Biochimica et Biophysica Acta, 708 (1982) 253–258
Elsevier Biomedical Press

BBA 31372

AMINO ACID SEQUENCE OF THE SMALLEST POLYPEPTIDE CHAIN CONTAINING HEME OF EXTRACELLULAR HEMOGLOBIN FROM THE POLYCHAETE *TYLORRHYNCHUS HETEROCHAETUS*

TOMOHIKO SUZUKI [a,*], TAKASHI TAKAGI [a] and TOSHIO GOTOH [b,**]

[a] *Biological Institute, Faculty of Science, Tohoku University, Sendai 980 and* [b] *Department of Biology, College of General Education, University of Tokushima, Tokushima 770 (Japan)*

(Received April 19th, 1982)

Key words: Extracellular hemoglobin; Amino acid sequence; (Polychaete)

The complete amino acid sequence of 139 residues of the smallest polypeptide chain of extracellular hemoglobin from the polychaete *Tylorrhynchus heterochaetus* was determined. Mainly owing to the absence of seven residues in the G-helical region, this polypeptide chain is the second shortest, after *Chironomus* III hemoglobin (136 residues), of all vertebrate and invertebrate globin sequences so far reported. The residues at 42 sites (30%) of the *Tylorrhynchus* chain are identical with those of *Lumbricus* extracellular hemoglobin (AIII), which was sequenced very recently (Garlick, R.L. and Riggs, A.F. (1982) J. Biol. Chem. 257, 9005–9015). The C-terminal 20 residues of the *Tylorrhynchus* chain also showed 35% homology with those of *Glycera* intracellular hemoglobin, whereas the C-terminal sequence of *Tylorrhynchus* chain showed little correspondence to that of *Lumbricus* chain. The molecular mass of the polypeptide chain was determined to be 16327 including the heme group.

Introduction

Annelid extracellular hemoglobins (erythrocruorins) have molecular weights of about $3 \cdot 10^6$ and consist of more than 100 subunits [1–4]. Previously [5], we suggested that common structural features of extracellular hemoglobins from polychaetes and oligochaetes include the presence of two types of subunit with molecular weights of 12000–16000 and 50000–54000, and that the main structural differences in these hemoglobins are the number and size of the middle-sized subunits with molecular weights of 22000–37000. Although there have been extensive studies on the arrangement of the subunits in extracellular hemoglobin molecules from various species [2–7], it is still uncertain whether the considerable differences between the models of the molecular architectures proposed by different works are due to differences between annelid species or to the use of different techniques [8]. In order to construct the model for the molecular asembly of extracellular hemoglobin, it is inevitably necessary to determine the molecular mass of each constituent polypeptide

* Present address: Department of Biology, Faculty of Science, Kochi University, Kochi 780, Japan.
** To whom correspondence should be addressed.

Data supplementary to this article are deposited with, and can be obtained from, Elsevier Biomedical Press, B.V., BBA Data Deposition, P.O. Box 1345, 1000 BH Amsterdam, The Netherlands. Reference should be made to BBA/DD/233/31372/708 (1982) 253. The supplementary information includes: experimental procedure for subtractive Edman degradation; Sephadex G-50 chromatography of cyanogen bromide fragments; amino acid composition of the peptides; amino acid sequences in the peptides.

Abbreviations: TPCK, L-1-tosylamide-2-phenylethylchloromethyl ketone; SDS, sodium dodecyl sulfate.

chain. Very recently, the amino acid sequence and molecular mass have been obtained from the oligochaete *Lumbricus terrestris* as the first example in the 'giant' extracellular hemoglobins of invertebrate animals [22].

In the present study, we determined the complete amino acid sequence and molecular mass of the smallest polypeptide chain of extracellular hemoglobin from the common polychaete *Tylorrhynchus heterochaetus* as a step in understanding its molecular architecture and the evolutionary relationship of extracellular hemoglobins with other types of hemoglobin.

Materials and Methods

Ultrogel AcA-44 was purchased from LKB, and Sephadex G-50 superfine from Pharmacia. TPCK-treated trypsin, chymotrypsin and carboxypeptidases A and B were from Worthington. Carboxypeptidase Y was obtained from Sigma. *Staphylococcus aureus* V-8 protease was from Miles and thermolysin from Daiwa-kasei. Pyrrolidone carboxylic acid peptidase was from Boehringer-Mannheim. Reagents for sequence determination were of 'sequanal' grade from Wako Pure Chemicals.

