Biochem. J. (1987) 241, 441–445 (Printed in Great Britain)



# Two globin strains in the giant annelid extracellular haemoglobins

Toshio GOTOH[a*], Fumio SHISHIKURA[a†], June W. SNOW[a], Khalil I. EREIFEJ[b], Serge N. VINOGRADOV[b] and Daniel A. WALZ[a‡]

Departments of [a]Physiology and [b]Biochemistry, Wayne State University, 540 East Canfield, Detroit, MI 48201, U.S.A.

The constituent polypeptide chains I, II, III and IV of the giant extracellular haemoglobin of the oligochaete *Lumbricus terrestris* were isolated by mono Q ion-exchange chromatography and $C_8$ reverse-phase chromatography. 2. The *N*-terminal amino acid sequences of *Lumbricus* chains I, III and IV were determined and aligned with those of *Lumbricus* chain II and the four chains of the extracellular haemoglobin of the polychaete *Tylorrhynchus heterochaetus*. 3. Three invariant amino acid residues, Cys-7, Val-15 and Trp-19, were found to occur in the *N*-terminal segments (17–22 residues) of the eight chains of *Lumbricus* and *Tylorrhynchus* haemoglobins. In addition, it was found that the eight sequences could be separated into two groups: 'A', consisting of *Lumbricus* chains I and II and *Tylorrhynchus* chains I and IIA, having invariant Lys-14 and Lys-16, and 'B', consisting of *Lumbricus* chains III and IV and *Tylorrhynchus* IIB and IIC, having invarient Cys-6, Ser-8 and Asp-11. This result suggests that there are two strains of globin chain in the annelid extracellular haemoglobins.

## INTRODUCTION

It is well known that there are three main families of globin chains, α- β- and myo-globin, among the vertebrate phyla (Goodman *et al.*, 1975). On the other hand, only limited information is available concerning the molecular phylogeny of invertebrate haemoglobins. From the viewpoints of molecular assembly and evolution, it is of particular interest to determine whether there exist families of globin chains in annelid extracellular haemoglobins which form one of four recognizable groups of invertebrate extracellular haemoglobins (Vinogradov, 1985).

The extracellular haemoglobin of the oligochaete *Lumbricus terrestris* consists of six types of chains (I–VI), and has a relative molecular mass of 3800000 (Vinogradov *et al.*, 1977; Vinogradov *et al.*, 1986). Chain I is 'monomeric', and chains II, III and IV form a disulphide-bonded 'trimer'. The $M_r$ values of these four chains are comparable with that of vertebrate myoglobin, whereas chains V and VI are about double the size of the other chains. Chain I has recently been sequenced (Shishikura *et al.*, 1986*b*) and chain II has been sequenced by Garlick & Riggs (1982). The amino acid sequences of the 'monomeric' chain I and 'trimeric' chains IIA, IIB and IIC of the extracellular haemoglobin of the polychaete *Tylorrhynchus heterochaetus* have been completed (Suzuki *et al.*, 1982; Suzuki *et al.*, 1985*a*,*b*; Suzuki & Gotoh, 1986*a*). The *Lumbricus* and *Tylorrhynchus* haemoglobins are very similar in electron microscopic appearance and probably consist of about 200 polypeptide chains (Vinogradov *et al.*, 1977; Mainwaring *et al.*, 1986; Suzuki & Gotoh, 1986*b*).

We report here the isolation of the constituent chains of *Lumbricus* haemoglobin and their *N*-terminal sequences. Alignment of the *Lumbricus* chains with the *Tylorrhynchus* chains clearly indicates that there are two distinct strains of globin chain in the annelid extracellular haemoglobins.

## MATERIALS AND METHODS

*Lumbricus* haemoglobin was prepared as described previously (Shlom & Vinogradov, 1973). Mono Q columns and reverse phase $C_8$ (Pro RPC HR5/10) columns were products of Pharmacia. [$^3$H]Iodoacetic acid was purchased from New England Nuclear.

