Linker Chains of the Gigantic Hemoglobin of the Earthworm *Lumbricus terrestris*: Primary Structures of Linkers L2, L3, and L4 and Analysis of the Connectivity of the Disulfide Bonds in Linker L1

Wen-Yen Kao,[1] Jun Qin,[2] Kenzo Fushitani,[3] Sandra S. Smith,[4] Thomas A. Gorr,[5] Claire K. Riggs,[1] James E. Knapp,[6] Brian T. Chait,[2] and Austen F. Riggs[1]*

[1]*Section of Neurobiology, School of Biological Sciences, and Institute of Cellular and Molecular Biology, University of Texas, Austin, Texas*
[2]*Laboratory for Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, New York*
[3]*Department of Biochemistry, Kawasaki Medical School, Kurashiki, Okayama, Japan*
[4]*Proteomics and Mass Spectrometry Facility, University of Texas, Austin, Texas*
[5]*Institute of Veterinary Physiology, Vetsuisse Faculty and Center for Integrative Human Physiology (CIHP), University of Zurich, Zurich, Switzerland*
[6]*Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts*

ABSTRACT **The extracellular hemoglobin (Hb) of the earthworm, *Lumbricus terrestris*, has four major kinds of globin chains: *a*, *b*, *c*, and *d*, present in equimolar proportions, and additional non-heme, non-globin scaffolding chains called linkers that are required for the calcium-dependent assembly of the full-sized molecule. The amino acid sequences of all four of the globin chains and one of the linkers (L1) have previously been determined. The amino acid sequences via cDNA of each of the three remaining linkers, L2, L3, and L4, have been determined so that the sequences of all constituent polypeptides of the hemoglobin are now known. Each linker has a highly conserved cysteine-rich segment of ~ 40 residues that is homologous with the seven ligand-binding repeats of the human low-density lipoprotein receptor (LDLR). Analysis of linker L1 shows that the connectivity of the three disulfide bonds is exactly the same as in the LDLR ligand-binding repeats. The presence of a calcium-binding site comprising one glutamyl and three aspartyl residues in both the LDLR repeats and in the linkers supports the suggestion that calcium is required for the folding and disulfide connectivity of the linkers as in the LDLR repeats. Linker L2 is markedly heterogeneous and contains unusual glycine-rich sequences near the NH_2-terminus and a polar zipper-like sequence with imperfect repeats of Asp-Asp-His at the carboxyl terminus. Similar Asp-Asp-His repeats have been found in a protein homologous to superoxide dismutase in the hemolymph of certain mussels. These repeats may function as metal-binding sites. Proteins 2006;63:174–187.** © 2006 Wiley-Liss, Inc.

Key words: low-density lipoprotein (LDL) receptor; calcium; scaffolding proteins; polar zipper; extracellular hemoglobin; hexagonal bilayer; annelid; invertebrate

INTRODUCTION

The classic sedimentation studies by Svedberg and Eriksson[3] showed that annelid extracellular hemoglobins (Hbs), erythrocruorins, are gigantic molecules with molar masses of over 3 MDa. Levin[4] and Roche[5] first observed the remarkable two-layered hexagonal images of the molecules by electron microscopy. Levin[4] found that the two layers in the images appeared to be held together by material forming part of a central hole. The axial-linking material of Levin was proposed by Chew and colleagues[6] to be a heme-free polypeptide component which they isolated from the hemoglobin (erythrocruorin) complex from the worm, *Marphysa*. In spite of these suggestive early observations, recognition that unique heme-deficient chains,

Abbreviations: bp, base pairs; DTT, dithiothreitol; Hb, hemoglobin; HPLC, high performance liquid chromatography; IAA, iodoacetamide; ITMS, ion trap mass spectrometry; LDL, low-density lipoprotein; LDLR, low-density lipoprotein receptor; MALDI, matrix assisted laser desorption/ionization; nt, nucleotide; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate

Preliminary accounts of the results have been presented.[1, 2]

The nucleotide sequences reported in this article have been submitted to the GenBank/EBI Data Bank with accession numbers DQ234597for L2, DQ234598 for L3, and DQ234599 for L4.

Grant sponsor: National Science Foundation; Grant numbers: MCB 951179, 972385, 0237651 (to AFR); Grant sponsor: NIH; Grant number; GM 35847(to AFR), NIH grant RR00862 (to BTC); Grant sponsor: New England Affiliate of the American Heart Association (JEK)

W. -Y. Kao's present address is #59, 11F, Guo-Guang Road, Panchiao, Taipei 220, Taiwan, ROC.

J. Qin's present address is the Department of Biochemistry, Baylor College of Medicine, Houston, TX 77030.

*Correspondence to: Austen F. Riggs, Section of Neurobiology, University of Texas, Austin, TX 78712-0252.
E-mail: riggs@uts.cc.utexas.edu

Published online 19 January 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20852

linkers, are important constituents of these Hbs was slow in coming.

Small quantities of components of 24 to 35 kDa, larger than the globin chains (14–17 kDa), were almost always observed but were often regarded as aggregates of smaller chains or as contaminants to be removed. Thus *Arenicola* Hb was reported[7] to have constituents of 13, 14, 26, and 28 kDa by SDS–polyacrylamide gel electrophoresis after reduction with mercaptoethanol but the two larger chains were thought nevertheless to be disulfide-linked dimers of the smaller ones, in spite of the fact that the 26 to 28 kDa bands persisted even after reduction and carboxymethylation or dialysis against 1% SDS and 1% 2-mercaptoethanol. Similarly, Garlick and Riggs[8] also attributed all components with masses greater than that of globin chains to ill-defined aggregation. Shlom and Vinogradov[9] were the first to show the presence of six electrophoretic components as constituents of *Lumbricus terrestris* hemoglobin. They isolated a 51 to 52 kDa subunit (trimer) which could be dissociated by reduction with mercaptoethanol into three unique chains. A fourth chain (monomer) was also isolated. These are the four major heme-containing globin chains, *a, b, c,* and *d*. They also found additional chains of apparent masses 24 to 33 kDa that were considered to be dimers of smaller polypeptides.