Purified extracellular hemoglobin was dissociated by incubation in 50 mM borate buffer (pH 9.0) containing 2 M sodium chloride and 1 mM EDTA at 4°C for 2 days as described previously [9]. The dissociated products were separated by gel filtration on an Ultrogel AcA-44 column. Three peaks of products were obtained, which were shown by SDS-polyacrylamide gel electrophoresis to contain undissociated proteins, and 54-kDa and 12-kDa subunits, in order of elution. Fractions containing the 12-kDa component, which is the smallest polypeptide chain of *Tylorrhynchus* hemoglobin, were pooled and used for sequence determination. The material appeared homogeneous on SDS-polyacrylamide gel electrophoresis in 0.1 M phosphate buffer (pH 7.2). Heme was removed by treatment with acidic acetone at -20°C to yield the apoprotein as insoluble material [10]. The apoprotein was reduced with 10 mM dithiothreitol and free cysteine was carboxymethylated by treatment with 15 mM iodoacetic acid in the presence of 6 M guanidine-HCl containing 10 mM EDTA and Tris-HCl buffer (pH 8.5) for 15 min.

CNBr cleavage was performed in 70% formic acid with a 200-fold excess of CNBr over methionine, at room temperature for 20 h, and the CNBr peptides were separated on a Sephadex G-50 column (2.6 × 150 cm) equilibrated with 5% formic acid. The CNBr fragments were then digested with trypsin, chymotrypsin, thermolysin and *S. aureus* V-8 protease in 0.1 M NH_4HCO_3 at 37°C for 4 h at an enzyme-to-substrate ratio of 1 : 50 (w/w). One CNBr fragment (CN2) was also subjected to partial acid-hydrolysis by treatment with 0.02 M HCl at 100°C for 18 h. Overlaps of CNBr fragments were obtained by digestion of the whole protein with thermolysin in 0.1 M NH_4CO_3 at 37°C for 10 min.

The peptides were mainly purified by high-voltage paper electrophoresis at pH 5.5 in the first dimension and descending paper chromatography in *n*-butanol/pyridine/acetic acid/water (15 : 10 : 3 : 12, v/v) in the second. The peptides were located under ultraviolet light after sparying the paper with 0.002% fluorescamine in acetone. Some peptides were purified further by high-voltage paper electrophoresis at pH 2.0. The peptides obtained by partial acid-hydrolysis were purified by high-performance liquid chromatography (Hitachi model 638-50) using a linear gradient concentration from 0 to 50% acetonitrile in 0.01 M ammonium formate buffer (pH 4.0).

Peptides were routinely hydrolyzed with 6 M HCl containing 0.2% phenol for 20 h in evacuated sealed tubes. Amino acid analysis was performed in a Hitachi Model 835-50 amino acid analyzer. Tryptophan in peptides was detected with Ehrlich stain [11].

The amino acid sequence of peptides was determined by the manual Edman method [12] or by subtractive Edman degradation [13]. The phenylthiohydantoin amino acid derivatives were identified by high-performance liquid chromatography [14] in a Hitachi Model 638-50 apparatus. If the identity of the phenylthiohydantoin derivatives was uncertain, the derivatives were hydrolyzed with alkali or acid [15] and the products were subjected to amino acid analysis. Carboxy-terminal residues were determined by digestion with carboxypeptidases A, B and Y.

The primary peptides are numbered from the amino-terminus, and the subpeptides are numbered in order within the parent peptide.

In this paper, the prefix CN indicates a cyanogen bromide peptide; T, a tryptic peptide; C, a chymotryptic peptide; H, a thermolytic peptide; A, a peptide obtained by partial acid-hydrolysis; and V, a peptide obtained after digestion with *S. aureus* V-8 protease.

Results and Discussion

Table I shows the amino acid composition of the smallest polypeptide chain of *Tylorrhynchus* extracellular hemoglobin. The N-terminal sequence of the whole protein was determined by the manual Edman method [12] as

1 5 10
Thr-Asp-CmCys-Gly-Ile-Leu-Gln-(Arg)-Ile-Lys-Val-Lys.