Fresh *Lumbricus* haemoglobin was converted to the cyanmet-form and reduced in the presence of dithiothreitol (DTT) (Suzuki *et al.*, 1985*a*). The dissociated chains were separated on a Mono Q column by using the system of fast protein liquid chromatography (FPLC; Pharmacia). Haem was removed by treatment with acidic methyl ethyl ketone (Teale, 1959) to yield the apoproteins. The apoproteins were reduced further with 10 mM-DTT at 50 °C for 2 h and free cysteine was carboxymethylated by treatment with 15 mM-monoiodoacetic acid containing 50 μCi of [$^3$H]iodoacetic acid in the presence of 6 M-guanidine–HCl, 10 mM-EDTA, 0.2 M-Tris, pH 8.5, for 15 min. Salts and contaminants were removed from the chain of interest by reverse-phase chromatography on a $C_8$ column. About 150 μg of freshly acidified protein was injected in 100 μl at 0% organic. The column was washed with 0.1% trifluoroacetic acid. A linear gradient of 0.15%/min from 30% to 42% acetonitrile at a flow rate of 0.2 ml/min was used for the elution of proteins. SDS-PAGE was carried out in a Hoeffer slab apparatus with the Laemmli (1970) buffer system on a 15–25% gradient gel. Samples were incubated in 2.5% SDS at 100 °C for 3 min and subjected to electrophoresis in the presence of 2-mercaptoethanol. The slab gel was

Abbreviations used: SDS-PAGE, sodium dodecyl sulphate/polyacrylamide-gel electrophoresis; DTT, dithiothreitol; f.p.l.c., fast protein liquid chromatography; h.p.l.c., high performance liquid chromatography.

* Present address: Department of Biology, University of Tokushima, Tokushima 770, Japan.

† Present address: Department of Biology, Nihon University School of Medicine, Tokyo 173, Japan.

‡ To whom correspondence should be addressed.

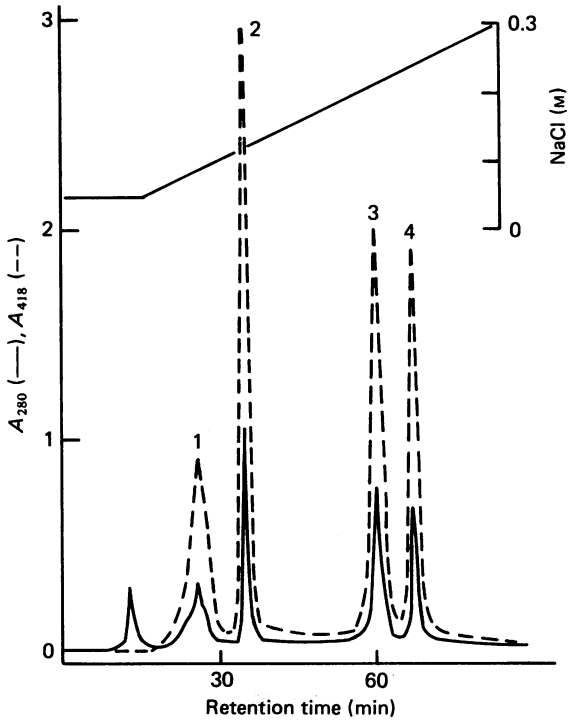


**Fig. 1. Chromatography of reduced *Lumbricus* cyanmethaemoglobin on a Mono Q column**

Haemoglobin solution (6 mg) was reduced with 5 mM-dithiothreitol (DTT) and applied to a Mono Q column (0.5 cm × 5 cm) equilibrated with 20 mM-Tris–HCl buffer (pH 8.0) containing 0.5 mM-DTT and 1 mM-EDTA. The column was washed with 50 mM-NaCl in the same buffer and the proteins were eluted with a linear gradient of 50–300 mM-NaCl in the buffer at a flow rate of 0.3 ml/min.

