

Structure covalente de la myoglobine de cheval

M. DAUTREVAUX, Y. BOULANGER, K. HAN et G. BISERTE

Laboratoire de Chimie Biologique de la Faculté de Médecine et de Pharmacie de Lille

(Reçu le 11 juin/19 juillet 1969)

The complete sequence (153 amino acids) of horse heart myoglobin has been established.

The sequence of a peptide isolated from chymotrypsin hydrolyzates of globin allowed to fill a gap in the tentative sequence previously proposed; this peptide was studied by means of pepsin hydrolysis, action of *N*-bromosuccinimide on histidyl bonds, and Edman degradation method associated with dansylation.

After hydrolysis of globin by chymotrypsin treated by TLCK, the same major peptides were found as after hydrolysis by non-treated chymotrypsin. These peptides were identified by ion-exchange resins and bidimensional paper chromatography, and also by the determination of their amino-acid composition.

The suppression of secondary cuts explains the presence of new peptides, whose the composition and N-terminal sequences were determined. It was then possible to confirm several sequences which could be previously deduced from the study of split products after cyanogen bromide treatment. The specificity of chymotrypsin towards certain types of peptide bonds is discussed.

Horse myoglobin, as compared with sperm whale myoglobin, contains 18 differences: 17 substitutions and one inversion. The major part of the substitutions are located in N- and C-terminal sequences; they are all punctiform, resulting from the replacement of only one base in each codon.

Dans des mémoires antérieurs [1,2], nous avons étudié les peptides libérés par hydrolyse chymotrypsique de la globine de myoglobine du cœur de cheval. Nous avons également publié [3—5] les séquences en amino-acides des trois segments libérés par l'action du bromure de cyanogène sur la globine. Ces résultats nous avaient déjà permis de proposer [6] une structure partielle de la myoglobine de cheval. La séquence publiée présentait un certain nombre de lacunes qu'il s'agissait de combler: certaines incertitudes ont déjà été levées par l'étude [2] des peptides libérés par action de la chymotrypsine (enchaînements 38—39, 52—53, 66—67, 102—113 et 125—129); il en est de même de l'enchaînement Lys-Lys-Lys- en position 77—78—79 en faveur duquel nous ne possédions que des arguments indirects [5]: l'étude de l'hydrolysats par la chymotrypsine du produit de guanidylation de la globine nous a récemment apporté des preuves directes [7].

Nous précisons dans ce mémoire l'enchaînement des résidus 18 à 28 par l'étude du peptide chymotrypsique S2E1S1P3. L'hydrolyse de la globine par

la chymotrypsine traitée par le TLCK nous permettra d'autre part de déterminer la nature du résidu en position 63. Les résultats confirment ceux que nous avons obtenus grâce aux peptides libérés par le bromure de cyanogène.

MATÉRIEL ET MÉTHODES

HYDROLYSE PAR LA CHYMOTRYPSINE TRAITÉE PAR LE TLCK

Traitement de la chymotrypsine par le TLCK

200 mg de chymotrypsine cristallisée [Worthington Biochemical Corporation (lot CD 537-40)] sont dissous dans 20 ml de tampon Tris 10 mM de pH 6,0, 5 mM en CaCl_2 et soumis à une ultrafiltration à 4° sur filtre Millipore 0,45 μ . Les pores laissent passer la chymotrypsine mais retiennent des impuretés insolubles représentant environ 10% du poids initial mis en jeu. Au filtrat on ajoute 20 mg de TLCK (Mann Research Laboratories) préalablement dissous dans 2 ml de méthanol et on laisse l'incubation se poursuivre pendant 2 1/2 heures à la température ambiante et sous agitation. La solution est ensuite à nouveau filtrée sur filtre Millipore 0,45 μ , afin d'éliminer l'excès d'insoluble de TLCK, et dialysée une nuit à 4° contre 2 litres d'acide chlorhydrique 0,01 N afin d'éliminer les ions métalliques bivalents, activateurs des carboxypeptidases A et B, qui

Abbreviations non usuelles. TLCK, tosyl-lysine-chloromethyl-ketone ou 1-chloro-3-tosylamido-7-amino-2-heptanone.

Enzymes. Trypsine (EC 3.4.4.4); chymotrypsine (EC 3.4.4.5); carboxypeptidase A (EC 3.4.2.1); leucine-amino-peptidase (EC 3.4.1.1).

peuvent éventuellement souiller les préparations de chymotrypsine. La solution d'enzyme est finalement lyophilisée et conservée à -15° .

Hydrolyse chymotrypsique

A une solution aqueuse de globine (1 g/100 ml) partiellement dénaturée [8], on ajoute du sesquicarbonate d'ammonium jusqu'à la concentration de 1 g/100 ml et le pH est ajusté à 9 par addition d'ammoniaque; on ajoute ensuite une quantité de chymotrypsine-TLCK correspondant à un rapport enzyme/substrat égal à 2% (en poids) et on laisse l'hydrolyse se poursuivre pendant 24 heures à 38° sous agitation. L'hydrolyse est arrêtée par simple congélation de la solution.

Techniques de fractionnement

Le fractionnement des mélanges peptidiques obtenus par hydrolyse enzymatique ou par coupure chimique a été réalisé par chromatographie sur une colonne de résine «Chromobeads P» (Technicon) de $130 \times 0,635$ cm (débit 36 ml/heure; température 50°) en utilisant un gradient de tampon volatil pyridine-acide formique ou acide acétique obtenu à l'aide du «Varigrad» décrit par Peterson et Sober [9]. L'effluent est collecté par fraction de 6 ml recueillies en 10 min et la courbe d'élution est établie automatiquement (Autoanalyseur Technicon) par réaction à la ninhydrine d'une aliquote de chaque fraction (0,32 ml). Après lyophilisation du contenu des tubes correspondant à un même pic de la courbe d'élution, l'homogénéité du peptide obtenu est vérifiée par électrophorèse et chromatographie sur papier. 100 mg du mélange peptidique simple peuvent être fractionnés en une seule fois par cette technique.

Si le mélange est très complexe (hydrolysats enzymatique de la globine), ce fractionnement sur colonne de résine à échange de cations est précédé d'un fractionnement de base par gel-filtration sur Sephadex et par électrophorèse en film liquide. La chromatographie de gel-filtration a été réalisée sur une colonne (105×3 cm) garnie de Sephadex G-25, éluée (36 ml/heure) au moyen d'un tampon pyridine 0,1 N-acide acétique de pH 4,9. Des fractions de 6 ml sont recueillies au moyen d'un collecteur automatique et la courbe d'élution est établie par dosage automatique des peptides de chaque fraction au moyen de la réaction de Lowry. Après lyophilisation des fractions correspondant aux différents groupes de peptides (acides aminés et peptides très aromatiques, S4; oligopeptides, S3; polypeptides, S2; gros polypeptides, S1), celles-ci sont soumises à une électrophorèse préparative en film liquide dans un appareil Elphor-Vap, selon la méthode de Hannig [10]; la courbe de distribution électrophorétique des peptides dans les 48 tubes de l'appareil est établie automatiquement par la réaction de Lowry.