The carboxy-terminal sequence of the whole protein was determined by carboxypeptidase A digestion as

Ile-Thr-Ala-COOH

Cyanogen bromide cleavage of the carboxymethylated apoprotein produced three fragments, CN1, CN2 and CN3, of 30 to 63 residues long. The N-terminal amino acid sequences of peptides CN1, CN2 and CN3 were determined by manual Edman degradation as Thr-Asp-CmCys-Gly-Ile-Leu, Val-(Arg)-Val-Phe-Ala and Phe-Gln-Thr, respectively. The C-terminal sequence of peptide CN3 was determined by carboxypeptidase A digestion as Ile-Thr-Ala-COOH. Therefore, the CNBr-cleavage peptides were arranged in the order

NH_2-CN1-CN2-CN3-COOH

with the aid of peptide H2 (Met-Val-Arg-Val) isolated from a thermolytic digest of the whole protein.

The amino acid sequence was mainly determined for the tryptic and chymotryptic peptides of CNBr fragments. The procedures used for complete sequence determination are summarized in Fig. 1, and detailed results are given as supplementary data. No overlap of peptides CN2T2 and CN2T3 was obtained, because trypsin cleaved unusually the tyrosyl bond completely. Therefore, the overlap was obtained with peptide CN2A1, which was obtained by partial acid-hydrolysis of peptide CN2 with 0.02 M HCl at 110°C for 18 h. Since the N-terminal glutamine residue of the CN2A1 peptide appeared to be cyclized to pyrrolidone carboxylic acid, the peptide was digested with pyrrolidone carboxylic acid peptidase before determination of its sequence.

Very recently, Garlick and Riggs separated three unique polypeptide chains (AI, AII, AIII) of extracellular hemoglobin of the earthworm *Lumbricus terrestris* on a CM-cellulose column in the presence of 8 M urea [8], and also determined the amino acid sequence of the AIII polypeptide chain

TABLE I

AMINO ACID COMPOSITION OF THE SMALLEST CHAIN OF *TYLORRHYNCHUS* EXTRACELLULAR HEMOGLOBIN

Carboxymethylated apoprotein was hydrolyzed with 6 M HCl containing 0.2% phenol at 110°C for 20, 48 and 72 h. Values shown are averages of 20-, 48- and 72-h hydrolases. n.d., not determined.

Amino acid	Number of Residues	
	From acid Hydrolysates	From the sequence
Aspartic acid	16.,27	16
Threonine [a]	7.52	8
Serine [a]	5.23	5
Glutamic acid	13.33	13
Proline	4.03	4
Glycine	9.99	9
Alanine	14.64	15
Cysteine [b]	1.97	2
Valine	10.32	11
Methionine	1.32	2
Isoleucine	10.17	11
Leucine	6.65	6
Tyrosine	2.39	4
Phenylalanine	11.65	13
Lysine	5.93	6
Histidine	4.64	5
Arginine	7.27	7
Tryptophan	n.d.	2
Total		139

[a] Values extrapolated to zero time.
[b] Determined as carboxymethylcysteine.

