Non-heme protein in the giant extracellular hemoglobin of the earthworm *Lumbricus terrestris*

(Kyoto 30609, Japan)

KENZO FUSHITANI* AND AUSTEN F. RIGGS†

Department of Zoology, University of Texas at Austin, Austin, TX 78712

Communicated by Lester J. Reed, September 12, 1988

**ABSTRACT** The protein/heme mass ratio for the extracellular hemoglobin of the earthworm *Lumbricus terrestris* has been reetermined. We find a value of 19,000 g of protein per mol of heme. Four major, heme-containing chains (a, b, c, and d), present in equal proportions, have a total molecular mass, with four hemes, of 69,664 Da based on their sequences. The intact hemoglobin comprises 12 subunits that form a two-layered hexagonal structure of about 3.8 MDa. This value, together with our determination of the protein/heme ratio, requires that 4 *abcd* units are present in each 1/12th subunit and that 192 heme-containing chains are present in the hemoglobin molecule. Our data indicate that ≈2200 g of non-heme protein is present for each mole of heme-containing, chain, or about 35,200 g per 1/12th subunit. This conclusion is consistent with the observation that chains of 31–37 kDa are present. On this basis the intact molecule would have 12 non-heme chains and 204 chains in all to give a total molecular mass of 3.77 MDa, close to that observed.

The extracellular hemoglobin of the earthworm *Lumbricus terrestris* is a giant molecule of ≈200 units and a molecular mass near 3.8 ± 0.15 MDa (1–3). The number of heme-containing chains in the whole assembly is uncertain. The reported iron content of extracellular annelid hemoglobins, 0.22–0.27%, is much lower than that of vertebrate hemoglobins, ≈0.35% (4, 5). The lower iron content corresponds to a minimal molecular mass of 21–27 kDa, but the actual masses of the four major chains of the hemoglobins in each of two species of annelid have been determined to be 16–17 kDa by amino acid sequencing (6–8). This discrepancy might be explained by the presence of non-heme protein, by heme loss, or by a combination of these factors. The carbohydrate content (9) is too small to explain the discrepancy. Three additional chains of 31–37 kDa have been reported for *L. terrestris* hemoglobin, at least some of which appear to be necessary for self-assembly and may lack heme (10, 11). Two bands at 33 kDa and 38 kDa occur in Western blots of in vitro translation products identified immunologically (12). The wide, unexplained differences in the published values of the iron or heme content of the hemoglobin of *L. terrestris* and other annelid hemoglobins (4) deserve careful scrutiny. The differences might be partially explained by variations in heme loss. Such loss can occur in human hemoglobin but requires oxidation to ferrihemoglobin and is blocked by heme ligands (13). Autooxidation and heme loss are accelerated in the unstable human sickle hemoglobin (14). Similar processes might occur in invertebrate hemoglobins. Recent measurement of the heme content of the hemoglobin of the nematode *Ascaris suum* has shown that the heme content varied by more than 40% but that samples differing in heme content could be titrated with heme to produce full binding and a reproducible content (15).

The determination of heme content has usually been made by coupling measurement of heme, done by the pyridine hemochromogen method or by iron determinations (16), with determination of the dry weight of the protein. The work of Wiechelman and Parkhurst (17), however, depended on a calibrated biuret assay for the protein. Their results provided the highest heme content reported for *Lumbricus* hemoglobin, with 1 mol of heme per 17,000 ± 1000 g of protein. All other measurements for this hemoglobin have given minimal molecular masses over 22 kDa. Here we report determinations of the heme/protein ratio that depend on the pyridine hemochromogen assay and protein assay by amino acid analysis. This procedure avoids all problems associated with dry-weight determination (18), such as presence of undialyzable salt or carbohydrate or inability to remove tightly bound water. We have performed these measurements on the intact hemoglobin, on the isolated subunits a, b, c, and d, and on the chromatographically isolated trimer composed of the disulfide-linked chains a, b, and c (8).