MÉTHODES D'ÉTUDE DES PEPTIDES

Composition en amino-acides

L'étude de la composition en amino-acides des peptides (0,2 μ mole) a été effectuée, après hydrolyse totale par l'acide chlorhydrique 5,6 N en tube scellé sous vide à 105° pendant 24 heures, au moyen d'un analyseur automatique (Autoanalyseur Technicon). Quand cette détermination révèle la présence d'acides aminés dicarboxyliques, une hydrolyse totale de 24 heures par la leucine-aminopeptidase est également effectuée; le dosage des acides aminés libérés permet alors la mise en évidence de l'asparagine et de la glutamine: dans certains cas, ces amides sont identifiées par couplage électrochromatographique (électrophorèse à pH 2,4 et chromatographie dans le système butanol-acétique).

Détermination des acides aminés terminaux

Les acides aminés N-terminaux des peptides ont été déterminés par la méthode des dansyl-amino-acides. La dansylation a été faite selon le protocole décrit par Woods et Wang [11], que nous avons adapté de la manière suivante: environ 1 nmole de peptide¹ en solution dans 20 μ l de solution de bicarbonate de sodium 0,1 M est placée dans un petit tube en verre de 6 mm de diamètre sur 15 mm de longueur². Après addition de 20 μ l de solution de 1-diméthylaminonaphthalène-5-sulfochlorure ou chlorure de dansyl (Sigma) (1 mg/ml dans l'acétone), le petit tube est fermé par du papier paraffiné (Parafilm) et laissé à la température ambiante pendant 2 à 16 heures. Après la réaction, le contenu du tube est évaporé sous vide sulfurique (12–17 mm de mercure) et le résidu est repris par 50 μ l de HCl 5,6 N.

L'hydrolyse est faite de façon commode en enfermant le petit tube de verre dans un tube plus grand (20 cm de longueur sur 17 mm de diamètre): un étranglement situé au 1/3 inférieur de la hauteur ménage dans ce tube une chambre inférieure contenant 2 à 3 ml de HCl 5,6 N et une chambre supérieure dont l'ouverture est scellée à chaud après introduction du petit tube contenant le dansyl-peptide. L'ensemble est placé verticalement dans une étuve à 105° pendant 6 heures. Après hydrolyse, le contenu du petit tube est desséché sous vide en présence de soude en paillettes.

L'identification des dansyl-amino-acides est effectuée par chromatographie sur couche mince de gel de silice en présence de témoins latéraux (témoins externes) selon la technique décrite par Seiler et Wiechmann [12].

La séquence N-terminale des peptides a été déterminée par la méthode de dégradation récurrente

¹ Cette donnée nous est fournie par la détermination de la composition en amino-acides.

² Tubes Teswells, réactifs Warner-Chilcott: Précibio, 22, rue des Fossés St. Jacques, Paris.

d'Edman. La technique de cyclisation à l'aide de phénylisothiocyanate est pratiquée dans les conditions décrites par Bahl et Smith [13]. Dans un tube conique rodé de 3 ml, on ajoute 500 μ l de solution de phénylisothiocyanate à 2 vol.-% dans la pyridine (pyridine Merck); après addition de 100 μ l de triméthylamine aqueuse (25 vol.-%), on laisse réagir à 40° pendant 2 heures sous atmosphère d'azote. Après une nouvelle addition de 2 μ l de phénylisothiocyanate pur, on laisse incubé encore pendant 2 heures à 40°. Le phénylisothiocyanate en excès est extrait par le benzène (5 fois 1 ml de benzène Merck, séparation par centrifugation). Après évaporation de la phase aqueuse sous vide, en présence de soude, la cyclisation est effectuée pendant 2 heures à 25° à l'aide d'acide trifluoroacétique anhydre (1 ml). On dessèche sous vide en présence de soude et le résidu est repris par 500 μ l de HCl N/100; la phénylthiohydantoïne est alors extraite par 1 ml d'éther exempt de peroxydes et préalablement saturé en eau.

La phase aqueuse est desséchée sous vide et représente le matériel de départ utilisable pour l'étape suivante; à chaque étape, une petite portion du peptide résiduel, correspondant à environ 1 à 2 nmole, est prélevée: l'acide aminé N-terminal est déterminé par dansylation.

Les acides aminés C-terminaux ont été identifiés par hydrazinolyse. Nous avons utilisé la technique de Niu et Fraenkel-Conrat [14]: le peptide (1 μ mole) est desséché pendant 48 heures à 20° dans un dessiccateur sous pression réduite et en présence de P_2O_5 . Il est ensuite repris dans 1 ml d'hydrazine à 98 vol.-% et maintenu en tube scellé à 70° pendant 16 heures. Après réaction, l'hydrazine est éliminée par évaporation dans un dessiccateur sous vide sulfurique; le résidu sec est repris par 2 ml d'eau et les hydrazides d'acides aminés sont éliminés par trois extractions successives au moyen de 2 ml de benzaldéhyde (agitation pendant 15 min, puis centrifugation); la phase aqueuse résiduelle est finalement extraite à l'éther de façon à éliminer le benzaldéhyde. L'acide aminé C-terminal est identifié par couplage électrochromatographique.

La cinétique d'action de la carboxypeptidase A [DFP-carboxypeptidase A (Sigma; lot 124 B 1850): suspension aqueuse à la concentration de 50 mg/ml] a été utilisée pour l'étude des séquences C-terminales: nous avons employé des enzymes préalablement traités par le di-isopropylfluorophosphate. Les suspensions commerciales de DFP-carboxypeptidase A sont, avant usage, diluées au 1/10e au moyen d'une solution de chlorure de lithium à 100 g/l. Après un séjour d'une nuit à 4°, la quasi-totalité de l'enzyme est dissoute. Le peptide (1 à 2 mg) est dissous dans 1 ml de tampon acétate d'ammonium 0,1 M de pH 7,8; on ajoute 8 μ l solution de DFP-carboxypeptidase A; l'hydrolyse est continuée à 38° et des prélèvements sont effectués à des temps variables,

déterminés en fonction de la séquence C-terminale probable: les acides aminés libérés sont dosés à l'Autoanalyseur Technicon et les calculs sont facilités par l'addition au peptide étudié d'une quantité déterminée de norleucine (10 nmole par prélèvement).

Nous avons en outre utilisé l'hydrolyse pepsique et la coupure chimique des liaisons histidyl pour simplifier les peptides isolés. L'hydrolyse pepsique a été conduite selon la technique préconisée par Chauvet et Acher [15]: le peptide (5 μ mole) est dissous dans 1 ml d'acide formique à 5 vol.-% (pH 1,8); on ajoute 40 μ l d'une solution de pepsine commerciale (Pepsine Light) à 50 mg/ml et on laisse l'incubation se poursuivre 24 heures à 37°; les hydrolysats sont enfin lyophilisés.

