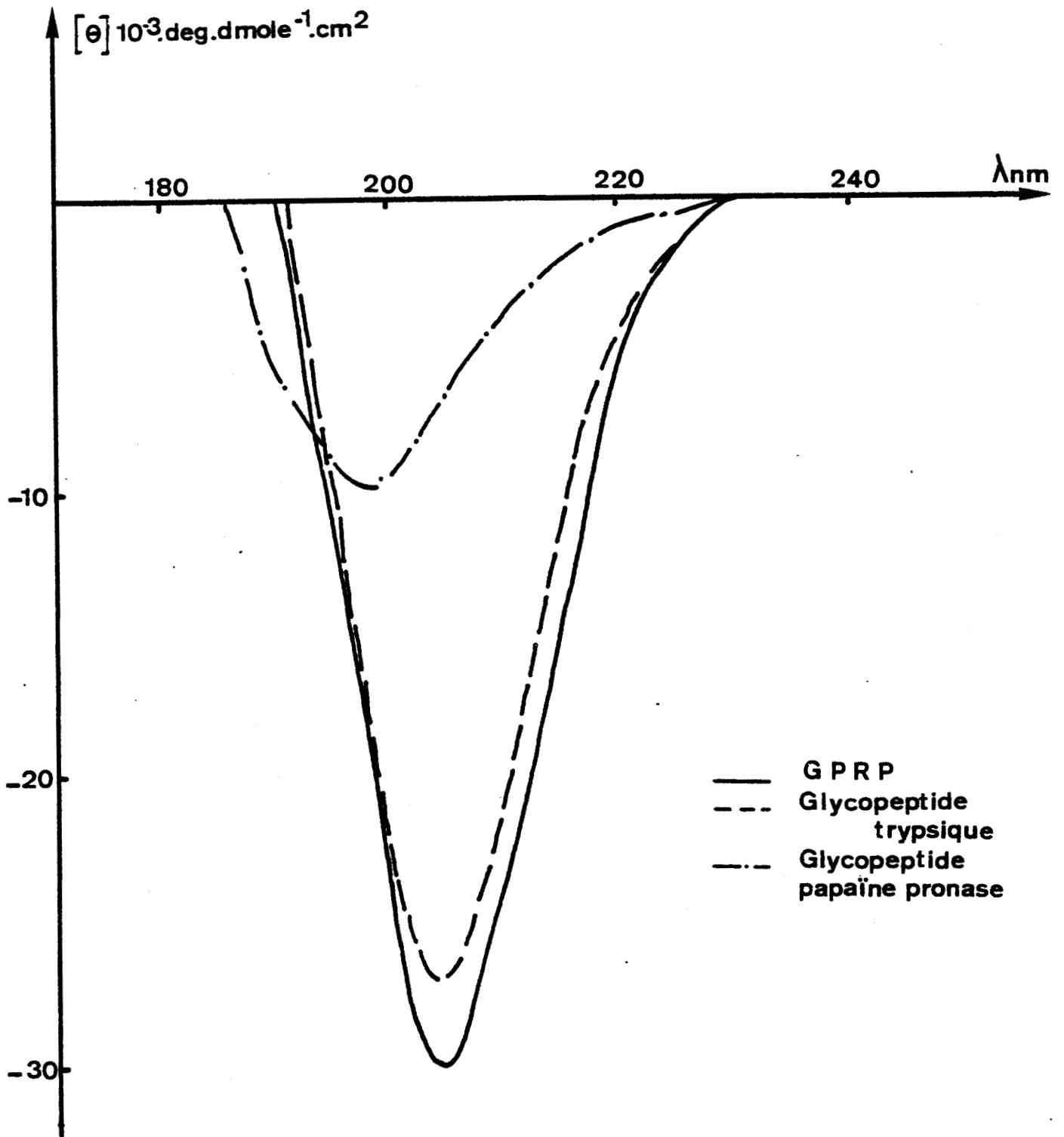
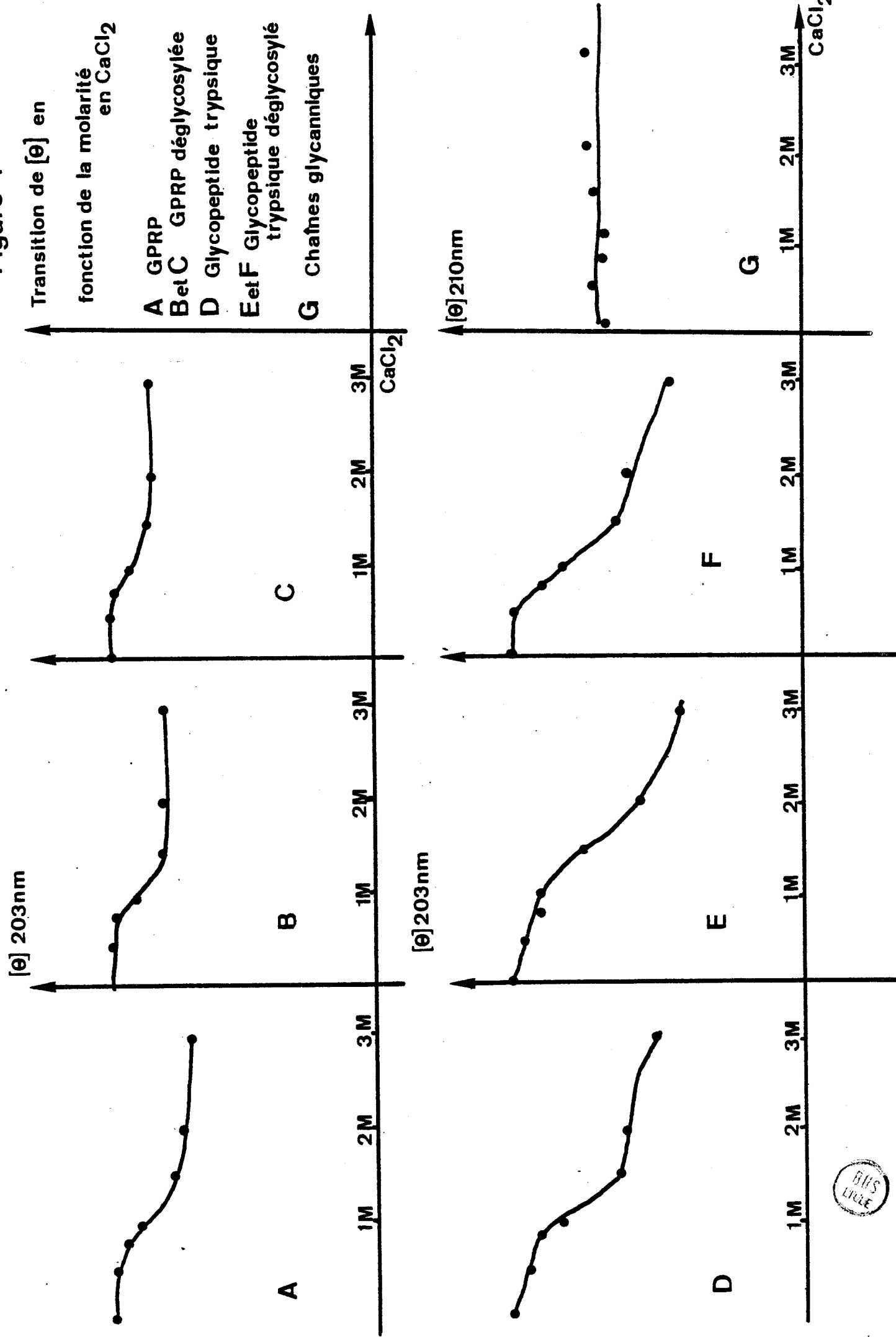


Figure 3



Figure 4



duit soit glycosylé ou pas et cela quelle que soit la longueur de la chaîne peptidique. Comme on pouvait s'y attendre, ceci n'est pas observé pour le tétrapeptide Gly-Asn-Gln-Ser qui, dans l'eau ne possède aucune organisation. Nous nous sommes de plus assuré que les chaînes glycaniques étaient insensibles à l'action du calcium.

4) Conclusions

-- -----

Dans cet exemple, nous avons donc confirmé qu'il était possible de déterminer la structure secondaire d'une glycoprotéine par dichroïsme circulaire.

Nous avons pu également observer que les chaînes glycaniques n'intervenaient en aucune façon dans la stabilité protéique vis-à-vis de l'action du calcium.

Comme la presque totalité de la structure primaire de la GPRP est à présent connue (Porchet et al 1981), il est possible de donner une représentation schématique de l'organisation de cette glycoprotéine (figure V). La GPRP serait formée d'un assemblage de glycopeptides séparés par les peptides "intercalaires". Chaque glycopeptide serait formé de deux zones adoptant une structure en hélice de type polyproline II séparées par un coude β sur lequel est fixé la chaîne glycanique. Le coude β est le tétrapeptide Gly-Gly-Asn-Gln. Les résidus 2 à 8 du côté N-terminal ainsi que les résidus 16 à 19 du côté C-terminal forment les deux zones hélicoïdales. Ceci représente 58% de structure polyproline II et 21% de coudes β . Il existe donc une très bonne corrélation entre l'estimation de structure secondaire que nous avons obtenue par dichroïsme circulaire (64% de structure polyproline II et 29% de coudes β). Etant donné que les peptides "intercalaires" sont formés de 50% de résidus proline et d'un grand nombre de résidus glycine, on peut penser que lorsque la structure primaire de la GPRP sera parfaitement établie, la corrélation entre les deux estimations sera parfaite.

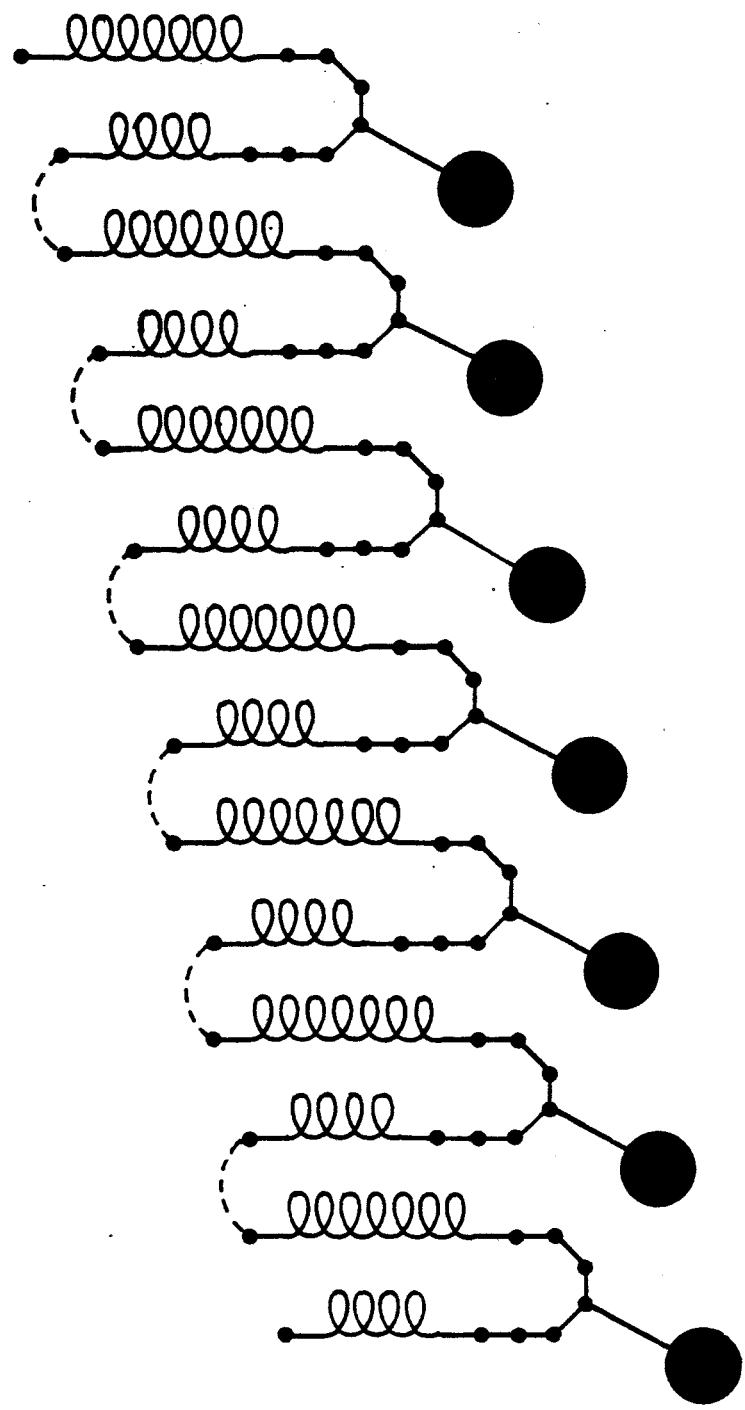
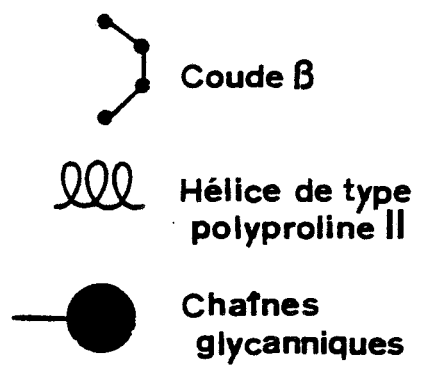


Figure 5

Structure primaire du glycopeptide tryptique

Pro Pro
 Gly Pro Gly Pro Pro Pro Gln Gly Gly Asn Gln Ser Gln Gly Pro Pro Pro
 Lys Glu



III) CONCLUSIONS GÉNÉRALES

Les résultats de ces deux études ont donc montré qu'il était possible d'étudier une glycoprotéine native par dichroïsme circulaire et de déterminer ainsi sa structure secondaire. Ceci implique que la chaîne glycanique n'amène aucune perturbation du signal dichroïque de la partie peptidique bien que dans les deux cas étudiés le pourcentage de sucres soit très élevé (40 et 50% respectivement). On peut donc penser qu'il n'existe aucune interaction forte entre les deux constituants d'une glycoprotéine. Ce résultat n'était pas évident à priori; il a cependant été confirmé par d'autres auteurs.

Wang et Hirs (Wang et Hirs 1977) ont étudié la ribonucléase de porc, soit native (3 chaînes glycaniques représentant 38% en poids de la glycoprotéine totale) soit partiellement déglycosylée par action d'exoglycosidases, technique qui ne leur a laissé qu'environ 25% des oses totaux. Ils étudient leurs échantillons dans le but d'observer une éventuelle modification de la conformation de la partie peptidique en fonction de la teneur en oses. D'autre part, l'étude de ces mêmes échantillons par absorption ultraviolette différentielle leur a permis d'obtenir des informations sur la localisation des résidus aromatiques, principalement des résidus tyrosine dont l'absorption ultraviolette varie selon qu'ils sont exposés ou non à la surface de la protéine. Ils ont ainsi pu montrer que les chaînes glycaniques n'avaient aucune influence sur la conformation globale de l'axe peptidique. D'autre part, ils ont pu affirmer que les interactions chaîne glycanique-chaîne protéique sont extrêmement localisées et ne concernent que quelques résidus situés près des liaisons sucre-protéine. En effet, lorsque les chaînes glycaniques sont partiellement hydrolysées, seules environ deux tyrosines supplémentaires sont exposées au solvant.

Ce résultat a été confirmé lors de l'étude d'une autre glycoprotéine réalisée par Wagner et al (Wagner et al 1978) par résonance magnétique nucléaire. Ces auteurs considèrent la conformation de deux inhibiteurs tryptiques: le premier étant glycosylé et provenant du colostrum de lait de vache et le second n'étant pas glycosylé et provenant du pancréas de boeuf. La structure de l'inhibiteur non glycosylé étant connu par cristallographie (Huber et al 1970; Deisenhorf et Steigemann 1975) et les structures primaires des deux inhibiteurs étant proches l'une de l'autre, ces auteurs ont comparé les modifications apportées par la chaîne glycanique à la structure spatiale des deux échantillons. Ils ont d'abord considéré la résonance des protons amide de l'axe peptidique et fait une cinétique de leur échange avec le deutérium. Ils ont ainsi pu démontrer que la structure globale était la même et que la stabilité de cette structure était identique vis-à-vis du pH et de la température. Ils étudient ensuite les résonances des protons aromatiques qu'ils peuvent parfaitement identifier. Les seules modifications qu'ils mettent en évidence concernent deux tyrosines, les résidus 21 et 23 donc situés près du site de la chaîne glycanique, le résidu asparagine 24 . A partir de la détermination des temps de relaxation, ils ont également pu démontrer que en solution, la chaîne glycanique est très étendue. De plus la détermination des rayons de giration montre qu'elle occupe un grand volume du à son haut pouvoir d'hydratation. Ce dernier résultat vient d'être confirmé par la technique de diffusion des neutrons (Li et al 1980) sur une autre glycoprotéine: la transcortine . Il a été possible de déterminer le volume occupé par les chaînes glycaniques en solution. Il est environ deux fois plus important que celui calculé sur la partie peptidique alors qu'en réalité les sucres ne représentent que 27% en poids de la glycoprotéine.

L'ensemble de ces résultats laisse à penser que

les chaînes glycaniques sont effectivement plantées sur le noyau peptidique et cela est en excellent accord avec ce que nous avons montré (Aubert et al 1976a): on trouve toujours la présence d'un coude β au niveau du point d'attache sucre-protéine. L'existence et l'étude de cette structure privilégiée feront l'objet du chapitre suivant.

ARTICLE N° 1

ARTICLE N° 1

ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS 175, 400-409 (1976)

Conformational Study of α_1 -Acid Glycoprotein

JEAN-PIERRE AUBERT AND MARIE H. LOUCHEUX-LEFEBVRE¹

*Institut de Recherches sur le Cancer de Lille et Unité n° 124 de l'Inserm, B.P. n° 3567,
59020 Lille Cédex, France*

Received December 23, 1975

The secondary structure of the protein moiety of the α_1 -acid glycoprotein (orosomucoid) was evaluated from its primary structure by using the Lim and Chou and Fasman predictions, and the corresponding dichroic spectrum was calculated. The experimental dichroic spectrum of the whole glycoprotein was compared with the summation of (i) the calculated dichroic spectrum due to the protein moiety and (ii) the experimental dichroic spectrum of the carbohydrate moiety. The results are in good agreement with the fact that the carbohydrate moieties do not produce any perturbation of the protein conformation. In addition, we observed that four out of five glycan chains are linked to Asn residues which are situated either in a reverse β -turn or in regions where charged and polar residues are numerous, that is, on the outside of the protein.

The conformation of the protein moiety of a glycoprotein is always difficult to determine from the native glycoprotein in solution. In particular, the circular dichroism (CD),² which is a versatile method for the determination of the secondary structure of a protein (1-3), cannot be used in the case of glycoprotein, since the contribution of the carbohydrate moiety in the 180-250 nm (where the dichroic bands due to the amide chromophore are observed) is generally unknown.

In some cases, the carbohydrate moiety can be separated from the protein, and therefore the dichroic spectra of glycans can be recorded. However, it is not certain that the spectrum of the free glycans is the same as those linked to the protein. For example, the glycans may adopt a secondary structure or an interaction may take place between carbohydrate and protein moieties. If the sequential amino acid structure is known, it is possible to obtain some information about the secondary structure by using methods which permit one to localize the α helical structure and β -sheet, and reverses β -turns in globular

proteins (4). We were interested in the case of α_1 -acid glycoprotein (orosomucoid) (5), the glycoprotein with the highest carbohydrate content in human serum (about 40%); its sequence of amino acids was recently completely established by Schmid *et al.* (6). Moreover, it was demonstrated that this glycoprotein contains five heteropolysaccharide groups which are linked *N*-glycosidically to asparaginyl residues, and the number of amino acids between two subsequent carbohydrate units differs considerably. The structure of glycans was almost perfectly established (7-9). The five glycan chains are linked to the protein moiety by asparaginyl residues 15, 38, 54, 75, and 85. For one glycan, an *N*-acetyl neuraminic acid is substituted by a fucose residue.

By hydrazinolysis the five glycans are separated from the protein moiety with a sugar composition exactly the same as that in native glycoprotein. Unfortunately, to date the separation of the carbohydrate moiety always has been accompanied by a degradation of amide bonds.

Some people have carried out a study of the α_1 -acid glycoprotein by using optical rotatory dispersion (ORD) neglecting the contribution of the carbohydrate to the ob-

¹ To whom correspondence should be addressed.

² Abbreviations used: CD, circular dichroism; ORD, optical rotatory dispersion.

served optical activity. In particular, 12 years ago, Schmid and Kamiyama (10) concluded from the optical rotation of α_1 -acid glycoprotein that its conformation was a nonhelical, unknown type with high stability. Then, a study of the glycoprotein using ORD was carried out (11) which admitted the possible existence of α -helical regions. Another study carried out using ORD (12) indicated a very small contribution of the carbohydrate moiety to the observed optical activity. But to date, it seems that for α_1 -acid glycoprotein no circular dichroic study (CD) has been performed.

In this work we will show that from (i) the experimental CD spectra of the glycoprotein and the carbohydrate moiety and (ii) the prediction of the secondary structure from the amino acid sequence in terms of α -helical structure (H), β -sheet (β), β -turns (T), and unordered conformation (U) and the subsequent calculations of their contribution to the CD spectrum that the carbohydrate moiety makes a small contribution to the overall CD spectra, that the predicted secondary structure gives satisfactory agreement with the observed CD spectrum, and that it involves some specific structure at the location of Asn where glycans are linked to the protein moiety. The plan adopted will be as follows: (1) determination of the CD of the protein moiety from its evaluated secondary structure; (2) experimental determination of CD of glycoprotein and glycan moiety; (3) determination of the CD due to the glycan moiety by subtracting the circular dichroism of the protein moiety calculated in 1 from the experimental dichroic spectra of the glycoprotein; and (4) comparison between the evaluated and experimental CD of glycan moiety.

PROCEDURES

First of all, the secondary structure of the protein moiety was determined by using two methods: the algorithm method of Lim (13) and the calculations given by Chou and Fasman (14). Both methods reveal the types of amino acid sequences in polypeptide chain regions of globular protein which form a regular (H or β) or unordered conformation (U) in the protein moiety. With Lim's method, this determination is made by taking into account general

"architectural" principles of packing of polypeptide chains in globular protein and considering the interactions of proteins with water molecules. With Chou and Fasman's method, H and β regions and also reverse β -turns (T) can be predicted by using conformational parameters P based on the frequency of each amino acid in the three possible structures. In some cases, and especially when Lim and Chou and Fasman determinations do not give the same results, we have used the predictions of Kabat and Wu (15) concerning the influence on secondary structure of nearest-neighbor amino acid residues. These authors determine in a sequence of three amino acids the frequency of occurrence of the middle amino acid (n) in relation to its nearest neighbors ($n - 1$) and ($n + 1$) in the α -helical domain, in the β -sheet domain, and outside these regions. In order to consider the stability of a helical sequence we have also taken into account in some cases the method given by Lewis and Bradbury (16) concerning the electrostatic attractions and repulsions from neighboring sidechains.

Secondly, the corresponding dichroic spectra were calculated. Indeed, since the optical activities of the conformations are assumed to be additive, at any fixed wavelength λ the ellipticity can be expressed by

$$[\theta] = f_H[\theta]_H + f_\beta[\theta]_\beta + f_T[\theta]_T + f_U[\theta]_U$$

where the f 's are the fractions of the four forms in the protein moiety and $[\theta]_H$, $[\theta]_\beta$, $[\theta]_T$, and $[\theta]_U$ are the reference values for pure helix, β form, β -turn, and unordered structure, respectively. Of course, since the secondary structure was evaluated from a predictive method, $\sum f = 1$ and each $f \geq 0$. To date, the regions of the protein which adopted neither an α -helical nor a β -form structure were considered as unordered structure and more especially, β -turns were included in this unordered structure. Recently, the dichroic spectrum of β -turns was treated theoretically by Woody (17). The flexibility in β -turn conformations leads to many variants, with many corresponding CD spectra, but it seems that the more likely shape of the β -turn CD spectrum is the so-called type II β -turn of Venkatachalam (18), which we have taken as a reference. For the three other reference structure CD spectra (H, β , U) we have taken the values computed by Yang *et al.* from the CD spectra of five proteins (1).

EXPERIMENTAL

Material. α_1 -acid glycoprotein from human plasma was prepared by B. Bayard, by the method of Schmid *et al.* (6). The carbohydrate moiety of α_1 -acid glycoprotein was isolated by hydrazinolysis as described by Bayard and Fournet (9). The *N*-deacetylated glycans were *N*-reacetylated by acetic anhydride. The monosaccharide composition of reacety-

lated glycans and that of glycoprotein was checked by gas-liquid chromatography after methanolysis of the carbohydrate moiety (19). Total hexoses were determined by the orcinol-sulfuric acid procedure (20), hexosamines by the Elson-Morgan reaction (21), and *N*-acetyl neuraminic acid by the diphenylamine method (22). The composition of monosaccharides in reacylated glycans and in native glycoprotein was exactly the same.

Circular dichroism study. The dichroic spectra were recorded with Jobin-Yvon dichrograph R. J. Mark III, between 180 and 250 nm. α_1 -acid glycoprotein was accurately weighed and dissolved in H₂O to give about 3 mg/ml solutions, but the exact concentration was determined by Lowry's method (23). The pH of these unbuffered solutions was 7.3 for a NaF concentration of 0.214 M. Changes in sodium fluoride concentration were made by adding small volumes of concentrated salt solutions.

For the reacylated glycans the concentration was 1.5 mg/ml and was determined by colorimetric dosage of the different monosaccharides. The CD measurements were made at different pH from an unbuffered 0.214 M NaF solution to which a little quantity of concentrated NaOH, and then HClO₄, was added. For each monosaccharide, the concentrations were adjusted to give optical densities between 0.8 and 1.5 in the spectral range explored with a given cell thickness. No Lorentz correction was performed.

For monosaccharides and glycans, molar ellipticities were calculated. For the glycoprotein, the molar ellipticities were not calculated and it is the experi-

mental spectrum which is shown; it was recorded for a concentration of 2.57 g/liter, a pathlength cell of 0.01 cm, and a sensibility of $5 \cdot 10^{-6}$.

Circular dichroism calculations. The calculations of the circular dichroism from the evaluated secondary structure of protein moiety was carried out each nm. The results are given in molar ellipticities and the mean residue weight was 119.