**MATERIALS AND METHODS**

The collection of the hemoglobin in the CO form in the presence of the protease inhibitor phenylmethylsulfonyl fluoride, precipitation with polyethylene glycol, pelleting twice by ultracentrifugation, and gel chromatography on Sepharose CL-6B (Pharmacia; exclusion limit, ≈4000 kDa) were performed as described (19). Native hemoglobin eluted from the Sepharose column was used for the heme and protein analyses. Aliquots of the native hemoglobin were chromatographed on a column of Ultrogel AcA 44 (LKB) in 0.1 M borate/1 mM EDTA, pH 9.3, to give two fractions, one comprising the disulfide-linked trimers of chains a, b, and c, and the other, monomeric chain d (20). Ultrogel AcA 44 is formed of beads made of a blend of 4% acrylamide and 4% agarose and has an exclusion limit of ≈200 kDa. Chromatography as described and referenced (8) of the native hemoglobin on diethylaminoethylcellulose (Whatman DE-52) in the presence of diithiothreitol yielded well-resolved peaks corresponding to chains a and b. Although the peaks for chains c and d were unresolved, fractions were isolated that contained single electrophoretic components. Both this chain-d preparation and that obtained from the AcA 44 chromatography were used and the results were averaged. All preparative steps were performed in the presence of CO. The isolated intact hemoglobin, the trimer, and the heme...
containing chains were then dialyzed exhaustively against CO-saturated water. The CO was then replaced with O2 and samples were taken for heme analysis by the pyridine hemochromogen method and for amino acid analysis. Each sample used for heme analysis was aliquoted into three tubes, and the results were averaged. A millimolar extinction coefficient at 557 nm of 32.0 was used for the pyridine hemochromogen. Each sample for amino acid analysis was aliquoted into six tubes, four of which were chosen randomly for analysis. Two sets of hydrolyses, comprising two tubes from each sample, were analyzed separately. The heme was not removed and the hemoglobin was not carboxymethylated. Aliquots were hydrolyzed in 6 M HCl in vacuo at 110°C for 24–29 hr. The amino acid analyses for each sample were averaged. Taurine and norleucine were added to the samples as internal standards prior to hydrolysis. The results with each standard were averaged. The quantity of protein determined was corrected for tryptophan lost during hydrolysis on the basis of the known amino acid sequence data (7, 8).

**RESULTS AND DISCUSSION**

The results, summarized in Table 1, clearly show that all four of the major chains (a, b, c, and d) contain 1 heme per chain. Only chain a shows a sizable discrepancy between analysis and sequence. Chain a is known to lose the first 16 NH2-terminal residues during preparation (8). The uncorrected value for the minimal mass for chain a was 13,866 Da. A correction of 1956 Da was therefore added. The remaining discrepancy of 10% might be explained by further proteolytic degradation, although no evidence for any other proteolysis was found. Such degradation may account for the occasional observation of globin components of 11−13 kDa in NaDodSO4/polyacrylamide gels (17). The close correspondence between the analytical and sequence-derived values for the trimer (disulfide-linked chains a, b, and c) indicates that no measurable proteolytic degradation or heme loss occurred during or prior to preparation of the trimer. A very small loss of heme (1–2%) may occur in the isolated chains. We believe that the results (Table 1) for the native hemoglobin and the trimer are the most accurate because they were subject to the least preparative manipulation. We conclude that our preparation is indeed that of the native hemoglobin because hemoglobin prepared in this way has all the qualitative and quantitative functional properties associated with the hemoglobin (19). Our results clearly show that the earlier conclusion (18) that one of three major chains (a, b, or c) lacks heme is incorrect. The argument had been based in part on A280/A540 absorbance ratios and on the assumption of a similar content of aromatic amino acids in human and earthworm hemoglobins, but sequencing (8) shows that the Trp/Tyr ratios are reversed: 12 Trp/8 Tyr per earthworm abcd unit versus 3 Trp/6 Tyr per human αβ pair. Our data also show that the conclusion (9) that chain a lacks heme cannot be correct.

Chains a, b, c, and d are present in equimolar proportions; the amino acid sequences provide a molecular mass of 69,664 Da for the abcd unit with four hemes. The native hemoglobin assembly is known to consist of 12 subunits. If 4 abcd units were present in each 1/12th subunit, the total molecular mass would be 2,507,904 Da. This value would require about 34% of the molecule to be nonheme protein if all the heme were in the abcd units. None of the NaDodSO4/polyacrylamide gels, nor any of the chromatographic separations, indicates that the abcd chains constitute only two-thirds of the total protein. Furthermore, the value is inconsistent with our analytical determination of the heme/protein ratio. If 5 abcd units were present in each 1/12th subunit, the total molecular mass would be 4,179,840 Da, which is considerably higher than any determinations of the actual molecular mass. We therefore conclude that 4 abcd units must be present in each 1/12th subunit. The molecular mass thus accounted for would be 3,343,872 Da, made up from 192 heme-containing chains, in agreement with the earlier estimate of Wiechelman and Parkhurst (17) and higher than a recent estimate that suggests only 156 hemoglobin sequences (11).