L'action de la *N*-bromosuccinimide a été conduite selon le protocole expérimental de Shaltiel et Patchornik [16], décrit et adapté par Han *et al.* [17]: 2 à 3 μ mole de peptides sont dissous dans 10 ml de tampon pyridine-acide acétique-eau de pH 3,3 (1:10:19, v/v/v). Une solution de *N*-bromosuccinimide dans le même tampon (35 mg/5 ml) est ajoutée à la solution précédente en plusieurs portions sous agitation magnétique continue. Après quinze min à la température ambiante, on détruit l'excès de *N*-bromosuccinimide par addition de cristaux d'imidazole jusqu'à réaction négative d'un papier réactif IK-amidon. L'hydrolysate est alors chauffé au bain-marie à 85–90° pendant une heure pour rompre les liaisons histidyl modifiées pendant la première étape. Après lyophilisation, les peptides sont séparés par électrophorèse sur papier à haute tention (appareil d'électrophorèse à haute tension modèle D de Gilson Medical Electronics) dans le mélange acide acétique—acide formique—eau (15:5:80, v/v/v) de pH 1,94; 100 volts/cm; 30 min. Après élution, les acides aminés N-terminaux des peptides sont déterminés soit par dansylation, soit par dinitrophénylation, et les acides aminés constitutifs sont identifiés par électrochromatographie sur papier.

RÉSULTATS

La désignation des peptides et des fractions peptidiques a été établie en fonction de leur mode d'obtention et de leur situation dans le plan de fractionnement. Les lettres majuscules correspondent, dans l'ordre de leur énoncé, à la méthode de fractionnement utilisé: gel-filtration sur Séphadex (S), électrophorèse en film liquide (E) chromatographie sur colonne de résine à échange d'ion (P). Les chiffres qui suivent chacune de ces lettres indiquent, pour chaque fractionnement, la position du peptide ou de la fraction peptidique dans la courbe d'élution ou la courbe de distribution obtenue; les lettres minuscules identifient les différents composants d'une fraction hétérogène, quand ceux-ci peuvent être isolés par chromatographie ou électrophorèse sur papier.

Fig.1. Séquence en amino-acides du peptide S2E1SIP3 et de ses produits d'hydrolyse par la pepsine (peptides «Pe») ou par la N-bromosuccinimide (peptides «Su»)

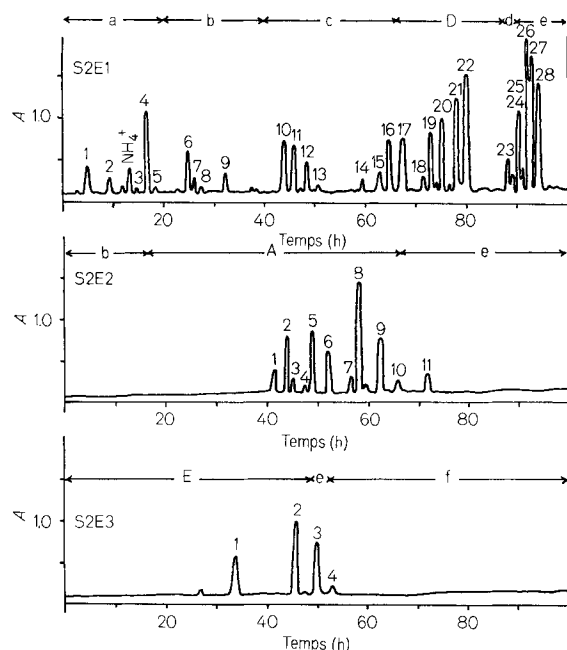


Fig. 2. Courbes d'élution sur colonne de résine Chromobeads P des hydrolysats de globine par la chymotrypsine-TLCK (fractions S2E1, S2E2, S2E3). Tampons: a, tampon pyridine 0,1 M—acide formique pH 2,90; b, tampon pyridine 0,1 M—acide formique pH 4,25; c, tampon pyridine 0,2 M—acide formique pH 4,25; d, tampon pyridine 1,0 M—acide formique pH 5,00; e, tampon pyridine 2,0 M—acide acétique pH 5,00; f, tampon pyridine 2,0 M—acide acétique pH 6,5. Gradient A: Varigrad à 9 cellules: cellule 1 à 4 garnies chacune de 200 ml de tampon b; cellules 5, 6, 7 garnies chacune de 200 ml de tampon d; cellules 8 et 9 garnies chacune de 200 ml de tampon e. Gradient D: Varigrad à 4 cellules: cellule 1, 200 ml de tampon c; cellule 2, 60 ml de tampon d et 140 ml d'eau distillée; cellule 3, 80 ml de tampon d et 120 ml d'eau distillée; cellule 4, 100 ml de tampon d et 100 ml d'eau distillée. Gradient E: Varigrad à 9 cellules: cellules 1 et 2 garnies chacune de 200 ml de tampon b; cellules 3, 4, 5 garnies chacune de 100 ml de tampon d et 100 ml d'eau distillée; cellules 6, 7, 8 garnies chacune de 200 ml de tampon d; cellule 9 garnie de 200 ml de tampon e

Le peptide Pe 3 correspond à une séquence connue: Gly-Lys-Val-Glu-Ala ou S2E2S2P3; le peptide Pe 2 possède le même acide aminé N-terminal, mais comporte un résidu d'acide aspartique supplémentaire (voir Fig. 1). A 37° pendant 24 heures, la carboxypeptidase A détache uniquement cet acide aminé (identifié par chromatographie sur papier) et libère un peptide résiduel ayant un comportement identique au peptide Pe 3: les peptides Pe 3 et Pe 2 correspondent donc à la séquence N-terminale du peptide S2E1S1P3.

Le peptide Pe 4 possède un acide aminé N-terminal difficilement identifiable: isoleucine ou valine; mais la localisation de l'unique résidu de valine dans la séquence C-terminale du peptide S2E1S1P3 lève l'ambiguïté et permet de conclure à la présence de

l'isoleucine. Par dégradation d'Edman, la séquence N-terminale Ile-Ala-Gly-His... est mise en évidence; le peptide Pe 1 possède un résidu supplémentaire d'acide aspartique en position N-terminale.

La séquence partielle représentée dans la Fig. 1 peut donc être attribuée à S2E1S1P3.

Action de la N-bromosuccinimide

Par chromatographie sur colonne de résine Chromobeads P, 6 peptides sont isolés du milieu de réaction. Leur composition et leur acide aminé N-terminal (dansyl) sont indiqués dans le Tableau 2.