The dimer hypothesis persisted for more than a decade.[10] However, Vinogradov and colleagues[11] found crucial evidence for the presence of constituent chains that are larger than globin chains and proposed that these chains, linkers, form an inner bracelet to which the globin subunits are attached. Subsequent studies have shown that these linker chains are non-globin constituents that are required for the assembly of the full-sized two-layered hexagonal molecule.[12,13] The determination of a 5.5 Å crystallographic structure of one form of the Hb definitely establishes the role of the linkers in the globin complex assemblage.[14a] Although the crystal structure clearly establishes a molar mass of ~3.6 MDa, substantial ultracentrifugal and light-scattering data suggest other forms with masses greater than 4.0 MDa.[14b] The basis for this difference remains unresolved. Only a few sequences of linker chains are so far known. Suzuki and colleagues[15–17] have determined the amino acid sequences of four linker chains from different annelid species. The sequences of linkers L1 and L3 from the polychaete, *Sabella spallanzanii,*[18] and linker L1 from the leech, *Macrobdella decora,*[19] are also known. Suzuki and Riggs[20] have previously elucidated the cDNA-derived amino acid sequence of linker L1 of *Lumbricus terrestris*. The present studies complete the determination of the primary structures of all the remaining linker chains, L2, L3, and L4 in the Hb of *Lumbricus terrestris*.

MATERIALS AND METHODS

Protein Analysis

Hemoglobin (Hb) was prepared as previously described.[21] The CO-saturated Hb in 50 m*M* Tris, pH 7.5, 1 m*M* EDTA, was stored at −80°C. NH_2-terminal sequences were determined with a pulsed liquid-phase protein sequencer (Beckman model 477A) of the Proteomics and Mass Spectrometry Facility of the University of Texas at Austin.

Linker L2, isolated from the Hb by HPLC,[21] was digested with trypsin. Approximately 0.5 mg of L2 was dissolved in 0.5 mL of 8*M* urea, 0.4*M* NH_4HCO_3. DTT (2.7 mg) was added and the solution was incubated for ~15 min at 3°C. Iodoacetamide was added (final concentration, ~100 m*M*) and the solution was incubated in the dark (room temperature, 1 h). The sample was then desalted on a C_{18} HPLC column (Äkta, Amersham Pharmacia Biotech, Piscataway, NJ), dried and dissolved in 100 μL 200 m*M* NH_4HCO_3. Trypsin, 10 μg (20 μL, 0.5 μg/μL sequencing grade modified trypsin, Promega, Madison, WI) was added and incubated overnight at 38°C. Additional trypsin (~5 μg) was added and the solution was incubated for an additional 3 h.

Linker L3, isolated from the Hb by native gel electrophoresis or HPLC, was purified by gel electrophoresis (4%–12% Bis-Tris NuPAGE Novex gel, Invitrogen Life Technology, Grand Island, NY). The L3 was reduced (10 m*M* DTT, 30 min, 37°C) and carbamidomethylated with the addition of iodoacetamide (50 m*M*, 1.5 h, 25°C). The reaction was terminated by the addition of excess DTT to 150 m*M* whereupon urea was added to 2*M* and the reaction underwent tryptic digestion (Promega sequencing grade modified trypsin, 1:20 w/w enzyme to protein, 37°C, 16 h). Peptides were recovered from the solution by solid-phase extraction (ZipTipμC18, Millipore Corp., Billerica, MA) according to manufacturer's protocol. Peptide sequence confirmation was made using tandem mass spectral analysis in parallel experiments on a Bruker Esquire-LC and a Thermo Electron LCQ (classic) with capillary chromatography as well as an Applied Biosystems 4700 Proteomics Analyzer (MALDI-TOF/TOF).

Isolation of mRNA and cDNA Synthesis

A FastTrack mRNA Isolation Kit (Invitrogen, San Diego, CA) was used to isolate Poly(A)$^+$ RNA) from 520 mg chloragogen cells of *L. terrestris* prepared previously[22] and stored at −80°C. The first-strand cDNA was synthesized from an aliquot of the mRNA preparation primed with 1 μL of random decaprimer (DECAprime, Ambion, Austin, TX) with 2-units of M-MuLV reverse transcriptase (U.S. Biochemical, Cleveland, OH). The resulting cDNA was ethanol precipitated in 0.3*M* sodium acetate, washed once with 70% ethanol, freeze-dried, redissolved in 100 μL sterile H_20, and stored at −80°C.

Amplification and Sequencing of cDNA for L2

The strategy for obtaining the sequence is given in Figure 1. Primers based on the previously reported NH_2-terminal sequence failed to amplify the L2 cDNA. Oligomer, #L2K5 (nt 270–nt 292, Table I), synthesized on the basis of an internal peptide, KIDPEHFV (data not shown), was used as a primer for amplification by PCR. The 100-μL reaction mixture contained two units of Taq polymerase (Promega) in the 1× buffer supplied, 2.0 m*M* $MgCl_2$, and 10 μg of acetylated bovine serum albumin (New England Biolabs, Beverly, MA). The amplification with the MJ

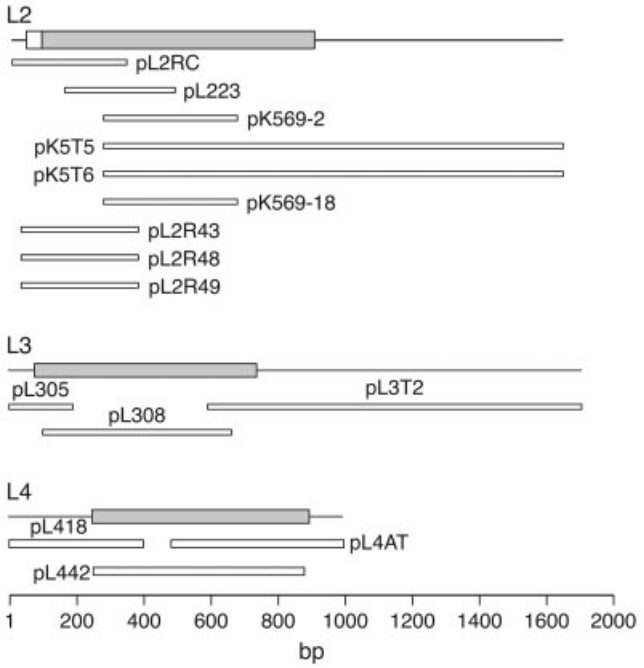

Fig. 1. Strategy for the determination of the nucleotide sequences of the cDNA for linkers L2, L3, and L4.