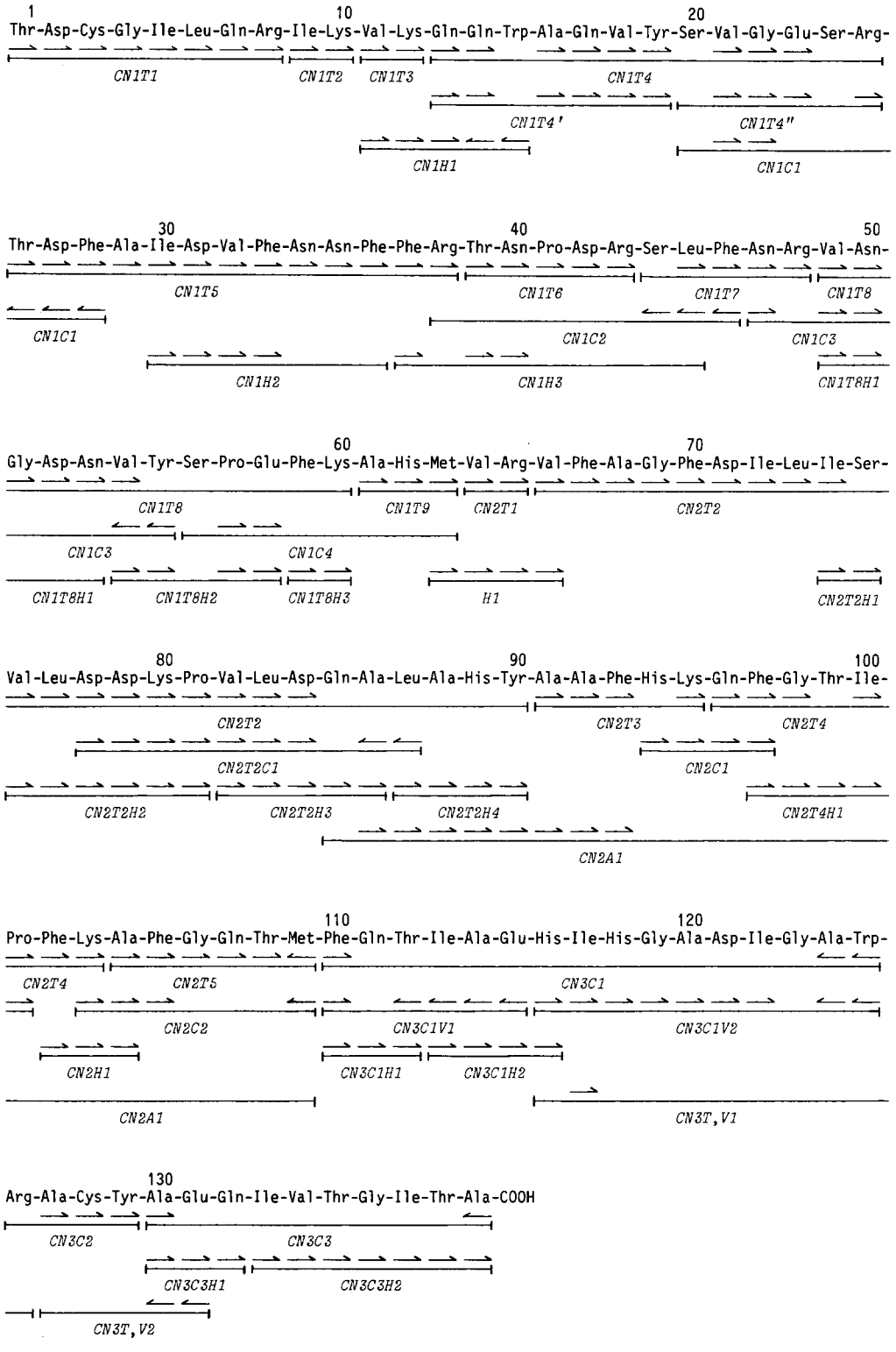

Fig. 1. Summary of data to establish the complete amino acid sequence for the smallest chain of *Tylorrhynchus* extracellular hemoglobin. Manual Edman degradation or subtractive Edman degradation (→) and carboxypeptidase digestion (←) were employed for sequence determination of peptides; CN, cyanogen bromide; T, trypsin; C, chymotrypsin; H, thermolysin; V, staphylococcal protease; A, partial acid-hydrolysis.

[22]. This is the first sequence obtained from one of the 'giant' extracellular hemoglobins of invertebrate animals.

The complete amino acid sequence of the smallest polypeptide chain of *Tylorrhynchus* extracellular hemoglobin is aligned with those of *Lumbricus* hemoglobin and *Glycera* intracellular hemoglobin [22,16] in Fig. 2. The apparent 'insertions' and 'deletions' used in Fig. 2 are used to obtain maximal correspondence between the sequences. The *Tylorrhynchus* chain is composed of 139 amino acid residues, of which 42 residues (30%) and 25 residues (18%) are identical with those in the corresponding positions in *Lumbricus* and *Glycera* hemoglobins. In the three globins, only eight residues appear to be invariant. It is noteworthy that in the *Tylorrhynchus* chain seven residues are missing from the G-helical region according to our alignment and thus this chain is the second shortest, after *Chironomus* III hemoglobin (136 residues), of all vertebrate and invertebrate globins so far sequenced. The presence of a distal histidine at position E7 of *Tylorrhynchus* and *Lumbricus* chains is in striking contrast with the case in the sequence of *Glycera* intracellular hemoglobin, in which the distal histidine is replaced by a leucine residue [16].