RESULTS AND DISCUSSION

Secondary Structure Predictions of the Protein Moiety of α_1 -Acid Glycoprotein

The complete linear acid sequence of the protein moiety of α_1 -acid glycoprotein was established by Schmid *et al.* (6) and is shown in Fig. 1. As seen in the sequential structure, many variants since exist 21 residues may show possible substitutions. The majority of these substitutions are conservative with respect to the charge, hydrophobicity, and size of the sidechains and therefore should have little effect on the secondary structure. The secondary structure of this protein moiety was investigated in two ways, by using the Lim and Chou and Fasman's methods and the result of our study is given in Table I, and this predicted conformation is presented in Fig. 2 by using the representation of Chou and Fasman: It shows 21% α -

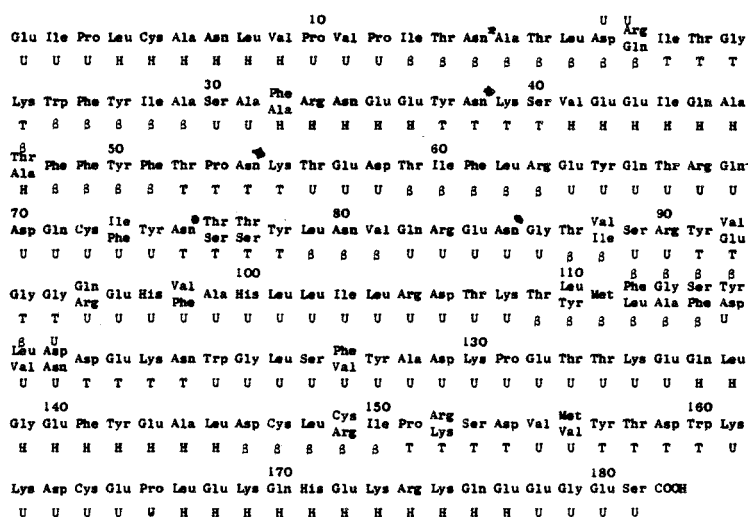


FIG. 1. The complete linear amino acid sequence of α_1 -acid glycoprotein from Schmid *et al.* (6). The five heteropolysaccharide units are linked *N*-glycosidically to the asparagine residues, which are marked with asterisks. The letters H, β , T, or U have been listed under each amino acid residue to locate the predicted helix, β -form, β -turn, β -turn, or unordered form, respectively (see text).

TABLE I
THE ESTIMATED SECONDARY STRUCTURE OF THE PROTEIN MOIETY OF α_1 -ACID GLYCOPROTEIN FROM THE PRIMARY SEQUENCE BY USING THE PREDICTIONS OF LIM (13) AND CHOU AND FASMAN (14)^a

	Chou and Fasman prediction	Lim prediction	Our choice
α helix	4-9; 30-36; 41-46 if 47 Thr; 41-47 if 47 Ala; 136-145; 167-177.	4-9; 32-47; 135-145; 168-177.	4-9; 32-36; 41-46 if 47 Thr; 41-47 if 47 Ala; 137-145; 167-177. 21, 55%
β sheet	13-20 if 20 Gln; 13-18 if 20 Arg; 25-29; 47-51 if 47 Thr; 48-51 if 47 Ala; 59-63; 65-69; 79-83; 87-90; 109-114 if 115 Asp; 109-116 if 115 Tyr; 146-150.	13-18; 19-22 if 20 Gln; 24-31; 59-64; 79-83; 108-116; 146-150.	13-18; 19-20 if 20 Gln; 25-29; 47-51 if 47 Thr; 48-51 if 47 Ala; 59-63; 79-81; 87-88; 109-114 if 115 Asp; 109-116 if 115 Tyr; 146-150. 21%
Unordered	1-3; 10-12; 19-20 if 20 Arg; 56-58; 64; 70-74; 84-86; 95-108; 115-117 if 115 Asp; 117 if 115 Tyr; 122-135; 155-156; 161-166; 178-181.	1-3; 10-12; 19-22 if 20 Arg; 23; 48-58; 65-78; 84-107; 117-134; 151-167; 178-181.	1-3; 10-12; 19-20 if 20 Arg; 30-31; 56-58; 64-74; 82-86; 89-90; 95-108; 115-117 if 115 Asp; 117 if 115 Tyr; 122-136; 155-156; 161-166; 178-181. 40, 33%
β turn	21-24; 37-40; 52-55; 75-78; 91-94; 118-121; 151-154; 157-160.		21-24; 37-40; 52-55; 75-78; 91-94; 118-121; 151-154; 157-160. 17, 68%

^a The third column gives the conformation chosen in this work. See text for discussion.

helix, 21% β -sheet, and eight reverse β -turns. Therefore, 40% of this protein moiety would adopt an unordered structure. Moreover, we have listed under each amino acid residue in Fig. 1 the letter H, β , T, or U to locate the predicted helices, β -form, β -turn, and unordered structure, respectively.

We shall only justify our choices in the case where the Lim and Chou and Fasman methods do not give exactly the same results. Of course, the determination of β -turns will not be discussed, since only one method (Chou and Fasman) gives this structure.

Residues 13-18. From the Chou and Fasman determination, the $\langle P_\beta \rangle > \langle P_\alpha \rangle$. Moreover, from the Lim determination it should have a β -form. Since no charged residues exists in this sequence, we have assigned it a β -form structure.

Residue 20. It may be either Arg or Gln. In the case where it is Gln, which is a β -sheet former (with $P_\beta = 1.23$), a β -sheet

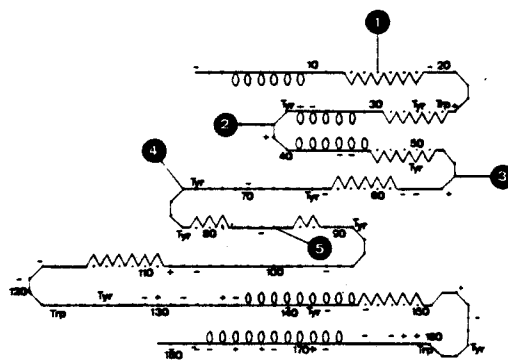


FIG. 2. Schematic diagram of helical, β -sheet, and reverse β -turn regions predicted in the protein moiety of α_1 -acid glycoprotein, based on Table I and following the representation of Chou and Fasman (14). In the present scale each spot represents a residue in its respective conformational state: helical (O), β -sheet (\wedge), coil (-). Each loop represents a single residue and not a single turn consisting of 3.6 residues. The charged residues are indicated (+, -) as well as the Tyr and Trp residues. In the scheme residues 20, 47, and 115 are the Arg, Thr, and Asp residues respectively.

structure can be nucleated for residues 20-22 ($\langle P_\beta \rangle = 1.34$).

Residues 25-29. From the Lim determination, it should be a 24-29 β -form sequence. But according to Chou and Fasman it is unlikely to find a charged residue in a β -sheet N-terminal (residue 24 is lysine). In this case, residue 24, lysine, is between Gly and Trp which, according to Kabat and Wu, is a position with no possibility for a residue to adopt a β structure. Therefore, only the 25-29 sequence adopts a β -sheet structure.

Residues 32-36 and 41-44. From the Lim calculations, the sequence 32-47 would adopt the α -helical form. From Chou and Fasman determinations the 30-36 and 41-46 sequences would be helical with a β -turn corresponding to residues 37-40. In application of rule 2 of Lim's calculations we have eliminated the 30 and 31 residues given by Chou and Fasman as helical. But in this sequence, the effect of electrostatic repulsions is too important and therefore we have assumed that it does not adopt a helical form. We have also assumed the existence of a β -turn for residues 37-40. Moreover, residue 47 is either Ala or Thr. If it is Ala, it adopts a helical structure, but if it is Thr, it belongs to the following β -form sequence. On the contrary, residue 32, which is either Phe or Ala, is always an helical former (with P_α respectively equal to 1.12 and 1.45).

Residues 59-63. From Lim's determination one finds that Glu 64 belongs to this β -sheet sequence. But it is unlikely that Glu, which is a charged residue, occurs at the C-terminal β -sheet end (Chou and Fasman).

Residues 65-69. According to the Chou and Fasman determination it would be a β -sheet sequence, but by Lim's method one finds an unordered structure. Residues 65, 67, 68, and 69 for Kabat and Wu have a very weak probability of adopting a β structure. Therefore, this sequence is taken to adopt an unordered structure.

Residues 79-81. From Lim's determination, residue 78 could adopt a β -sheet structure, but it belongs to a β -turn. Moreover, residues 82 and 83 could also adopt a

β -sheet structure. But by the Kabat and Wu laws, residue 82, between Val and Arg, and residue 83, between Gln and Glu, have no chance to adopt a β -form structure. Therefore, only the sequence 79-81 would adopt a β -sheet structure.

Residues 87-88. From the Chou and Fasman calculations, the 87-90 sequence could adopt a β -form structure, but it could adopt neither a helical nor a β -form structure according to Lim. Let us now consider the Kabat laws: Residues 87 and 88 have a weak probability of adopting a β -form conformation, but residues 89 and 90 have no chance to adopt a β -form conformation. Therefore we think that only residues 87 and 88 can adopt a β -form conformation.

Residues 109-116. From the Lim determination, residue 108 (Lys) could adopt a β -sheet structure, but charged residues occur rarely at the N-terminal β -sheet end. Moreover, this residue, which is situated between two residues Thr, has only a weak probability of adopting a β -sheet structure. Therefore, it will be considered as adopting an unordered form. Most of the residues which can be substituted occur in the sequence 109-116. Let us consider them one by one. 112 is either Phe or Leu, 113 is Gly or Ala: for these two residues, the probability of adopting a β -sheet conformation is about the same. On the contrary, for the following residues, that is, 114 (Ser or Phe) and 115 (Tyr or Asp), the structure can be changed with the residue considered. Indeed, for Ser and Phe, $\langle P_\beta \rangle$ is 0.72 and 1.28, respectively, and for Tyr and Asp, 1.29 and 0.8, respectively. If it is possible to admit Ser in the β -sheet sequence (because, even in this case, the $\langle P_\beta \rangle$ of the entire sequence is greater than 1) the probability of having Asp at the C-terminal end is near zero. Therefore we assume that this β -sheet sequence is 109-114 if we have 115_{Asp}, but 109-116 if it is 115_{Tyr}.

Residues 137-145. This sequence, 135-147, has $\langle P_\alpha \rangle = 1.18$, but residue 135, lysine, is unlikely in an N-terminal sequence. Then residue 136, glutamic acid, between Lys and Ala has no possibility to adopt a helical form (Kabat and Wu). There remains a sequence 137-145. To ob-

tain more information concerning the stability of such a helical sequence, we took into account the effect of electrostatic interactions by using the method given by Lewis and Bradbury concerning the electrostatic attractions and repulsions from neighboring sidechains. Since only one repulsion between 140_{Glu} and 143_{Glu} is present, we assume that the sequence 137-145 as helical.

Residues 146-150. This is a β -sheet sequence (with $\langle P_\beta \rangle = 1.25$). Although 146_{Asp} is a charged residue, we admit it in the β -sheet sequence, because it is at the N-terminal end (Chou and Fasman). Moreover, the residue 149 is either Lys ($P_\beta = 1.3$) or Arg ($P_\beta = 0.9$). In both cases the corresponding $\langle P_\beta \rangle$ is greater than 1 since it is, respectively, 1.25 and 1.17.

Residues 168-177. From the Chou and Fasman predictions, the sequence 167-177 would adopt a helical structure, but from the Lim determination it would be only the case of the sequence 168-177. Therefore, we used the Kabat and Wu laws to determine if the residue 167, leucine, situated between Pro and Glu has a possibility of adopting a helical structure; since this possibility is weak, residue 167_{Leu} has been eliminated from the helical sequence. On the contrary, although there exists in the sequence 168-177 a C-terminal residue Glu and an N-terminal residue, Lys, the sequence 168-177 is considered helical because in this case the electrostatic attractions from neighboring sidechains are very numerous. From this secondary structure, it is interesting to note that the β -sheets (eight strands) are almost devoid of charged residues and thus the α_1 -acid glycoprotein has a hydrophobic nucleus. On the other hand, the sequence 112-117, that is, the sequence where the substitutions of amino acids are very important, always adopts an unordered structure. The major part of the charged and polar residues are found in the C-terminal sequence (150-180) and in the sequence 34-90, where four of the five glycans are linked (Asn 38, 54, 75 and 85). Since Kendrew (24) observed that polar and nonpolar residues occupy, respectively, the outside and inside of proteins, it is interesting to note that for four

out of five glycans the linking Asn's are situated in a region where the charged residues are numerous. Of course, this conformation allows the glycan easy access to the outside of the protein. On the contrary, Asn 15 (that is the residue to which the glycan called 1 in Fig. 3 is linked) belongs to a nonpolar and uncharged β -sheet region. We do not have any explanation for this conformation: We have only checked that the determinations of Lim and Chou and Fasman are in excellent agreement for this 13-18 region (although $\langle P_\alpha \rangle = 1$). From this predicted secondary structure we may now see whether the conclusions of Yamagami and Schmid (12) and Yamagami *et al.* (25) concerning the existence of "free" and "buried" residues of tryptophan and tyrosine are in good agreement with our results. These authors indicate that 2 out of 4 tryptophan residues and 5 out of 12 of tyrosine residues were "free," 3 tyrosine slightly buried, and 4 tyrosine completely buried. In fact, in the α_1 -acid glycoprotein there exist only 3 tryptophan residues (25, 122, and 160) and 10 tyrosine residues (27, 37, 50, 65, 74, 78, 91, 127, 142, and 157). For two variants tyrosine is also present in positions 110 and 115). The relative proportion between free, slightly buried, and buried residues can be explained from Fig. 2. Residues Tyr 27, Tyr 50, Tyr 110, and Tyr 142 situated in β -sheet and nonpolar

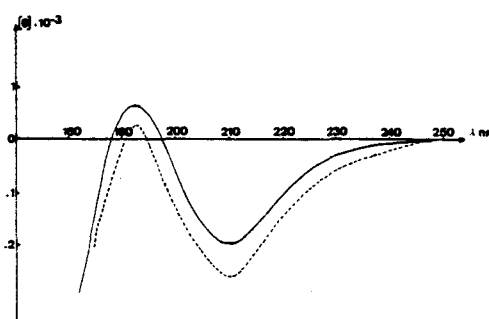


FIG. 3. Circular dichroic spectra of reacylated glycan chains (---) between 180 and 250 nm are expressed in molar ellipticities in degree $\text{cm}^2 \text{dmole}^{-1}$. The mean residue weight is 205. The concentration was about 5 g/liter in water. The full line represents the dichroic spectra calculated by linear combination of ellipticities of each monosaccharide corresponding to its proportion in glycan chains.

regions would be buried, while residues Tyr 37, Tyr 78, Tyr 91, Tyr 157, and Trp 160 situated in bends or near polar residues are free. Tyr 65, Tyr 74, Tyr 115, Tyr 127, Tyr 25, and Trp 122, in intermediate regions, would be slightly buried.

Experimental CD Spectra

α_1 -acid glycoprotein. The dichroic spectrum of α_1 -acid glycoprotein was recorded at different ionic strengths. The spectrum is always exactly the same for solutions between 5 mM and 1 M NaF.

Glycan moiety. The dichroic spectrum of reacylated glycan chains was recorded between 180 and 250 nm (Fig. 3). It is characterized by two dichroic bands centered at 210 nm (negative) and at 192 nm (positive). We have also recorded the spectra corresponding to each component of the glycan chains and in Fig. 4 one can see the dichroic spectra corresponding to D-mannose, D-galactose, D-N-acetyl glucosamine,

D-fucose, and N-acetyl neuraminic acid. As expected, for D-mannose and D-galactose, the CD spectrum shows only the tail of the CD band, which probably arises from an electronic transition in the ring oxygen (26). The opposite signs of the CD bands may be attributed to the overlap, with a second electronic transition occurring at somewhat lower wavelength, which is believed to originate in the hydroxyl group (27) and whose sign would depend on the position of the hydroxyl groups. On the contrary, N-acetyl neuraminic acid and N-acetyl glucosamine present dichroic bands between 180 and 250 nm. The CD curve of N-acetyl neuraminic acid indicates the presence of a transition centered at 195 nm (strong positive), 240 nm (weak negative), and 265 nm (weak positive) and for N-acetyl glucosamine at 207 nm (strong negative). It is tentatively assumed that these bands correspond to the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions in the N-acetyl group, though other possibilities exist (28). The curve of the glycan moiety is reminiscent of that calculated by a linear combination of the circular dichroism of each monosaccharide according to its relative proportion in glycan chains (Fig. 3). Therefore, it is evident that the transitions at 210 and 192 nm do not change significantly on going from the N-acetyl glucosamine and N-acetyl neuraminic acid to the glycan chain. Even the molar ellipticities are the same, indicating that interactions between the active chromophores can be neglected. But since it is known that oligo- and polysaccharides can adopt a secondary structure (29), which is revealed by a variation in ellipticity on going from monomer to polymer, we can claim that no organization exists in the "free" glycan chains. Indeed, no variation of dichroic spectrum was observed from reacylated glycans when pH was varied from 4 to 8. Of course, it is not possible to be certain that this lack of organization exists for the carbohydrate moiety linked to protein moiety in the glycoprotein.

Therefore, we tried to fit a calculated spectrum with the experimental spectrum of α_1 -acid glycoprotein as follows. An experimental dichroic spectrum is recorded

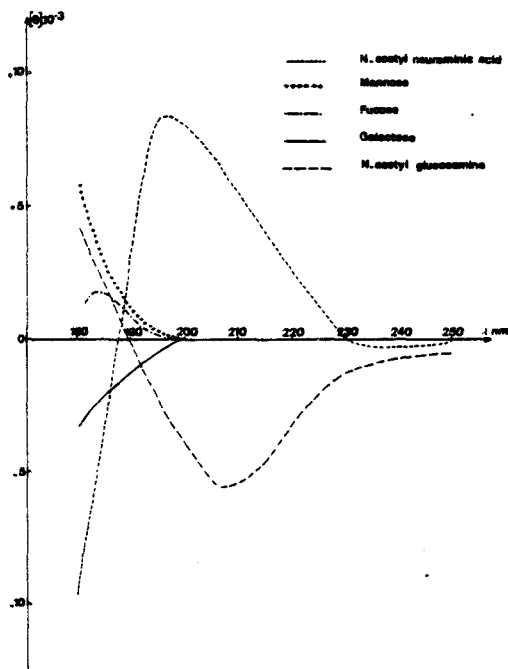


FIG. 4. Circular dichroic spectra expressed in molar ellipticities in degree $\text{cm}^2 \text{dmole}^{-1}$ of N-acetyl neuraminic acid (---), D-mannose (++++), D-fucose (·-·-·-·-), D-galactose (—), and N-acetyl-glucosamine (- -) between 180 and 250 nm. The concentrations are about 9 g/liter in water.

for α_1 -acid glycoprotein under well-known conditions, that is $c = 2.57$ g/liter, $l = 0.01$ cm, and the sensibility of the dichrograph $s = 5 \cdot 10^{-6}$ (Fig. 5). This spectrum is compared with the sum of the two following spectra:

(i) The protein contribution obtained from the evaluated secondary structure as explained above and by using the additive properties of circular dichroism. This contribution is obtained starting from the calculated spectrum expressed in molar ellipticities and normalizing to a protein concentration corresponding to the protein moiety in the above conditions, i.e., 1.42 g/liter for 2.57 g of α_1 -acid glycoprotein. The precision obtained in the determination of the evaluated structure is about 20%; then the percentages of each secondary structure are: $18.1 < H < 27.1$, $18.6 < \beta < 27.8$, $14.2 < T < 21.1$, and $29.2 < U < 43.8$ and evidently the dichroic spectra will vary from one composition to another. The two dichroic spectra corresponding to these extreme percentages were calculated (Fig. 5).

(ii) The carbohydrate contribution obtained from the glycan spectrum, expressed in molar ellipticities and normalized for the carbohydrate concentration corresponding to the carbohydrate moiety in the above conditions, i.e., 1.15 g/liter for 2.57 g of α_1 -acid glycoprotein.

As shown in Fig. 5, the fitting between the experimental and the calculated curves is rather good between 200 and 250 nm. It is not as good below 200 nm due to the error on the CD reference spectra in the 180–200 nm domain. In addition, it appears that the contribution of the carbohydrate moiety to circular dichroism between 180 and 250 nm is very small, and a determination of the secondary structure of the protein moiety from the dichroic spectra recorded for the glycoprotein in this region would be possible. On the contrary, the determination from the visible ORD spectrum is not possible, especially use the Moffitt–Yang plot. Indeed, this last determination may be carried out for the determination of the secondary structure of a protein only if the amide chromophore is present in the molecule. In the

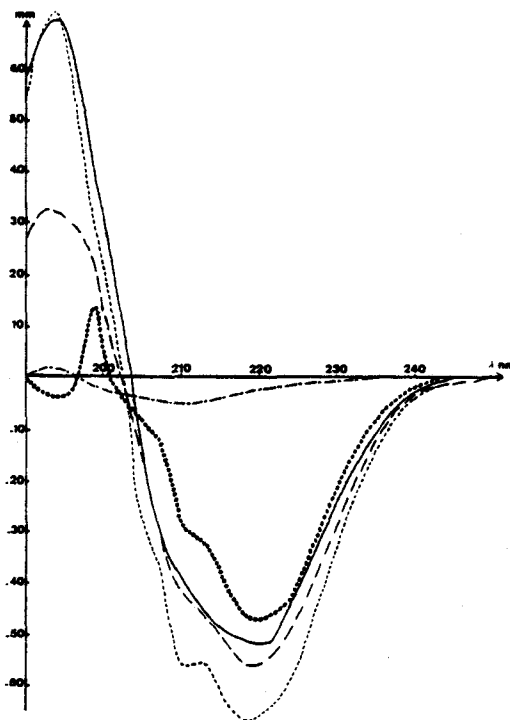


FIG. 5. Circular dichroic spectra of α_1 -acid glycoprotein under the following conditions: $C = 2.57$ g/liter, $l = 0.01$ cm, $s = 5 \cdot 10^{-6}$, in 0.214 M NaF, pH 7.3 (—). Circular dichroic spectra of the carbohydrate moiety: $C = 1.15$ g/liter, $l = 0.01$ cm, $s = 5 \cdot 10^{-6}$ in 0.214 M NaF, pH 7.3 (---). Calculated dichroic spectrum corresponding to the summation of contribution of carbohydrate and protein moieties, taking for the secondary structure of protein moiety: 21% H, 21% β , 18% T, and 40% U (— —). The calculated spectrum was normalized under conditions in which the experimental spectra of α_1 -acid glycoprotein was recorded, that is, with $l = 0.01$ cm, $s = 5 \cdot 10^{-6}$, and $C = 1.42$ g/liter for the protein moiety and 1.15 g/liter for the carbohydrate moiety. Extreme circular dichroic spectra were calculated under the conditions given above for 26% H, 27% β , 14% T, and 33% U (— · — · — ·), and for 19% H, 20% β , 21% T, and 40% U (·····).

case of glycoprotein, a band certainly exists below 180 nm due to the carbohydrate moiety and this "hidden" band makes the use of the Moffitt–Yang law impossible. It is probably due to this band situated in the far ultraviolet region that previous workers have obtained a smaller value in the experimental determination of α -helical composition.

CONCLUSION

In conclusion we can claim that this work is the first one in which the comparison between calculated and experimental circular dichroic spectra corresponding to a biopolymer with two types of macromolecular chains have been tried.

Of course, the determination of such dichroic spectra is known with a certain error which is very difficult to appreciate since the evaluated secondary structure is known with a precision of about 20%.

The contribution of the glycan chains to the optical activity in the range of 180–250 nm, where the amide chromophore shows dichroic bands is very weak and can be neglected in order to determine the secondary structure of the protein moiety from the circular dichroism of the whole glycoprotein.

The experimental dichroic spectrum of the α_1 -acid glycoprotein and the summation of the dichroic spectra due to protein and carbohydrate moieties are in good agreement with the fact that if an interaction takes place between protein and carbohydrate moieties, it does not produce any perturbation of the protein conformation, nor does it induce a new asymmetry due to their interaction. In other words, a major difference would exist between interactions in glycoproteins and in globular proteins, in which several peptidic chains are in interactions.

Four out of five glycan chains are linked to Asn residues which are situated in a reverse β -turn and in regions were charged and polar residues are numerous, that is, on the outside of the protein. We will see in the following paper that this situation of glycan chains in reverse β -turns is often observed in glycoproteins.

APPENDIX

We have seen that the ellipticity $\{\theta\}$ at any fixed wavelength of the protein moiety of the α_1 -acid glycoprotein was determined from the following expression

$$\{\theta\} = f_H\{\theta\}_H + f_B\{\theta\}_B + f_T\{\theta\}_T + f_U\{\theta\}_U$$

in which f_H , f_B , f_T , and f_U are the predicted fractions of the pure helix, β -form, β -turn, and unordered structure, respectively, and

TABLE II

ELLIPTICITIES CALCULATED BY WOODY (17)
BETWEEN 240 AND 192 nm FOR THE
TYPE II β -TURN.

λ (nm)	$\{\theta\}$ (degree cm^2 dmol^{-1})	λ	$\{\theta\}$	λ	$\{\theta\}$
240	0	220	1,400	200	20,000
238	-200	218	4,000	198	9,600
236	-400	216	8,000	196	-6,000
234	-800	214	14,800	194	-15,200
232	-1,200	212	19,200	192	-20,400
230	-1,600	210	23,600		
228	-2,000	208	28,000		
226	-2,200	206	29,600		
224	-1,600	204	29,600		
222	-400	202	26,000		

$\{\theta\}_H$, $\{\theta\}_B$, $\{\theta\}_T$, and $\{\theta\}_U$ the reference values of the ellipticity for these four forms.

Since it is the first time that such a calculation is carried out by using β -turn references, one of the referees asked to us to give the corresponding values. Therefore, you will find in Table II the values of the ellipticities calculated by Woody (17) between 240 and 192 nm for the type II β -turn.

ACKNOWLEDGMENTS

We thank Dr. B. Bayard who has carried out the preparation of *N*-reacetylated glycans used in this study. We are indebted to Professors M. Daune and G. Weill for a critical reading of the manuscript and to Professor G. Biserte whose helpful discussions were extremely valuable to us. We also thank Mrs. M. P. Hildebrand for her excellent technical assistance. The authors also thank Dr. G. D. Fasman for his helpful advice.

REFERENCES

- CHEN, J. H., YANG, J. T., AND CHAU, K. H. (1974) *Biochemistry* 13, 3350-3359.
- CHEN, J. H., YANG, J. T., AND MARTINEZ, H. M. (1972) *Biochemistry* 11, 4120-4131.
- SAXENA, V., AND WETLAUFER, D. B. (1971) *Proc. Nat. Acad. Sci. USA* 68, 969-972.
- SCHULZ, G. E., BARRY, C. D., FRIEDMAN, J., CHOU, P. Y., FASMAN, G. D., FINKELSTEIN, A. V., LIM, V. I., PTITSYN, O. B., KABAT, E. A., WU, T. T., LEVITT, M., ROBSON, B., AND NAGANO, K. (1974) *Nature (London)* 250, 140-142. This article gives a critical review concerning different methods used to determine the secondary structure of a protein from its primary structure: adenylyl kinase.