Minicycler (MJ Research, Waltham, MA) was performed for 35 cycles (94°C × 45 s/50°C × 45 s/72°C × 1 min). Electrophoresis of the products in a 2%-agarose gel in 1 × TAE buffer showed a 0.4 kb product which was cloned into the pUC19 *Sma*I site as pK569 and sequenced completely by the dideoxy method using Sequenase (USB, Cleveland, OH).

The 3′ flanking sequence of L2 was obtained by amplification as described above with oligomer L2K5 and oligo-dT (Xba-Primer Adapter, Promega, Madison WI) and cloned into the pUC19 *Sma*I site. Two independent clones were obtained from this amplification, pK5T5 and pK5T6, each of which was sequenced completely. The sequences correspond to the last 195 residues of the polypeptide plus 747 bp of untranslated 3′ sequence.

The 5′ flanking sequence of L2 was first obtained from a previously constructed *Lumbricus* λgt10 cDNA library[22] by using a λgt10 reverse primer and oligomer #L2C5 (nt 461–nt 485, Table I). The PCR amplification protocol was the same as used above. The largest of seven fragments observed on a 2% agarose gel was inserted into the pUC19 *Sma*I site as pL223, cloned and sequenced completely. The sequence was confirmed and extended by using the 5′ RACE System (Life Technologies, Gaithersburg, MD) with two gene-specific primers, GSP-1 (nt 356–nt 378, Table I) and GSP-2 (nt 325–nt 347, Table I), for PCR amplification in conjunction with the anchor primer supplied with the 5′ RACE System kit. Each amplification was performed in a 50 μL reaction volume with 2 units of Taq polymerase (Promega, Madison, WI), 20 m*M* $MgCl_2$, and 5 μg acetylated bovine serum albumin (New England Biolabs, Beverly, MA) in 1× buffer A supplied with the kit. The amplification protocol used 42 cycles (94°C × 30 s/55°C × 45 s/72°C × 1 min) plus an accumulative 2-s extension per cycle. The 372-bp fragment obtained was cloned as pL2RC and sequenced. Analysis of the sequence showed that it includes 41 nt of 5′ untranslated sequence followed by the Met start codon and a 15-residue signal sequence. The following 6 residues, DHHQPS, correspond exactly to the NH_2-terminal sequence found by Fushitani and colleagues.[23] The NH_2-terminal sequence obtained by Ownby and colleagues[21] corresponds to residues 26 to 40 and is evidently the result of the proteolytic removal of the first 25 residues. (Oxidation of the hemes is accompanied by partial dissociation of globin subunits and the exposure of bonds sensitive to proteolysis.[12])

Further confirmation of the 5′ end was obtained by PCR amplification with a 20-mer sense oligomer, #L2RC1 (nt 23–nt 42, Table I), together with GSP-1 with the cDNA target (94°C × 30 s/60°C × 45 s/72°C × 1 min) for 30 cycles. The single fragment of 350 bp was cloned (pL2R43, pL2R48, and pL2R49) and sequenced completely.

Amplification and Sequencing of cDNA for L3

The strategy for the sequence determination is shown in Figure 1. Two oligomers, #L3M1B (nt 96–nt 116) and #L3M7R (nt 639–nt 660, Table I), based on L3 peptide sequences HDEIIDK and EFDGYNF, respectively, from Fushitani and colleagues,[23] were used to amplify part of the coding region of the cDNA for L3. The cDNA was amplified for 35 cycles (94°C × 45 s/50°C × 45 s/72°C × 1 min). The resulting 562-bp cDNA fragment was inserted into the *Sma*I site of pUC19, cloned as pL308 and sequenced.

The 5′ flanking sequence of L3 was obtained from a λgt10 cDNA library constructed previously[22] by using a λgt10 forward primer and oligomer #L3LDDR (nt 173–nt 192, Table I), based on the cDNA sequence of pL308 corresponding to the amino acid sequence LDDRLDP of peptide L3a.[23] The amplification reaction conditions were 94°C × 45 s/50°C × 45 s/72°C × 1 min for 35 cycles. Two resulting fragments were inserted into the *Sma*I site of pUC19, cloned as pL305 (nt 1–nt 192) and pL320(nt 51–nt 191) separately, and sequenced. The sequences of the three clones agreed completely. pL305 includes 14 bp of 5′ untranslated sequence, the signal peptide sequence (nt 13 – nt 74), and 117 bp corresponding to the first 39 residues at the NH_2-terminus of the mature L3 polypeptide. The prominent cysteine-rich segment (residues 60 – 96) is very similar to those of L1[15] and L2 (see above), except that it is two residues shorter, with the last two cysteines separated by 8 residues rather than 10.

The 3′ flanking sequence of L3 was obtained by amplification with oligo dT (*Xba*I-primer Adapter, Promega) paired with a nonredundant oligomer #L3ADHR (nt 584–nt 602), based on the cDNA sequence of pL308 that codes for the amino acid sequence ADHRLTI in L3 peptide d.[23] Amplification (94°C × 45 s/50°C × 45 s/72°C × 1 min) for 35 cycles produced a 1.1-kb DNA fragment that was cloned in the *Sma*I site as pL3T2 and sequenced.

TABLE I. List of Oligomers

L2	
#L2K5	5′.AARATHGAYCCNGARCAYTTYGT.3′(384-fold redundant)
oligo-dT	5′.GTCGACTCTAGATTTTTTTTTTTTTTT.3′
#L2C5	5′.GAAGACGTTTCCTGCTTTGACCAC.3′
λgt10 Reverse	5′.GGTGGCTTATGAGTATTTCTTCC.3′
Anchor Primer	5′.CUACUACUACUAGGCCACGCGTCGACTAGTACGGGIGGGIIGGGIIG.3′
GSP-1: #L2-221	5′.ACTCCTGTTCGTTGCCTCCGCAC.3′
GSP-2: #L2-190	5′.TCTCTTTTCACAGTGGGTCCCTT.3′
#L2-RC1	5′.GCGCGAGTATACGTTAAGCA.3′
L3	
#L3M1B	5′.CAYGAYGARATHATHGAYAAR.3′(288-fold redundant)
#L3M7R	5′.CTYAARCTRCCNATRTTRAAR.3′(256-fold redundant)
#L3ADHR	5′.CGACCACCGGTTGACCATC.3′
oligo-dT	5′.GTCGACTCTAGATTTTTTTTTTTTTTT.3′
#L3LDDR	5′.TTGGGTCAAGGCGATCGTCT.3′
λgt10 Forward	5′.AGCAAGTTCAGCCTGG.3′
L4	
#L4M1	5′.GCNGCNGARGARGAYAAYMG.3′(512-fold redundant)
#L4M5	5′.GCRCANGGSAGNCCNGTNCC.3′(1024-fold redundant)
#L4(230)F	5′.TGCGATGGAATCACAGATTGC.3′
#L4KARR	5′.ATTGCATCTACGCGAGCCTT.3′
λgt10 Forward	5′.AGCAAGTTCAGCCTGG.3′
λgt10 Reverse	5′.GGTGGCTTATGAGTATTTCTTCC.3′