It is also of great interest that the N-terminal sequence of 105 residues of *Tylorrhynchus* chain shows high homology (37%) with that of *Lumbricus* chain, but there is no significant homology in the C-terminal sequence of 30 residues. On the other hand, the C-terminal sequence of 20 residues of *Tylorrhynchus* chain shows strong homology (35%) with that of *Glycera* hemoglobin. The amino acid sequence of *Tylorrhynchus* chain showed 14% homology with that of sperm whale myoglobin.

Previously [9] we reported that *Tylorrhynchus* hemoglobin is composed of two types of subunit with molecular weights of approx. 12000 and 54000, and that the smaller subunit consists of a single polypeptide chain, while the larger subunit has four polypeptide chains with identical molecular weights of approx. 13500. Our proposed sequence for the smaller subunit (the smallest chain) of *Tylorrhynchus* hemoglobin gives a molecular weight of 16327 including the heme group. Therefore, it is evident that the value of 12000 estimated by SDS-gel electrophoresis [9] is an underestimate. The value of 16327 for the smallest polypeptide chain in *Tylorrhynchus* hemoglobin is close to that of 16000 for the smallest subunit of *Lumbricus* hemoglobin estimated by Vinogradov et al. [18] by heme analysis, but higher than most values for the

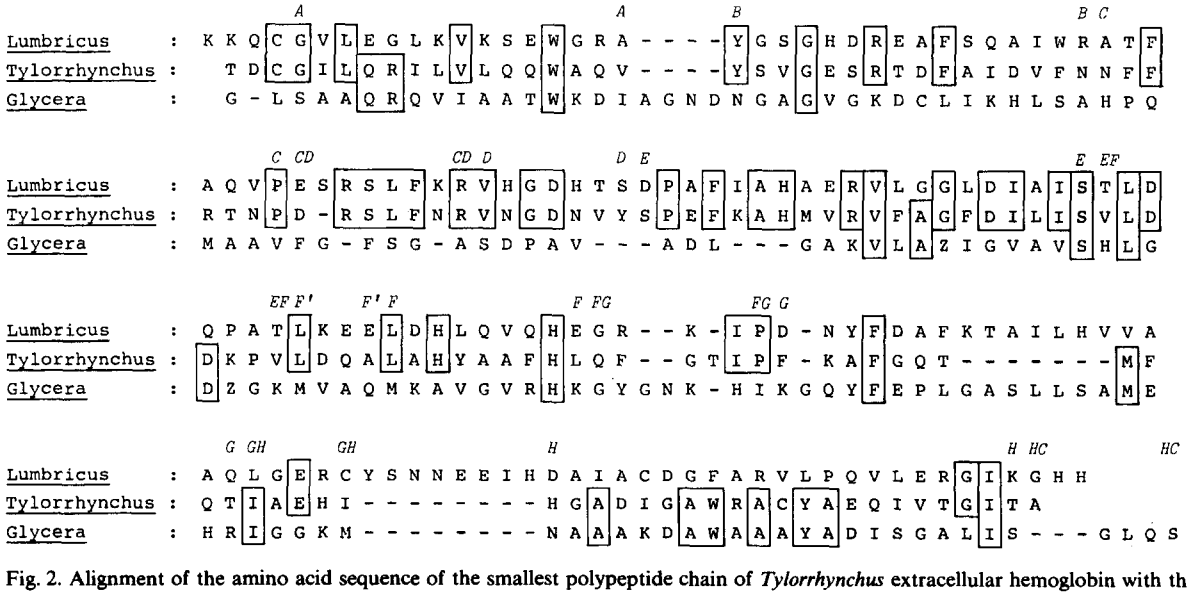

Fig. 2. Alignment of the amino acid sequence of the smallest polypeptide chain of *Tylorrhynchus* extracellular hemoglobin with those of *Lumbricus* and *Glycera* hemoglobins [22,16].