5. JEANLOZ, R. W. (1972) in *Glycoproteins* (Gottschalk, A., ed.), Vol. 5, Part A, p. 564, Elsevier, New York. In this article a review concerning the properties of α_1 -acid glycoprotein is given.
6. SCHMID, K., KAUFMAN, H., ISEMURA, S., BAUER, F., EMURA, J., MOTOYAMA, T., ISHIGURO, M., AND NANNO, S. (1973) *Biochemistry* 12, 2711-2724.
7. SCHWARZMANN, G., REINHOLD, V., AND JEANLOZ, R. W. (1973) *Colloq. Int. Centre Nat. Rech. Sci.* 221, 85-94.
8. HATCHER, V. B., AND JEANLOZ, R. W. (1973) *Colloq. Int. Centre Nat. Rech. Sci.* 221, 329-338.
9. BAYARD, B., AND FOURNET, B. (1975) *Carbohydrate Res.* 46, 75-86.
10. SCHMID, K., AND KAMIYAMA, S. (1963) *Biochemistry* 2, 271-275.
11. JIRGENSONS, B. (1965) *J. Biol. Chem.* 240, 1064-1071.
12. YAMAGAMI, K., AND SCHMID, K. (1967) *J. Biol. Chem.* 242, 4176-4181.
13. LIM, V. I. (1974) *J. Mol. Biol.* 88, 873-894.
14. CHOU, P. Y., AND FASMAN, G. D. (1974) *Biochemistry* 13, 211-245.
15. KABAT, E. A., AND WU, T. T. (1973) *Proc. Nat. Acad. Sci. USA* 70, 1473-1477.
16. LEWIS, P. N., AND BRADBURY, E. M. (1974) *Biochem. Biophys. Acta* 336, 165-177.
17. WOODY, R. W. (1974) in *Peptides, Polypeptides and Proteins* (Blout, E. R., Bovey, F. A., Goodman, M., and Lotan, N., eds.), Wiley Interscience, New York.
18. VENKATACHALAM, C. M. (1968) *Biopolymers* 6, 1425-1436.
19. ZANETTA, J. P., BRECKENRIDGE, W. C., AND VINCENDON, G. (1972) *J. Chromatogr.* 69, 291-304.
20. RIMINGTON, C. (1940) *Biochem. J.* 34, 931-940.
21. BELCHER, R., NUTTEN, A. J., AND SAMBROOK, C. M. (1954) *Analyst* 79, 20.
22. WERNER, I., AND ODIN, L. (1952) *Acta Soc. Med. Upsala* 57, 230-241.
23. LOWEY, O. H., JEBBOUGH, N. J., FARR, A. L., AND RANDALL, R. J. (1951) *J. Biol. Chem.* 193, 265-275.
24. KENDREW, J. C. (1962) *Brookhaven Symp. Biol.* 15, 216.
25. YAMAGAMI, K., LABAT, J., PANDEY, R. S., AND SCHMID, K. (1968) *Biochemistry* 7, 2873-2879.
26. PICKETT, L. W., HOEFELICK, N. J., AND LIU, T. C. (1951) *J. Amer. Chem. Soc.* 73, 4865-4869.
27. HARRISON, A. J., CEDERHOLM, B. J., AND TERWILLIGER, M. A. (1959) *J. Chem. Phys.* 30, 355.
28. WALTON, A. G., AND BLACKWELL, J. (1973) in *Biopolymers*, p. 464, Academic Press, New York/London.
29. STONE, A. L., AND MOSS, H. (1967) *Biophys. Acta* 136, 56.

C) MISE EN EVIDENCE DU ROLE DU COUDE β DANS LA N-GLYCO-
SYLATION DES PROTEINES

I) TECHNIQUES UTILISÉES.

Dans un premier temps, nous avons évalué la structure secondaire de la partie peptidique autour du point d'attache chaîne glycanne-chaîne protéique en utilisant la méthode de Chou et Fasman (Voir chapitre B § IA)

Nous avons ensuite étudié la structure secondaire de peptides synthétiques contenant la séquence code Asn-X-Ser/Thr par dichroïsme circulaire (Voir chapitre B § I B) et essayé d'établir une corrélation entre cette structure secondaire et la N-glycosylation in vitro .

Nous avons choisi comme technique de N-glycosylation la méthode de Pless et Lennarz (Pless et Lennarz 1975). Les détails techniques qui y correspondent sont donnés dans la publication n° 3.

II) ÉVALUATION DE LA STRUCTURE SECONDAIRE DE LA PARTIE

PEPTIDIQUE AU NIVEAU DE LA LIAISON CHAÎNE GLYCANNIQUE-CHAÎNE

PROTÉIQUE

Nous joignons à cette thèse l'article n°2 qui correspond à ce paragraphe.

A) Introduction

Nous avons défini au début de ce travail les deux types de liaisons N- et O-glycosidiques que l'on rencontre dans les glycoprotéines. (page 4)

Nous avons donc recherché les séquences primaires de la partie peptidique situées au niveau de ces deux types de liaisons et évalué la structure secondaire de celles-ci à l'aide de la méthode de Chou et Fasman.

L'examen des structures primaires confirme l'existence d'une séquence code Asn-X-Ser/Thr pour les liaisons N-glycosidiques alors que pour les liaisons O-glycosidiques, aucun codage ne peut être mis en évidence. On peut simplement dire dans ce dernier cas qu'il existe toujours un résidu proline aux environs immédiats de la liaison.

Nous avons donc essayé de voir s'il n'existait pas une contrainte structurale de l'axe peptidique qui dirigerait dans l'espace la position des chaînes glycaniques.

Dans ce cas, pour les liaisons N-glycosidiques, la séquence code ne serait qu'une condition nécessaire alors que pour les liaisons O-glycosidiques, ce pourrait être la structure secondaire qui serait responsable uniquement de la glycosylation (Bause et Lehle 1979).

B) Résultats

Les résultats des prédictions de conformation sont donnés dans les tableaux I et II (article n° 2).

- Liaisons O-glycosidiques

Dans ce cas, tous les résidus sérine ou thréonine impliqués sont situés dans un coude β .

- Liaisons N-glycosidiques

Dans la majorité des cas, le résidu d'asparagine impliqué est situé dans un coude β . De plus, le résidu asparagine occupe très souvent la position $i+2$ dans ce coude ce qui impose que le résidu hydroxylé est en dehors de cette conformation.

C) Conclusions et discussion

Les techniques d'évaluation de structures secondaires s'étant affinées grâce à la publication de nouveaux résultats cristallographiques et le nombre d'exemple de séquences primaires de glycoprotéines ayant augmenté, nos résultats ont été parfaitement confirmés par d'autres auteurs (Beeley 1977 ; Nagarajan et Rao 1977). Il apparait maintenant certain qu'au moins 99% des chaînes glycaniques contractant une liaison N-ou O-glycosidique sont situées dans un coude β

Aucune contrainte structurale autre que l'existence d'un tel coude n'a pu être mise en évidence. Il en est de même en ce qui concerne l'hydrophilie ou l'hydrophobicité relative autour du point d'attache.

La première fonction qui a été attribuée aux coudes β est de permettre le repliement de la chaîne protéique, donc son organisation. De ce fait, ils sont localisés à la surface de la protéine et de plus, on les trouve très souvent dans des domaines riches en acides aminés chargés. Il apparait donc naturel de penser que les chaînes glycaniques sont situées à l'extérieur de la protéine. Comme de plus, nous avons démontré dans le chapitre B qu'il n'existait aucune interaction forte entre les chaînes glycaniques et la protéine, nous pouvons donc admettre que les glycannes sont plantés comme des antennes sur le noyau protéique. Une telle organisation est en bon accord avec le modèle proposé par Montreuil (Montreuil 1975).

On a longtemps pensé que le rôle des glycoprotéines était directement lié à la nature des oses constituant les chaînes glycaniques et qu'à chaque fonction correspondait une séquence différente. Or il n'en est rien comme cela a pu être montré à partir de la détermination de structures complè-

tes de nombreuses chaînes glycaniques (Strecker et Montreuil 1979). On s'aperçoit en effet, que celles-ci répondent à deux classes précises, appelées soit " simple " et " complexe " par Kornfeld et Kornfeld (Kornfeld et Kornfeld 1976) soit " oligomannosidique " et " N-acétyllactosaminidique " par Strecker et Montreuil (Strecker et Montreuil 1979). A l'intérieur de chacune de ces classes, on peut trouver des glycoprotéines qui, bien qu'ayant des fonctions biologiques tout à fait différentes, sont caractérisées par des chaînes glycaniques identiques.

Ceci implique que si les rôles des copules glucidiques sont liés à la nature des oses, ils dépendent surtout de la position dans l'espace de ces derniers.

L'organisation de la glycoprotéine telle que nous l'avons supposée, permettrait aux chaînes glycaniques d'acquiescer dans l'espace leur orientation correcte par rapport à leur cible en leur permettant ainsi de remplir leur rôle biologique.

Ayant démontré la présence des coudes β au niveau des liaisons glycannes-protéine, nous avons essayé de voir si cette conformation était également importante lors de la biosynthèse des glycoprotéines et plus spécialement au cours de l'addition de la chaîne glycanique à la protéine en formation. L'utilisation de peptides modèles pouvant adopter en solution une structure en coude β est donc intéressante pour étudier ce mécanisme d'addition. C'est pour cette raison que nous avons entrepris une étude de N-glycosylation *in vitro* de peptides synthétiques (Chapitre C§ III)

L'importance des coudes β dans les protéines a aussi été reconnue dans de nombreux autres cas. Pour une revue on peut consulter deux articles récents (Loucheux-Lefebvre 1980 ; Smith et Pease 1980).

Nous allons dans ce paragraphe discuter du rôle des coudes β dans les protéines d'un point de vue beaucoup plus général que celui de la seule glycosylation.

En particulier, leur importance a été reconnue dans :

- La phosphorylation des protéines (Small et al 1977 et Mercier 1981).
- La conservation des coudes β au cours de l'évolution (Chou et Fasman 1978).
- L'hydroxylation des résidus proline dans le collagène (Brahmachari et Ananthanarayanan 1979).
- La coupure permettant le passage préprotéine-protéine (Austen 1979).

Les caséines kappa sont un très bon exemple pour illustrer certains des rôles des coudes β : ce sont des glycoprotéines qui possèdent un site de phosphorylation; nous en avons prédit la conformation pour deux espèces différentes : la Vache et le Mouton (Loucheux-Lefebvre et al 1978).

La figure VI montre les résultats obtenus. On voit que les sites de phosphorylation (sérine 149 pour la Vache et 151 pour le Mouton) ainsi que les sites de glycosylation (thréonine 133 et 131 pour la Vache et le Mouton respectivement) sont toujours situés dans un coude β . De plus, les modifications intervenant dans la structure primaire se font de façon à conserver les coudes β comme on peut le voir en comparant les deux schémas de la figure VI.

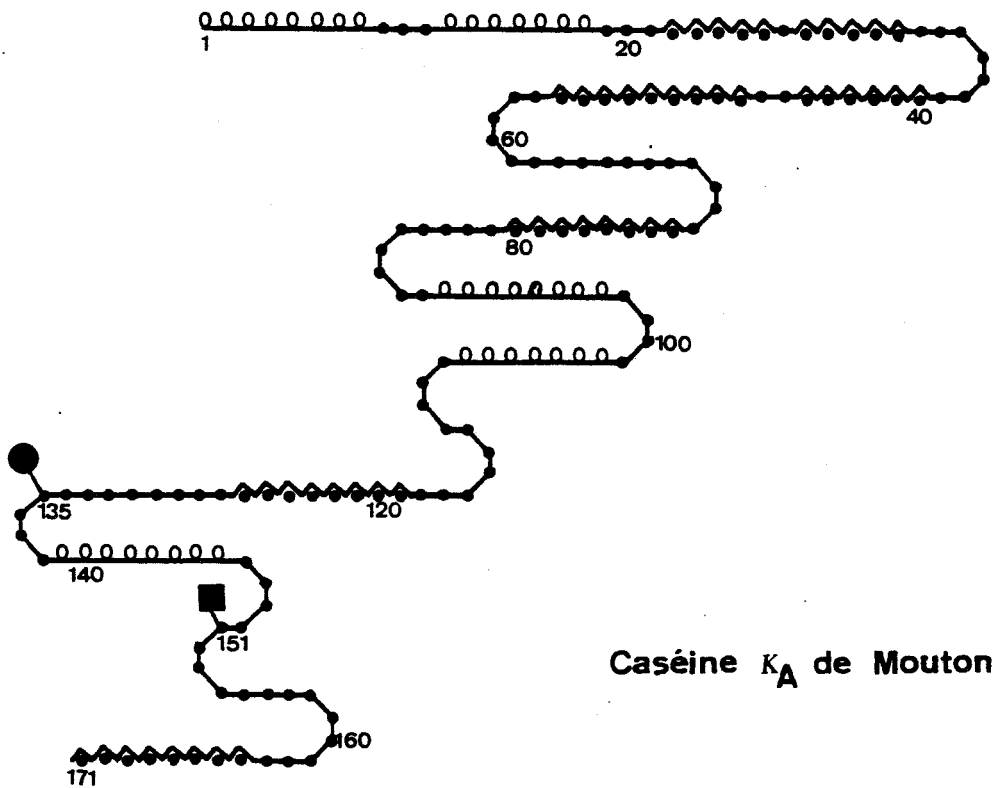
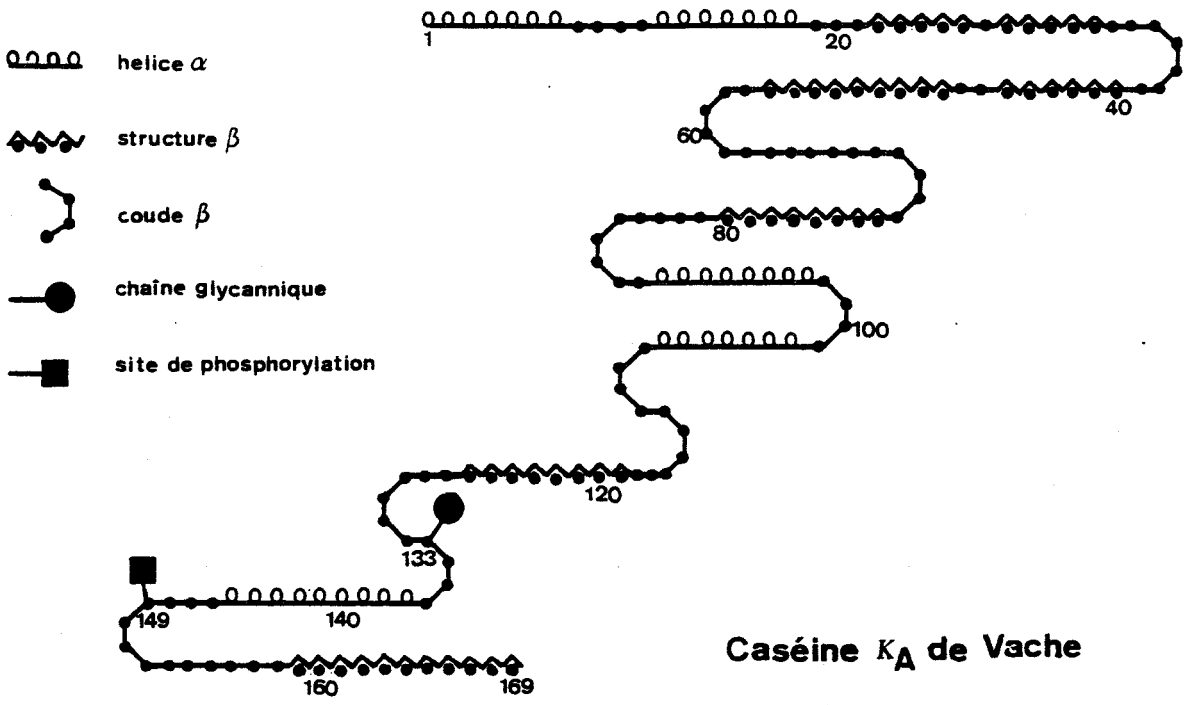


Figure 6



ARTICLE N° 2

ARTICLE N° 2
Carbohydrate-Peptide Linkage in Glycoproteins

JEAN-PIERRE AUBERT, GÉRARD BISERTE, AND
MARIE-H. LOUCHEUX-LEFEBVRE¹

*Institut de Recherches sur le Cancer de Lille et Unité 124 de l'I.N.S.E.R.M., B.P. 3567,
59020 Lille Cédex, France*

Received December 23, 1975

The secondary structure of the peptide segment around the carbohydrate-peptide linkage in glycoproteins was predicted by using the Chou and Fasman determination. Such a study was carried out for 9 *O*-glycosidically linkages and 28 *N*-glycosidically linkages. In the case of *O*-glycosidically linkages, the residue Ser or Thr involved in the linkage always belongs to a β -turn. In the case of *N*-glycosidically linkages, 19 out of the 28 Asn studied belong to a β -turn. A predicted determination concerning the whole protein moiety of 9 glycoproteins in order to obtain some information concerning the spatial organization of the entire glycoprotein was carried out also. It seems that carbohydrate moiety takes place outside the glycoprotein.

The type of carbohydrate-peptide linkage in glycoproteins is generally established from the knowledge of monosaccharide and amino acid residues which are involved in the linkage: A review concerning the methods used for such a determination was carried out by Neuberger *et al.* (1). Two different modes of linkage between carbohydrate and protein have been recognized: (i) Either the glycan chain is *N*-glycosidically linked to the protein moiety, and asparagine is involved in the carbohydrate-peptide bond, or (ii) it is serine and/or threonine as well as the modified hydroxylysine or hydroxyproline which are involved in the carbohydrate-peptide bond and the glycan chain is *O*-glycosidically linked to the protein moiety. Glycoproteins containing more than one carbohydrate group may have more than one type of carbohydrate-protein linkage. The knowledge of the sequence of monosaccharide residues and that of amino acid residues and their mode of linkage is required to describe the glycoprotein adequately: Such determinations are rather difficult and to date only a few glycoproteins are entirely known. Generally, the

most useful first approach is to get an accurate analysis of the sugar components of the whole glycoprotein molecule. Indeed, it is often possible to obtain the carbohydrate-containing moieties in association with an amino acid residue and much effort has been directed at preparing glycopeptides which contain only one amino acid residue. But it is obvious that isolation of glycopeptides with a peptide length of four to six amino acid residues can serve for the identification of the carbohydrate unit too, and therefore conditions of proteolysis were selected such that larger peptides remain attached to the carbohydrate unit. Therefore the amino acid sequence of a number of glycoproteins in the neighborhood of carbohydrate moieties is known from glycopeptides isolated from them. In particular, the amino acid directly adjoining the glycosylated asparagine residue on both the carboxyl and amino sides differ from one protein to another. On the contrary, the frequency with which threonine and serine occur in the vicinity of the asparagine residue which is covalently bound to *N*-acetylglucosamine seems to be a more specific feature and has been commented upon by Eylar (2). In this case, it was suggested that the amino acid se-

¹ To whom correspondence should be addressed.

quence X-Asn-Y-Thr (or Ser) in many cases may be a necessary but not a sufficient condition for glycosylation to occur (3-5). The frequent need for the particular threonine or serine may be because it acts as a recognition signal for the sugar transferase involved in glycosylation.

It is interesting to discuss these observations in the light of the recent work concerning the folding of polypeptide chains. In particular, the relative frequency of occurrence of various amino acid residues involved in turn regions has been evaluated: It has been observed that asparagine, glycine, serine, threonine, tyrosine, proline, and cysteine are especially abundant in these regions (6, 7).

These bends or turns consist of only four amino acids, that is, three peptide bonds, enabling a polypeptide chain to reverse itself by nearly 180°. The conformational analyses of possible β -turns were carried out by Venkatachalam (8) who surveyed the conformations which are sterically feasible for L-amino acids. He found that for these so-called β -turns three types of conformations which contain NH---O hydrogen bonding between the first and the third peptide units may exist. More recently, Crawford *et al.* (7) characterized the β -turns in terms of dihedral angles and analyzed the composition of the turn regions: They arrived at conclusions similar to those of Lewis *et al.* (6) as to the preferences of certain amino acids.

With the recent work of Chou and Fasman (9), which describes a predictive model for the secondary structure of globular proteins, that is, α helix, β -sheet, and β -turns, the prediction of chain reversal and tertiary folding in proteins can be obtained by the use of the β -turn conformational parameters computed for all 20 amino acids. The determination of the conformational parameters for the β -turns, $p(t)$, is based on the frequent occurrence of each amino acid into the β -turn structure.

Therefore, we have compiled a series of known amino acid sequences around either asparagine or threonine (or serine) found in literature (10) in order to see if it is possible to find a correlation between such a sequence and its secondary structure by using the structural parameters of

Chou and Fasman. In other words, we tried to see if Asn or Thr (or Ser) where glycans are linked, occupy a privileged situation in a reverse β -turn or in the neighboring of a β -turn. The plan adopted in this work is as follows.

First, we have estimated the secondary structure of several amino acid sequences around Thr (or Ser) and Asn in the cases of threonyl (or seryl)-O-glycosidic and asparaginyl-N-glycosidic linkage, respectively. Then, in the second part, from six glycoproteins in which the amino acid sequence is entirely known, we have estimated its secondary structure and tried to predict the spatial organization of the glycan group.

PROCEDURES

In all cases, that is, either for short sequences of amino acids or for a whole protein moiety, the determination of the secondary structure was carried out as shown in the case of α -acid glycoproteins (11). More especially, the probability that a tetrapeptide forms a reverse β -turn was given by Lewis *et al.*

$$p(t) = f_i \cdot f_{i+1} \cdot f_{i+2} \cdot f_{i+3}$$

In this expression, the f 's represent the frequency of occurrence for a certain residue at the first, second, third, and fourth position of a β -turn and were evaluated by Chou and Fasman (8) for the 20 amino acids. These authors took $p(t) = 0.5 \cdot 10^{-4}$ as a cut-off value in predicting the β -turns. Since they claim that for 108 β -turns studied, the average of the helical parameter $\langle P(\alpha) \rangle$ is approximately 0.9 and the relation $\langle P(t) \rangle > \langle P(\beta) \rangle$ always exists in which $\langle P(t) \rangle$ and $\langle P(\beta) \rangle$ are the structural parameters for the β -turn and β -sheet, respectively, we admitted that a tetrapeptide has a high probability of occurrence as a β -turn if the following three conditions are met:

- (1) $\langle P(\alpha) \rangle < 0.9$; (2) $\langle P(t) \rangle > \langle P(\beta) \rangle$; (3) $p(t) > 0.5 \cdot 10^{-4}$

RESULTS

The results are given in Tables I and II. These tables concern the study of O-glycosidically and N-glycosidically linked glycan, respectively. The amino acid residue where the glycan chain is linked on the protein moiety is always called "n" and the $n - 4$, $n + 4$ amino acid sequence is shown. The determination of structural parameters was carried out with these nine residues and the values $\langle P(\beta) \rangle$, $\langle P(t) \rangle$, and $p(t)$

TABLE I
 O-GLYCOSIDICALLY LINKED GLYCANS^a

	$n - 4$	$n - 3$	$n - 2$	$n - 1$	n	$n + 1$	$n + 2$	$n + 3$	$n + 4$	$\langle P(\beta) \rangle$	$\langle P(t) \rangle$	$p(t)10^4$	References
Heavy chain disease protein Zuc		Pro i	Gly $i + 1$	Gly $i + 2$	Ser $i + 3$	Ser	Glu	Pro	Lys	0.74	1.61	1.09	12
Chondroitin sulfate from pig costal cartilage			Glu i	Gly $i + 1$	Ser $i + 2$	Gly $i + 3$				0.65	1.34	1.06	13
Ox submaxillary glycoprotein		Ser i	Thr $i + 1$	Gly $i + 2$	Ser $i + 3$					0.81	1.45	1.53	14
Ox submaxillary glycoprotein				Ser i	Thr $i + 1$	Gly $i + 2$	Ser $i + 3$			0.81	1.45	1.53	14
Rabbit γ G immunoglobulin		Ser	Lys	Pro i	Thr $i + 1$	Cys $i + 2$	Pro $i + 3$	Pro	Pro	0.94	1.31	0.5	15
Hormono cortico gonadotropin (HCG) β subunit	Asp	Ser i	Ser $i + 1$	Ser $i + 2$	¹²¹ Ser $i + 3$	Ser	Lys	Ala		0.72	1.56	0.94	16
HCG β subunit	Ala	Pro	Pro i	Pro $i + 1$	¹²⁷ Ser $i + 2$	Leu $i + 3$	Pro	Ser	Pro	0.8	1.29	0.97	16
HCG β subunit	Leu	Pro	Ser i	Pro $i + 1$	¹³² Ser $i + 2$	Arg $i + 3$	Leu	Pro		0.74	1.41	2.61	16
HCG β subunit	Leu	Pro	Gly i	Pro $i + 1$	¹³⁶ Ser $i + 2$	Thr $i + 3$	Pro	Ile		0.84	1.51	1.59	16

^a The letters $n, n + 1$, etc., give the positions of the amino acid residues in the sequential structure of the protein moiety. The letters $i, i + 1$, etc., give their position in a β -turn (if it exists). Glycan chains are always bounded to the so-called " n " residue.

are given for the four amino acids, and in the case where four amino acids form a β -turn they are indicated as $i, i + 1, i + 2$, and $i + 3$. In all cases, we verified that the three necessary conditions to obtain a high probability of occurrence in a β -turn for these four amino acids are observed.

The $\langle P(\alpha) \rangle$ values are not shown in Tables I and II. In all cases we verified that the condition $\langle P(\alpha) \rangle < 0.9$ is always observed.

O-Glycosidically Linked Glycans

The results concerning this study are given in Table I, where the prediction of β -turns concerning nine glycan chains are shown. The presence of one or more residues of proline in the vicinity of the Thr or Ser linked also seems to be a characteristic of these linkages and could explain the

high probability of obtaining a β -turn in these cases. Indeed, Pro is a very good β -turn former. In these cases, it is interesting to note that the residue Thr or Ser always belongs to a β -turn.

N-Glycosidically Linked Glycans

The determination of the secondary structure around the asparagine was carried out for 28 glycan chains. One can see from Table II that the asparagine is often situated either in or near a reverse β -turn. Especially, in 2 cases, Asn occupies the position i of the β -turn, in 4 cases the position $i + 1$, in 12 cases the position $i + 2$ and in 1 case the position $i + 3$.

That is, in 19 out of 28 cases, the asparagine residue is localized on a β -turn, and more likely in the position $i + 2$.