N: A/G/C/T; R: A/G; Y: C/T; H: A/T/C; M: A/C

Amplification and Sequencing of cDNA for L4

The strategy for obtaining the sequence of the L4 cDNA is shown in Figure 1. Two redundant primers were used: #L4M1 (nt 248–nt 267) and #L4M5 (nt 848–nt 867, Table I), corresponding to the NH_2-terminal sequence AAEEDNR and the internal peptide GTGLPCA,[23] respectively. Amplification for 30 cycles (94°C × 45 s/55°C × 45 s/72°C × 1 min) resulted in a 620-bp fragment that was cloned in the *Sma*I site of pUC19 as pL442 and sequenced.

The 5′ flanking sequence of the L4 cDNA was obtained from a previously constructed λgt10 cDNA library[22] by using oligomer #L4KARR (nt 380–nt 399) with λgt10 forward or reverse primers for the amplification of L4 cDNA (94°C × 45 s/50°C × 45 s/72°C × 1 min, for 35 cycles). Two fragments, 199 and 399 bp, produced from the forward and reverse primer reactions were cloned in the pUC19 *Sma*I site as pL417 and pL418, respectively, and sequenced.

The 3′ flanking sequence was obtained by amplification with a nonredundant oligomer #L4(230)F (nt 476–nt 496, Table I) paired with oligo-dT (*XbaI*-primer Adapter, Promega, Madison, WI) in a 100-μL reaction mixture with Taq polymerase. The 35-cycle amplification (94°C × 45 s/50°C × 45 s/72°C × 1 min) produced a 533-bp product which was cloned in the *Sma*I site of pUC19 as pL4AT and sequenced.

Sequence Analysis

Sequence alignment using ClustalW[24] was done online at the NPS site (Network Protein Sequence Analysis, http://pbil.ibcp.fr)[25] or in MacVector 6.5. Sequence homology searches with the BLASTp program[26] used the NCBI server (http://www.ncbi.nlm.nih.gov/BLAST or http://www.ebi.ac.uk/blastall/index.html).

Disulfide Connectivity

Linker chain L1, purified by HPLC as previously described,[21] was digested with endoproteinase Asp-N at a high enzyme to protein ratio of 1:1 (w/w). The high enzyme-to-protein ratio was crucial to ensure complete digestion of the compact, closely connected disulfide-linked polypeptide. The total digest was measured by MALDI spectrometry as described[27] and applied directly to the ion-trap mass spectrometer[28] without chromatographic purification.

Modeling of LDLR-Like Sequences

The structures of the LDLR-like segments of the linker chains were approximated by superimposing the sequence of each linker chain onto the backbone coordinates of the ligand-binding repeat #5 module[29] from the human LDL receptor (PDB code: 1AJJ), using program 0.[30] Side-chains were then adjusted to remove bad contacts with other groups and to improve the geometry as judged by Procheck.[31] All four linker chains include a two-residue insertion in a loop that precedes the third cysteine of the LDLR motif. Linker chains L1 and L2 have an additional two-residue insertion before the last cysteine residue of the motif. The insertions were added to the model on the basis of the sequence alignment with care taken to avoid alteration of the positions of the conserved disulfide bonds and the positions of the residues that form the conserved calcium-binding site. The first insertion had almost no effect on the overall structure whereas the second insertion in linkers L1 and L2 resulted in a reorientation of the last disulfide bond. The surface potentials of each LDLR-like sequence in the linkers were calculated with the program GRASP.[32]

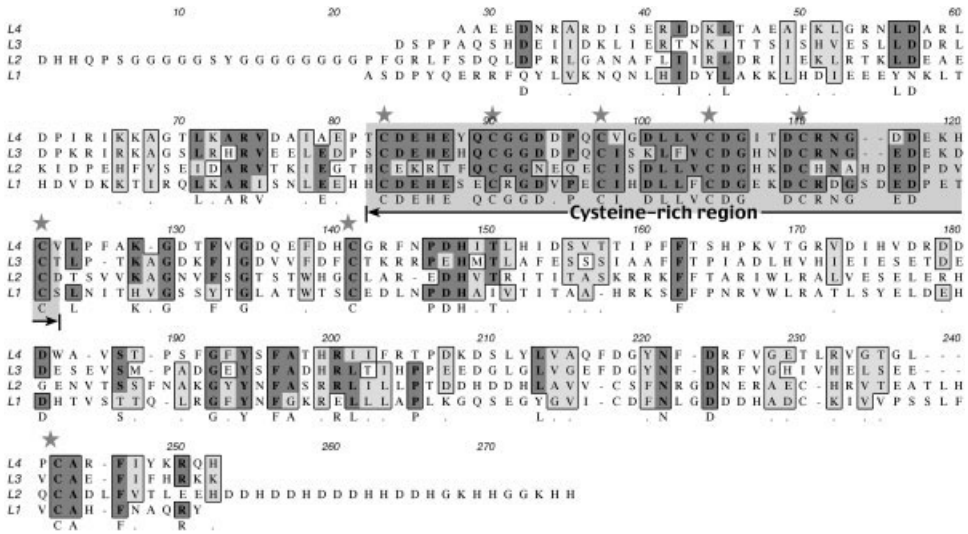

Fig. 2. The cDNA-derived amino acid sequences for linkers L1, L2, L3, and L4.