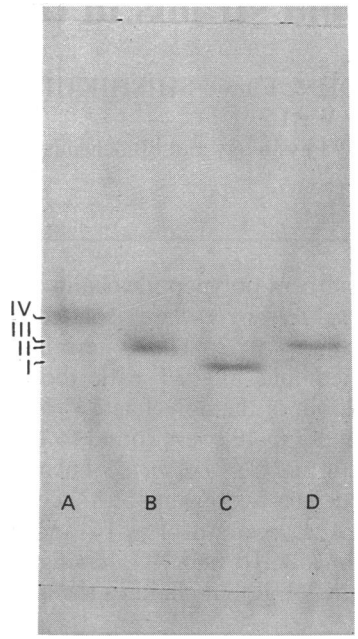


**Fig. 2. SDS-PAGE patterns of the isolated chains of *Lumbricus* haemoglobin**

A, Peak 1 (chain IV) in Fig. 1; B, peak 2 (chain II); C, peak 3 (chain I)3 D, peak 4 (chain III).

stained with Coomassie Brilliant Blue R250 and destained in a solution containing 25% methanol and 7.5% acetic acid.

The amino acid sequence was determined with a Beckman model 890M sequencer and Applied Biosystem 470A protein sequencer equipped with a model 120 PTH analyser. Phenylthiohydantoin amino acid derivatives were also identified in a Waters h.p.l.c. system and a Beckman Ultrasphere $C_{18}$ column. Carboxymethylcysteine was identified by using a Packard beta liquid scintillation counter. The minimum numbers of amino acid replacements and base substitutions were calculated by the method described by Dayhoff *et al.* (1972).

## RESULTS AND DISCUSSION

Since *Lumbricus* cyanmethaemoglobin dissociates almost completely into each constituent chain in the presence of a reducing agent such as DTT or 2-mercaptoethanol without protein denaturants, the reduced materials were subjected to anion-exchange chromatography to separate the polypeptide chains. A typical elution profile of reduced *Lumbricus* cyanmethaemoglobin on a Mono Q column is shown in Fig. 1. The reduced materials separated into four fractions. The area proportions of peaks 1, 2, 3 and 4 were about 1.0:1.1:1.1:1.0. Peaks 1–4 were identified as chains IV, II, I and III, respectively, by SDS-PAGE (Fig. 2). Each isolated chain showed a typical spectrum of haemoglobin, indicating the presence of the haem group. When the column was washed with 1 M-NaCl, chains V and VI were eluted as one peak. This may reflect the characteristic 'stickiness' of the two chains to chromatographic matrices. The recovery of protein was about 70%. The reduced chains did not sequence well. However, the respective carboxymethylated chains, recovered after reverse phase chromatography, yielded sequence data for each chain.

Although extensive studies have been done on the remarkable structures of annelid extracellular haemoglobins, there is little consensus among investigators about their molecular assembly, mainly because each group uses a different haemoglobin and has proposed a unique model from results obtained by different methods (see references for reviews: Antonini & Chiancone, 1977; Chung & Ellerton, 1979; Vinogradov *et al.*, 1980; Vinogradov, 1985). One step toward understanding the common structure is to investigate the homology between the oligochaete and polychaete haemoglobins in terms of amino acid sequence. In this respect, we have determined the *N*-terminal sequences of four chains of the haemoglobin of the earthworm *Lumbricus* and compared them with those corresponding chains of the haemoglobin of the marine worm *Tylorrhynchus*.