Tableau 2. Composition et acide aminé N-terminal des peptides libérés par action de la N-bromosuccinimide sur le peptide S2E1S1P3

L'histidine est présente dans ces peptides sous une forme modifiée par l'action de la N-bromosuccinimide; elle échappe à notre analyse

	S2E1S1P3	Su 1	Su 2	Su 3	Su 4	Su 5	Su 6
Asp	1,03	—	0,99	1,04	0,96	0,97	—
Glu	3,03	1,99	1,10	1,12	1,06	1,02	0,88
Gly	2,88	1,02	2,08	2,11	1,03	2,00	1,27
Ala	2,10	—	1,98	1,95	1,04	2,02	1,18
Val	2,04	1,01	1,08	0,96	1,00	1,03	0,97
Ile	0,99	—	0,88	0,84	—	0,99	—
Leu	1,00	0,98	—	—	—	—	—
Lys	1,03	—	0,89	0,96	0,90	0,98	0,70
His	0,86	—	—	—	—	—	—
DNS	Gly	Gly	Gly	Gly	Gly	Gly	Gly

Seul, le peptide Su 1 a pu être récupéré en quantité suffisante pour permettre son étude: par dégradation récurrente d'Edman, couplée avec la méthode des dansyl-amino-acides, la séquence Gly-Gln... a été précisée; à la 2ème étape la phénylthiohydantoïne libérée est identifiée à celle de la glutamine. L'action de la carboxypeptidase A permet de libérer au bout de 8 heures de la valine et de la leucine identifiables en électrochromatographie sur papier. L'hydrolyse totale (24 heures) par la leucine-aminopeptidase met en évidence un résidu de glutamine: la séquence du peptide Su 1 est donc Gly-Gln-Glu-Val-Leu.

La composition en amino-acides des autres peptides confirme les résultats obtenus par l'hydrolyse pepsique (Fig. 1).

ÉTUDE DE L'HYDROLYSAT CHYMOTRYPSIQUE-TLCK DE GLOBINE

Une quantité relativement faible de globine (500 mg) a été hydrolysée; de ce fait, le protocole d'isolement des peptides a été simplifié par rapport à nos études précédentes [18]. Par gel-filtration sur Sephadex, on récupère environ 250 mg de fraction S2

(polypeptides). Par électrophorèse préparative en film liquide à pH 3,9, trois fractions S2E1, S2E2, S2E3, de caractère de plus en plus basique, sont isolées. Chacune des fractions S2E1, S2E2 et S2E3 a été enfin fractionnée sur colonne de résine Chromobeads P (Fig. 2).

Les peptides ont été dans la plupart des cas identifiés à des peptides connus grâce à leur comportement dans les divers systèmes chromatographiques et électrophorétiques utilisés pour leur isolement ou le contrôle de leur homogénéité; la vérification qualitative par électrophorèse et chromatographie sur papier de leur composition en amino-acides a confirmé l'identification. Dans quelques cas, les peptides isolés n'étaient pas connus: ils ont alors été étudiés par les méthodes habituelles.

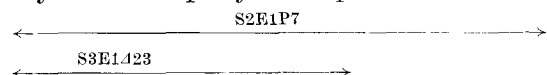
Tableau 3. Identification des peptides de la fraction S2E1 (première partie)

Peptide	Structure	Peptides de référence
S2E1P1	Gly-Ala-Asp-Ala-Gln-Gly-Ala-Met	S2E1S2P1 [2]
S2E1P2	Gly-Thr-Val-Val-Leu	S2E1S2P2 [2]
S2E1P3	Peptide hétérogène	—
S2E1P4	Thr-Ala-Leu	S3E1A14 [1]
S2E1P5	Thr-Ala-Leu-Gly-Gly-Ile-Leu	S2E1S2P4 [2]
S2E1P6	Gly-Gly-Ile-Leu	S3E1A16 [1]
S2E1P7	Gly-Leu-Ser-Asp-Gly-Glu-Trp-Gln-Gln-Val-Leu	voir texte
S2E1P8	Ala-Leu-Glu-Leu-Phe	S2E1S2P7 [2]
S2E1P9	Ile-Ser-Asp-Ala-Ile-Ile-His	S2E1S2P8 [2]
S2E1P10	Lys-Thr-Glu-Ala-Glu-Met	S2E1S2P9 [2]
S2E1P11	Lys-Ala-Ser-Glu-Asp-Leu	S2E1S2P10 [2]
S2E1P12	Thr-Gly-His-Pro-Glu-Thr-Leu	S2E1S2P11 [2]
S2E1P13	Ile-Ser-Asp-Ala-Ile-Ile-His-Val-Leu	voir texte
S2E1P14	Peptide hétérogène	—
S2E1P15	Lys-Glu-Leu-Gly-Phe-Glu-Gly	S2E1S2P13 [2]

La structure des peptides libérés par hydrolyse chymotrypsique-TLCK est indiquée dans les Tableaux 3, 4 et 5; les peptides S2E1P7, S2E1P13, S2E1P27, S2E2P5, S2E2P9, S2E2P10 et S2E2P11 ont été étudiés plus particulièrement.

Peptide S2E1P7

Gly-Leu-Ser-Asp-Gly-Glu-Trp-Gln-Gln-Val-Leu



Composition. Après hydrolyse totale acide: Asp (1,05), Ser (0,99), Glu (2,87), Gly (2,02), Val (1,04), Leu (2,01); après hydrolyse totale par la leucine-aminopeptidase: Asp (0,98), Ser (0,86), Gln (1,86), Glu (1,02), Gly (1,89), Val (0,96), Leu (2,14), Trp (0,90).

Tableau 4. Identification des peptides de la fraction S2E1 (fin)

Peptide	Structure	Peptides de référence
S2E1P16	Gly-Lys-Val-Glu-Ala-Asp-Ile-Ala-Gly-His-Gly-Gln-Glu-Val-Leu	S2E1S1P3 [2]
S2E1P17	Lys-Glu-Leu-Gly-Phe-Gln-Gly	S2E1S2P14 [2]
S2E1P18	Thr-Lys-Ala-Leu	S2E2S2P7 [2]
S2E1P19	Lys-Glu-Leu-Gly-Phe	S3E1A26 [1]
S2E1P20	Peptide hétérogène	—
S2E1P21a	Thr-Lys-Ala-Leu-Glu-Leu-Phe	S2E1S2P17 [2]
S2E1P21b	Arg-Asp-Asp-Ile-Ala-Ala-Lys-Tyr	S2E1S2P16 [2]
S2E1P22	Lys-Ala-Leu-Glu-Leu-Phe	S2E1S2P18 [2]
S2E1P23a	Thr-Gly-His-Pro-Glu-Thr-Leu-Glu-Lys-Phe	S2E1S1P2 [2]
S2E1P23b	Arg-Asn-Asp-Ile-Ala-Ala-Lys-Tyr	S2E1S2P19b [2]
S2E1P24	Peptide hétérogène	—
S2E1P25	His-Ser-Lys-His-Pro-Gly-Asn-Phe-Gly-Ala-Asp-Ala-Gln-Gly-Ala-Met	S2E1S1P6 [2]
S2E1P26	Arg-Asn-Asp-Ile-Ala-Ala-Lys-Tyr	S2E1S2P20 [2]
S2E1P27	Lys-Ala-Ser-Glu-Asp-Leu-Lys-Lys-His-Gly-Thr-Val-Val-Leu	voir texte
S2E1P28	Lys-His-Leu-Lys-Thr-Glu-Ala-Glu-Met	S2E3S1P5 [2]

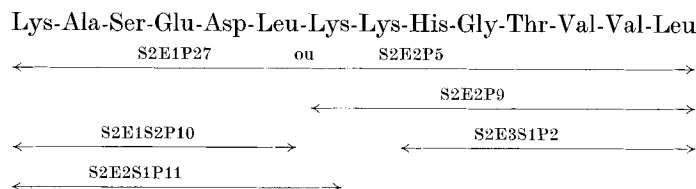
Tableau 5. Identification des peptides des fractions S2E2 et S2E3