It is also interesting to consider the posi-

TABLE II
N-GLYCOSIDICALLY LINKED GLYCANS^a

	<i>n</i> - 4	<i>n</i> - 3	<i>n</i> - 2	<i>n</i> - 1	<i>n</i>	<i>n</i> + 1	<i>n</i> + 2	<i>n</i> + 3	<i>n</i> + 4	<i>P</i> (β)	<i>P</i> (<i>t</i>)	<i>p</i> (<i>t</i>)/10 ⁴	Refer- ences
Human lactotransferrin			Gly <i>i</i>	Ser <i>i</i> + 1	⁶ Asn <i>i</i> + 2	Val <i>i</i> + 3	Thr	Asp		0.96	1.31	0.62	17
Human parotid saliva			Gly <i>i</i>	Pro <i>i</i> + 1	Asn <i>i</i> + 2	Gln <i>i</i> + 3	Ser			0.83	1.37	5.44	18
α ₁ -acid glycoprotein	Val	Pro	Ile	Thr	²² Asn <i>i</i> - 6	Ala <i>i</i> - 5	Thr <i>i</i> - 4	Leu <i>i</i> - 3	Asp <i>i</i> - 2				19
α ₁ -acid glycoprotein	Asn	Glu	Glu	Tyr <i>i</i>	⁴⁵ Asn <i>i</i> + 1	Lys <i>i</i> + 2	Ser <i>i</i> + 3	Val	Glu	0.85	1.38	0.82	19
α ₁ -acid glycoprotein	Tyr	Phe	Thr <i>i</i>	Pro <i>i</i> + 1	⁶¹ Asn <i>i</i> + 2	Lys <i>i</i> + 3	Thr	Glu	Asp	0.8	1.31	2.66	19
α ₁ -acid glycoprotein	Gln	Cys	Ile	Tyr <i>i</i>	³² Asn <i>i</i> + 1	Ser <i>i</i> + 2	Ser <i>i</i> + 3	Tyr	Leu	1.03	1.51	1.15	19
α ₁ -acid glycoprotein	Val <i>i</i> + 4	Gln <i>i</i> + 5	Arg <i>i</i> + 6	Glu <i>i</i> + 7	⁹² Asn <i>i</i> + 8	Gly <i>i</i> - 5	Thr <i>i</i> - 4	Val <i>i</i> - 3	Ser <i>i</i> - 2				19
Avidin	Asp	Leu	Gly <i>i</i>	Ser <i>i</i> + 1	¹⁷ Asn <i>i</i> + 2	Met <i>i</i> + 3	Thr	Ile	Gly	0.96	1.39	1.49	20
Bromelain	Val	Pro <i>i</i>	Arg <i>i</i> + 1	Asn <i>i</i> + 2	²⁰ Asn <i>i</i> + 3	Glu	Ser	Ser	Met	0.7	1.47	1.32	21
Human fibrogen γ chain	His	Gln	Val	Glu	³² Asn <i>i</i>	Lys <i>i</i> + 1	Thr <i>i</i> + 2	Ser <i>i</i> + 3	Glu	0.82	1.31	0.48	22
Human haptoglobin β chain	Leu	His	Gln	Asx <i>i</i>	³² Asn <i>i</i> + 1	Ser <i>i</i> + 2	Thr <i>i</i> + 3	Ala	Lys	0.82	1.41	0.67	10
Sheep luteinizing hormone α chain	Leu	Val	Pro <i>i</i>	Lys <i>i</i> + 1	⁵⁶ Asn <i>i</i> + 2	Ile <i>i</i> + 3	Thr	Ser	Glu	0.9	1.20	0.65	23
Sheep luteinizing hormone α chain	Val	Arg	Val	Glu	⁹² Asn	His	Thr	Glu	Cys				23
Chicken transferrin	Gly	Leu	Ile	His <i>i</i>	⁶ Asn <i>i</i>	Arg <i>i</i> + 1	Thr <i>i</i> + 2	Gly <i>i</i> + 3	Thr	0.89	1.34	0.81	24
Chicken transferrin	Val	Leu	Cys <i>i</i>	Met <i>i</i>	³³⁶ Asn <i>i</i> + 1	Asn <i>i</i> + 2	Ser <i>i</i> + 3	Phe		0.92	1.40	1.35	24
Waldenstrom's macroglobulin Ou μ chain	Tyr	Ser	Asn <i>i</i>	Asn <i>i</i> + 1	¹⁷⁰ Asn <i>i</i> + 2	Lys <i>i</i> + 3	Ser	Ile	Ser	0.67	1.51	1.35	25
Waldenstrom's macroglobulin Ou μ chain	Thr	Phe	Gln	Gln	³³² Asn	Ala	Ser	Ser	Met				25

TABLE II—Continued

	$n - 4$	$n - 3$	$n - 2$	$n - 1$	n	$n + 1$	$n + 2$	$n + 3$	$n + 4$	$(P(\beta))$	$(P(t))$	$p(t)10^3$	References
Waldenstrom's macroglobulin Ou μ chain	Lys	Thr	His <i>i</i>	Thr <i>i + 1</i>	⁷⁹⁵ Asn <i>i + 2</i>	Ile <i>i + 3</i>	Ser	Glu	Ser	1.04	0.99	0.85	25
Waldenstrom's macroglobulin Ou μ chain	Glu	Ser	His <i>i</i>	Pro <i>i + 1</i>	⁴⁰² Asn <i>i + 2</i>	Ala <i>i + 3</i>	Thr	Phe	Ser	0.74	1.12	3.74	25
Waldenstrom's macroglobulin Ou μ chain	Pro	Thr	Leu	Tyr	³⁶³ Asn	Val	Ser	Leu	Val				
Pig pancreatic ribonuclease	Ser	Ser	Ser <i>i</i>	Ser <i>i + 1</i>	²¹ Asn <i>i + 2</i>	Ser <i>i + 3</i>	Ser	Asn	Tyr	0.7	1.59	2.14	26
Pig pancreatic ribonuclease	Met	Ser	Arg <i>i</i>	Arg <i>i + 1</i>	³⁴ Asn <i>i + 2</i>	Met <i>i + 3</i>	Thr	Gln	Gly	1.03	1.09	0.98	26
Pig pancreatic ribonuclease	Cys	Tyr	Gln <i>i</i>	Ser <i>i + 1</i>	⁷⁶ Asn <i>i + 2</i>	Ser <i>i + 3</i>	Thr	Met	His	0.83	1.34	1.07	26
Bovine pancreatic ribonuclease	Met	Lys	Ser <i>i</i>	Arg <i>i + 1</i>	³⁴ Asn <i>i + 2</i>	Leu <i>i + 3</i>	Thr	Lys	Gly	0.87	1.19	1.41	27
Hormono cortico gonadotropin (HCG) α subunit	Leu	Val	Gln	Lys	⁵² Asn <i>i - 5</i>	Val <i>i - 4</i>	Thr <i>i - 3</i>	Ser <i>i - 2</i>	Glu <i>i - 1</i>				16
HCG α subunit	Phe	Lys	Val	Glu	⁷⁸ Asn	His	Thr	Ala	Cys				16
HCG β subunit	Arg <i>i + 2</i>	Cys <i>i + 3</i>	Pro <i>i + 4</i>	Ile <i>i + 5</i>	¹³ Asn <i>i + 6</i>	Ala	Thr	Leu	Ala				16
HCG β subunit	Cys	Ile	Thr	Val	³⁰ Asn <i>i - 4</i>	Thr <i>i - 3</i>	Thr <i>i - 2</i>	Ile <i>i - 1</i>	Cys <i>i</i>				16

* The significance of n and i is the same as for Table I.

tion of the threonine (or serine) residue in the vicinity of the asparagine residue since the sequence Asn-X-Thr (or Ser) seems to play a major role in the glycosylation (2-5). As evidenced, if Asn occupies the i or $i + 1$ position in the β -turn, Thr (or Ser)

also will be on the β -turn either in $i + 2$ or $i + 3$ positions, respectively. But in 12 out of 19 cases considered previously, Thr (or Ser) will be situated out the β -turn in position $i + 4$.

Moreover, in Table II, it was not possi-



ble to localize a reverse β -turn around Asn in nine cases which correspond to following glycoproteins: α_1 -acid glycoprotein, sheep luteinizing hormone α chain, Waldenstrom's macroglobulin Ou μ chain, hormono cortico gonadotropin α and β subunits, and chicken transferrin.

Therefore, for these six glycoproteins, in which the amino acid sequence is well known, the secondary structure was evaluated by using the Chou and Fasman's determination in order to localize the carbohydrate linkage.

α_1 -Acid Glycoprotein

The results of the study of this glycoprotein are given in the companion paper (11), in which the complete evaluated secondary structure is given. More especially, three out of five glycan chains are linked to an Asn residue situated in a reverse β -turn.

The so-called glycan chain "5" is not linked to a β -turn, but in a region which probably is situated outside of the protein moiety. On the contrary, the so-called glycan chain "1" is linked in a β -sheet region, that is, hydrophobic domain, which is generally situated inside the protein moiety.

Sheep Luteinizing Hormone α -Chain

The amino acid sequence of this glycoprotein is known (23), and Asn 56 and Asn 82 have bound carbohydrate. The evaluated secondary structure of protein moiety is given in Fig. 1. The glycan chain bound to the Asn 56 is situated in a β -turn. These are bound to the Asn 82, which is not

linked in a β -turn but is situated in a random region where the charged residues are numerous. Moreover, two β -sheet sequences, 73-80 and 86-93, are situated near this short-charged random sequence. Therefore it seems reasonable to fold these chains in the fashion indicated in Fig. 1 and in this case the bend observed would be situated outside the protein moiety.

Waldenstrom's Macroglobulin Ou μ Chain

As it is well known this IgM is constituted by the μ heavy chains and the κ light chains and the carbohydrate moieties are linked to asparagine residues belonging to the heavy chain.

Therefore the secondary structure of the μ chain was evaluated since its sequential structure is well known (26). Asparagine residues 170, 332, 395, 402, and 563 carry the carbohydrate moieties. Many β -turns exist in this structure and more especially the asparagine residues 170, 395, and 402 are localized in such β -turns.

The two other asparagine residues which have bound carbohydrate, that is, residues 332 and 563, are localized in regions other than β -turns and belong to the hinge region and the C-terminal end. Therefore, in order to obtain a better idea of the whole IgM, the evaluated secondary structure of the light chain was drawn also. The N-terminal ends of the light and heavy chains adopt essentially a β -sheet secondary structure; indeed, the hydrophobic amino acid residues are numerous in these regions.

Such a structure allows us to explain the

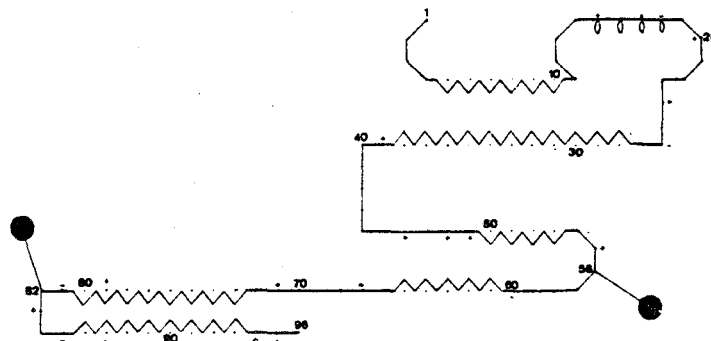


FIG. 1. Evaluated secondary structure of sheep luteinizing hormone α -chain. Residues are represented in their respective conformational state: helical (O), β -sheet (\wedge), coil (-), β -turn tetrapeptides are denoted Z. Asn 56 and Asn 82 carry a carbohydrate moiety.

specificity of the IgM and the rigidity of the hinge region. It is in this last region, near a β -sheet sequence, that a carbohydrate moiety is linked to the asparagine residue 332. On the contrary, the fifth carbohydrate moiety is linked to the asparagine residue 563, that is, in the C-terminal end where the helical structure seems to play a major role. But the 22 last C-terminal residues adopt a β -structure in which one finds asparagine residue 563.

Hormono Cortico Gonadotropin α Subunit

The evaluated secondary structure of the protein moiety whose amino acid sequence is well known (16) is given in Fig. 2; Asn 52 and Asn 78 have bound carbohydrates.

It is interesting to note that there does not exist any helical region in this protein moiety. The charged residues are localized in the random structure parts of the protein. Therefore, it seems that a hydrophobic core exists inside the protein constituted by the residues in β -sheet structure. The charged residues would be situated outside of the protein. Although there exist five β -turns, the glycan chain is not

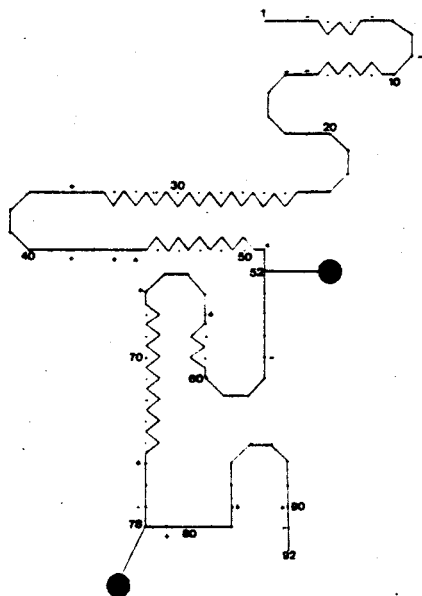


FIG. 2. Evaluated secondary structure of hormono cortico gonadotropin α subunit, with the same representation as in Fig. 1. Asn 52 and Asn 78 carry a carbohydrate moiety.

linked in such a β -turn, but in two charged random structure regions, that is, in regions which may be situated outside the proteins.

Hormono Cortico Gonadotropin β Subunit

This glycoprotein contains two *N*-glycosidically and four *O*-glycosidically linked carbohydrates. The evaluated secondary structure obtained from the sequential structure (16) is given in Fig. 3.

Ten β -turns are obtained. The four *O*-glycosidically linked carbohydrates are situated at the C-terminal end of the protein moiety. More especially, Ser 121, Ser 127, Ser 132, and Ser 138 carry a glycan chain. As expected from Table I, all of them are situated in a β -turn, either in the $i + 2$ or $i + 3$ position, as shown in Fig. 3.

Let us now consider the *N*-glycosidically linked carbohydrate in Asn 13 and Asn 30. Both do not belong to a β -turn but are situated near a β -turn in a random structure region and in a β -sheet structure, respectively.

Asn 13, situated near a charged region and near a β -turn may be outside of the protein. On the contrary, the position of Asn 30 in a β -sheet region is more difficult

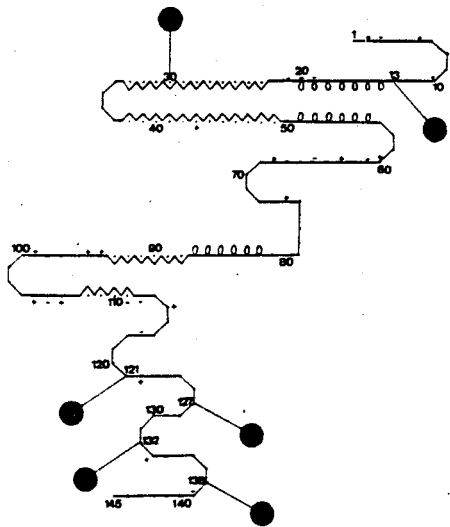


FIG. 3. Evaluated secondary structure of hormono cortico gonadotropin β subunit, with the same representation as in Fig. 1. Asn 13, Asn 30, Ser 121, Ser 127, Ser 132, and Ser 138 carry a carbohydrate moiety.

to understand. Perhaps, due to the neighboring of a β -turn (Asn 30 is situated in $i - 4$ position), this part of the protein moiety should be outside of the molecule.

Chicken Transferrin

The sequential structure of the protein moiety is not perfectly known (24) but the sequences around the two *N*-glycosidical linkages are well defined. The secondary structure was evaluated by taking the order given in Ref. (10). Asn 6 and Asn 336 have bound carbohydrates, that is, at the *N*- and *C*-terminal ends, respectively. Both Asn 6 and Asn 336 are situated in two β -turns, in positions i and $i + 1$, respectively.

In conclusion, from the evaluated secondary structures of protein moieties of six glycoproteins, the 4-*O*-glycosidically glycan chains are linked, without any exception, to a serine or threonine residue belonging to a β -turn, and this result is in good agreement with that previously observed from Table I. The results are more difficult to interpret for the *N*-glycosidically linked carbohydrate moiety. Indeed only 9 out of 18 asparagine residues which carry carbohydrate moiety are situated in a β -turn.

But 6 out of 18 such asparagine residues are situated in charged random region and therefore may be outside the protein. Then only 3 asparagine residues out of 18 are situated in a β -sheet region, that is, in a likely hydrophobic region which could be inside the protein.

We have seen that it is well known now that the tripeptide sequence Asn-X-Ser or Asn-X-Thr, where X is any amino acid is a necessary but not a sufficient condition for glycosylation to occur (1-5), and Hunt and Dayhoff (28) have searched proteins containing the tripeptide regions and lacking bound carbohydrate. In these cases, no carbohydrate moiety linked to the protein exists and therefore it is impossible to know if the lack of glycan chain is due to either the absence of the enzyme of glycosylation, or the local structure of the protein moiety, or something else. Therefore we have to try to find some glycoproteins in which both glycosylated and unglycosylated tripeptide sequences exist. This is

the case for the Waldenstrom's macroglobulin Ou μ chain since, besides five glycosylated asparagine residues one finds the following sequence (71-80):

Ile-Ser-Lys-Asn-Asp-Ser-Lys-Asn-Gln-Val
71 72 73 74 75 76 77 78 79 80.

with the tripeptide Asn-X-Ser (X = Asp) but where Asn 74 has not bound carbohydrate. It is interesting to observe the evaluated secondary structure of this sequence: It consists of two adjacent β -turns in 72-75 and 76-79. Moreover, it is generally admitted (29) that in the sequence Asn-X-Ser (or Thr) in which Asn carries a carbohydrate moiety, the hydroxyl group of the β -hydroxy amino acid residues is hydrogen bonded to the amide oxygen of the asparagine residue of the carbohydrate-peptide linkage. That is, in the present case, Ser 76 would be hydrogen bonded to the Asn 74 by its hydroxyl group, but would also be bonded to Gln 79 by its amide oxygen. Due to the steric hindrance it is not sure that such bonds are possible.

Another unusual case seems to be interesting to consider: It is this one, exceptional, in which a glycoprotein contains a carbohydrate peptide linkage but where the sequence Asn-X-Ser (or Thr) does not exist.

Indeed in the hen egg yolk phosphovitin, the glycosylated asparagine is situated in the following sequence (30):

-Ser-Asn-Ser-Gly-Psr-Arg-Ser-Val-
Ser-His-His-

(in which Psr represents phosphorylated serine residue). Although this example is an exception, it is interesting to note that the sequence Ser-Asn-Ser-Gly certainly adopts a β -turn structure (with $p(t) = 0.923 \cdot 10^{-4}$).

CONCLUSION

The main results obtained in this study concern the fact that the carbohydrate moiety of a glycoprotein linked to an amino acid residue is on the outside of the protein moiety and often situated in a β -turn, that is, in a domain geometrically well defined and therefore would allow an easy approach of the sugar during the bio-

synthesis of the glycan chains. In addition, such a spatial organization would be in good agreement with a model given by Montreuil in a recent review (31): The carbohydrate moiety would consist of an "aerial" zone, or more exactly "antennae" planted on the core of the glycoprotein, and this part of the molecule would support the glycan specificity. In addition, it is interesting to do a comparison with the results obtained by biochemists as recently expressed by Montreuil (31): "It appears more and more evident that numerous glycoproteins play two important roles. The first one is of a physicochemical order and concerns the protein moiety conformation. The second is of biological order, the basis of which is essentially the notion of recognition signals brought by glycan groups." It seems that our present study and our previous results about α_1 -acid glycoprotein are in excellent agreement with such conclusions.

ACKNOWLEDGMENTS

The authors are grateful to Professor G. D. Fasman for a critical reading of their manuscript.

REFERENCES

- NEUBERGER, A., GOTTSCHALK, A., MARSHALL, R. D., AND SPIRO, R. G. (1972) in *Glycoproteins: Their Composition, Structure and Function* (Gottschalk, A., ed.), Vol. 5, Part A, p. 450, Elsevier, Amsterdam/London/New York.
- EYLAR E. H. (1965) *J. Theoret. Biol.* 10, 89-113.
- MARSHALL, R. D., AND NEUBERGER, A. (1970) *Advan. Carb. Chem. Biochem.* 25, 407-478.
- NEUBERGER, A., AND MARSHALL, R. D. (1968) in *Carbohydrates and Their Roles* (Schultz, H. W., Cain, R. F. and Wrolstao, R. W., eds.), 5th Symp. on Foods at Oregon State University, p. 115, AVI Publishing, Westport.
- SOX, H. C., AND HOOD, L. (1970) *Proc. Nat. Acad. Sci. USA* 66, 975-982.
- LEWIS, P. N., MOMAMY, F. A., AND SCHERAGA, H. A. (1970) *Biochim. Biophys. Acta* 303, 211-229.
- CRAWFORD, J. L., LIPSCOMB, W. N., AND SCHELLMAN, C. G. (1973) *Proc. Nat. Acad. Sci. USA* 70, 538-542.
- VENKATACHALAM, C. N. (1968) *Biopolymers* 6, 1425-1436.
- CHOU, P. Y., AND FASMAN, G. D. (1974) *Biochemistry* 13, 211-245.
- DAYHOFF, M. O. (1972) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., ed.), National Biomedical Research Foundation, Washington, D.C.
- AUBERT, J. P., AND LOUCHEUX-LEFEBVRE, M. H. (1976) *Arch. Biochem. Biophys.* 175.
- FRANGIONE, B., AND MILSTEIN, C. (1969) *Nature (London)* 224, 597-599.
- KATSURA, N., AND DAVIDSON, E. A. (1966) *Biochim. Biophys. Acta* 121, 120-127.
- OZEKI, T., AND YOSIZAWA, Z. (1971) *Arch. Biochim. Biophys.* 142, 177-183.
- SMYTH, D. S., AND UTSUMI, S. (1967) *Nature (London)* 216, 332-335.
- MORGAN, F. J., BIRKEN, S., AND CAMFIELD, R. E. (1973) *Mol. Cell. Biochem.* 2, 97.
- CHARET, P., MONSIGNY, M., SPIK, G., AND MONTREUIL, J. (1969) *C. R. Acad. Sci.* 269, 1019-1022.
- DEGAND, P., personal communication.
- SCHMID, K., KAUFMAN, H., ISEMURA, S., BAUER, F., EMURA, J., MOTOYAMA, T., ISHIGURO, M., AND NANNO, S. (1973) *Biochemistry* 12, 2711-2724.
- DELANGE R. J., AND HUANG, T. S. (1971) *J. Biol. Chem.* 246, 698-709.
- MURACHI, T., AND TAKAHASHI, N. (1970) in *Structure-Function Relationships of Proteolytic Enzymes* (Desnuelle, P., Neurath, H., and Ottesem, M., eds) p. 298, Munkogaard, Copenhagen.
- IWANAGA, S., BLOMBACK, B., GRONDAHL, N. J., HESSEL, B., AND WALLEN, P. (1968) *Biochim. Biophys. Acta* 160, 280-283.
- LIU, W. K., NAHM, H. S., SWEENEY, C. M., BAKER, H. N., LAMKIN, W. M., AND WARD, D. N. (1971) *Res. Comm. Chem. Path. Pharmacol.* 2, 168-176.
- ELLEMAN, T. C., AND WILLIAMS, J. (1970) *Biochem. J.* 116, 515-532.
- SHIMIZU, A., PAUL, C., KÖHLER, H., SHINODA, T., AND PUTNAM, F. W. (1971) *Science* 173, 629-633.
- JACKSON, R. L., AND HIRS, C. H. W. (1970) *J. Biol. Chem.* 245, 637-653.
- SMYTH, D. G., STEIN, W. H., AND MOORE, S. (1963) *J. Biol. Chem.* 238, 227-234.
- HUNT, L. T., AND DAYHOFF, M. O. (1970) *Biochim. Biophys. Res. Commun.* 39, 757-765.
- MARSHALL, R. D. (1972) *Annu. Rev. Biochem.* 41, 673-702.
- SHAIKIN, R., AND PERLMANN, G. E. (1971) *J. Biol. Chem.* 246, 2278-2284.
- MONTREUIL, J. (1975) *Pure and Applied Chem.* 42, 431-477.

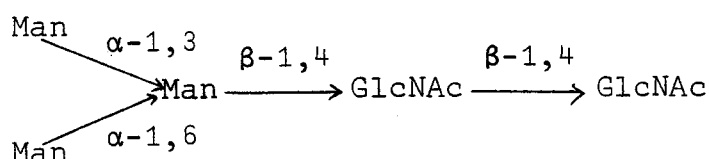
III) RELATION STRUCTURE-N-GLYCOSYLATION *in vitro* DE PEPTIDES SYNTHÉTIQUES. (ARTICLE N°3)

A) Introduction

Dans ce travail nous ne parlerons que des liaisons N-glycosidiques.

Les glycoprotéines lorsqu'elles sont N-glycosylées le sont donc toujours au niveau d'un résidu d'asparagine inclus dans une séquence code Asn-X-Ser/Thr (Marshall et Neuberger 1970). C'est toujours un résidu de N-acétylglucosamine qui est lié à l'asparagine.

La partie interne des motifs N-glycosidique est toujours constituée d'un noyau pentasaccharidique (Montreuil 1975).



A partir de ce noyau de base on peut définir deux grandes classes:

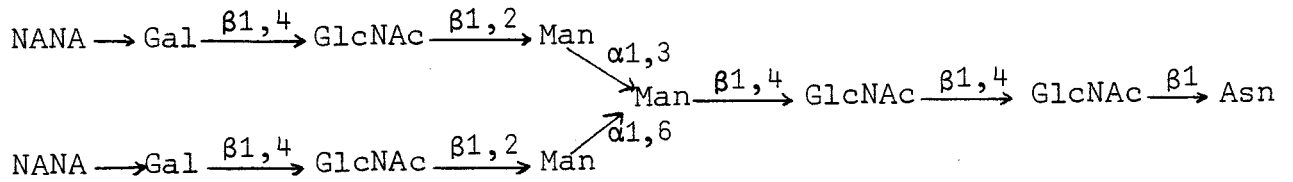
- Classe oligomannosidique (Strecker et Montreuil 1979) ou simple (Kornfeld et Kornfeld 1976).

Dans ce cas seuls des résidus de mannose sont ajoutés au pentasaccharide.

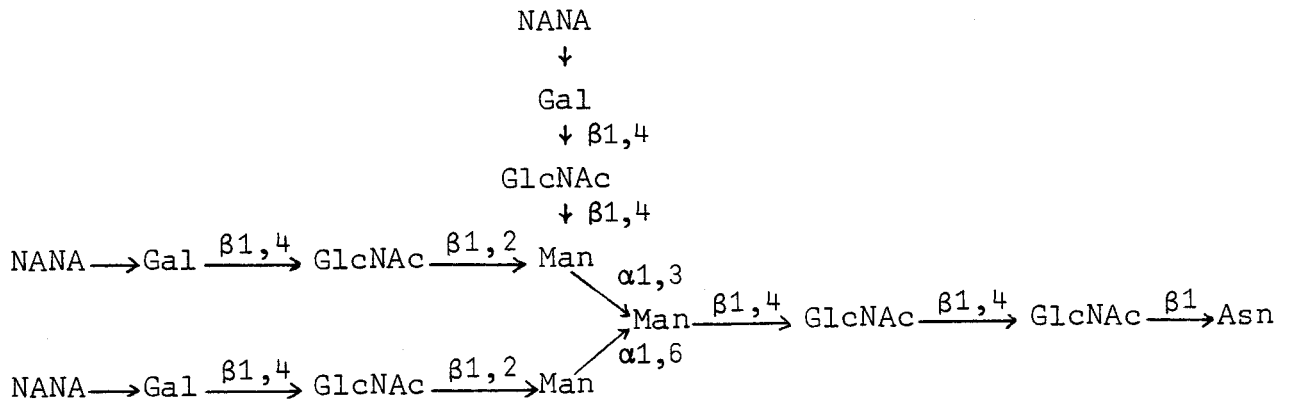
- Classe N-acétyllactosaminidique (Strecker et Montreuil 1979) ou complexe (Kornfeld et Kornfeld 1976).