RESULTS AND DISCUSSION

Amino Acid Sequences

Sequences have now been determined for all eight polypeptide constituents of the Hb of *L. terrestris*: globin chains *a*, *b,* and *c,*[33] chain *d,*[34, 35] linkers L1,[20] and L2, L3, and L4 (present work). Each of the four linker sequences contain three domains which are based in part on sequence homology and on the 5.5 Å crystal structure.[14] Figure 9(a) shows that each linker chain includes a coiled-coil helical domain, a middle LDLR-like domain, and a C-terminal domain. The middle domain of each linker chain has an LDLR-like fold that contains all three of the disulfide bonds that were initially proposed from the correspondence of the cysteine-rich segment of the *Lumbricus* L1 linker chain with the LDLR domain.[20] The middle LDLR domain interacts both with its own C-terminal domain and with the C-terminal domain of a threefold related subunit within the trimeric linker assemblage. Each LDLR-like domain within the trimeric assemblage interacts with a globin *b* subunit such that the quasi-threefold symmetry is maintained. The C-terminal domain has more extensive interactions with the globin dodecamer, containing the *a, b,* and *c* globin components. A detailed description of each linker sequence is given below.

Linker L2

The amino acid sequence of L2 (Fig. 2), derived from one cDNA clone (#1), has 272 residues and a mass of 30,330.2 Da after subtracting the 10 hydrogens removed in forming five disulfide bonds. However, there is substantial evidence for heterogeneity. Analysis of three additional cDNA clones show that the glycine-rich NH_2-terminal segment (residues 7–21) is heterogeneous. The cDNA clones showed three different sequence patterns in which two strings of 4 to 8 glycines are separated by Ser-Tyr as follows:

$$(G)_4\,SY(G)_6,\ (G)_4SY(G)_8,\quad \text{and}$$
$$(G)_5\,SY(G)_8\ (\text{Clone \#1}). \quad (1)$$

A BLASTp search for such sequences located several proteins with Ser-Tyr between strings of glycines. Examples are the eukaryotic initiation factor 4B from *Arabidopsis* and cytoplasmic keratin complexes from the mouse. In addition, four proteins were identified with Tyr alone separating strings of glycines. Possible functions for this curious motif are unknown.

Additional heterogeneity was found elsewhere in the molecule. The sequencing of pK5T6 shows Lys → Ala, Ser → Gly, and Leu → Ser at positions 77, 197, and 212, respectively, compared with pK5T5.

Amino acid sequencing of the NH_2-terminal segment proved difficult because ~90% of the NH_2 termini are blocked. However, sequencing of the remaining unblocked chains showed additional heterogeneity: a Gly → Ala substitution at position #7 in one preparation of Hb and both Ala *and* Gly at positions #7 and #11 in a second preparation. In addition, the amino acid composition of a Lys-C peptide (data not shown), corresponding to residues 1 to 51, showed 1.9 additional Ala and 1.5 fewer Gly compared with the sequence shown in Figure 2. Mass spectrometric MALDI analysis of tryptic peptides of the major chromatographic peak of L2 (Ref. 36, Fig. 2, fraction 5 excluding the shoulder) shows that the major component has a mass of 2168.9 ± 0.1 Da (2169.2 Da expected for residues 1–25 of L2; Fig. 2). MALDI–MS also showed peaks at 2183.3 and 2225.9 Da, consistent with one Gly → Ala substitution without an acetylated NH_2-terminus (2183.3 Da calculated and 2225.9 Da with acetylation). This result is consistent with an acetyl group at the NH_2-termini of most molecules of L2.

A second unique feature of the L2 sequence is the presence of a polar zipperlike 26-residue C-terminal sequence with imperfect repeats of DDH (residues 250–275 in Fig. 2). (The imperfect repeats of Asp-Asp-His in L2 can be compared with the polar zipper repeat sequence described for *Ascaris* hemoglobin, Glu-Glu-His-Lys, for which a purely structural role has been proposed.[37]) This se-

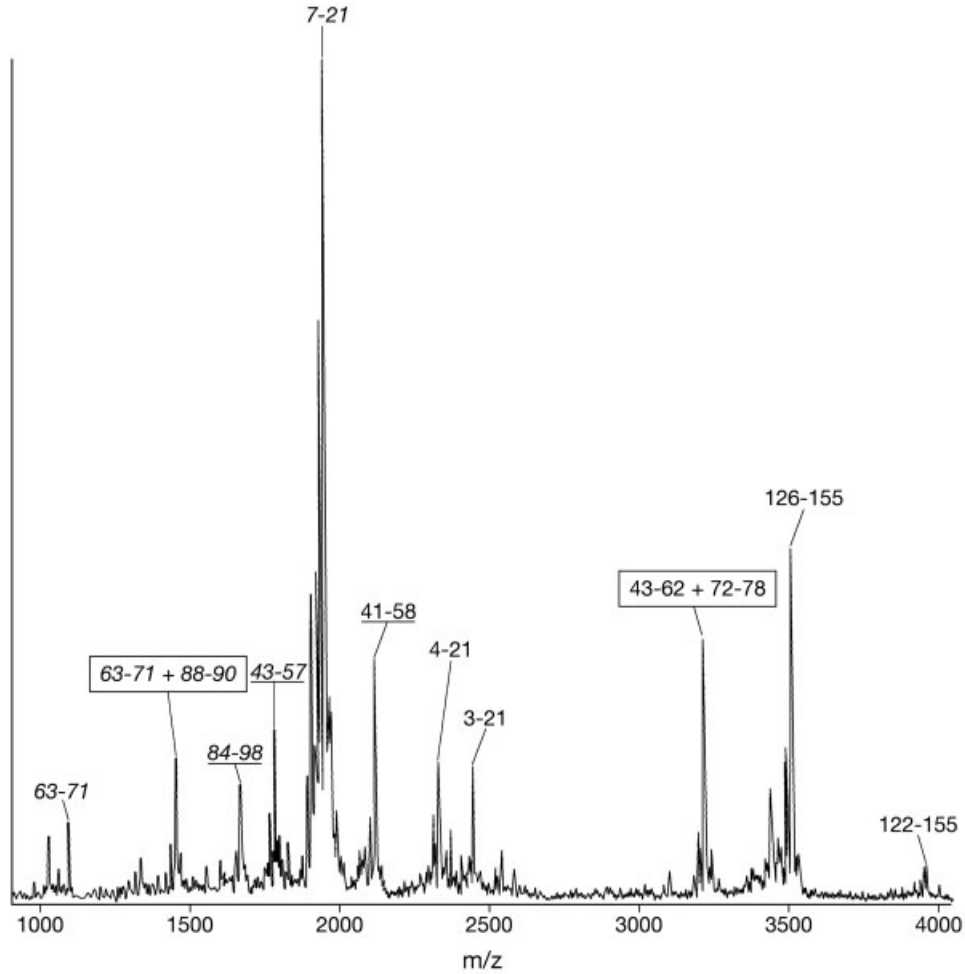

Fig. 3. Mass spectrum of the Asp-N digest of the linker chain L1 at a high enzyme to protein ratio (1:1 w/w). Boxed labels represent the disulfide-containing peptides. Underlined labels represent fragments produced by cleavage at Glu residues. The cleavages at Glu result from the very high enzyme-to-substrate ratio.

quence also occurs as the C-terminal sequence of L2 sequences of the Hbs of *Tylorrhynchus*[16] and *Neanthes.*[15] Curiously, a similar motif, DHH, is the NH_2-terminus of L2 in *Lumbricus* and of L2 sequences in both *Neanthes* and *Tylorrhynchus* which begin with DD.