Peptide	Structure	Peptides de référence
S2E2P1	Lys-Pro-Leu	S3E2A8 [1]
S2E2P2	Ala-Gln-Ser-His-Ala-Thr-Lys	S2E3S2P6 [2]
S2E2P3	Peptide hétérogène	—
S2E2P4	Lys-His-Leu-Lys-Thr-Glu-Ala-Glu-Met	S2E3S1P5 [2]
S2E2P5	Lys-Ala-Ser-Glu-Asp-Leu-Lys-Lys-His-Gly-Thr-Val-Leu-Leu	voir texte
S2E2P6	Peptide hétérogène	—
S2E2P7	His-Ser-Lys-His-Pro-Gly-Asn-Phe	S2E3S2P13 [2]
S2E2P8	Lys-Ile-Pro-Ile-Lys-Tyr	S2E3S2P14b [2]
S2E2P9	Lys-Lys-His-Gly-Thr-Val-Val-Leu	voir texte
S2E2P10	His-Lys-Ile-Pro-Ile-Lys-Tyr	voir texte
S2E2P11	Lys-Lys-His-Gly-Thr-Val-Val-Leu	voir texte
S2E3P1	Ala-Thr-Lys-His	S2E3S2P7 [2]
S2E3P2	Ala-Gln-Ser-His-Ala-Thr-Lys-His	S2E3S2P10 [2]
S2E3P3	Lys-His-Leu	S3E3A21 [1]
S2E3P4	Ser-His-Ala-Thr-Lys-His	S2E2S2P15 [2]

Acides aminés terminaux. Dansylation: DNS-Gly. Hydrazinolyse: Leu. Action de la carboxypeptidase A: après 15 min, libération de Leu et Val; après 30 min, libération de Leu, Val, Gln.

Séquence. La séquence se déduit de ces résultats, ainsi que de la structure du peptide S3E1A23 [1].

Peptides S2E1P27 et S2E2P9

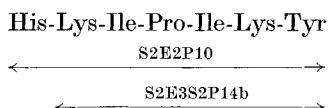


S2E2P9: *Lys-Lys-His-Gly-Thr-Val-Val-Leu*. Composition en amino-acides: Thr (0,92), Gly (1,19), Val (1,96), Leu (1,07), Lys (1,96), His (0,99). Acide aminé N-terminal (dansyl): lysine. Action de la carboxypeptidase A (24 heures à 37°): Leu (0,89), Val (1,78) (en nombre de résidus par mole de peptide). La séquence de S2E2P9 se déduit également de celle du peptide S2E3S1P2 (His-Gly-Thr-Val-Val-Leu) [2].

Il est à noter que le peptide S2E2P11 possède une composition en acides aminés identique à celle de S2E2P9: Thr (0,93), Gly (1,17), Val (1,82), Leu (1,11), Lys (1,96) et His (1,01).

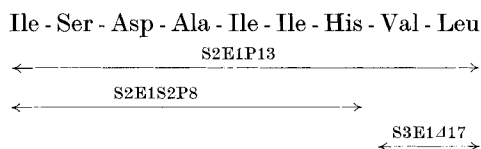
S2E1P27: *Lys-Ala-Ser-Glu-Asp-Leu-Lys-Lys-His-Gly-Thr-Val-Val-Leu*. Composition en amino-acides: Asp (1,07), Thr (1,00), Ser (1,19), Glu (1,12), Gly (1,15), Ala (1,12), Val (1,74), Leu (1,92), Lys (2,74), His (0,93). Amino-acide N-terminal (dansyl): lysine. La séquence de S2E1P27 se déduit de celle des peptides Lys-Ala-Ser-Glu-Asp-Leu-Lys (S2E2S1P11) et du peptide précédent (S2E2P9). Il est à noter qu'elle est en contradiction avec celle du peptide Lys-Ala-Ser-Glu-Asp-Leu-Lys-Leu-His (S2E2S2P4) qui a été trouvé dans l'un de nos hydrolysats chymotrypsique (2 et 6): comme nous l'avons indiqué précédemment, le peptide S2E2S2P4 pourrait correspondre à une variante très mineure de la séquence majeure du peptide S2E1P27.

Peptide S2E2P10



Composition en amino-acides: Pro (1,04), Ile (2,10), Tyr (0,85), Lys (2,22), His (0,79). Acide aminé N-terminal (dansyl): histidine. La séquence de ce peptide se déduit de celle de S2E3S2P14b [2].

Peptide S2E1P13



Composition en amino-acides: Asp (1,02), Ser (0,84), Ala (1,07), Val (0,98), Ile (2,06), Leu (1,00), His (0,36). Nous savons, grâce au peptide S2E1S2P8, que ce peptide contient une séquence Ile-Ile-His difficilement hydrolysable: cela explique le déficit en isoleucine et en histidine.

Amino-acide N-terminal (dansyl): isoleucine. La séquence du peptide se déduit de celle des peptides précédemment décrits: S2E1S2P8 [2] et S3E1A17 [1].

DISCUSSION

L'étude du peptide S2E1S1P3 conduit à la structure complète du segment N-terminal de la myoglobine de cheval, qui présentait une lacune entre les résidus 18 et 28; il est à noter que ce peptide comporte à lui seul 5 différences par rapport au segment correspondant de la myoglobine de cachalot.

L'hydrolysats de la globine par la chymotrypsine dont l'activité trypsique résiduelle est inhibée par le TLCK donne des résultats intéressants malgré l'étude relativement sommaire que nous en avons faite. Les peptides isolés de cet hydrolysats recouvrent la quasi-totalité de la chaîne peptidique (Fig. 3): seules n'ont pas été retrouvées de courtes séquences représentées par des oligopeptides normalement isolés dans les fractions S4 (Asn-Val-Trp) ou S3 (Ile-Arg-Leu-Phe, Asp-Lys-Phe ou Leu-Glu-Phe). La séquence Lys-Lys-Lys-Gly-His-His-Glu-Ala-Glu-Leu que nous avons recherchée dans cet hydrolysats n'a pas non plus été retrouvée, malgré sa présence obligatoire dans la fraction S2; il est probable que, du fait de son caractère basique très prononcé, son élution des colonnes de résine à échange de cations n'a pas eu lieu dans les conditions mises en oeuvre. Il faut noter que le peptide correspondant isolé des hydrolysats chymotrypsiques de globine guanidylée n'a pas non plus été retrouvé en chromatographie sur colonne de résine et a dû être isolé par électrophorèse à haut voltage sur papier [7].

Le fractionnement de l'hydrolysats chymotrypsique-TLCK est beaucoup plus simple que celui que nous avons utilisé initialement: en raison de la limitation des coupures secondaires, le nombre des peptides est en effet réduit; l'étude en est cependant plus délicate en raison de leur taille et du taux plus faible de recouvrement des séquences.