Dans ce cas on trouve sur le motif de base des résidus de N-acétylglucosamine, de galactose, de fucose et d'acides sialiques. Ces structures peuvent être bi, tri ou tétra-

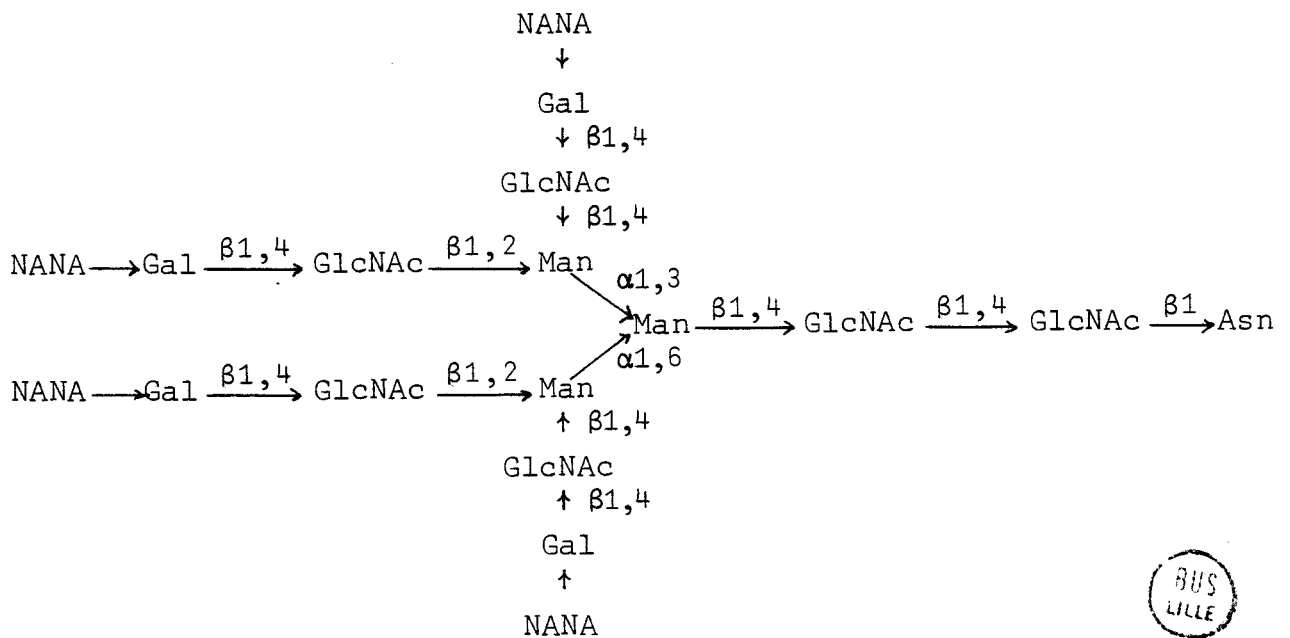
Tableau IV



STRUCTURE BIAN TENNEE



STRUCTURE TRIANTENNEE



STRUCTURE TETRAANTENNEE

antennées comme par exemple: (Tableau IV).

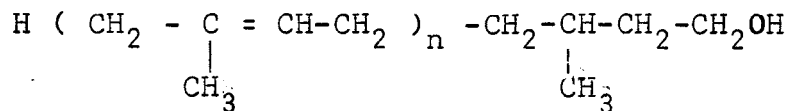
Vers 1970, un certain nombre de travaux ont permis d'établir avec certitude que le réticulum endoplasmique rugueux était le siège de l'activité N-acétylglucosaminyl (Tetas et al 1970) et mannosyl (Tetas et al 1970 ; Caccam et al 1969) transférases.

Plus précisément, des travaux récents ont montré que la glycosylation ne peut avoir lieu qu'au cours de l'élongation de la chaîne protéique (Lingappa et al 1978; Bielinska et Boime 1978; ; Toneyuzzo et Ghosh 1978; Katz et al 1978).

Comment se fait cette glycosylation?

On a longtemps cru que la biosynthèse des chaînes glycaniques se faisait par additions successives de monosaccharides à partir des nucléotides correspondants grâce à des transférases spécifiques; or on sait maintenant que cette glycosylation met en jeu des intermédiaires lipidiques glycosylés (Waechter et al 1973; Ronin et Bouchilloux 1976; Ronin et Bouchilloux 1978) suivant le schéma de biosynthèse général suivant:(Tableau V).

Dans cet intermédiaire lipidique glycosylé, Dol représente la partie lipidique appelée dolichol. Ce dolichol est un polyisoprénol en $C_{80}-C_{100}$ de formule:



Nous avons donc donné la formulation générale de la biosynthèse de l'oligosaccharide-lipide. Le détail de cette biosynthèse est maintenant connu (Tabas et Kornfeld 1978) et fait comme suit: (Tableau VI).

C'est ce motif qui est transféré en bloc sur la

Tableau V

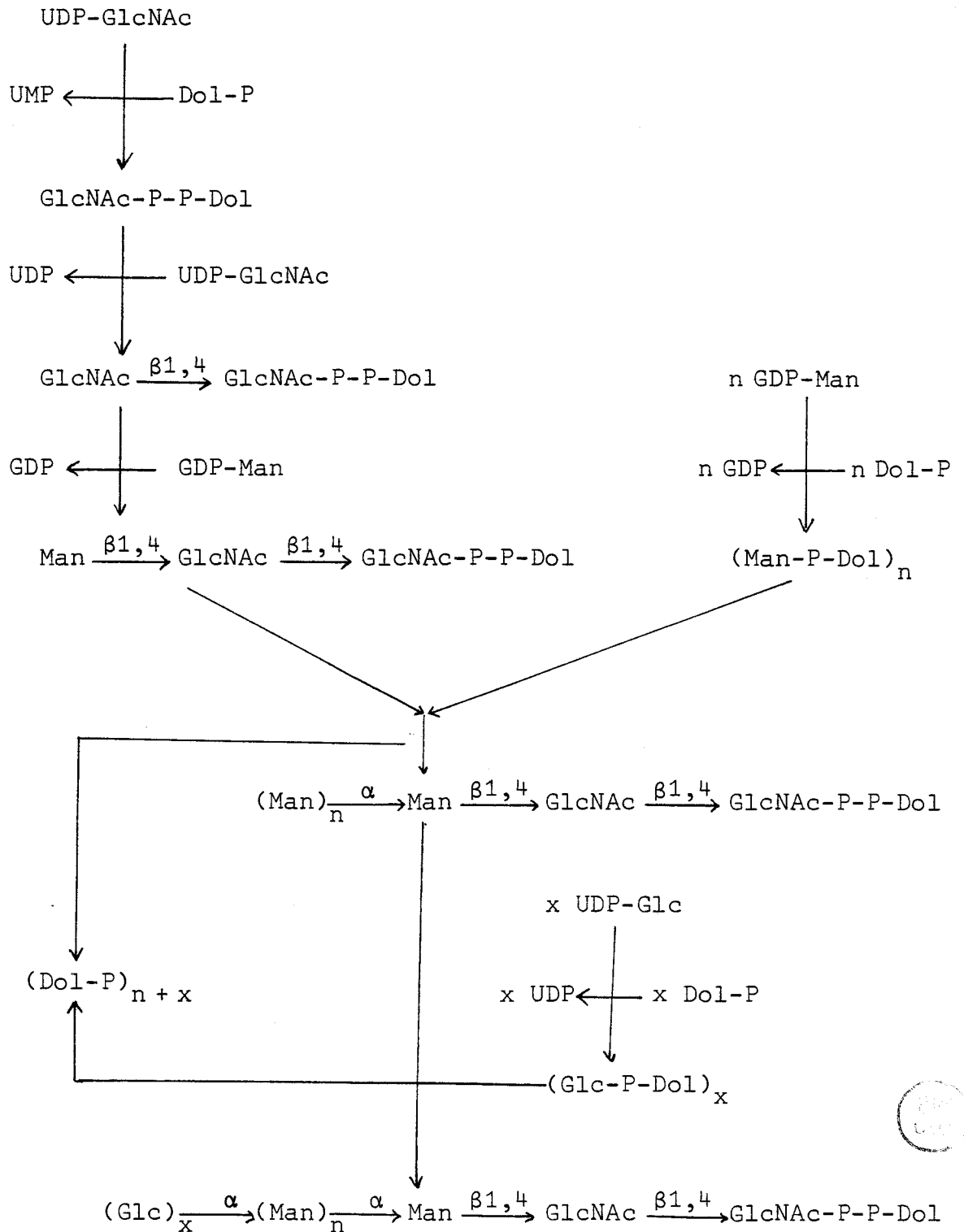
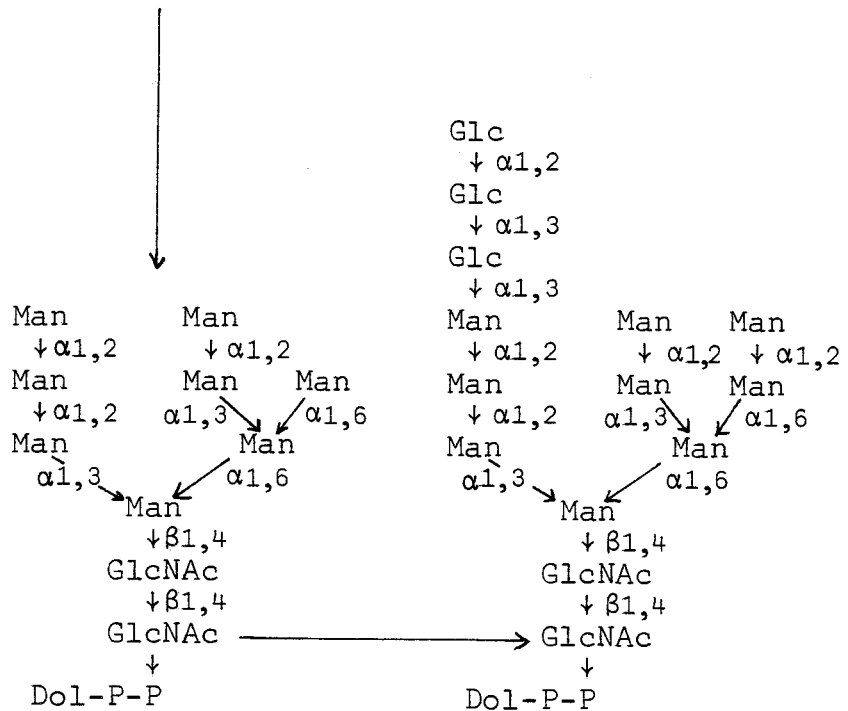
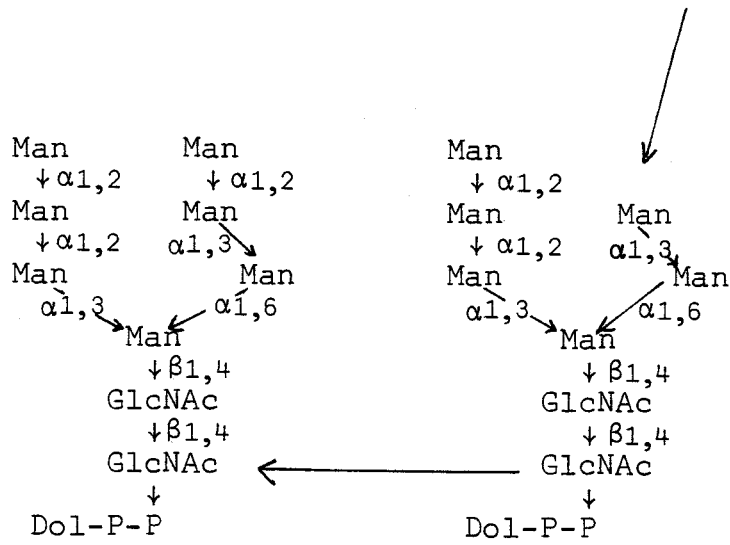
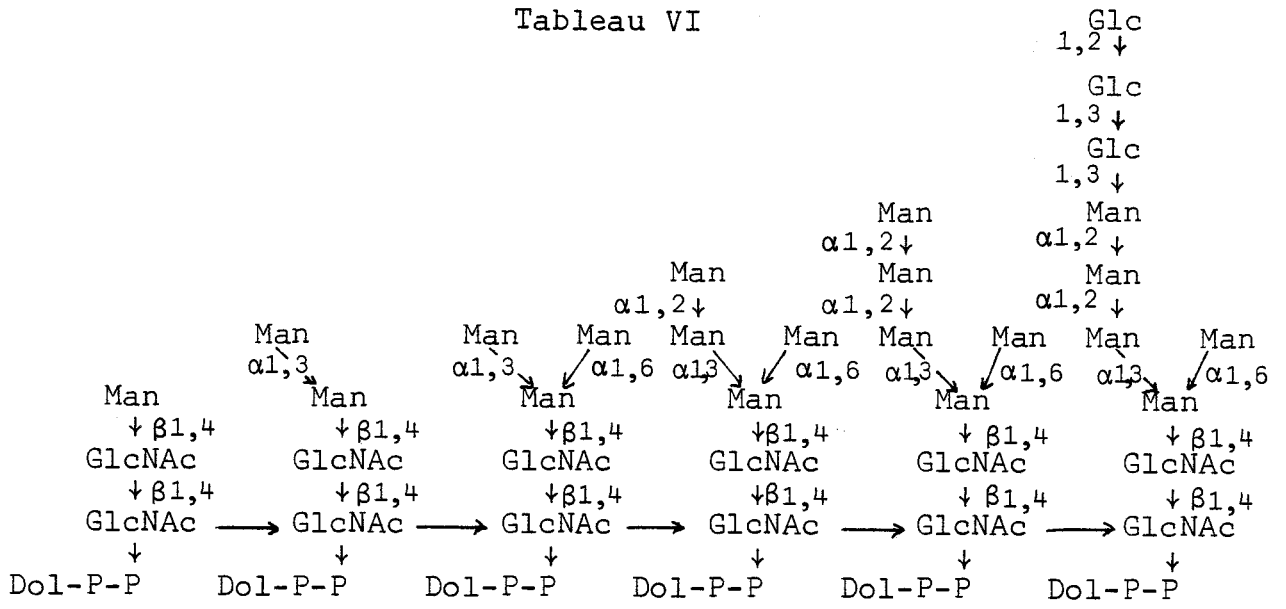


Tableau VI



protéine. Comme nous l'avons vu précédemment, la structure des chaînes glycaniques de type N-acétyllactosaminidique est totalement différente de celle qui est fixée en premier lieu. Un élagage est donc nécessaire et doit être suivi d'une autre biosynthèse. Au niveau du réticulum endoplasmique agit une glucosidase (Ugalde et al 1978) puis les résidus de mannose sont libérés au niveau du Golgi (Tabas et Kornfeld 1979) et enfin agissent des transférases spécifiques également au niveau du Golgi. Le schéma généralement admis est le suivant (Tableau VII) (Briles et Kornfeld 1978; Harpaz et Schachter 1980).

La liste des cellules et tissus dans lesquels la présence d'un dolicholpyrophosphate oligosaccharide a été mise en évidence est déjà très longue aussi bien chez les mammifères que chez les oiseaux, les insectes, les plantes, les algues etc... (Parodi et Leloir 1979). On connaît donc parfaitement maintenant les conditions qui permettent de réaliser in vitro une telle glycosylation.

B) Méthodes utilisées

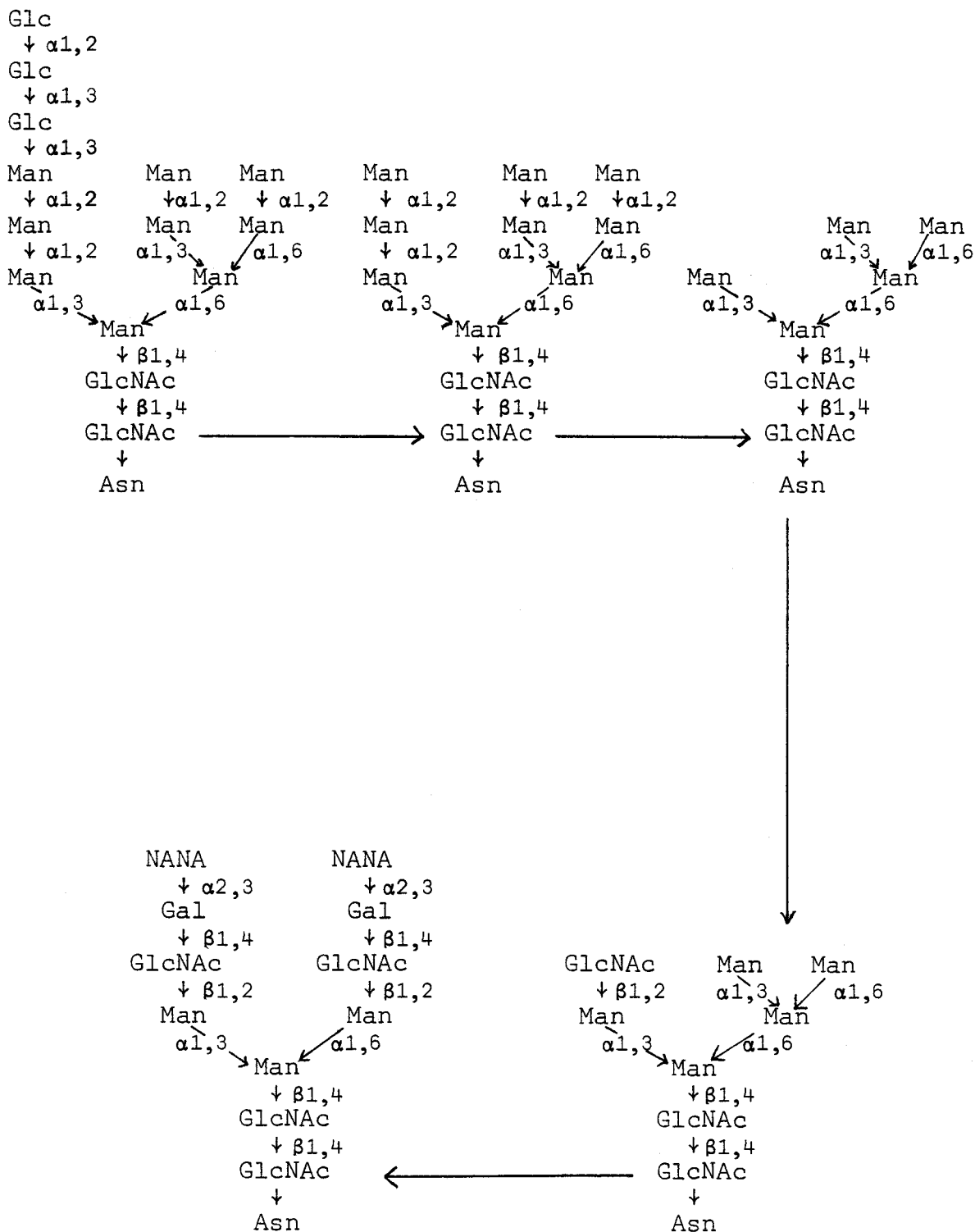
Ce travail a été réalisé en collaboration avec Mlle N. Helbecque.

Nous disposons de cinq peptides synthétiques

I	Gly-Asn-Gln-Ser	G-N-Q-S
II	Pro-Gly-Asn-Gln-Ser	P-G-N-Q-S
III	Thr-His-Thr-Asn-Ile-Ser-Glu-Ser-His-Pro-Asn-Ala-Thr-Phe	T-H-T-N-I-S-E-S-H-P-N-A-T-F
IV	Thr-His-Thr-Asn-Ile-Ser	T-H-T-N-I-S
V	His-Pro-Asn-Ala-Thr-Phe	H-P-N-A-T-F

Les peptides I et II sont les modèles d'un glycopeptide obtenu après hydrolyse par un mélange de papaïne et de pronase de la glycoprotéine riche en proline de la salive

Tableau VII



parotidienne humaine (Degand et al 1975). Le peptide III est le modèle d'un glycopeptide diglycosylé obtenu par hydrolyse trypsique et chymotrypsique des chaînes lourdes d'immunoglobuline M (IgM). Les peptides IV et V sont deux hexapeptides correspondants aux parties N et C terminales du tétradécapeptide III.

Nous avons tout d'abord essayé de glycosyler chaque peptide en utilisant comme source enzymatique les membranes d'oviducte de poule. Nous avons ensuite étudié ces peptides dans différentes conditions par dichroïsme circulaire pour essayer d'établir une corrélation entre leur structure secondaire et la N-glycosylation.

C) Résultats

1) Glycosylation

Les membranes d'oviducte de poule catalysent la fixation du mannose à partir du GDP - [^{14}C] mannose sur des peptides exogènes ayant la séquence tripeptidique code Asn-X-Ser/Thr. Comme nous l'avons vu, cette incorporation se fait par l'intermédiaire d'un dolichyldiphosphate oligosaccharide (Lennarz 1975; Chen et al 1975). Les protéines endogènes bénéficient aussi de l'activité enzymatique et sont donc capables d'être glycosylées (Pless et Lennarz 1977; Struck et al 1978).

Nous avons donc testé les cinq peptides en notre possession. Pour cela il suffit d'incuber à 37°C ces derniers en présence de membranes d'oviducte de poule et de GDP- [^{14}C] mannose. Après centrifugation de chaque essai, les surnageants sont soumis à une électrophorèse sur papier. Si les peptides ne sont pas accepteurs, ils restent dans le surnageant. Au contraire, s'il y a glycosylation, aucune trace de peptide n'est décelée et cela est bien observé sur la figure I (article n°3) où sont donnés en exemple les résultats corres-

pondants aux essais de glycosylation des peptides I et II qui sont respectivement non glycosylé et glycosylé. Pour les peptides de la série IgM, les résultats obtenus sont identiques: le peptide IV qui n'est pas glycosylé est observé dans le surnageant alors que les peptides III et V restent dans le culot. Les résultats sont résumés dans le tableau II (article n° 3).

Les glycopeptides sont obtenus par dissolution du culot membranaire dans du SDS 3% et passage sur Biogel P 6. La figure II (article n° 3) présente les profils d'élu-tion typiques pour chaque peptide. Les masses moléculai-res des nouvelles espèces radioactives sont compatibles avec la formation d'un glycopeptide. Le tétradécapeptide III qui possède 2 sites potentiels de glycosylation peut être mono ou diglycosylé. On constate en effet que les deux chaînes glycaniques sont présentes mais seulement dans 30% des cas. Pour le peptide monoglycosylé qui représente 70% du total, il est impossible de préciser s'il s'agit de la glycosylation d'un seul des sites ou des deux sites en quan-tité égale.

Ces glycopeptides ont été soumis à une chromato-graphie d'affinité sur concanavoline A qui permet la fixa-tion des chaînes polymannosidiques.

Tout ce qui n'est pas glycosylé est exclu de la colonne et après élu-tion par des solutions de molarité crois-sante d' α - méthyl glucoside, différents glycopeptides sont ainsi récupérés. Ces derniers ne diffèrent que par la quanti-té de résidus mannose composant la chaîne glycanique. Les courbes d'élu-tion sont données sur la figure III (article n° 3) Chaque pic obtenu a été soumis à une chromatographie en phase gazeuse pour déterminer les rapports molaires en oses (Table II article n° 3). Les compositions sont en très bon accord avec ce que l'on connaît de la biosynthèse des glycoprotéines.

En effet, pour 2 résidus de N-acétylglucosamine, on trouve 3 à 9 résidus de mannose et ceci correspond parfaitement à la fixation d'une chaîne polymannosidique.

Afin de s'assurer que le peptide de départ n'a subi aucune dégradation, les glycopeptides ont aussi été caractérisés par leur composition en acides aminés (Table I'article n° 3). On voit qu'ils correspondent parfaitement à la formule attendue.

2) Dichroïsme circulaire

Nous avons tout d'abord étudié les glycopeptides dont la partie peptidique a servi de modèle lors de la synthèse peptidique; ils ont été obtenus par coupure enzymatique.

L'étude de ces glycopeptides par dichroïsme circulaire a été faite en solution aqueuse. Dans ce solvant, aucun des spectres obtenus n'est caractéristique d'une structure organisée. De plus, comme cela avait été fait précédemment pour l'orosomucoïde, la chaîne glycannique des glycopeptides a été coupée par hydrazinolyse et après N-réacétylation, son étude dichroïque a été réalisée. On constate que la différence entre les spectres dichroïques correspondant au glycopeptide et à la chaîne glycannique se superpose parfaitement avec le spectre du peptide correspondant obtenu par synthèse. Ceci confirme donc de nouveau qu'aucune interaction forte n'existe entre la partie peptidique et la partie glycannique. Sur la figure IV de l'article n° 3, l'exemple donné correspond au glycopeptide Gly-Asn-Gln-Ser.

Les cinq peptides synthétiques ont également été étudiés par dichroïsme circulaire en solution aqueuse. Leurs spectres sont tous caractéristiques d'une structure inorganisée et montrent une intense bande négative centrée à environ 200 nm.

Par contre, si les spectres dichroïques sont réalisés en solution aqueuse contenant des lipides ou du diméthyl sulfoxyde (DMSO), d'importantes modifications sont observées pour certains peptides.

La figure V (article n° 3) montre les spectres obtenus pour les cinq peptides en présence de 10% de DMSO. Les spectres des peptides I et IV restent inchangés alors que pour les peptides II, III et V une nouvelle bande positive apparaît. Cette bande est centrée à 217 nm et elle reflète la formation d'une nouvelle structure.

Les résultats obtenus sont identiques lorsqu'on utilise des lipides à la place du DMSO (Figure VI article n° 3). Ces nouveaux spectres dichroïques montrent certaines ressemblances avec ceux calculés par Woody (Woody 1974) pour différents types de coude β . Ils sont aussi semblables à un spectre dichroïque publié par Howard et al (Howard et al 1975) concernant un tétrapeptide synthétique correspondant à une séquence de l'alpha-chymotrypsine qui a été identifiée par RX comme adoptant, dans la protéine, une structure en coude β .

Nous pouvons donc penser qu'une fraction de la structure adoptée par nos peptides est constituée par des coudes β ; malheureusement cette fraction n'est peut être pas très importante car les peptides considérés sont petits et par conséquent leur flexibilité est très grande. Cependant, l'addition de lipides ou de DMSO réduit cette flexibilité et permet la formation d'une structure privilégiée.

Nous pensons qu'il est très important de noter que quel que soit le peptide glycosylable et quelles que soient les conditions utilisées, le spectre dichroïque obtenu présente toujours les mêmes caractéristiques. Ceci veut dire qu'il y a bien formation d'une structure privilégiée, que cette structure privilégiée est unique et qu'elle pourrait bien être un coude β .