Linker L3

The translated amino acid sequence of the mature L3 (Fig. 2) shows that it has 220 residues and a calculated polypeptide mass of 24,913.4 Da. This calculated mass is on the basis of Val and Phe at positions 111 and114, respectively. These identifications are based on the sequence of the corresponding peptide determined earlier by Fushitani and colleagues.[23] The cDNA sequencing was ambiguous for position 111 but gave a clear Val in sequencing the peptide. The cDNA gave a clear His at position 114 indicating heterogeneity. The close correspondence of the calculated and observed masses suggests that the peptide with 114His is a minor component. The sequence was further confirmed by tandem mass spectral analysis (see Materials and Methods). An alkylated tryptic peptide, corresponding to residues 106-117(FIGDVVFDFCTK), had an observed mass of 1447.73 Da by MS (1447.69 Da expected) and the sequence was verified by MS/MS.

Linker L4

The translated amino acid sequence (Fig. 2) has a 24-residue signal sequence and a mature L4 polypeptide of 215 residues and a calculated polypeptide mass of 24,248.0 Da. The translated 20-residue NH_2-terminal sequence corresponds exactly to that determined directly by protein analysis.[23]

Cysteine-Rich Segment

Disulfide connectivity of linker L1

Linker L1 was digested with Asp-N protease at the high enzyme to protein ratio of 1:1 (w/w) (see Materials and Methods). Figure 3 shows the spectrum obtained from the mixture, and Table II lists the mass assignments. The peak of m/z 1452.5 can be assigned to the peptides comprising residues 63 to 71and 88 to 90 in which C69 is connected to C89. The peak at m/z 1061.1 can be assigned

TABLE II. Mass Assignments for Peptides after Asp-N Digestion of Linker Chain L1

Measured Mass (Da)[a]	Calculated Mass (Da)	Sequence Assignment
1060.1	1060.1	64–71
1451.5	1451.5	63–71 + 88–90
1666.9	1666.6	84–98
1784.0	1784.1	43–57
1958.2	1958.3	7–21
2127.3	2127.5	41–58
2345.4	2346.7	4–21
2418.9	2418.8	43–62
2461.6	2461.8	3–21
3229.5	3229.7	43–62 + 72–78
3514.2	3514.1	126–155
3953.7	3953.5	122–155

[a]Values shown are for the unprotonated masses.

to sequence 64 to 71. It is evident from the triplet peak around m/z 1061.1 that this group of peaks arises from the dissociation of a disulfide bond, which is consistent with the disulfide assignment as C69:C89. Another peak at m/z 3230.5 can be attributed to sequence (43–62) + (72–78), but the peak corresponding to the mass of the sequence of 43 to 62 is too weak to permit a definitive assignment.

The putative assignment of the peak at m/z 3230.5 as (43–62) + (72–78) was tested by performing tandem ion-trap mass spectrometry on this ion. The MS/MS spectrum is shown in Figure 4. All the observed fragments can be assigned to preferential cleavage of the amino terminal Asp residues[38] and to preferential cleavage of the disulfide bond.[39] A peak at m/z 2419.2 is observed that matches the sequence 43 to 62. An intense group of peaks is also observed to cluster around the ion that corresponds to (43–62) − Asp. The characteristic triplet of peaks centered