L'hydrolysats que nous avons étudié conduit à des séquences qui se superposent très exactement à

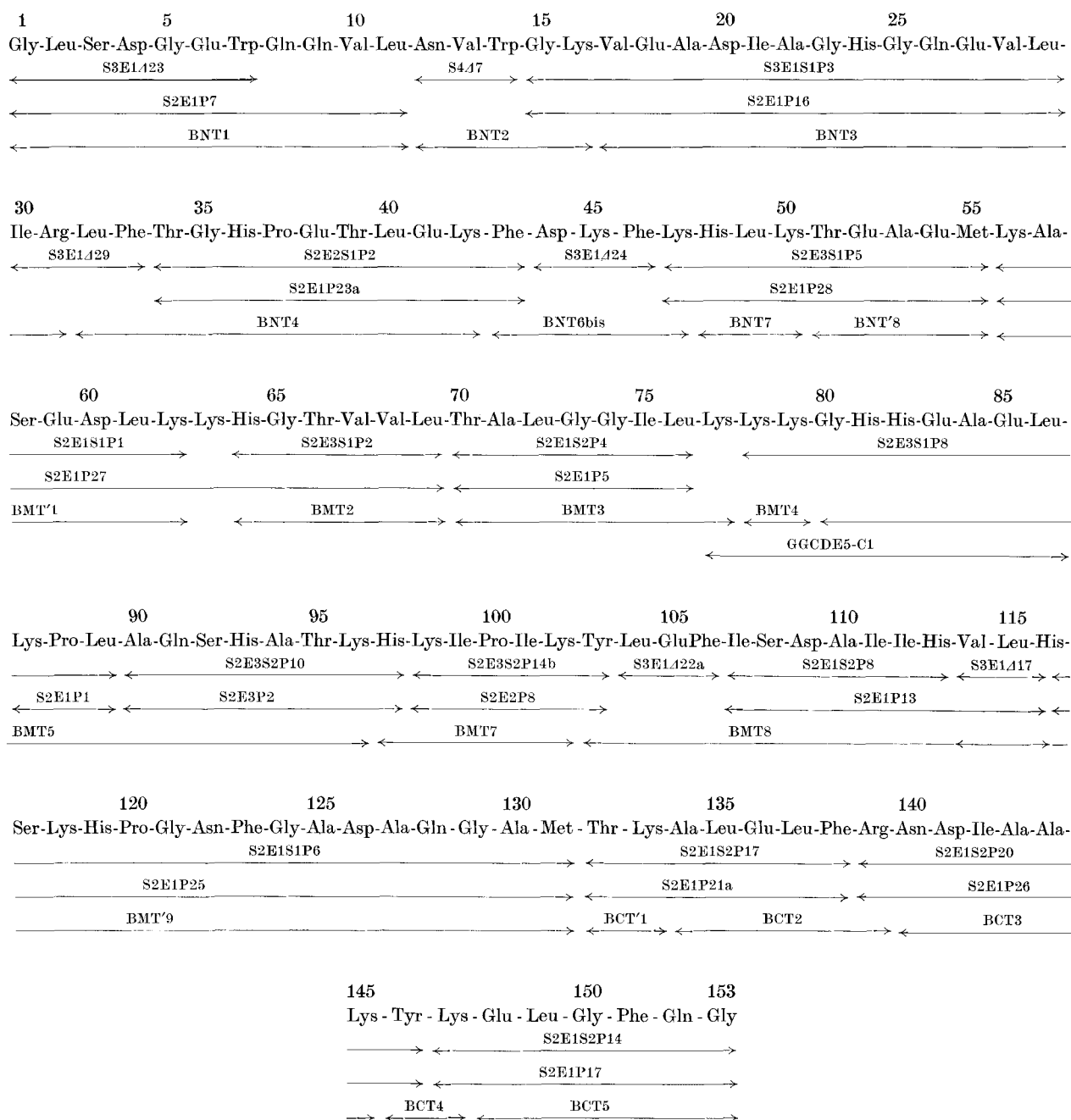


Fig. 3. Structure de la myoglobine de cheval en fonction des résultats des différents types d'hydrolyse. BNT = peptide tryptique du fragment N-terminal obtenu par coupure au bromure de cyanogène; BMT = *i. d.* fragment médian; BCT = *i. d.* fragment C-terminal; A = chromatographie sur Dowex 50 X 2; GGCD = peptide isolé de l'hydrolysate chymotrypsique de myoglobine guanidylée

celles que nous avons déduites de l'hydrolysate chymotrypsique ordinaire par le jeu des recouvrements de séquence (Fig. 3); il ne laisse apparaître que les points de rupture primaires de la protéine sous l'action de l'enzyme. L'inhibition de l'activité tryptique

résiduelle par le TLCK apparaît très efficace, car on ne retrouve pratiquement pas de peptides correspondant à une spécificité de type tryptique: seul le peptide Ala-Gln-Ser-His-Ala-Thr-Lys (S2E2P2) est le témoin de cette activité; la coupure

Lys-His (86—97), qui a permis la formation de ce peptide, est cependant mineure (23%) par rapport à celle de la liaison His-Lys (97—98) qui lui fait suite: 1,08 μ mole du peptide S2E2P2 a été en effet isolée, contre 2,36 μ moles de S2E3P2, 0,56 μ moles de S2E3P4 et 0,66 μ moles de S2E3P1, ces trois peptides (voir Tableau 5) représentant des fragments libérés par hydrolyse de cette liaison His-Lys (3,6 μ moles en tout).

Le peptide Ala-Leu-Glu-Leu-Phe (S2E1P8) est également libéré par hydrolyse d'une liaison lysyl (Lys-Ala, 134—138), mais il est aussi en faible quantité (Fig. 2) par rapport aux autres peptides de cette séquence (S2E1P18, S2E1P22, S2E1P21a).

Il est évidemment impossible d'établir un bilan complet des coupures principales effectuées par la chymotrypsine-TLCK car nous n'avons pas étudié les fractions S4 et S3; mais celles-ci étant beaucoup plus simples que S2, nos observations gardent toute leur valeur.

On constate ainsi que les deux liaisons méthionyl et les deux liaisons tyrosyl sont totalement hydrolysées par l'enzyme; au contraire, si l'une des liaisons tryptophyl est totalement coupée, l'autre (liaison 7—8) ne l'est que partiellement (environ 70%): on retrouve en effet dans S2 1,2 μ moles du peptide S2E1P7.

De même, 5 sur 7 des liaisons phénylalaninyl sont totalement hydrolysées; la liaison Phe-Gln (151—152) est hydrolysée à 50% environ, comme semble le montrer dans la Fig. 2 la comparaison des pics S2E1P17 (Lys-Glu-Leu-Gly-Phe-Gln-Gly) et S2E1P19 (Lys-Glu-Leu-Gly-Phe). La liaison Phe-Gly (123 à 124) est également peu hydrolysée (moins de 30%), car le pic S2E1P25 (116 à 131) est plus important que les pics S2E1P1 (124 à 131) et S2E2P7 (116 à 123).