The difference between the observed minimal molecular mass and that calculated for the abcd unit (Table 1) indicates that about 2200 g of non-heme protein is present for each mole of heme-containing chain. Since we have concluded that 192 heme-containing chains are present, the total contribution of non-heme protein to the molecular mass should be ~422 kDa. This brings the total molecular mass to 3.77 MDa, which is close to the reported values of 3.85 MDa (1) and 3.75 MDa (2). The reported carbohydrate content (9, 29) should add about 50 kDa to the estimated 3.35 MDa. The 422,000 g of non-heme protein per mol corresponds to ~35,200 Da for each 1/12th subunit. This suggests that the 31- to 37-kDa subunits of Lumbricus hemoglobin (10) are present in the proportion of one per 1/12th subunit. This proposal is consistent with the conclusion that these subunits are required for assembly (10). Each 1/12th subunit might be anchored to adjacent 1/12th subunits in part by an ~35- ± 3 kDa protein. The fact that more than one kind of 31- to 37-kDa protein is present suggests that all 1/12th subunits may not be identical. Our suggestion is consistent with the proposal (10) that the 31- to 37-kDa subunits might be linkers between complexes of abcd units. A linker is clearly not required for joining the trimer to subunits. The linker molecules spontaneously form a complex with highly cooperative 02-binding properties close to those of the native hemoglobins (19, 20).

The nature of the 31- to 37-kDa proteins (chains D1A, D1B, and D2 in the nomenclature of ref. 23) is uncertain. The reported NH2-terminal sequences of the first 13 residues of DIA and DIB show no obvious relationship with other globins (23). We have extended the sequencing of chain D1A to 28 residues: Ala-Ser-Asp-Pro-Tyr-Gln-Glu-Arg-Arg-Phe-Gln-Tyr-Leu-Val-Lys-Asn-Gln-Asn-Leu-Leu-Ile-Asp-Tyr-Leu-Ala-Lys-Lys-Leu- (unpublished data). Although this sequence does not appear to resemble any known annelid (or other) globin sequence, chains D and the monomer chain d (chain M in the nomenclature of ref. 23) appear to be related, since a monoclonal antibody has been isolated that binds only to these chains and not to the trimer (24).

Our value for the molecular mass of protein associated with each mole of heme is much lower than those of most other studies. Several reasons may exist for this. The values reported previously for L. terrestris hemoglobin vary between 17,000 and 27,200 Da (17, 18). Such huge differences might be explained by heme loss, by the variable removal of non-heme protein, or by a combination of these factors. Electron microscopy has shown that some molecules appear

### Table 1. Comparison of the quantity of protein per mole of heme determined by analysis with that calculated from the amino acid sequences

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Protein, g</th>
<th>Analysis</th>
<th>Sequence</th>
<th>Difference</th>
<th>% deviation*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain a</td>
<td>15,822</td>
<td>17,525</td>
<td>-1703</td>
<td>-10.8</td>
<td></td>
</tr>
<tr>
<td>Chain b</td>
<td>16,553</td>
<td>16,254</td>
<td>299</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>Chain c</td>
<td>17,642</td>
<td>17,289</td>
<td>353</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Chain d</td>
<td>16,341</td>
<td>16,130</td>
<td>211</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Trimer abc</td>
<td>17,130</td>
<td>17,023</td>
<td>127</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Native Hb</td>
<td>18,995+</td>
<td>16,800+</td>
<td>2195</td>
<td>11.6</td>
<td></td>
</tr>
</tbody>
</table>

*Difference/analytical value × 100.

†Not corrected for the unknown tryptophan content of the non-heme protein.

‡Average of the values for chains a, b, c, and d.
to have additional material in the central cavity (25). This material might contribute to the high, variable values obtained for the protein/heme ratio. Our purification procedure might have removed material present in earlier preparations. Studies of human hemoglobin have shown that heme loss can occur provided that the iron is oxidized and no ligand is present (13). Most preparations of *Lumbricus* hemoglobin have not included any protection against methemoglobin formation. Human CO-hemoglobin is particularly stable (14). Our preparation of *Lumbricus* hemoglobin was carried out in the presence of saturating concentrations of CO. The importance of ligand state is indicated by the observation (17) that what pellets in ultracentrifugal purification depends on the ligand state of the hemoglobin. Most estimates of protein quantity have been made by determining the dry weight (18, 26). This procedure requires that all ions, small molecules, and water be completely removed. A common procedure has been to dialyze the protein exhaustively against distilled water (18), but this cannot be expected to be completely effective. For example, dialysis of human hemoglobin against distilled water does not remove any significant quantity of 2,3-bis(phosphoglycerate) (27, 28). We have avoided these problems by determining the amino acid composition of the globins. We believe that use of dry weight combined with some heme loss may have contributed to the measurement of low heme content.

We wish to acknowledge the expert assistance of Sandra Smith of the University of Texas Protein Sequencing Center. The work was supported by National Science Foundation Grant DMB-8502857, Welch Foundation Grant F-213, and National Institutes of Health Grant GM-28410.