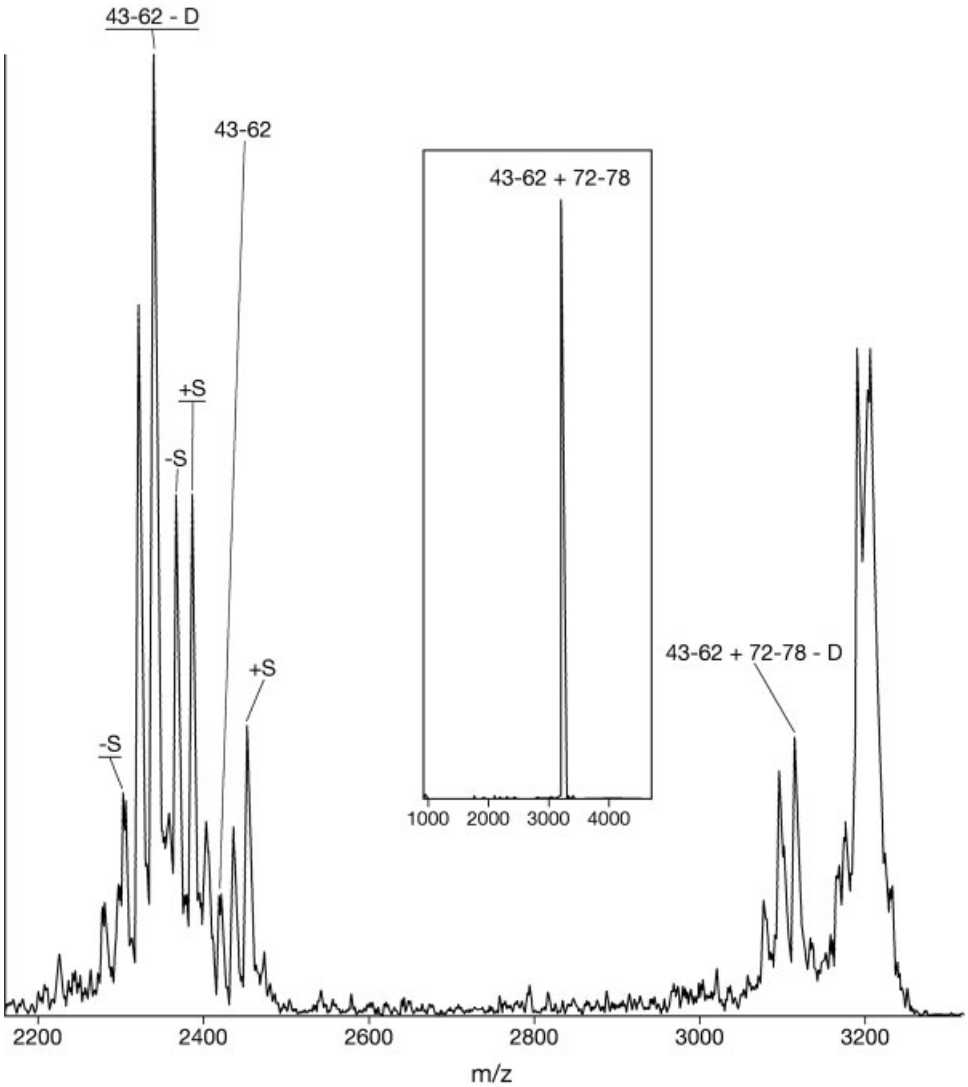

Fig. 4. MS/MS spectrum of the peptide ions corresponding to the sequence of peptides (43–62) + (72–68) (shown in the insert). Only two types of dominant dissociation channels are observed — cleavage at the disulfide linkage and at the C-terminal of Asp. The symbols, S and D, stand for sulfur and aspartic acid.

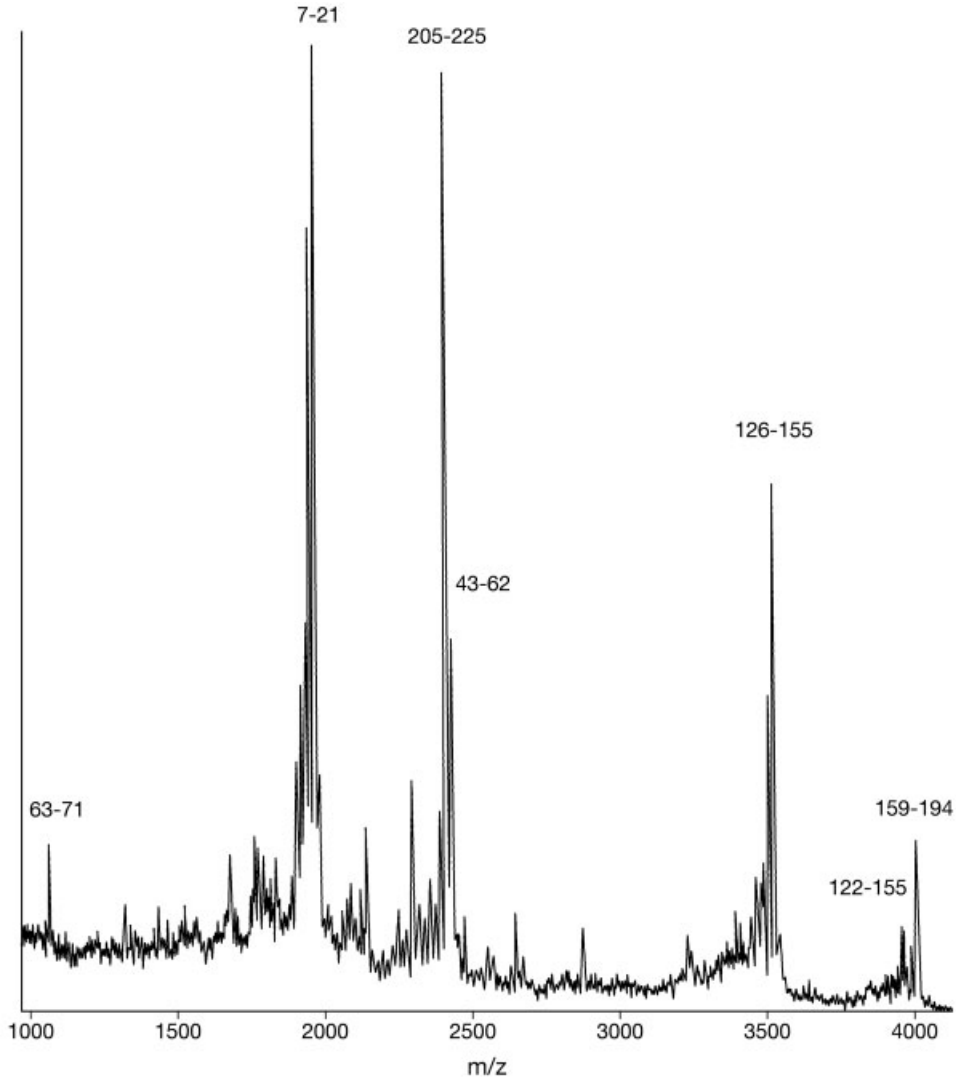

Fig. 5. MS spectrum of the Asp-N digestion products of the linker chain L1 (as shown in Fig. 4) after reduction with DTT.

on (43–62) − Asp confirm the existence of a disulfide bond in this peptide. Thus, we can assign another disulfide bond to C62:C76.

As a further test of these disulfide assignments, the total Asp-N digest was reduced with DTT. The resulting spectrum is shown in Figure 5 with the mass assignments listed in Table III. Peaks corresponding to sequences (63–71) + (88–90) and 43 to 62 and 72 to 78 disappear from the spectrum whereas peaks corresponding to sequences 63 to 71 and 43 to 62 appear (compare Fig. 5 with Fig. 3). This observation confirms the disulfide bond assignments. Two additional peaks are observed at m/z 2397.8 and 4000.6 that presumably arise from the reduction of C-terminal polypeptides that contain disulfide bonds. The intact disulfide-bonded C-terminal polypeptide is outside the mass range of the ion-trap mass spectrometer.

We summarize the disulfide linkages in the linker L1 in Figure 6. Two disulfide links can be definitely established as C62:C76 and C69:C89. If we assume that the third disulfide linkage is internal to the region of homology with the LDL receptor repeats, then another disulfide bond can be inferred as C83:C100. The power of disulfide mapping by MALDI-ITMS is apparent because these experiments

TABLE III. Mass Assignments of Peptides Reduced with Dithiothreitol from Asp-N Digestion of Linker Chain L1

Measures Mass (Da)[a]	Calculated Mass (Da)	Sequence Assignment
1061.1	1061.1	63–71
1957.6	1958.2	7–21
2396.8	2397.8	205–225
2418.3	2419.8	43–62
3513.5	3514.0	126–155
3953.7	3953.5	122–155
3999.6	3999.6	159–194

[a]Values shown are for the unprotonated masses.