smallest subunits of annelid extracellular hemoglobins of various species [2–7] estimated by SDS-gel electrophoresis to be below 14000. This discrepancy seems to be due to the inaccuracy of the method of SDS-gel electrophoresis for determination of the molecular weights of proteins in the low molecular weight ranges [19,20]. In the present study, it has also become evident that the value of 13500 for the constituent polypeptide chains of the larger subunit in *Tylorrhynchus* erythrocruorin estimated by SDS-gel electrophoresis [9] is an underestimate. Very recently, a value of 17600 for these chains was obtained by gel filtration after reduction of the larger subunit with dithiothreitol [21].

In any event, determination of the exact molecular masses of constituent polypeptide chains would be useful in arranging the polypeptide chains in order in the native molecule, particularly as annelid extracellular hemoglobin consists of approximately 200 polypeptide chains. Further work on the structure of the larger subunit should provide definite information on the topology of *Tylorrhynchus* extracellular hemoglobin.

Acknowledgements

We thank Dr. Robert L. Garlick and Dr. Austin F. Riggs of the University of Texas for sending us their manuscript before publication. We are indebted to Dr. Takahiro Furukori of Kochi University for supplying specimens of *Tylorrhynchus heterochaetus*. We are also grateful to Dr. Kazuhiko Konishi, Dr. Keiji Shikama and Mr. Takayuki Niemoto of Tohoku University for their interest and encouragement during this work.

References

1 Svedberg, T. (1933) J. Biol. Chem. 103, 311–325
2 Antonini, E. and Chiancone, E. (1977) Annu. Rev. Biophys. Bioeng. 6, 239–271
3 Chung, M.C.M. and Ellerton, H.D. (1979) Prog. Mol. Biol. 35, 53–102
4 Vinogradov, S.N., Shlom, J.M., Kapp, O.H. and Frossard, P. (1980) Comp. Biochem. Physiol. 67B, 1–16
5 Gotoh, T. and Kamada, S. (1980) J. Biochem. (Tokyo) 87, 557–562
6 Ochiai, T. and Enoki, Y. (1981) Comp. Biochem. Physiol. 68B, 275–280
7 Pionetti, J.M. and Pouyet, J. (1980) Eur. J. Biochem. 105, 131–138
8 Garlick, R.L. and Riggs, A.F. (1981) Arch. Biochem. Biophys. 208, 563–575
9 Gotoh, T. (1980) J. Sci. Univ. Tokushima 13, 1–7
10 Antonini, E. and Burunori, M. (1971) Hemoglobin and Myoglobin in their Reactions with Ligands, pp. 123–125, North-Holland, Amsterdam
11 Smith, I. (1953) Nature 171, 43–44
12 Edman, P. and Henschen, A. (1975) in Protein Sequence Determination (Needleman, S.D., ed.), 2nd Edn., pp. 232–279, Springer-Verlag, Berlin
13 Konigsberg, W. (1972) Methods Enzymol. 25, 326–332
14 Zimmerman, C.L., Appella, E. and Pisano, J.J. (1977) Anal. Biochem. 77, 569–573
15 Smithies, O., Gibson, D., Fanning, E.M., Goodfliesch, R.M., Gilman, J.G. and Ballantyne, D.J. (1971) Biochemistry 10, 4912–4921
16 Imamura, T., Baldwin, T.O. and Riggs, A. (1972) J. Biol. Chem. 247, 2785–2792
17 Buse, G., Braig, S. and Braunitzer, G. (1969) Z. Physiol. Chem. 350, 1686–1690
18 Vinogradov, S.N., Shlom, J.M., Hall, B.C., Kapp, O.H. and Mizukami, H. (1977) Biochim. Biophys. Acta 492, 136–155
19 Fish, W.W., Reynolds, J.M. and Tanford, C. (1970) J. Biol. Chem. 245, 5166–5168
20 Williams, J.G. and Gratzer, W.B. (1971) J. Chromatogr. 57, 121–125
21 Gotoh, T. (1982) J. Sci. Univ. Tokushima 15, 1–10
22 Garlick, R.L. and Riggs, A.F. (1982) J. Biol. Chem. 257, 9005–9015