L'examen de l'hydrolyse des liaisons leucyl se révèle très instructif: sur les 17 présentes dans la globine, 7 sont totalement hydrolysées (voir Tableau 6), 5 le sont partiellement (5 à 20%) et 5 ne semblent pas coupées (de petits fragments peuvent cependant se trouver dans les fractions S3 et S4). Parmi les liaisons leucyl totalement hydrolysées, nous trouvons les 5 liaisons impliquant des enchaînements à la fois fortement hydrophobes (Val-Leu-X ou Ile-Leu-X) et d'autre part non attaqués par l'enzyme (liaisons valyl, isoleucyl): nous avons déjà montré que les liaisons de ce type étaient également scindées par la trypsine traitée par le TPCK [4,5]. Les liaisons leucyl non hydrolysées par la chymotrypsine sont situées soit à proximité des acides aminés terminaux (résidus 2 et 149), soit au voisinage immédiat de liaisons particulièrement sensibles à l'action de la chymotrypsine (leucines 32 et 137, impliquées dans des enchaînements Leu-Phe, ou leucine 105, dans la séquence Tyr-Leu-Glu-Phe).

Tableau 6. *Hydrolyse chymotrypsique des liaisons leucyl de la myoglobine de cheval*

Numéro du résidu	Environnement de la liaison leucyl	Taux d'hydrolyse
2	Gly-Leu-Ser-Asp-	0
11	-Val-Leu-Asn-Val-	100
29	-Val-Leu-Ile-Arg-	100
32	-Arg-Leu-Phe-Thr-	0
40	-Thr-Leu-Glu-Lys-	< 20
49	-His-Leu-Lys-Thr-	< 20
61	-Asp-Leu-Lys-Lys-	< 20
69	-Val-Leu-Thr-Ala-	100
72	-Ala-Leu-Gly-Gly-	< 5
76	-Ile-Leu-Lys-Lys-	100
86	-Glu-Leu-Lys-Pro-	100
89	-Pro-Leu-Ala-Gln-	100
104	-Tyr-Leu-Glu-Phe-	0
115	-Val-Leu-His-Ser-	100
135	-Ala-Leu-Glu-Leu-	< 10
137	-Glu-Leu-Phe-Arg-	0
149	-Glu-Leu-Gly-Phe-	0

Les liaisons histidyl sont relativement peu hydrolysées: seule la liaison His-Lys (97—98) est rompue avec un rendement de l'ordre de 90%.

En dehors des précisions nouvelles, la structure covalente de la myoglobine de cheval que nous présentons dans la Fig. 3 comporte au niveau du résidu 63 une différence minime par rapport à la structure partielle publiée précédemment [6]: la séquence ... Leu-Lys-Leu-His ... n'a en effet été retrouvée que dans un seul hydrolysats et elle doit correspondre à une variante de la forme plus générale représentée par la séquence ... Leu-Lys-Lys-His ... On sait en effet que la myoglobine de cheval, — comme d'ailleurs toutes les myoglobines, — est hétérogène en électrophorèse en gel d'amidon [19]. La cause de ce polymorphisme a été discutée ailleurs [20] et, parmi les raisons possibles, on peut envisager les variations dans le taux d'amidification, bien que les dosages effectués par Åkeson et Theorell [21] donnent les mêmes résultats pour les différentes formes de la myoglobine de cheval; mais si l'on considère les différents peptides isolés des hydrolysats de globine par la chymotrypsine, traitée ou non par le TLCK, on constate que deux amides seulement peuvent se retrouver sous forme d'acide dans des peptides: il s'agit du résidu d'asparagine 140 (peptide chymotrypsique-TLCK S2E1P26 et sa forme désamidée S2E1P21b) et du résidu de glutamine 152 (peptide S2E1P17 et sa forme désamidée S2E1P15). Des calculs de rendement effectués à partir des hydrolysats chymotrypsiques normaux, il résulte que la forme désamidée du résidu 140 représente environ 75% contre 25% pour la forme asparagine; la forme désamidée du résidu 152 ne représente au contraire que 40% contre 60% pour la forme glutamine. L'existence de ces formes désamidées peut être simplement due à une hydrolyse partielle, au

	1	5	10	15	20	25
Cheval	Gly-Leu-Ser-Asp-Gly-Glu-Trp-Gln-Gln-Val-Leu-Asn-Val-Trp-Gly-Lys-Val-Glu-Ala-Asp-Ile-Ala-Gly-His-Gly-Gln-					
Cachalot	Val ———— Glu ———— Leu ———— His ———— Ala ———— Val ————					
	30	35	40	45	50	
Cheval	Glu-Val-Leu-Ile-Arg-Leu-Phe-Thr-Gly-His-Pro-Glu-Thr-Leu-Glu-Lys-Phe-Asp-Lys-Phe-Lys-His-Leu-Lys-Thr-					
Cachalot	Asp-Ile ———— Lys-Ser ———— Arg ————					
	55	60	65	70	75	
Cheval	Glu-Ala-Glu-Met-Lys-Ala-Ser-Glu-Asp-Leu-Lys-Lys-His-Gly-Thr-Val-Val-Leu-Thr-Ala-Leu-Gly-Gly-Ile-Leu-					
Cachalot	————— Val-Thr ———— Ala ————					
	80	85	90	95	100	
Cheval	Lys-Lys-Lys-Gly-His-His-Glu-Ala-Glu-Leu-Lys-Pro-Leu-Ala-Gln-Ser-His-Ala-Thr-Lys-His-Lys-Ile-Pro-Ile-Lys-					
Cachalot	—————					
	105	110	115	120	125	
Cheval	Tyr-Leu-Glu-Phe-Ile-Ser-Asp-Ala-Ile-Ile-His-Val-Leu-His-Ser-Lys-His-Pro-Gly-Asn-Phe-Gly-Ala-Asp-Ala-Gln-					
Cachalot	————— Glu ———— Arg ————					
	130	135	140	145	150	
Cheval	Gly-Ala-Met-Thr-Lys-Ala-Leu-Glu-Leu-Phe-Arg-Asn-Asp-Ile-Ala-Ala-Lys-Tyr-Lys-Glu-Leu-Gly-Phe-Gln-Gly					
Cachalot	————— Asn ———— Lys ———— Tyr ————					

Fig. 4. Comparaison des séquences des acides aminés dans les myoglobines de cheval et de cachalot

Tableau 7. Substitutions entre les myoglobines de cheval et de cachalot

Pyr, base pyrimidique (U, C); Pur, base purique (A, G). Transition (T), substitution Pyr-Pyr ou Pur-Pur; transversion (t), substitution Pyr-Pur. C, substitution conservatrice