Alors que nous rédigeons cette thèse, une étude par RMN du ^{13}C et par dichroïsme circulaire de la tuftsine et de ses analogues a été réalisée (Siemion et al 1980). La tuftsine est un tétrapeptide Thr-Lys-Pro-Arg qui stimule le phénomène de phagocytose dans les granulocytes. Les résultats obtenus par ces chercheurs, à partir d'un certain nombre d'homologues, montrent que l'activité biologique du peptide est directement liée à son aptitude à adopter une conformation en coude β . De plus, les spectres dichroïques correspondants à ces peptides sont tout à fait semblables aux nôtres et montrent la présence d'une bande dichroïque positive centrée à 218 nm.

D) Conclusions

Nous avons donc vu que les peptides I et IV ne sont pas glycosylables alors que les peptides II, III et V le sont. Nous avons vu également que, parallèlement, les peptides I et IV sont insensibles aux lipides et au DMSO alors que les peptides II, III et V le sont. Il y a donc un rapport évident structure-glycosylation.

Les calculs de conformation de Chou et Fasman appliqués à nos peptides montrent que G-N-Q-S (I) présente une faible probabilité d'adopter une structure en coude β à la différence de P-G-N-Q-S (II). De même les peptides T-H-T-N-I-S-E-S-H-P-N-A-T-F (III), T-H-T-N-I-S (IV) et H-P-N-A-T-F (V) présentent tous trois une probabilité élevée d'adopter une structure en coude β . Les résultats sont résumés dans le tableau suivant.

Peptides	Coude choisi	P (t) 10^4	$P < T >$
G-N-Q-S (I)		0,534	1,38
P-G-N-Q-S (II)	P-G-N-Q	4,1	1,40
T-H-T-N-I-S-E-S-H-P-N-A-T-F (III)	H T N I et H P N A	4,55 6,51	0,99 1,00
T-H-T-N-I-S (IV)	H-T-N-I	4,55	0,99
H-P-N-A-T-F (V)	H-P-N-A	6,51	1,00

Le peptide IV n'est pourtant pas glycosylable contrairement au peptide V . Il est intéressant de remarquer qu'il est moins hydrophobe que le peptide V. Une des explications possible pourrait donc être que lorsqu'un peptide est petit, l'hydrophobicité jouerait un grand rôle; cela serait surtout important lors de l'interaction avec les lipides. En effet, la présence des résidus hydrophobes est moins importante lors de la biosynthèse. Il ne faut pas oublier que, *in vivo* , ce qui est inséré dans la membrane n'est pas un petit peptide mais une chaîne protéique beaucoup plus longue; cette dernière adopte une structure secondaire donnée qui stabiliserait le coude β créé au niveau de la séquence code Asn-X-Ser/Thr, et favoriserait la glycosylation.

Il est intéressant de comparer nos résultats à ce qui se passe lors de la O-glycosylation à la lumière des résultats récemment publiés, qui laissent à penser qu'elle est directement liée à l'accessibilité du site réactionnel.

En effet, Hanover et al (Hanover et al 1980) ont montré que la O-glycosylation avait lieu au niveau du Golgi et qu'elle se faisait sans intermédiaire lipidique sur une protéine ayant déjà sa structuration complète. Il faut admettre dans ce cas que la glycosylation est directement reliée à l'accessibilité des résidus hydroxylés. Peu de temps auparavant Bause et Lehle (Bause et Lehle 1979) avaient tenté de trouver une séquence code nécessaire à la O-glycosylation en réalisant des essais *in vitro* sur des peptides synthétiques. Ils n'ont pu établir aucune corrélation entre structure primaire du peptide et rendement de glycosylation. Leur conclusion est donc que la clé de la O-glycosylation est bien l'accessibilité du site.

Ces résultats confirment donc dans leur ensemble la nécessité de la présence d'un coude β aussi bien lors de la N- que de la O-glycosylation.

ARTICLE N° 3

Circular Dichroism Studies of Synthetic Asn-X-Ser/Thr-Containing Peptides: Structure-Glycosylation Relationship

J. P. AUBERT, N. HELBECQUE, AND M. H. LOUCHEUX-LEFEBVRE

Institut de Recherches sur le Cancer de Lille et Unité 124 de l'INSERM, BP 311, 59020 Lille Cédex, France

Received August 7, 1980

Many proteins are glycosylated and widely distributed. Knowledge of the biosynthetic pathway is of great interest in elucidating the exact role of the oligosaccharide chain of the glycoproteins. It is now possible to glycosylate carbohydrate-depleted or chemically denatured glycoproteins or synthetic Asn-X-Thr/Ser-containing peptides in a cell-free system. The results obtained showed that it is possible to glycosylate a tripeptide but that the yield (of glycosylation) increases as the length of the peptide increases. We have shown that in native glycoproteins, the oligosaccharide chains are always situated in a β -turn. It seemed so very interesting to see if this condition was necessary for the glycosylation. To elucidate this problem, synthetic peptides were tested as glycosyl acceptors using hen oviduct membranes as enzyme source. These peptides were studied by circular dichroism in aqueous lipid mixtures and the results showed that the presence of a secondary structure in the lipid state promotes greatly the yield of glycosylation.

Asparagine-linked glycoproteins are always glycosylated at an asparagine residue in the sequence Asn-X-Thr or Asn-X-Ser (1). The biosynthetic pathway of the asparagine-linked oligosaccharides of glycoproteins is extremely complex but recent biochemical experiments have revealed that it involves a dolichyldiphosphate oligosaccharide containing glucose, mannose, and *N*-acetylglucosamine. A great variety of tissues either in eucaryotes or in bacteria catalyze the glycosylation of proteins (for review see Ref. (2)) and the reactions involved in the assembly of the lipid-linked intermediate that participate in the synthesis of glycoproteins are now well understood (for review see Ref. (2)).

Recently, a number of authors have directed their efforts toward the delineation of the structural features of the proteins and peptides that are glycosylated *in vitro*. More specifically, it has been demonstrated that a variety of exogenous peptides and unfolded proteins containing the sequence Asn-X-Ser/Thr in their primary structure could be glycosylated *in vitro*. Concerning the structural features of the

amino acid sequences of glycoproteins, we have shown (3, 4), and, following us Beeley (5) and Nagarajan and Rao (6), that the most likely peptide conformation in the region of *N*-glycosylation is the β -turn. So we decided to test different peptides in order to make a comparison of their *in vitro* acceptor activity as a function of the probability with which they could adopt a β -turn secondary structure. At first, their ability to exhibit a β -turn was determined using the predictive method of Chou and Fasman (7). The peptides were chosen as models of natural glycopeptides obtained by hydrolysis of two glycoproteins: a proline-rich glycoprotein of human parotid saliva (GPRP)¹ and the heavy chain of immunoglobulin M (IgM).

In addition, it was suggested by Hart *et al.* that the accessibility of the sequence Asn-X-Ser/Thr to the oligosaccharide transferase is the dominant structural feature of a protein that determines

¹ Abbreviations used: GPRP, proline-rich glycoprotein of human parotid saliva; SDS, sodium dodecyl sulfate; Con A, concanavalin A; glc, gas-liquid chromatography; DMSO, dimethyl sulfoxide.

TABLE I
AMINO ACID COMPOSITION OF SYNTHETIC PEPTIDE SUBSTRATES

Peptide	Gly	Phe	Ala	Pro	Glu	Ser	Ile	Asp	His	Thr
I Gly-Asn-Gln-Ser	1.10 (1)				1.04 (1)	0.90 (1)		0.98 (1)		
II Pro-Gly-Asn-Gln-Ser	1.12 (1)			0.96 (1)	1.07 (1)	0.88 (1)		0.97 (1)		
III Thr-His-Thr-Asn-Ile-Ser-Glu-Ser-His-Pro-Asn-Ala-Thr-Phe		0.95 (1)	0.99 (1)	1.15 (1)	1.02 (1)	1.85 (2)	0.91 (1)	2.00 (2)	2.00 (2)	2.74 (3)
IV Thr-His-Thr-Asn-Ile-Ser						1.09 (1)	0.90 (1)	0.99 (1)	0.84 (1)	2.18 (2)
V His-Pro-Asn-Ala-Thr-Phe		0.97 (1)	0.99 (1)	0.94 (1)				1.00 (1)	1.17 (1)	0.92 (1)

Note. Integer values are shown in parentheses.

whether or not it is glycosylated (8). Therefore, since the oligosaccharide transferase is located in the membranes, it would be interesting to determine if the membrane plays a role in the organization of the secondary structure of the peptide acceptor. This problem was approached by a circular dichroism study of these peptides in aqueous solution in the presence of lipid mixtures, or of organic solvent.

In the present study we give the results we have obtained on the *in vitro* glycosylation of the following peptides, using hen oviduct membranes as enzyme source: Gly-Asn-Gln-Ser (I); Pro-Gly-Asn-Gln-Ser (II); Thr-His-Thr-Asn-Ile-Ser-Glu-Ser-His-Pro-Asn-Ala-Thr-Phe (III); Thr-His-Thr-Asn-Ile-Ser (IV); and His-Pro-Asn-Ala-Thr-Phe (V). A comparison was made between the ability of these peptides to exhibit an organized secondary structure in lipid solutions or in dimethyl sulfoxide and their glycosylation yield.

MATERIALS AND METHODS

Products. Lipids and nonionic surfactant used in this study and UDP-*N*-acetylglucosamine were purchased from Sigma. GDP- ^{14}C mannose (146 mCi/mmol) was obtained from Amersham Radiochemical Centre. All other materials were of the highest grade commercially available and were used without further purification.

Peptide synthesis. The peptides were obtained by the solid-phase method. Their synthesis, purification, and purity control will be described elsewhere. In Table I are listed the different peptides we synthesized and their corresponding amino acid analyses.

Preparation of hen oviduct membranes. The membranes were prepared according to a procedure published by Pless and Lennarz (9) and slightly modified by us (see below). Excised magnum sections of hen oviduct added to 2 vol of 50 mM Tris-HCl, pH 7.2, containing 5% sucrose, 0.9% NaCl, and 1 mM EDTA, were minced in a Waring Blendor for 1 min. Then the tissue was homogenized by 12 strokes in a Potter-Elvehjem. The membrane preparation was used on the same day for the assays of glycosylation.

Assay for glycosylation of peptides. Incubation mixtures containing 500 μg of membranes, 50 mM Tris-HCl, pH 7.5, 10 mM MgCl_2 , 140 mM sucrose, 25 mM NaCl, 200 μM UDP-*N*-acetylglucosamine, 25 μM GDP- ^{14}C mannose, and 100 to 300 nmol of exogenous acceptor peptide in a final volume of 400 μl were incubated at 37°C for 40 min. After incubation the mixture was centrifuged at 5000g for 5 min at 4°C.

ST
SR

The supernatant contained the unglycosylated peptides; the glycosylated peptides were in the pellet. These conditions were at optimal UDP-*N*-acetylglucosamine and GDP-mannose concentrations, and the incorporation of radioactivity into glycopeptide was linearly dependent upon the exogenous peptide concentration.

Characterization of glycosylated peptides. The supernatant of the previous centrifugation was subjected to paper electrophoresis at pH 1.9 (formic acid) and 12 V/cm for 5 h. The unglycosylated peptides were characterized by their electrophoretic mobility. The pellet containing the glycosylated peptides was dissolved in 3% SDS and applied to a Bio-Gel P6 column (Bio-Rad Laboratories, 1.5 × 100 cm) in 0.214 M NaF. The radioactivity of the peaks was determined in a Nuclear Chicago 300 scintillator counter. The labeled peaks were subjected to a concanavalin A-affinity chromatography (Con A-Sepharose from Pharmacia or Con A-Sepharose as described in (10)). The Con A column was equilibrated with 0.05 M Tris-HCl buffer (pH 7.6) containing 1 M NaCl, 1 mM MnCl₂, 1 mM MgCl₂, 1 mM CaCl₂. The samples were passed through the column at a 12 ml/h flow rate. After 10 fractions of 1.6 ml had been collected, elution was carried out with 0.25 M *O*-methyl- α -D-glucose and finally with 0.5 M *O*-methyl- α -D-glucose. Each peak of the elution profile was then dialyzed against distilled water and lyophilized. The glycopeptides obtained were subjected to amino acid analysis and to gas-liquid chromatography (glc).

The amino acid analysis was carried out on a Beckman Multichrom 4255 after hydrolysis of the sample for 24 h at 110°C in 5.6 N HCl. The carbohydrate composition of the glycopeptides was determined after treatment with 0.5 M methanol/HCl for 20 h at 80°C and pertrifluoroacetylation. This procedure allowing the determination of the molar ratio of neutral and amino sugars was carried out on Girdel 300 apparatus equipped with a flame-ionization detector and a glass column 0.3 × 320 cm, 3% OV-210 on Chromosorb HP, nitrogen as carrier gas at 20 ml/min flow rate, and a temperature program from 90 to 220°C at 2.5°C/min.

Preparation of "native" glycopeptides. The glycopeptides obtained by enzymatic digestion of the corresponding glycoproteins will be called "native."

In the first case, GPRP, the glycopeptide was obtained by digestion with papain- β ronase for 48 h at 37°C. The peptide moiety of this glycopeptide is always Gly-Asn-Gln-Ser and the carbohydrate moiety is also well known (11).

In the second case, IgM, the glycopeptide was obtained by digestion of reduced IgM with trypsin and chymotrypsin (12). It corresponds to the heavy-chain 392-405 sequence in which two glycan chains are linked to Asn residues 395 and 402, respectively. The formula of the peptide moiety of this glycopeptide

is Thr-His-Thr-Asn-Ile-Ser-Glu-Ser-His-Pro-Asn-Ala-Thr-Phe.

Preparation of the "native" glycan moiety. The glycan moieties were obtained by hydrazinolysis of glycopeptides following a procedure previously described (3).

Circular dichroism. The circular dichroic spectra were recorded on a Jobin-Yvon R. J. Mark III dichrograph in cells of 0.01- to 0.2-cm path lengths at 25°C. The "native" glycopeptides, synthetic peptides, and sugar moieties were studied in distilled water. Synthetic peptides were also studied in lipid solutions. Solutions were prepared following a procedure given by Wu *et al.* (13). Cerebrosides, β -L-phosphatidylserine, α -L-phosphatidic acid, and dipalmitoyl phosphatidylcholine were first dissolved in 10-tridecylpolyoxyethylene ether used as a nonionic surfactant and the peptides were then added in water. The ellipticity $[\theta]$ is reported in deg·dmol⁻¹·cm² taking the mean residue weight of each peptide.

RESULTS

Hen oviduct membranes catalyze the incorporation of GDP-[¹⁴C]mannose into exogenous acceptor peptides containing the tripeptide sequence Asn-X-Ser/Thr (8, 14, 15). This incorporation was made with dolichyldiphosphate oligosaccharide as intermediate (16, 17). Significant radioactivity was also incorporated into the endogenous protein acceptor (14, 15). Using the conditions described under Materials and Methods we have tested the five peptides in our possession. All experiments were performed at least five times with different membrane preparations. After incubation at 37°C, the supernatants were subjected to paper electrophoresis. It was found that if the peptides were not acceptors, they remained in the supernatant. On the contrary, if glycosylation occurred, no trace of peptide was detectable by this technique. On Fig. 1 are given, as an example, the results corresponding to the assay of glycosylation of Gly-Asn-Gln-Ser and Pro-Gly-Asn-Gln-Ser which are, respectively, unglycosylated and glycosylated. Similar results were obtained in the IgM series where peptide IV remains in the supernatant whereas peptides III and V are undetectable after the glycosylation assays. In Table II are given the results obtained for the five peptides studied here.

To isolate glycopeptides, the membrane



FIG. 1. Paper electrophoresis at pH 1.9 and 12 V/cm for 5 h of: (a) assay of glycosylation of peptide II; (b) peptide II as reference; (c) assay of glycosylation of peptide I; (d) peptide I as reference. The five upper bands for a and c come from membrane preparations.

pellet must be dissolved in 3% SDS. After centrifugation, the supernatant was subjected to chromatography on Bio-Gel P6. On Fig. 2 are presented typical elution

patterns obtained for the five peptides studied here. As observed from Fig. 2, the most important information from these gel filtrations was the appearance of a new radioactive peak when glycosylation occurred. This peak intensity increased with increasing peptide concentration in the initial incubation mixture, whereas in control incubations, or in case of nonglycosylation, elution profiles remained flat in the domain of M_r 4500-1500. The molecular weights of these new molecular species were compatible with the formation of glycopeptides. For the tetradecapeptide III which possesses two Asn residues where glycosylation can take place, two new labeled peaks were obtained. Their molecular weights correspond to two glycopeptides: a dicarbohydrate-chain-linked peptide and a monocarbohydrate-chain-linked peptide. The percentages of each species are 30 and 70, respectively.

Characterization of the products. Having used the enzyme source and the techniques described by Lennarz and co-workers (15), we have accepted, as shown in their work, that an asparagine-glucosamine bond had been effectively obtained; therefore we have studied only sugar composition, amino acid analysis, and electrophoretic mobility (paper electrophoresis) to characterize our glycopeptides.

To obtain more information on the carbohydrate moiety of our synthetic glycopeptides, these products were subjected to Con A-affinity chromatography. Typical

TABLE II

ACCEPTOR OR NONACCEPTOR ACTIVITIES FOR THE GLYCOSYLATION OF THE SYNTHETIC PEPTIDES

Peptides	Acceptor ^a	Nonacceptor
I Gly-Asn-Gln-Ser		+ (0)
II Pro-Gly-Asn-Gln-Ser	+ (1)	
III Thr-His-Thr-Asn-Ile-Ser-Glu-Ser-His-Pro-Asn-Ala-Thr-Phe	+++ (3.62)	
IV Thr-His-Thr-Asn-Ile-Ser		+ (0)
V His-Pro-Asn-Ala-Thr-Phe	++ (1.55)	

Note. The acceptor activity of peptide II was taken as 1 and the values in parentheses give the relative acceptor activity for other peptides.

^a We have considered peptides as acceptors only if no traces of these peptides subsisted in the supernatant. Nevertheless it was possible to glycosylate peptide I with a preparation of microsomal thyroidea in a very poor yield (C. Ronin, personal communication).

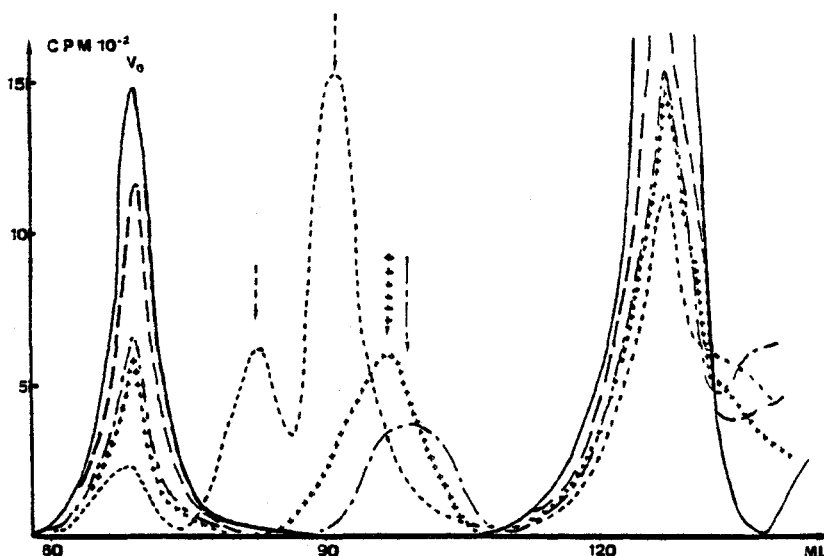


FIG. 2. Elution patterns on a Bio-Gel P6 column (1.5×100 cm) of membrane pellets dissolved in 3% SDS: —, Control; ---, incubation mixtures of peptides I and IV; - - -, incubation mixture of peptide II; + + +, incubation mixture of peptide V; — · —, incubation mixture of peptide III. The arrows indicate the estimated molecular weights: peptide II, — · —, M_r 1900; peptide V + + +, M_r 2100 daltons; peptide III, — — —, M_r 2800 and 4300, respectively.

elution curves obtained for the three glycosylated peptides are presented in Fig. 3.

Each peak obtained was then subjected to glc. In Table III are summarized the molar ratios of sugars taking two *N*-acetylglucosamine residues as a reference. This composition in saccharides is in fair agreement with the knowledge and the results obtained on the biosynthetic pathway of asparagine-linked oligosaccharides.

A great number of glucose residues appear in our gas-liquid chromatograms. This is not surprising since the Con A-Sepharose column is eluted with *O*-methyl- α -D-glucose. Unfortunately, in the case of the dicarbohydrate-linked tetradecapeptide, passage on the Con A-Sepharose column is inauspicious since all of the radioactivity remains on the column. One hypothesis to explain such a result is the presence of two polymannosidic chains on the same peptidic moiety. The glycopeptides were also characterized by their electrophoretic mobility relative to "native" glycopeptides and by autoradiography (data not shown). Moreover we have to

underline that we never found traces of synthetic peptides in the membrane pellet. This is probably due to the action of peptidases present in the membrane.

The synthetic glycopeptides were also characterized by their amino acid composition. The results are given in Table IV.

Circular dichroic study. First we studied by circular dichroism the "native" glycopeptides dissolved in water and then the oligosaccharide chains obtained by hydrazinolysis. The spectra of the peptide moieties were obtained by subtraction, and they could then be compared with the spectra of the synthetic peptides (following a procedure given in (3)). Such a comparison was carried out with the two "native" glycopeptides, that is, Gly-Asn-Gln-Ser and the tetradecapeptide. The dichroic spectra are characteristic of an unordered structure and in the two cases the superimposition was perfect (Fig. 4). In this way, we confirmed that the peptide moiety of the glycopeptides has no special organization in water and that no interactions take place between the oligosaccharide and the peptide moieties.

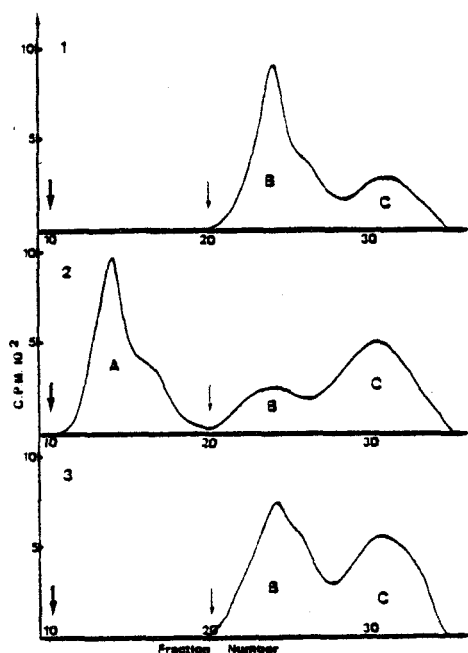


FIG. 3. Con A-affinity chromatography of glycosylated peptides: (1) peptide II exhibits two peaks, B and C; (2) peptide III exhibits three peaks, A, B, and C; (3) peptide V exhibits two peaks, B and C. The two arrows, — and - - , correspond to the elution with 0.25 and 0.5 M *O*-methyl- α -D-glucose, respectively.

All the synthetic peptides were also studied by circular dichroism in water: their spectra were always characteristic of an unordered structure and exhibited an intense negative band located near 200 nm. But if the dichroic spectra are recorded in aqueous solutions containing lipids or an organic solvent, such as dimethyl sulfoxide, significant modifications are observed for a few peptides.

On Fig. 5 are shown the dichroic spectra obtained in 10% DMSO for the five peptides presented in this study.

For I and IV, the shape of the spectrum looks similar to that in aqueous solution. But for III, and particularly for II and V, a new positive band is observed at 217 nm, reflecting the formation of a favored secondary structure.

Moreover, studies of the five peptides in aqueous lipid mixtures, and especially in the presence of α -L-phosphatidylserine, α -L-phosphatidic acid, and cerebrosides,

have demonstrated important changes in the shape of the dichroic spectrum of peptides II, III, and V (Fig. 6). As in DMSO solutions, the appearance of a new band at 217 nm is consistent with the organization of our peptides in lipid mixtures.

It is interesting to compare these results with those obtained from the glycosylation assays: peptides II, III, and V were glycosylated, while I and IV were not.

DISCUSSION

In the present study we tried to see if a structural requirement was necessary for glycosylation to take place on a peptide bearing the Asn-X-Ser/Thr sequence. Indeed our results seem to show that organization of the acceptor peptide increases the yield of glycosylation. However, this organization does not constitute a necessary condition, since other authors (8, 18, 19) have succeeded in glycosylating peptides as small as tri- or tetrapeptides. Yet all these authors had always established that the yield of glycosylation increased with the length of the peptides, and it is not impossible to think that the structural organization of a peptide increases with its length. Moreover, our results show that

TABLE III
MOLAR RATIO IN SUGARS OF THE
GLYCOSYLATED PEPTIDES

Glycosylated peptide	N-Acetyl-glucosamine	Mannose
Peptide II		
Peak B (75%)	2	5 \pm 1
Peak C (25%)	2	7 \pm 1
Peptide III		
Peak A (45%)	2	3
Peak B (20%)	2	5 \pm 1
Peak C (35%)	2	7 \pm 1
Peptide V		
Peak B (55%)	2	6 \pm 1
Peak C (45%)	2	8 \pm 1

Note. Obtained by gic as described under Materials and Methods, after Con A-affinity chromatography. The relative percentage of each peak is given.

61
52

TABLE IV
AMINO ACID ANALYSIS OF THE GLYCOSYLATED PEPTIDES

Glyco-peptide	Pro	Gly ^a	Asp	Glu	Ser	Thr	His	Ile	Ala	Phe	N-Acetyl-gluco-samine
II	1.10(1)	1.50(1)	0.99(1)	1.20(1)	1.00(1)						1.90(2)
III	1.17(1)	0.50(0)	2.15(2)	1.15(1)	2.20(2)	2.91(3)	1.84(2)	0.75(1)	1.20(1)	1.07(1)	2.10(2)
V	0.93(1)	0.40(0)	1.00(1)			1.14(1)	1.03(1)		1.00(1)	0.92(1)	1.72(2)

Note. Integer values are shown in parentheses.

^a The excess Gly was eluted first in the Con A-affinity chromatography (10).

an interaction takes place between hen oviduct membranes and acceptor peptides. This seems to be the first report of this phenomenon, since no one has ever pointed out such a result. This interaction seems to be a general characteristic of glycosylated peptides, since, with all our hen oviduct membrane preparations, the peptide-membrane interaction was always observed, even if, as expected, slight differences exist from one preparation to another.