Fig. 3 compares the *N*-terminal sequences of the four constituent chains of *Lumbricus* haemoglobin (chain I: Shishikura *et al.*, 1986*b*; chain II: Garlick & Riggs, 1982) and those of the *Tylorrhynchus* haemoglobin (Suzuki & Gotoh, 1986*a*). Three residues appeared to be invariant in the eight chains thus far sequenced. Two of them, Val-15 and Trp-19, are also conserved in the human $\beta$ chain. It is noteworthy that Cys-7 is conserved in the eight chains since these annelid extracellular haemoglobins differ from

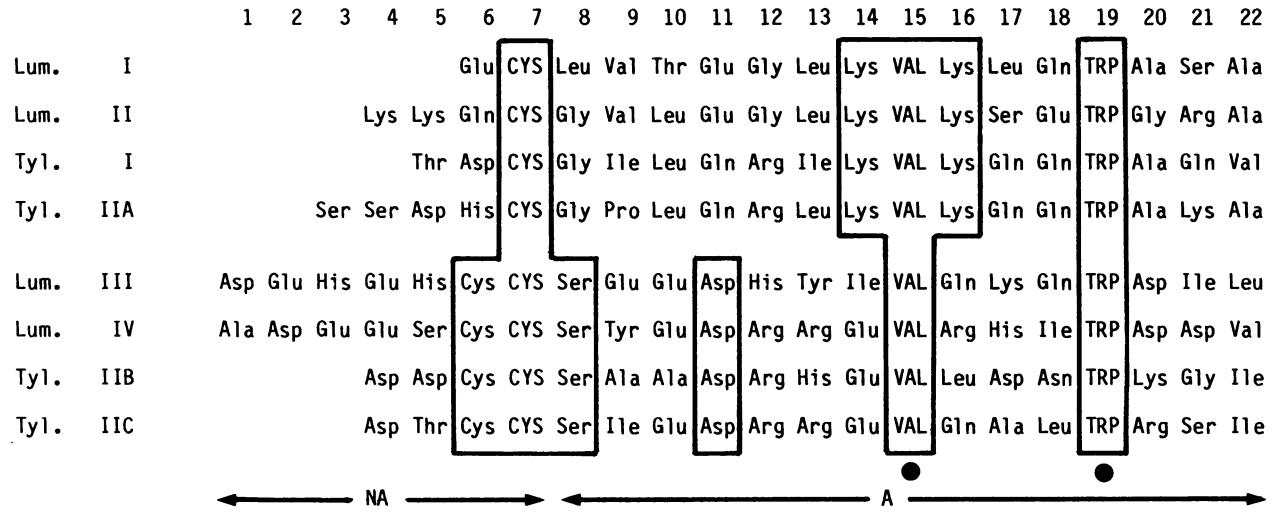


**Fig. 3. Alignment of the *N*-terminal amino acid sequences of *Lumbricus* chains I (Shishikura *et al.*, 1986*b*), II (Garlick & Riggs, 1982), III and IV with those of *Tylorrhynchus* chains I, IIA, IIB and IIC (Suzuki & Gotoh, 1986*a*)**

*Lumbricus* chain IV showed heterogeneity at position 15, where two residues, Val and Ile, were identified. The boxed residues indicate invariant residues in the eight chains and either the upper strain (strain A) or the lower strain (strain B). The residues indicated by ● are homologous with those of human β-haemoglobin chain. NA and A represent the *N*-terminal segments for the globin chain. The amino acid sequences are as aligned by Suzuki & Gotoh (1986*a*).

**Table 1. Difference matrix for *N*-terminal sequences of *Lumbricus* and *Tylorrhynchus* chains given in Fig. 3**

Below the diagonal, amino acid; above the diagonal, nucleotide. The minimum numbers of amino acid replacements and base substitutions were calculated according to Dayhoff *et al.* (1972).