Numéro du résidu	Différences		Codons		Numéro de la base substituée	Transition (T), transversion (t)	Nature de la substitution	Localisation segment:
	Cheval	Cachalot	Cheval	Cachalot				
1	Gly	Val	GGX	GUX	2	t		AN
4	Asp	Glu	GA Pyr	GA Pur	3	t	C	A
9	Gln	Leu	CA Pur	CU Pur	2	t		A
12	Asn	His	AA Pyr	CA Pyr	1	t		A
15	Gly	Ala	GGX	GGX	2	t	C	A
21	Ile	Val	AU Pyr	GU Pyr	1	T	C	B
27	Glu	Asp	GA Pur	GA Pyr	3	t	C	B
28	Val	Ile	GU Pyr	AU Pyr	1	T	C	B
34	Thr	Lys	AC Pur	AA Pur	2	t		B
35	Gly	Ser	GG Pyr	AG Pyr	1	T		B
45	Lys	Arg	AA Pur	AG Pur	2	T	C	CD
74	Gly	Ala	GGX	GGX	2	t	C	E
109	Asp	Glu	GA Pyr	Ga Pur	3	t	C	G
118	Lys	Arg	AA Pur	AG Pur	2	T	C	G
132	Thr	Asn	AC Pyr	AA Pyr	2	t		H
140	Asn	Lys	AA Pyr	AA Pyr	3	t		H
151	Phe	Tyr	UU Pyr	UA Pyr	2	T	C	HC

cours du fractionnement de ces peptides particulièrement fragiles. Mais il est permis d'émettre l'hypothèse que la forme majeure de la myoglobine de cheval possède en 140 un résidu d'acide aspartique et en 152 un résidu de glutamine, alors que l'une des formes mineures aurait en 140 un résidu d'asparagine et en 152 un résidu d'acide glutamique. L'isolement des différentes formes, notamment de la forme ma-

jeure (environ 80% du total) permettra peut-être de résoudre le problème; de toute façon, l'hétérogénéité ne constitue pas un obstacle à l'étude structurale du fait de la proportion élevée de la forme majeure.

Si l'on compare la myoglobine de cheval avec celle du cachalot, déterminée par Edmundson [22], on constate 18 différences, soit 17 substitutions et 1 inversion (66—67); on peut discuter des substitutions

en fonction soit de leur nature, soit de leur localisation dans la chaîne peptidique (Fig. 4).

Toutes les substitutions observées sont punctiformes et affectent dans le DNA du génome le remplacement d'une seule base par codon (code génétique de Nirenberg [23]); sur 17 remplacements de bases (Tableau 7), 9, soit 50%, affectent la deuxième base des codons, contre 4 pour la première et 4 pour la troisième; il s'agit en outre, dans la majorité des cas, de « transversion » par substitution d'une base purique par une base pyrimidique ou inversement (11 sur 17, soit 60%). D'un autre point de vue, les substitutions d'acides aminés sont dans une large proportion de type conservatif (Val-Ile, Glu-Asp, Lys-Arg...): 10 cas sur 17, alors que 8 autres cas sont douteux (Gly-Ser, Thr-Asn).

Si l'on découpe la chaîne polypeptidique, d'après les données de Perutz [15], en segments spiralés (de A à H en partant de l'acide aminé N-terminal) et en segments non spiralés, on constate que 10 différences sur 18 sont localisées dans la séquence N-terminale (segments spiralés A et B) et 5 sur 18 dans la séquence C-terminale (segments G, H et HC); trois modifications seulement sur 18 sont situées dans la partie médiane de la chaîne: l'importance des liaisons secondaires entre l'hème et cette partie de la chaîne peptidique doit être rapprochée de cette analogie structurale chez deux espèces aussi différentes que le cachalot et le cheval.

Ce travail a été réalisé avec l'aide du Centre National de la Recherche Scientifique dans le cadre de l'Equipe de Recherches Associée n° 32 et de la R.C.P. n° 23. Nous remercions vivement le Docteur Yves Moschetto, qui a effectué les fractionnements de base par électrophorèse en film liquide. Nous remercions Mmes M. F. Demouveau-Clineckx, M. C. Caner-Desprière et Melle F. Dausse de leur collaboration technique.

BIBLIOGRAPHIE

1. Dautrevaux, M., Boulanger, Y., Han, K., Moschetto, Y., et Biserte, G., *Bull. Soc. Chim. Biol.* 49 (1967) 371.
2. Boulanger, Y., Dautrevaux, M., Han, K., et Biserte, G., *Bull. Soc. Chim. Biol.* 50 (1968) 1651.
3. Dautrevaux, M., Han, K., Boulanger, Y., et Biserte, G., *Bull. Soc. Chim. Biol.* 49 (1967) 1073.
4. Dautrevaux, M., Han, K., Boulanger, Y., et Biserte, G., *Bull. Soc. Chim. Biol.* 50 (1968) 505.
5. Han, K., Dautrevaux, M., Boulanger, Y., et Biserte, G., *Bull. Soc. Chim. Biol.* (sous presse).
6. Dautrevaux, M., Boulanger, Y., Han, K., Moschetto, Y., et Biserte, G., *Bull. Soc. Chim. Biol.* 49 (1967) 1409.
7. Han, K., Dautrevaux, M., Boulanger, Y., et Biserte, G., *FEBS Letters*, 3 (1969) 141.
8. Holleman, J. W., et Biserte, G., *Biochim. Biophys. Acta*, 33 (1959) 143.
9. Peterson, E. A., et Sober, H. A., *Anal. Chem.* 31 (1959) 857.
10. Hannig, K., *Z. Physiol. Chem.* 338 (1964) 211.
11. Woods, K. R., et Wang, K. T., *Biochim. Biophys. Acta*, 133 (1967) 369.
12. Seiler, M., et Wiechmann, J., *Experientia*, 20 (1964) 559.
13. Bahl, O. P., et Smith, E. L., *J. Biol. Chem.* 240 (1965) 3585.
14. Niu, C. I., et Fraenkel-Conrat, H., *J. Am. Chem. Soc.* 77 (1955) 5882.
15. Chauvet, J., et Acher, R., *Bull. Soc. Chim. Biol.* 49 (1967) 985.
16. Shaltiel, S., et Patchornik, A., *J. Am. Chem. Soc.* 85 (1963) 2799.
17. Han, K., Dautrevaux, M., et Biserte, G., *Ann. Pharm. Franc.* 24 (1966) 649.
18. Dautrevaux, M., Boulanger, Y., Han, K., Moschetto, Y., et Biserte, G., *Bull. Soc. Chim. Biol.* 50 (1968) 1637.
19. Theorell, H., et Åkeson, A., *Ann. Acad. Sci. Fennicae*, 60 (1955) 303.
20. Dautrevaux, M., et Boulanger, Y., *Bull. Soc. Chim. Biol.* 49 (1967) 949.
21. Åkeson, A., et Theorell, H., *Arch. Biochem. Biophys.* 91 (1960) 319.
22. Edmundson, A. B., *Nature*, 205 (1965) 883.
23. Nirenberg, M. W., Leder, P., Bernfeld, M., Brimacombe, R., Trupin, J., Rottman, F., et O'Neal, C., *Proc. Natl. Acad. Sci. U. S.* 53 (1965) 1161.
24. Perutz, M. F., *Dansk. Tidsskr. Farm.* 38 (1964) 113.

M. Dautrevaux, K. Han et G. Biserte
Laboratoire de Chimie Biologique
Faculté de Médecine et de Pharmacie
Place de Verdun, F-59 Lille, France

Adresse actuelle de Y. Boulanger:
Laboratoire de Chimie Biologique, Faculté des Sciences
Rue Descartes, Esplanade, F-67 Strasbourg, France