Let us now try to compare these results with the ones obtained by circular dichroism. As we have observed, the spectra obtained in aqueous solution for the three acceptor peptides differ from those of the same peptides in lipid mixtures or DMSO aqueous solution. More specifically, a new positive band appears at 217 nm in addition to the intense negative band centered at about 200 nm. Are these spectra characteristic of a β -turn conformation? First, it is interesting to observe that they bear some resemblance to those computed by Woody (20) for various β -bends (see, e.g., Fig. 2b in Ref. (20)). But comparison with previously published β -turn spectra is not easy. In a recent work devoted to the study of proteins by vacuum ultraviolet circular dichroism, Brahms and Brahms (21) published a dichroic spectrum which seems representative of different β -reverse turns. They are characterized by a strong negative band and positive bands centered in the regions 182-189 and 202-206 nm, respectively, and a weak positive band with a maximum at about 225 nm. Our spectra

do not present these features, but they are reminiscent of two spectra published by Howard *et al.* (22). These authors studied four peptides exhibiting different theoretical probabilities of occurrence of β -bends. Two of them present a high probability to adopt a bend conformation in agreement

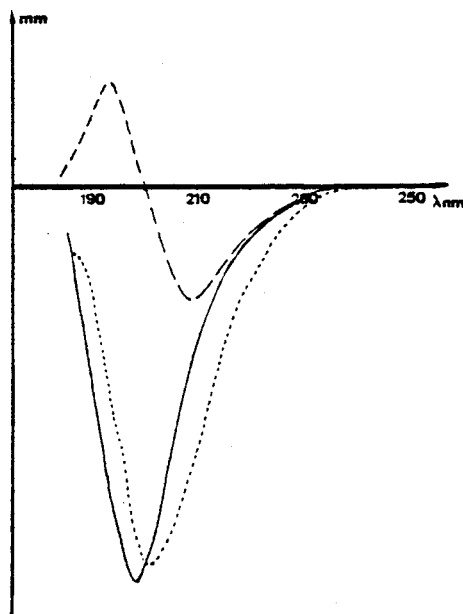


FIG. 4. Dichroic spectra in water of: —, "native" glycopeptide Gly-Asn-Gln-Ser; ---, carbohydrate moiety corresponding to the "native" glycopeptide Gly-Asn-Gln-Ser; ···, difference between "native" glycopeptide and carbohydrate moiety spectra. This difference spectrum is exactly superimposed with those of the synthetic peptide Gly-Asn-Gln-Ser. The conditions used were the same described in Ref. (3).

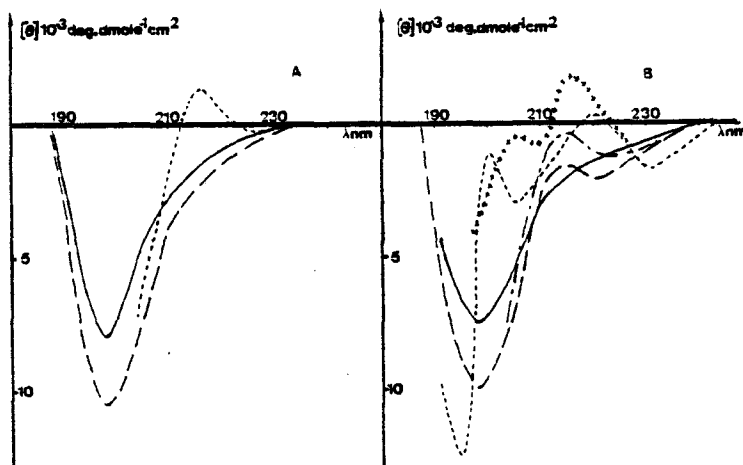


FIG. 5. Dichroic spectra in a water-10% DMSO mixture of: (A) —, peptide I in water and in a water-10% DMSO mixture; ---, peptide II in water; —, peptide II in a water-10% DMSO mixture. (B) —, Peptide IV in water and in a water-10% DMSO mixture; ---, peptide III in water; —, peptide III in a water-10% DMSO mixture; ---, Peptide V in water; + + +, Peptide V in a water-10% DMSO mixture.

with the results of nuclear magnetic resonance and conformational energy calculations. Therefore, it is not impossible to think that their corresponding dichroic spectra are characteristic of a bend conformation with a positive band at 213-218 nm and a strong negative band at about 200 nm.

It is more reasonable to expect that a fraction of our peptides can exist in the bend conformation; since the flexibility inherent in short peptides may be important, it is possible to think that this flexibility decreases when an interaction exists between the lipids—or the membrane—and the peptides.

It is interesting to note that in the case of Howard *et al.* (22) as well as in our work, the peptides studied were also the models of the so-called "native" peptides. In both cases, whatever the sequences of the peptides and the solution conditions used to perform the dichroic studies, the spectra always exhibited the same shape with the new band located at the same wavelength and the intensities in the same range. Therefore, organization of the peptides would correspond to a favored structure.

In the case of peptides I and II, Chou and Fasman's (7) calculations show the pos-

sibility of a β -turn conformation adopted by Pro-Gly-Asn-Gln-Ser, whereas any ordered secondary structure is not expected for the tetrapeptide. This result was experimentally confirmed. Circular dichroism studies show the onset of an ordered structure in the case of the pentapeptide when placed in the presence of lipids; on the other hand, no change was observed for the tetrapeptide under similar conditions. Moreover, Pro-Gly-Asn-Gln-Ser can be glycosylated, but Gly-Asn-Gln-Ser cannot.

If we now consider the tetradecapeptide and its two subhexapeptides, we see that the tetradecapeptide (III) can be glycosylated. This result is very surprising since, if we consider the predictive secondary structure determination according to Chou and Fasman (7), peptide IV, like peptide V, can adopt a β -turn structure. However peptide V is much more hydrophobic than peptide IV and, in the case of such small peptides, the peptide-lipid interaction is certainly a function of their relative hydrophobicity. For tetradecapeptide III, it is probable that the Asn residue corresponding to peptide IV can be glycosylated only because the C-ter-

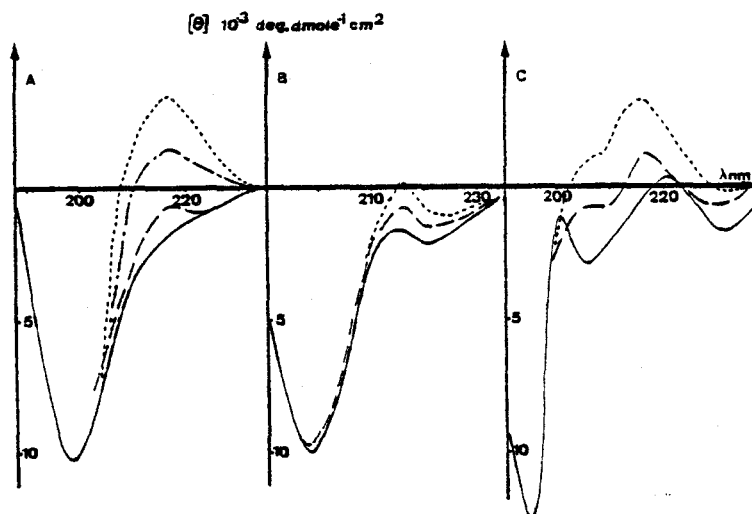


FIG. 6. Dichroic spectra in water and in aqueous lipid mixture of: (A) —, peptide II in water; ---, 0.4 μM peptide II in 8 μM $\alpha\text{-L}$ -phosphatidic acid dissolved in 900 μM 10-tridecylpolyoxyethylene ether; ----, 0.4 μM peptide II in 4 μM $\alpha\text{-L}$ -phosphatidylserine dissolved in 900 μM 10-tridecylpolyoxyethylene ether; —·—, 0.4 μM peptide II in 8 μM $\alpha\text{-L}$ -phosphatidylserine dissolved in 900 μM 10-tridecylpolyoxyethylene ether. (B) —, Peptide III in water; ---, 0.15 μM peptide III in 1.5 μM cerebrosides dissolved in 900 μM 10-tridecylpolyoxyethylene ether; ----, 0.15 μM peptide III in 3 μM cerebrosides dissolved in 900 μM 10-tridecylpolyoxyethylene ether. (C) —, Peptide V in water; ---, 0.3 μM peptide V in 3 μM $\alpha\text{-L}$ -phosphatidylserine dissolved in 900 μM 10-tridecylpolyoxyethylene ether; ----, 0.3 μM peptide V in 6 μM $\alpha\text{-L}$ -phosphatidylserine dissolved in 900 μM 10-tridecylpolyoxyethylene ether.

minal part (corresponding to the peptide V) interacts with the membrane. Such a hypothesis could explain why, as observed from Fig. 2, one out of the two Asn residues was glycosylated with a poor yield. It is interesting to note that, *in vivo*, the tetradecaglycopeptide has two kinds of oligosaccharide chains (23, 24), one polymannosidic and one *N*-acetylactosaminidic chain. The action of α -mannosidase and the glycosyltransferases after folding of the glycoprotein was certainly due to steric hindrance.

It appears certain, therefore, that the spatial organization of an acceptor peptide (β -turn or not) plays a very important role in the yield of glycosylation. We believe, as stated by Hart *et al.* (8), that "the accessibility of the tripeptide sequence to the oligosaccharide transferase is what determines whether or not the asparagine residue is glycosylated." Since it has been shown (3-6) that oligosaccharide chains

are always linked to a β -turn, we may think that glycosylation is favored *in vivo* by the presence of β -turns allowing accessibility to the asparagine residue. Therefore, we believe that the presence of a β -turn is not the condition necessary for oligosaccharide transferase activity, but, since the β -turn must exist in the finished glycoprotein to orient the oligosaccharide chain, this conformational feature is used in the glycosylation period too.

ACKNOWLEDGMENTS

We wish to thank Dr. J. P. Le Breton for the gift of immunoglobulin M, Professors P. Degand and A. Boersma for the gift of GPRP, Drs. B. Bayard and J. P. Kerckaert for the use of (and help with) the Girdel 300 apparatus, Dr. J. Brison (University of Mons, Belgium) and Dr. S. St-Pierre (University of Sherbrooke, Quebec, Canada) for allowing us to perform peptide synthesis in their laboratories. We thank Mr. G. Lecocq for his excellent technical assistance. This work was supported by RCP 529 (CNRS) and by CRL 30.30.18 INSERM.

REFERENCES

1. MARSHALL, R. D., AND NEUBERGER, A. (1970) *Advan. Carbohydr. Chem. Biochem.* 25, 407-478.
2. PARODI, A. J., AND LELOIR, L. F. (1979) *Biochim. Biophys. Acta* 559, 1-37.
3. AUBERT, J. P., AND LOUCHEUX-LEFEBVRE, M. H. (1976) *Arch. Biochem. Biophys.* 175, 400-409.
4. AUBERT, J. P., BISERTE, G., AND LOUCHEUX-LEFEBVRE, M. H. (1976) *Arch. Biochem. Biophys.* 175, 410-418.
5. BEELEY, J. G. (1977) *Biochem. Biophys. Res. Commun.* 76, 1051-1055.
6. NAGARAJAN, M., AND RAO, V. S. R. (1977) *Curr. Sci. India* 46, 395-401.
7. CHOU, P. Y., AND FASMAN, G. D. (1974) *Biochemistry* 13, 211-245.
8. HART, G. W., BREW, K., GRANT, G. A., BRADSHAW, R. A., AND LENNARZ, W. J. (1979) *J. Biol. Chem.* 254, 9747-9753.
9. PLESS, D. D., AND LENNARZ, W. J. (1975) *J. Biol. Chem.* 250, 7014-7019.
10. KERCKAERT, J. P., AND BAYARD, B. (1980) *Biochem. Biophys. Res. Commun.* 92, 95-101.
11. DEGAND, P., BOERSMA, A., ROUSSEL, P., RICHEL, C., AND BISERTE, G. (1975) *FEBS Lett.* 54, 189-192.
12. AUBERT, J. P., manuscript in preparation.
13. WU, C. S. C., LEE, N. M., LOH, H. H., YANG, J. T., AND LI, C. H. (1979) *Proc. Nat. Acad. Sci. USA* 76, 3656-3659.
14. PLESS, D. D., AND LENNARZ, W. J. (1977) *Proc. Nat. Acad. Sci. USA* 74, 134-138.
15. STRUCK, D. K., LENNARZ, W. J., AND BREW, K. (1978) *J. Biol. Chem.* 253, 5786-5794.
16. LENNARZ, W. J. (1975) *Science* 188, 986-991.
17. CHEN, W. W., LENNARZ, W. J., TARENTINO, A. L., AND MALEY, F. (1975) *J. Biol. Chem.* 250, 7006-7013.
18. RONIN, C., BOUCHILLOUX, S., GRANIER, C., AND VAN RIETSCHOTEN, J. (1978) *FEBS Lett.* 96, 179-182.
19. BAUSE, E. (1979) *FEBS Lett.* 103, 296-299.
20. WOODY, R. W. (1974) in *Peptides, Polypeptides and Proteins* (Blout, E. R., Bovey, F. A., Goodman, M., and Lotan, N., eds.), pp. 338-350, Wiley, New York.
21. BRAHMS, S., AND BRAHMS, J. (1980) *J. Mol. Biol.* 138, 149-173.
22. HOWARD, J. C., ALI, A., SCHERAGA, H. A., AND MOMANY, F. A. (1975) *Macromolecules* 8, 607-622.
23. TARENTINO, A. L., PLUMMER, T. H., JR., AND MALEY, F. (1975) *Biochemistry* 14, 5516-5523.
24. HICKMAN, S., KORNFIELD, R., OSTERLAND, C. K., AND KORNFIELD, S. (1972) *J. Biol. Chem.* 247, 2156-2163.

D) CONCLUSIONS ET PERSPECTIVES

Le travail que nous venons d'exposer nous a donc permis de démontrer certaines notions concernant l'organisation spatiale des glycoprotéines, en particulier au niveau du point d'attache sucre-protéine. De plus, il nous a ouvert d'autres voies de recherche.

Il apparaît tout d'abord qu'aucune interaction forte n'existe entre la partie peptidique et la partie glycanique d'une glycoprotéine. Cette constatation est en accord avec le modèle proposé précédemment: les chaînes glycaniques sont plantées sur le noyau protéique. Ce serait donc la partie peptidique qui dirigerait l'orientation spatiale des chaînes glycaniques grâce à l'existence d'un coude β au niveau du point d'attache. Cette structuration permet à la chaîne glycanique d'être toujours localisée à l'extérieur du " core " peptidique . De plus, le coude β constitue une structure locale assez rigide puisqu'il a été montré (Chou et Fasman 1977) qu'il est le plus souvent stabilisé par des interactions $\alpha\alpha$, $\beta\beta$ ou $\alpha\beta$ entre les régions situées de part et d'autre de ce coude. Cette structuration impose à la chaîne glycanique une localisation précise dans l'espace par rapport au noyau protéique, ce qui permettrait de comprendre pourquoi des glycoprotéines ayant des chaînes glycaniques identiques peuvent avoir des fonctions différentes.

L'étude des conditions de N-glycosylation *in vitro* nous a permis de mettre en évidence l'importance des coudes β et nous avons pu établir une corrélation entre la structure secondaire des peptides accepteurs et la possibilité de les glycosyler.

Ces résultats ainsi que les techniques utilisées nous ouvrent deux voies nouvelles de recherche:

- La première concerne l'enzyme de glycosylation.

Puisqu'il est spécifique à la fois de la séquence code Asn-X-Ser/Thr et de sa structuration en coude β , peut-on

imaginer une chromatographie d'affinité dans laquelle le ligand serait un peptide adoptant dans certaines conditions de solvant une structure en coude ? .

Les premiers essais effectués nous ont montré qu'il était possible de préparer la dolichyldiphosphate-protéine oligosaccharide transférase par cette technique. Dans ce cas, le tétradécapeptide T-H-T-N-I-S-E-S-H-P-N-A-T-F (le peptide III de notre étude de glycosylation) est accroché de façon covalente sur une colonne de Sépharose. Les membranes " solubilisées " dans un détergent sont éluées en présence de 10% de DMSO . Dans ces conditions, l'enzyme est retenue. L'utilisation du même tampon mais sans DMSO permet ensuite de décrocher cette enzyme par destruction de la structure secondaire du peptide.

Les résultats de ce travail feront l'objet du prochain D.E.A. de M. Chiroutre.

Ceci nous semble très important car il existe certainement dans l'organisme d'autres enzymes qui sont spécifiques à la fois de la structure primaire et de la structure secondaire du substrat.

- La deuxième voie de recherche découle de la première. En effet, l'étude physicochimique de cette enzyme de glycosylation, qui est une protéine membranaire, devrait permettre notre initiation à l'étude des protéines et glycoprotéines membranaires en général. Comme nous l'avons rappelé dans l'introduction, les membranes jouent un rôle primordial dans de nombreux phénomènes biologiques normaux et surtout pathologiques, une étude de ces membranes de cellules normales et transformées s'imposera donc en essayant d'appliquer des techniques physicochimiques.

E) APPENDICE TECHNIQUE

I) ÉNONCÉ DES RÈGLES PERMETTANT DE RECHERCHER LES STRUCTURES

ORGANISÉES SELON CHOU ET FASMAN (CHOU ET FASMAN 1974)

Dans ce paragraphe, nous allons énoncer les règles permettant d'évaluer la structure secondaire complète d'une protéine.

A) Recherche des régions α hélicoïdales

Il faut découper la chaîne peptidique en groupes de six résidus et il faut que sur ces six résidus quatre au moins soient initiateurs d'hélice α (c'est à dire H_α ou h_α) La propagation de l'hélice est arrêtée par les tétrapeptides suivants: b_4 , b_3i , b_3h , b_2i_2 , b_2ih , b_2h_2 , bi_3 , bi_2h et i_4 . Le résidu proline ne peut exister ni au milieu ni à l'extrémité C-terminale d'une région hélicoïdale.

Les résidus proline, acide aspartique, acide glutamique préfèrent l'extrémité N-terminale alors que les résidus histidine, lysine et arginine préfèrent l'extrémité C-terminale.

En conclusion, tout segment de six résidus ou plus satisfaisant aux conditions précédentes et dont $P \langle \alpha \rangle \gg 1,03$ $P \langle \beta \rangle$ est considéré comme hélicoïdal.

B) Recherche des feuillets plissés β

On découpe cette fois la chaîne peptidique en groupes de cinq résidus et il faut que sur ces cinq résidus, trois au moins soient initiateurs de structure β (c'est à dire H_β ou h_β) La propagation de cette structure est soumise aux mêmes règles que l'hélice α . Les résidus proline et acide glutamique apparaissent rarement dans cette structure.

En conclusion, tout segment de cinq résidus ou plus

satisfaisant aux conditions précédentes et dont $P \langle \beta \rangle \gg 1,05$
 $P \langle \alpha \rangle$ est prévu comme adoptant une structure en feuillets plis-
 sés β

C) Recherche des coudes β

Tout segment de quatre résidus dont $P \langle \beta T \rangle > P \langle \alpha \rangle$
 et $P \langle \beta T \rangle > P \langle \beta \rangle$ et dont $p (\beta T) > 10^{-4}$ est considéré comme
 coude β .

II) DICHROÏSME CIRCULAIRE.

Dans ce paragraphe, nous allons faire quelques rappels
 concernant cette technique.

A) Origine phénoménologique.

Lorsqu'une onde électromagnétique linéairement pola-
 risée traverse un milieu possédant un ou plusieurs chromophores
 optiquement actifs, en plus de la différence d'indice de réfrac-
 tion, on observe une différence d'absorption entre la lumière
 circulairement polarisée droite et la lumière circulairement
 polarisée gauche.

La lumière devient donc elliptiquement polarisée et
 il y a apparition du dichroïsme circulaire.

L'angle Ψ (Figure VII) appelé ellipticité est relié
 aux indices d'absorption K_G et K_D des lumières circulairement
 polarisées droite et gauche par la relation:

$$\Psi = \frac{\pi}{\lambda} (K_G - K_D) l$$

Ψ : ellipticité en radians

λ : longueur d'onde

l : épaisseur de l'échantillon.

On a introduit deux grandeurs macroscopiques pour

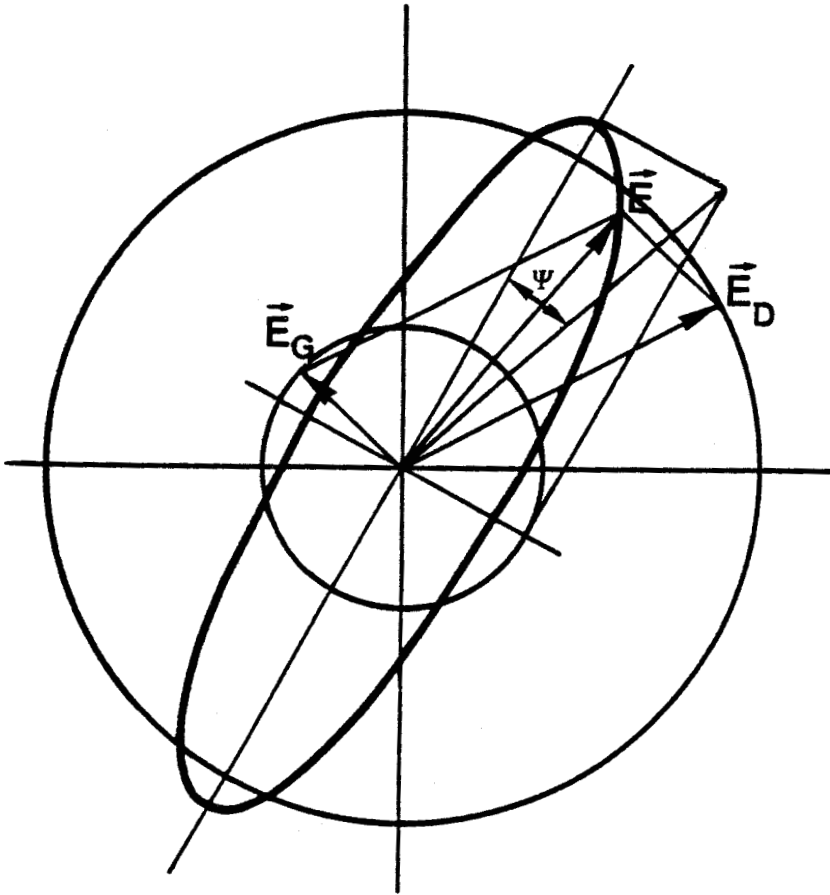


Figure 7



exprimer l'ellipticité d'un composé d'une manière commode:

- l'ellipticité spécifique $[\Psi]$

$$[\Psi] = \frac{\Psi}{lc}$$

Ψ : ellipticité exprimée en degrés.

l : épaisseur de l'échantillon en dm

c : concentration en g/ml

- l'ellipticité molaire $[\Theta]$

Par définition $[\Theta] = [\Psi] \frac{M_0}{100}$

Sachant que l'indice d'absorption K est relié au coefficient d'extinction molaire ϵ par la relation:

$$K = \log 10 \frac{\lambda}{4\pi} C \epsilon \text{ avec } C \text{ en mole / l}$$

On montre que:

$$[\Theta] = 3300 \Delta\epsilon$$

$\Delta\epsilon = \epsilon_G - \epsilon_D$ = absorption différentielle dichroïque

$[\Theta]$ s'exprime par convention en $\text{deg. cm}^2 \cdot \text{dmole}^{-1}$ qui est une unité hors système.

Le dichrographe fournit directement la différence de densité optique $A_G - A_D$

$$\Delta\epsilon = \frac{M}{Cl} (A_G - A_D) s$$

M : Masse moléculaire du soluté (en g) ou masse moyenne du résidu dans les protéines

C : Concentration (en g/l)

l : épaisseur de la cuve en cm

$A_G - A_D$: mesure en mm sur le spectre

s : sensibilité de l'appareil en mm^{-1}

B) Dichroïsme circulaire des peptides, polypeptides

et protéines. Conformation des chaînes peptidiques

Les peptides, polypeptides et protéines sont par définition un enchaînement linéaire d'acides aminés liés de façon covalente par des liaisons amide.



La figure VIII nous montre un tel enchaînement. On peut considérer que la liaison CO-NH est plane (Figure IX) et ceci entraîne qu'une rotation est possible autour des liaisons NH-C α et C α -CO.

Il faudra donc, pour pouvoir préciser la structure d'un peptide, connaître les positions respectives de chaque groupement constituant l'édifice moléculaire.

On désigne par ϕ l'angle de torsion autour de la liaison NH-C α et par ψ l'angle de torsion autour de la liaison C α -CO.

Par convention (IUPAC-IUB 1970), pour une chaîne complètement étendue on aura $\psi = \phi = 180^\circ$

C) Rappels concernant les chromophores optique-

ment actifs

On peut ranger les chromophores optiquement actifs en deux groupes:

- Les chromophores intrinsèquement dissymétriques: l'activité optique dépend directement de la géométrie de la molécule

- Les chromophores intrinsèquement symétriques: l'activité optique dépend de l'environnement du chromophore.

C'est dans ce deuxième groupe que l'on trouve le chromophore CO-NH ainsi que les chromophores tyrosine, phénylalanine, tryptophane, histidine

L' activité optique du groupement CO-NH est directement reliée à la structure secondaire du polypeptide ou de la protéine.