| | | Lum. I | Lum. II | Tyl. I | Tyl. IIA | Lum. III | Lum. IV | Tyl. IIB | Tyl. IIC | |
|---|---|---|---|---|---|---|---|---|---|---|
| Amino acid replacements | Lum. I | | 10 | 14 | 16 | 26 | 25 | 23 | 21 | Nucleotide replacements |
| | Lum. II | 9 | | 14 | 15 | 28 | 27 | 26 | 22 | |
| | Tyl. I | 11 | 12 | | 10 | 24 | 23 | 25 | 19 | |
| | Tyl. IIA | 11 | 11 | 8 | | 26 | 26 | 25 | 26 | |
| | Lum. III | 18 | 19 | 18 | 18 | | 20 | 21 | 20 | |
| | Lum. IV | 19 | 19 | 17 | 18 | 13 | | 17 | 16 | |
| | Tyl. IIB | 16 | 16 | 15 | 15 | 16 | 14 | | 12 | |
| | Tyl. IIC | 15 | 16 | 13 | 16 | 13 | 12 | 9 | | |

other haemoglobins in possessing disulphide-bonded 'trimers' (Shishikura *et al.*, 1986*a*; Suzuki & Gotoh, 1986*a*).

Even more important, there are two distinct groups of chains for the amino acid sequence homology, as clearly seen in Fig. 3. One group consists of *Lumbricus* chains I and II and *Tylorrhynchus* chain I and IIA. This strain has conserved Lys-14 and Lys-16 as well as the three invariant residues. Another group consists of *Lumbricus* chains III and IV and *Tylorrhynchus* chains IIB and IIC. This strain has conserved Cys-6, Ser-8 and Asp-11. It is easy to locate three of four identical residues in strain A (upper group, Fig. 3), for instance, Gly-8, Leu-10, Leu-13, Gln-18, Ala-20 and Ala-22, and, for strain B (lower group, Fig. 3), Glu-10, Arg-12 and Glu-14. It is noteworthy that strain A involves not only the 'monomeric' chains, *Lumbricus* I and *Tylorrhynchus* I, but also the 'trimeric' chains, *Lumbricus* II and *Tylorrhynchus* IIA. A difference matrix for the sequences of *Lumbricus* and *Tylorrhynchus* chains shown in Table 1 also demonstrates that there are two strains of globin chains in the two annelid haemoglobins. For instance, nucleotide substitutions between *Lumbricus* chain II and *Tylorrhynchus* chain I represent only 14 base-pair differences in the segments NA and A, in contrast with 28 differences between *Lumbricus* chains II and III. Table 2 compares some additional characteristics of *Lumbricus* and *Tylorrhynchus* chains. *Lumbricus* chain I and *Tylorrhynchus* chain I are both the smallest and 'monomeric' chain in each haemoglobin. Suzuki & Gotoh (1986*a*) reported that there were 61 identical residues (42%) between *Lumbricus* chain II and *Tylorrhynchus* chain IIA. This value is comparable with that (44%) between human α and β chains. At present, however, it is difficult to deduce the correspondences among the other two chains of *Lumbricus* and *Tylorrhynchus* haemoglobins. As far as *N*-terminal sequence is concerned, both of the *Tylorrhynchus* chains IIB and IIC apparently resemble *Lumbricus* chain IV more than *Lumbricus* chain III (Table 1). In any case, it is surprising that the sequences of *Lumbricus* and *Tylorrhynchus* haemoglobins are very similar, especially in each strain.

**Table 2. Characteristics of the chains of *Lumbricus* and *Tylorrhynchus* haemoglobins**

Abbreviation used: UK, exact number of amino acid residues is unknown. Note that *Lumbricus* chains were named in the order of mobility in SDS-PAGE (Shlom & Vinogradov, 1973).

| | | Molecular structure | No. of amino acids | $M_r$* | Reference |
|---|---|---|---|---|---|
| *Lumbricus* | I | Monomeric | 142 | 16750 | Shishikura *et al.* (1986*b*) |
| | II | Trimeric | 157 | 18114 | Garlick & Riggs (1982) |
| | III | Trimeric | UK | 16000 | Shlom & Vinogradov (1973) |
| | IV | Trimeric | UK | 19000 | Shlom & Vinogradov (1973) |
| *Tylorrhynchus* | I | Monomeric | 139 | 16327 | Suzuki *et al.* (1982) |
| | IIA | Trimeric | 146 | 17268 | Suzuki & Gotoh (1986*a*) |
| | IIB | Trimeric | 148 | 17236 | Suzuki *et al.* (1985*b*) |
| | IIC | Trimeric | 149 | 17415 | Suzuki *et al.* (1985*a*) |

* Deduced from the amino acid sequences, including haem group, except for *Lumbricus* chains III and IV, whose $M_r$ values were obtained by SDS-PAGE.