Dimension d'une chaîne polypeptidique
complètement étendue

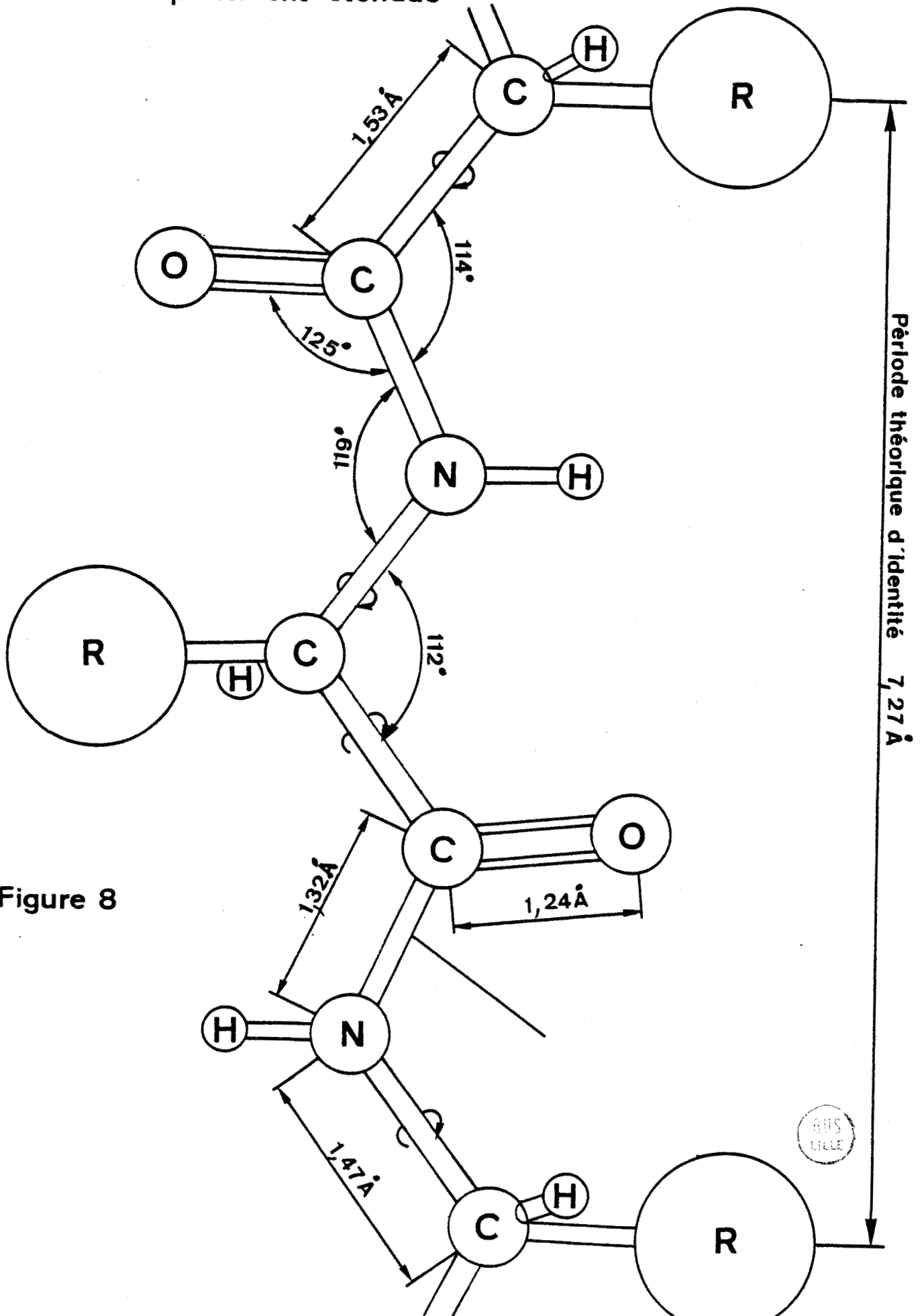


Figure 8

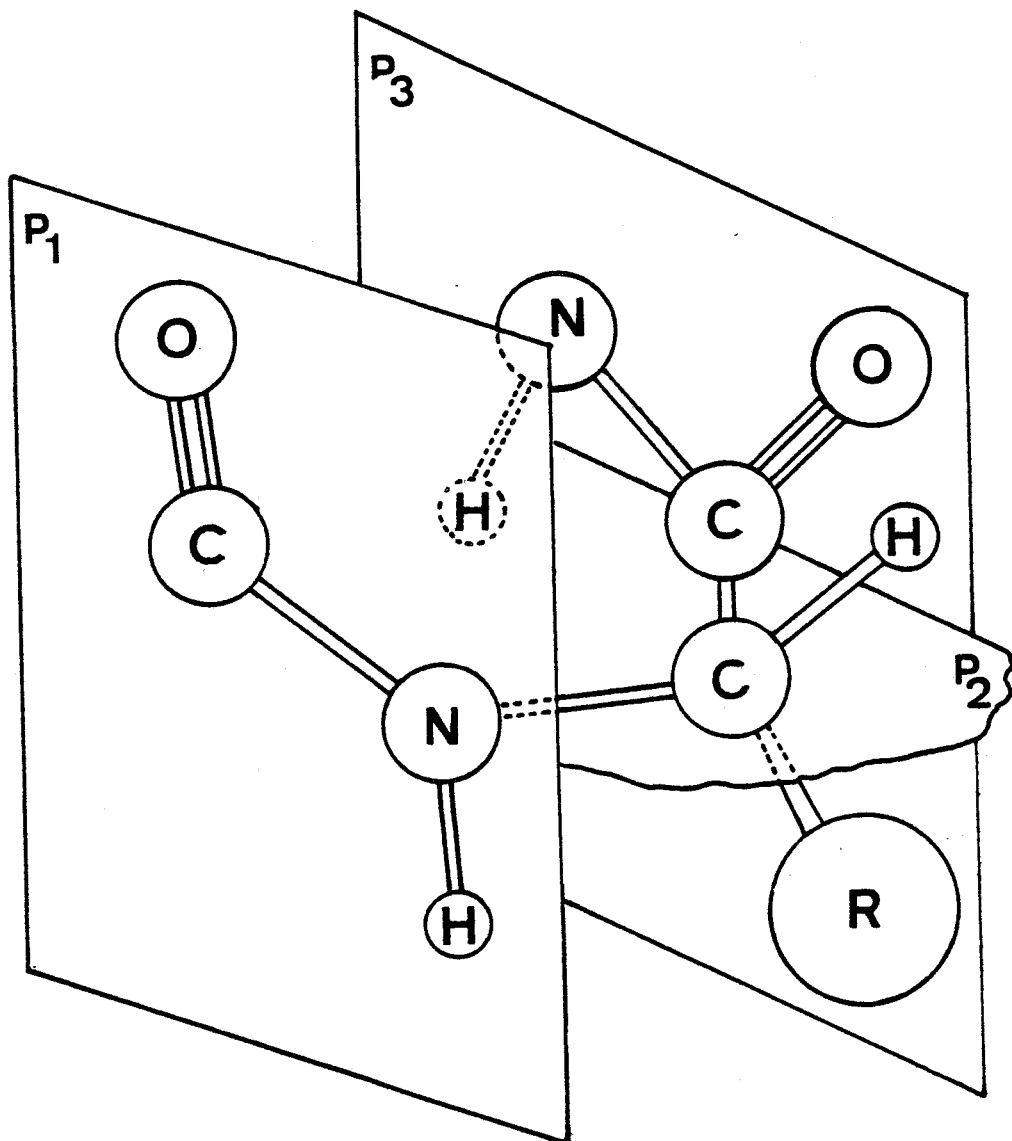
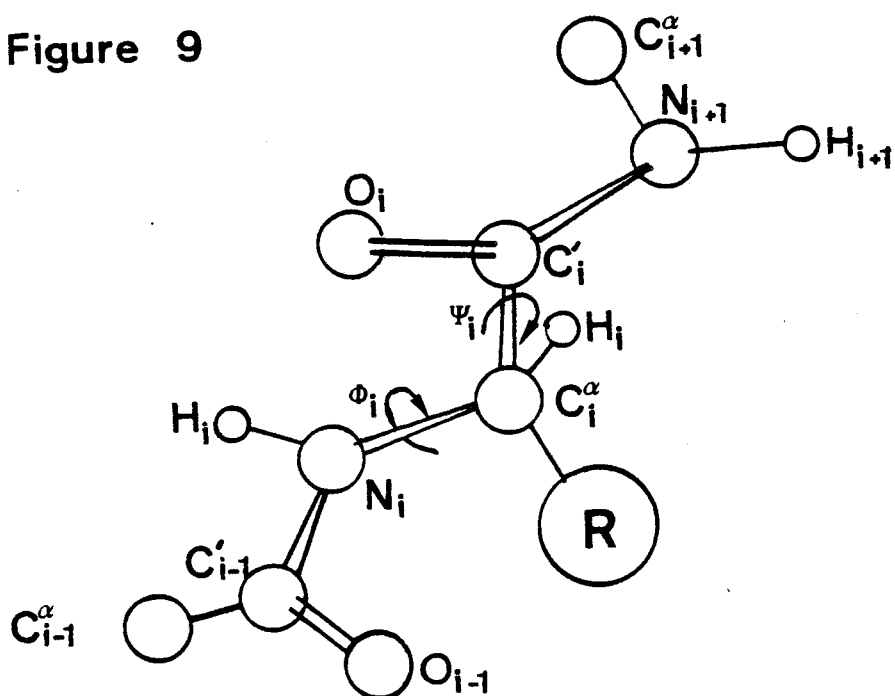


Figure 9



Les effets Cotton correspondants sont appelés " intrinsèques " et se situent dans le domaine 150-240 nm.

- Vers 150 nm: la transition $n \rightarrow \sigma^*$
- Vers 165 nm: la transition $\pi \pi_2 \rightarrow \pi^*$
- Vers 190 nm: la transition $\pi_1 \rightarrow \pi^*$
- Vers 220 nm: la transition $n \rightarrow \pi^*$

Expérimentalement, seules les deux dernières transitions sont accessibles.

En ce qui concerne les chromophores des acides aminés aromatiques, les effets Cotton sont appelés " extrinsèques " et se situent dans le domaine 240-300 nm.

BIBLIOGRAPHIE

- ASHWELL, G. et MORELL, A.G.
(1974) Adv. Enzymol. Relat. Areas Mol. Biol. 41, 99-128(p 2)
- AUBERT, J.P., BOERSMA, A., LOUCHEUX-LEFEBVRE, M.H., DEGAND, P. et BISERTE, G.
(1975) FEBS Lett. 56, 263-267. (p 16, 21, 22)
- AUBERT, J.P. et LOUCHEUX-LEFEBVRE, M.H.
(1975) Biopolymers 14, 2429-2433. (p 16)
- AUBERT, J.P. et LOUCHEUX-LEFEBVRE, M.H.
(1976) Arch. Biochem. Biophys. 175, 400-409. (p 16, 19)
- AUBERT, J.P., BISERTE, G. et LOUCHEUX-LEFEBVRE, M.H.
(1976 a) Arch. Biochem. Biophys. 175, 410-418 (p 10, 26)
- AUBERT, J.P., BAYARD, B. et LOUCHEUX-LEFEBVRE, M.H.
(1976 b) Carbohyd. Res. 51, 263-268. (p 10)
- AUBERT, J.P., KERCKAERT, J.P., BAYARD, B. et LOUCHEUX-LEFEBVRE, M.H.
(1977 a) FEBS Lett. 78, 194-198 (p 16)
- AUBERT, J.P., DRUPT, F., ROUSSEAU, J. et LOUCHEUX-LEFEBVRE, M.H.
(1977 b) FEBS Lett. 84, 375-378 (p 10, 16)
- AUBERT, J.P., HELBECQUE, N. et LOUCHEUX-LEFEBVRE, M.H.
(1981 a) Arch. Biochem. Biophys. sous presse. (p 16)
- AUBERT, J.P., PORCHET, N., BOERSMA, A., LOUCHEUX-LEFEBVRE, M.H. et DEGAND, P.
(1981 b) Biochim. Biophys. Acta Soumis pour publication. (p 21)
- AUSTEN, B.M.
(1979) FEBS Lett. 103, 308-313. (p 32)
- BABCSZINSKI, P.
(1980) FEBS Lett., 117, 207-211. (p 3)

- BAUSE, E. et LEHLE, L.
(1979) Eur. J. Biochem. 101, 531-540. (p 29,42)
- BAYARD, B. et FOURNET, B.
(1975) Carbohyd. Res. 46, 75-86. (p 18)
- BAYARD, B. et ROUX, D.
(1975) FEBS Lett. 55, 206- 207. (p 17)
- BAYARD, B. et KERCKAERT, J.P.
(1980) Biochem. Biophys. Res. Comm. 95, 777-784.(p 4)
- BAYARD, B. et KERCKAERT, J.P.
(1981) Eur. J. Biochem. 113, 405-414. (p 4)
- BEELEY, J.G.
(1977) Biochem. Biophys. Res. Comm. 76, 1051-1055. (p 30)
- BIELINSKA, M. et BOIME, I.
(1978) Proc. Nat. Acad. Sci. USA 75, 1768-1772. (p 34)
- BOERSMA, A.
(1977) Thèse Biologie Humaine Lille. (p 21)
- BRAHMACHARI, S.K., ANANTHANARAYANAN, V.S., BRAHMS,S., BRAHMS,J.
RAPAKA, R.S. et BHATNAGAR, R.S.
(1979) Biochem. Biophys. Res. Comm. 86, 605-612. (p 13)
- BRAHMACHARI, S.K. et ANANTHANARAYANAN, V.S.
(1979) Proc. Nat. Acad. Sci. USA 76, 5119-5123. (p 32)
- BRAHMS, S.,BRAHMS,J., SPACH, G. et BRACK, A.
(1977) Proc. Nat. Acad. Sci. USA 74, 3208-3212. (p 13)
- BRAHMS, S. et BRAHMS, J.
(1980) J. Mol. Biol. 138, 149-178. (p 15)

- BRETTAUER, R.K. et WU, S.
(1975) Arch. Biochem. Biophys. 167, 151-160. (p 3)
- BRILES, E.B. et KORNFELD, S.
(1978) Trends. Biochem. Res 3, 223-226. (p 35)
- BURGESS, A.W., PONNUSWAMY, P.K. et SCHERAGA, H.A.
(1974) Isr. J. Chem. 12, 239-286. (p 9)
- BUSH, C.H., SARKAR, S.K. et KOPPLE, K.D.
(1978) Biochemistry 17 4951-4954. (p 13)
- CACCAM, J.F., JACKSON, J.J. et EYLAR, E.M.
(1969) Biochem. Biophys. Res. Comm. 35, 505-511. (p 34)
- CHANG, C.T., WU, C.S.C. et YANG, J.T.
(1978) Anal. Biochem. 91, 13-31. (p 14)
- CHEN, Y.H., YANG, J.T. et MARTINEZ, H.M.
(1972) Biochemistry 11, 4120-4131. (p 13)
- CHEN, Y.H., YANG, J.T. et CHAU, K.H.
(1974) Biochemistry 13, 3350-3359. (p 19)
- CHEN, W.W., LENNARZ, W.J., TARENTINO, A.L. et MALEY, F.
(1975) J. Biol. Chem. 250, 7006-7013. (p 36)
- CHOU, P.Y. et FASMAN, G.D.
(1974) Biochemistry 13, 211-249. (p 8, 47)
- CHOU, P.Y. et FASMAN, G.D.
(1977) J. Mol. Biol. 115, 135-175. (p 9, 11, 44)
- CHOU, P.Y. et FASMAN, G.D.
(1978) Ann. Rev. Biochem. 47, 251-276. (p 9, 11, 32)

CODINGTON, J.F., VAN DER EIJNDEN, D.H. et JEANLOZ, R.W.
(1978) Cell Surface carbohydrate chemistry (R.E. HARMON
Ed.) Academic press, New-York and London. 49-66. (p 2)

DEGAND, P., BOERSMA, A., ROUSSEL, P., RICHEL, C. et
BISERTE, G.
(1975) FEBS Lett. 54, 189-192. (p 22, 36)

DEGAND, P., AUBERT, J.P., BOERSMA, A., RICHEL, C., LOUCHEUX-
LEFEBVRE, M.H. et BISERTE, G.
(1976) FEBS Lett. 63, 137-141. (p 16)

DEISENHOFER, J. et STEIGEMANN, W.
(1975) Acta Crystallogr. B 31, 238-250. (p 25)

DORLAND, L., SCHUT, B.L., VLIENGENTHART, J.F.G., STRECKER, G.,
FOURNET, B., SPIK, G. et MONTREUIL, J.
(1977a) Eur. J. Biochem. 73, 93-97. (p 5)

DORLAND, L., HAVERKAMP, J., SCHUT, B.L., VLIENGENTHART, J.F.G.,
SPIK, G., STRECKER, G., FOURNET, B. et MONTREUIL, J.
(1977 b) FEBS Lett. 77, 15-20. (p 5)

DORLAND, L., HAVERKAMP, J., VLIENGENTHART, J.F.G., FOURNET, B.,
STRECKER, G., SPIK, G. et MONTREUIL, J.
(1978 a) FEBS Lett. 89, 149-152. (p 5)

DORLAND, L., HAVERKAMP, J., VLIENGENTHART, J.F.G., STRECKER, G.,
MICHALSKI, J.C., FOURNET, B., SPIK, G. et MONTREUIL, J.
(1978 b) Eur. J. Biochem. 87, 323-329. (p 5)

EYLAR, H.
(1965) J. Theor. Biol. 10, 89-112. (p 3)

FIAT, A.M., JOLLES, J., AUBERT, J.P., LOUCHEUX-LEFEBVRE, M.H.
et JOLLES, P.
(1980) Eur. J. Biochem. 111, 333-339. (p 10)

FINKELSTEIN, A.V. et PTITSYN, O.B.

(1971) J. Mol. Biol. 62, 613. (p 8)

GILBERT, E.F., DAWSON, G., RHEIN, G.M., OPITZ, J.M. et SPANGER, J.W.

(1973) Z. Kinderheilk 114, 259-292. (p 3)

GOLD, M.H. et HAHN, H.J.

(1976) Biochemistry 15, 1808-1814. (p 3)

GUZZO, A.V.

(1965) Biophys. J. 5, 809. (p 7)

HANOVER, J.A., LENNARZ, W.J. et YOUNG, J.D.

(1980) J. Biol. Chem 255. 6713-6716. (p 42)

HARPAZ, N. et SCHACHTER, H.

(1980) J. Biol. Chem. 255, 4885-4902. (p 35)

HATCHER, V.B. et JEANLOZ, R.W.

(1973) Colloq. Int. Centre Nat. Rech. Sci. 221,
329-338. (p 18)

HOWARD, J.C., ALI, A. , SCHERAGA, H.A. et MOMANY, F.A.

(1975) Macromolecules 8, 607-622. (13, 39)

HUBER, R., KUKLA, D., RUHLMANN, A., EPP, O. et FORMANEK, H.

(1970) Naturwissenschaften 57, 389-392. (p 25)

IANNINO, N.I. de., STANELONI, R.J., BEHRENS, N.H. et DANKERT, M.A.

(1979) FEBS Lett. 99, 73-76. (p 3)

IUPAC-IUB Commission on Biochemical Nomenclature

(1970) Biochemistry 9, 3471. (p 50)

- KABBAT, E.A. et WU, T.T.
(1974) Proc. Nat. Acad. Sci. USA 71, 4217-4220. (p 8)
- KATZ, F.N., ROTHMAN, T.E., LINGAPPA, V.R., BLOBEL, G. et
LODISH, H.F.
(1978) Proc. Nat. Acad. Sci. USA 74, 3278-3282. (p 34)
- KAWAI, M. et FASMAN, G.D.
(1978) J. A. C. S 100, 3630-3632. (p 13)
- KAWASAKI, T. et ASHWELL, G.
(1976) J. Biol. Chem. 251, 5292-5299. (p 2)
- KORNFELD, R. et KORNFELD, S.
(1976) Ann. Rev. Biochem. 45, 217-238. (p 31, 33)
- LE GAILLARD, F., AUBERT, J.P., DAUTREVAUX, M. et LOUCHEUX-
LEFEBVRE, M.H.
(1976) FEBS Lett. 64, 278-284. (p 16)
- LENARD, J.
(1969) Chem. Rev. 69, 625-638. (p 17)
- LENNARZ, W.J.
(1975) Science 188, 986-991. (p 36)
- LENSTRA, J.A.
(1977) Biochim. Biophys. Acta 491, 333-338. (p 9)
- LEWIS, P.N. et BRADBURY, E.M.
(1974) Biochim. Biophys. Acta 336, 165-177. (p 10)
- LI, Z.Q., JACROT, F., LE GAILLARD, F. et LOUCHEUX-LEFEBVRE M.H.
(1980) FEBS Lett. 122, 203-206. (p 5, 25)
- LIM, V.I.
(1974) J. Mol. Biol. 88, 873-895. (p 8)

LINGAPPA, V.R., LINGAPPA, J.R., PRASAD, R., EBNER, K.E. et BLOBEL, G.

(1978) Proc. Nat. Acad. Sci. USA 75, 2338-2342. (p 34)

LODISH, H. et ROTHMAN, J.

(1979) Scientific American 239, 38-53. (p 3)

LOUCHEUX-LEFEBVRE, M.H. et AUBERT, J.P.

(1976) C.R. Acad. Sci. Paris 282, 585-587. (p 10)

LOUCHEUX-LEFEBVRE, M.H., AUBERT, J.P. et JOLLES, P.

(1978) Biophys. J. 23, 323-336. (p 10, 16, 32)

LOUCHEUX-LEFEBVRE, M.H.

(1980) Actualité Chimique 1, 12-27. (p 31)

MARCHESI, V.T., FURTHMAYR, H. et TOMITA, M.

(1976) Ann. Rev. Biochem. 45, 217-238. (p 3)

MARSHALL, R.D. et NEUBERGER, A.

(1970) Advan. Carb. Chem. Biochem. 25, 405-478. (p 5, 33)

MARTEL, P. KIM, S.M. et POWELL, B.M.

(1980) Biophys. J. 31, 371-380. (p 5)

MAZURIER, J. AUBERT, J.P. LOUCHEUX-LEFEBVRE, M.H. et SPIK, G.

(1976) FEBS Lett. 66, 238-242. (p 16)

MERCIER, J.C.

(1981) Biochimie 63, 1-17. (p 32)

MONTREUIL, J.

(1975) Pure and Applied Chem. 42, 431-477. (p 30, 33)

MORT, A.J. et LAMPORT, D.T.A.

(1977) Anal. Biochem. 82, 289-309. (p 17)

NAGANO, K.

(1977) J. Mol. Biol. 109, 251-274. (p 9)

NAGARAJAN, M. et RAO, V.S.R.

(1977) Current Science of India 46, 395-401. (p 30)

NEMETHY, G. et SHERAGA, H.A.

(1962) J. Phys. Chem. 66, 1773. (p 7)

PALAU, J. et PUIGDOMENECH, P.

(1974) J. Mol. Biol. 88, 457-469. (p 8)

PARODI, A.J. et LELOIR, L.F.

(1979) Biochim. Biophys. Acta 559, 1-37. (p 35)

PENA, S.D.J. et HUGUES, R.C.

(1978) Nature 276, 80-83. (p 2)

PLESS, D.D. et LENNARZ, W.J.

(1975) J. Biol. Chem. 250, 7014-7019. (p 28)

PLESS, D.D. et LENNARZ, W.J.

(1977) Proc. Nat. Acad. Sci. USA 74, 134-138. (p 36)

PORCHET, N., RICHET, C., DEMEYER, D., ROUSSEL, P. et
DEGAND, P.

(1981) Biochim. Biophys. Acta Soumis pour publication
(p 21, 23)

PROTHERO, J.W.

(1966) Biophys. J. 6, 367. (p 7)

PTITSYN, O.B.

(1969) J. Mol. Biol. 42, 501-510. (p 8)

PTITSYN, O.B.

(1974) J. Mol. Biol. 88, 287-300. (p 8)

- READING, C.L., PENHOET, E.E. et BALLOU, C.E.
(1978) J. Biol. Chem. 253, 5600-5612. (p 17)
- RINAUDO, M et DOMARD, A.
(1976) J.A.C.S. 98, 6360-6362. (p 13)
- ROBSON, B. et SUSUKI, E.
(1976) J. Mol. Biol. 107, 327-356. (p 9)
- RONIN, C. et BOUCHILLOUX, S.
(1976) Biochim. Biophys. Acta 428, 445-455. (p 34)
- RONIN, C. et BOUCHILLOUX, S.
(1978) Biochim. Biophys. Acta 539, 470-480. p (34)
- SAXENA, V.P. et WETLAUFER, D.B.
(1971) Proc. Nat. Acad. Sci. USA 68, 969-972. (p 13)
- SCHIFFER, M. et EDMUNSON, A.B.
(1967) Biophys. J. 7, 121-135. (p 7)
- SCHMID, K. et KAMIYAMA, S.
(1963) Biochemistry 2, 271-275. (p 18)
- SCHMID, K., KAUFMAN, H. ISEMURA, S., BAUER, F., EMURA, J.
MOTOYAMA, T. ISHIGURO, M. et NANNO, S.
(1973) Biochemistry 12, 2711-2724. (p 18, 19)
- SCHULZ, G.E., BARRY, C.D., FRIEDMAN, J., CHOU, P.Y., FASMAN,
G.D., FINKELSTEIN, A.V., LIM, V.I., PTITSYN, O.B., KABBAT, E.A.
WU, T.T., LEVITT, M., ROBSON, B. et NAGANO, K.
(1974) Nature 250, 140-142. (p 9)
- SCHWARZMANN, G., REINHOLD, V. et JEANLOZ, R.W.
(1973) Colloq. Int. Centre Nat. Rech. Sci. 221, 85-94.(p18)

- SIEMION, I.Z., LISOWSKI, M., KONOPINSKA, D. et NAWROCKA, E.
(1980) Eur. J. Biochem. 112, 339-343. (13, 40)
- SINGH, V.N., DAVE, S.S. et VENKITASUBRAMANIAN, T.
(1967) Biochem. J. 104, 48c-50c. (p 4)
- SMALL, D., CHOU, P.Y. et FASMAN, G.D.
(1977) Biochem. Biophys. Res. Comm. 79, 341-346 (p 32)
- SMITH, J.A. et PEASE, L.G.
(1980) Cri. Rev. Biochem. 8, 315-399. (p 13, 31)
- SOLIDAY, C.L. et KOLATTUKUDY, P.E.
(1979) Arch. Biochem. Biophys. 197, 367-378. (p 3)
- STERNBERG, M.J.E. et THORNTON, J.M.
(1978) Nature 271, 15. (p 7)
- STRECKER, G., FOURNET, B., MONTREUIL, J., DORLAND, L.,
HAVERKAMP, J., Vliegenthart, J.F.G. et DUBESSET, D.
(1978) Biochimie 60, 725-734. (p 5)
- STRECKER, G. et MONTREUIL, J.
(1979) Biochimie 61, 1199-1246. (p 31, 33)
- STRUCK, D.K., LENNARZ, W.J. et BREW, K.
(1978) J. Biol. Chem. 253, 5786-5794. (p 36)
- TABAS, I. et KORNFELD, S.
(1978) Fed. Proc. 37, 169. (p 34)
- TABAS, I. et KORNFELD, S.
(1979) J. Biol. Chem. 254, 11655-11663. (p 35)
- TAKAHASHI, N.
(1977) Biochem. Biophys. Res. Comm. 76, 1194-1201. (p 17)

- TETAS, M. CHAO, H. et MOLNAR, J.
(1970) Arch. Biochem. Biophys. 138, 135-146. (p 34)
- TIFFANY, M.L. et KRIMM, S.
(1968) Biopolymers 6, 1767-1770. (p 22)
- TONNEGUZZO, F. et GHOSH, H.P.
(1978) Proc. Nat. Acad. Sci. USA 75, 715-719. (p 34)
- TURKINGTON, R.W. et TOPPER, Y.J.
(1966) Biochim. Biophys. Acta 127, 366-372. (p ')
- ULGADE, R.A., STANELONI, R.J. et LELOIR, L.F.
(1978) FEBS Lett. 91, 209-212. (p 35)
- VAHERI, A. et MOSHER, D.F.
(1978) Biochim. Biophys. Acta 516, 1-25. (p 2)
- VENKATACHALAM, C.M.
(1968) Biopolymers 6, 1425. (p 12, 19)
- WAECHTER, C.J., LUCAS, J.J. et LENNARZ, W.J.
(1973) J. Biol. Chem. 248, 7570- 7579. (p 34)
- WAGNER, G., WUTHRICH, K. et TSCHESCHE, H.
(1978) Eur. J. Biochem. 86, 67-76. (p 25)
- WANG, F.F.C. et HIRS, C.H.W.
(1977) J. Biol. Chem. 252, 8358-8364. (p 24)
- WOODY, R.W.
(1974) Peptides, polypeptides and proteins. Proceeding of the Rehovot symposium on poly (amino acids), polypeptides and proteins and their biological implications.
Edited by E.R. BLOUT, F.A. BOVEY, M. GOODMAN and N. LOTAN.
338-350. (p 13, 39)