From the viewpoint of molecular phylogeny, it is also of great interest to consider the sequence of evolutionary events for the annelid haemoglobins. Our results strongly suggest that the formation of a disulphide-bonded trimer took place after the separation of strains A and B by gene duplication. Furthermore, the ancestral species of the polychaetes and oligochaetes must have had the two strains A and B. Suzuki & Gotoh (1986*a*) have already found such grouping of the amino acid sequences in *Tylorrhynchus* chains. However, it should be emphasized that the present finding is an extension of their original idea to two very different species of the annelid phylum.

There is some controversy concerning the subunit structure of *Lumbricus* haemoglobin between two groups: Garlick & Riggs (1981, 1982) have argued that it consists of only three chains whereas Vinogradov *et al.* have maintained that there were at least six chains (Shlom & Vinogradov, 1973; Vinogradov *et al.*, 1977; Kapp *et al.*, 1984). This has been resolved by sequence analyses of the *N*-terminal residues of the whole molecule, the trimer subunit and the isolated chains. Recently, Shishikura *et al.* (1986*a*) have found that *Lumbricus* chain II corresponds to chain AIII (nomenclature according to Garlick & Riggs, 1981). In the present study, we have reported that chains III and IV correspond to the chains AI and AII respectively of Garlick & Riggs (1981). On the other hand, Fushitani *et al.* (1985) have confirmed the existence of the 'monomeric' chain I, which was missing in their original preparations of *Lumbricus* haemoglobin (Garlick & Riggs, 1981, 1982). From the present study we can conclude, from the results shown in Figs. 1 and 3, that *Lumbricus* chains I–IV contain haem. There is now little disagreement as to the constituents of *Lumbricus* haemoglobin except the amounts of the dimeric subunits, namely, chains V and VI.

Very recently, Mainwaring *et al.* (1986) have proposed a novel 'bracelet' model for the molecular assembly of *Lumbricus* haemoglobin and stressed the role of chains V and VI as a scaffolding for the complexes of the 'monomeric' chain I and 'trimeric' chains II, III and IV. On the other hand, Suzuki & Gotoh (1986*b*) have proposed a 'symmetrical' model for the molecular assembly of *Tylorrhynchus* haemoglobin, which consists of 48 tetramers of the 'monomeric' chain I and 'trimeric' chains IIA, IIB and IIC. Although there is some disagreement between these models, we believe that the gap will be overcome in part by sequencing *Lumbricus* chains V and VI in the near future. The challenging puzzle to construct a common model for the molecular architectures of the giant annelid haemoglobins has continued ever since Svedberg's (1933) finding that the $M_r$ of *Lumbricus* haemoglobin was about 3 million.

In conclusion, comparison of the *N*-terminal sequences of four chains each of the oligochaete *Lumbricus* and polychaete *Tylorrhynchus* haemoglobins revealed that there are two distinct strains, termed A and B, in the giant annelid haemoglobins. This is the first finding of two strains of globin chain in the invertebrate multi-subunit haemoglobins. The two strains may have a relationship similar to each other, as do the $\alpha$ and $\beta$ families of vertebrate haemoglobins. *Lumbricus* and *Tylorrhynchus* haemoglobins are very similar in electron microscopic appearance, and consist of about 200 polypeptide chains including the characteristic 'monomeric' and 'trimeric' chains (Mainwaring *et al.*, 1986; Suzuki & Gotoh, 1986*b*). In addition, we have found a good correspondence in terms of the amino acid sequences for the monomeric and trimeric chains between these two haemoglobins. These findings strongly suggest that the oligochaete and polychaete haemoglobins should have similar subunit structures.

This work was supported by grants from the National Institutes of Health, AM 30382 and HL 25952.

## REFERENCES

Antonini, E. & Chiancone, E. (1977) Annu. Rev. Biophys. Bioeng. **6**, 239–271

Chung, M. C. M. & Ellerton, H. D. (1979) Prog. Biophys. Mol. Biol. **35**, 53–102

Dayhoff, M. O., Hunt, L. T., McLaughlin, P. J. & Jones, D. D. (1972) in Atlas of Protein Sequence and Structure (Dayhoff, M. O., ed.), vol. 5, pp. 17–30, National Biomedical Research Foundation, Washington

Fushitani, K., Imai, K. & Riggs, A. F. (1985) in Invertebrate Oxygen Carriers (Linzen, B., ed.), in the press

Garlick, R. L. & Riggs, A. (1981) Arch. Biochem. Biophys. **208**, 563–575

Garlick, R. L. & Riggs, A. F. (1982) J. Biol. Chem. **257**, 9005–9015

Goodman, M., Moore, G. W. & Matsuda, G. (1975) Nature (London) **253**, 604–608
Kapp, O. H., Polidori, G., Mainwaring, M. G., Crewe, A. V. & Vinogradov, S. N. (1984) J. Biol. Chem. **259**, 628–639
Laemmli, U. K. (1970) Nature (London) **227**, 680–685
Mainwaring, M. G., Lugo, S. D., Fingal, R. A., Kapp, O. H. & Vinogradov, S. N. (1986) J. Biol. Chem., in the press
Shishikura, F., Mainwaring, M. G., Yurewicz, E. D., Lightbody, J. J., Walz, D. A. & Vinogradov, S. N. (1986*a*) Biochim. Biophys. Acta **896**, 314–321
Shishikura, F., Snow, J. W., Gotoh, T., Vinogradov, S. N. & Walz, D. A. (1986*b*) J. Biol. Chem., in the press
Shlom, J. M. & Vinogradov, S. N. (1973) J. Biol. Chem. **248**, 7904–7912
Suzuki, T. & Gotoh, T. (1986*a*) J. Biol. Chem. **261**, 9257–9267
Suzuki, T. & Gotoh, T. (1986*b*) J. Mol. Biol. **190**, 119–123
Suzuki, T., Takagi, T. & Gotoh, T. (1982) Biochim. Biophys. Acta **708**, 253–258
Suzuki, T., Furukohri, T. & Gotoh, T. (1985*a*) J. Biol Chem. **260**, 3145–3154
Suzuki, T., Yasunaga, H., Furukohri, T., Nakamura, K. & Gotoh, T. (1985*b*) J. Biol. Chem. **260**, 11481–11487
Svedberg, T. (1933) J. Biol. Chem. **103**, 311–325
Teale, F. W. G. (1959) Biochim. Biophys. Acta **35**, 543
Vinogradov, S. N. (1985) Comp. Biochem. Physiol. B **82**, 1–15
Vinogradov, S. N., Shlom, J. M., Hall, B. C., Kapp, O. H. & Mizukami, H. (1977) Biochim. Biophys. Acta **492**, 136–155
Vinogradov, S. N., Shlom, J. M., Kapp, O. H. & Frossard, P. (1980) Comp. Biochem. Physiol. B **67**, 1–12
Vinogradov, S. N., Lugo, S. D., Mainwaring, M. G., Kapp, O. H. & Crewe, A. V. (1986) Proc. Natl. Acad. Sci. U.S.A., in the press

Received 8 July 1986/12 August 1986; accepted 23 September 1986