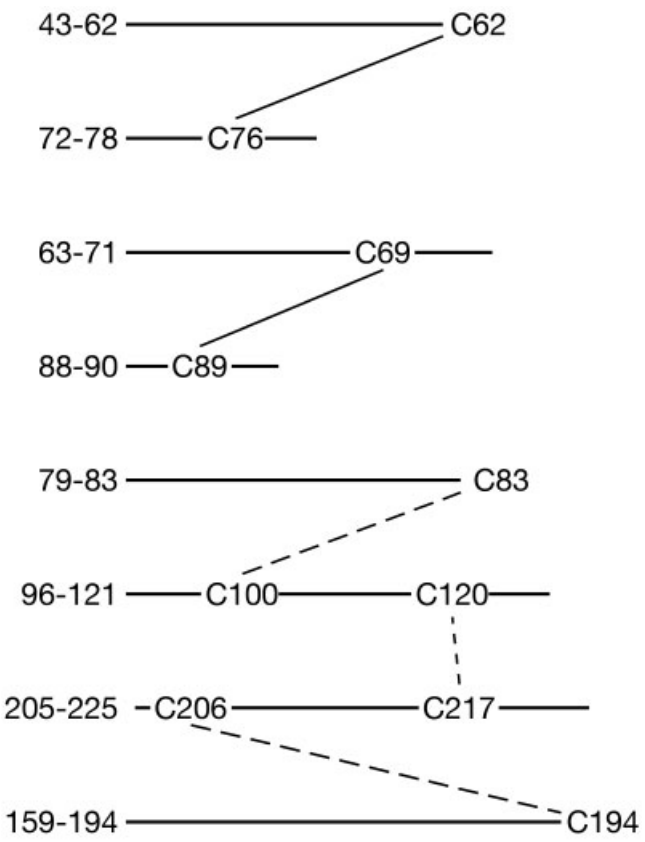

Fig. 6. Summary of the disulfide linkage determined by MALDI-ITMS in linker chain L1. The dashed lines indicate inferred linkages. The C83:C100 linkage was inferred from the homology with the LDL receptor.

reduced the number of possible disulfide linkages from 945 to 2. The C62:C76 and C69:C89 disulfide assignments, together with the inferred C83:C100 disulfide, is identical to the connectivity found for the cysteine-rich ligand-binding domain of the LDL receptor.[40]

Modeling of the cysteine-rich segment

The linker chain sequences reveal that each linker has a highly conserved cysteine-rich segment of ~40 residues (Fig. 7) that is homologous with the ligand-binding domains in the human low-density lipoprotein receptor (LDLR).[41] This relationship, first found for L1,[20] is shown to include the connectivity of the three disulfide bonds. Although the connectivity has been determined only for L1, we conclude from the high degree of correspondence that the LDLR-like segments of all linkers have the same disulfide connectivity. Each ligand-binding repeat of the LDL receptor contains an absolutely conserved calcium binding site comprising one glutamyl and three aspartyl residues which are also conserved in the linker sequences. These residues in the LDLR repeats confer a negative charge on the surface for binding apo-lipoprotein T3.[42] The much greater net negative charge in the linker-chain LDLR-like segments should enable binding of these polypeptides to positively charged residues of globin chains. Suzuki and Riggs[20] previously suggested that the NH_2-terminal segment of globin chain *b* may be such a binding site because it has five positively charged residues in exactly the same positions as those in helix 4 of apolipoprotein E known to be involved in the LDL particle-receptor binding.[43,44] Interaction of chain *b* with the cysteine-rich module has been confirmed by X-ray crystallography of *Lumbricus* Hb (W. E. Royer, personal communication).

We have generated models of the four LDLR-like segments of the linker chains to help us to understand the role played in the assembly of the hemoglobin. Models were made by using the crystal structure of ligand-binding domain five of the human LDLR.[29] These models were then used to determine the surface charge distribution of the LDLR domain from each linker chain. The surfaces of the four LDLR domains are predominantly negative (Fig. 8). The LDLR-like domain of L1 is the most negative whereas those of L3 and L4 are the least negative. The close sequence similarity between the L3 and L4 LDLR-like segments is reflected in the near identity of the surface charge patterns on both the front and back sides of the domains (Fig. 8). (The close similarity of L3 and L4 suggests polymorphism. Although the structure of *Lumbricus* Hb has frequently been described as having both these linkers no evidence exists that each normally occurs in *individual* molecules. L3 and L4 appear to be interchangeable alternatives: alleles or polymorphs.[36] The relative proportions of the different linkers vary in individual worms.[21] Reassembly studies have shown that the four linkers can replace one another and that at least two different kinds of linkers may be required.[45] These studies do not specify, however, which pattern of linkers may be kinetically optimal in assembly and function or favorable for long-term stability.) The electrostatic surface potential of the LDLR-like domain of L2 shows a dramatic difference in charge distribution when compared with those of the other three LDLR-like domains. The top half of the surface of both the front and back of the L2 domain is densely packed with negative charges as it is in that of L1. However, the bottom half of the surface of the L2 segment includes a strong positively charged patch due to Lys at position 9 and Arg at position 10 of the LDLR-like domain sequence using the residue numbering shown in Figure 7. These two residues appear as a V-shaped protrusion when the surface of L2 is viewed from the front or back.

In contrast to the rest of the molecule, the top half of the LDLR-like domains are remarkably similar when viewed from the side. From this view, the top right corner of each linker chain domain is characterized by a large patch of negative charge that is located above a set of conserved, hydrophobic residues. This negatively charged patch arises from residues that form the calcium-binding site as indicated by the black arrows in Figure 8. The calcium-binding site includes the main-chain carbonyl oxygen atoms from residues 26 and 31 as well as the side-chain oxygen atoms of Asp 29, Asp 33, Asp 39, and Glu 40. These four acidic residues coordinate the calcium atom and are invariant in the LDLR-like modules of all known linker chain sequences. Although the hydrophobic patch located below

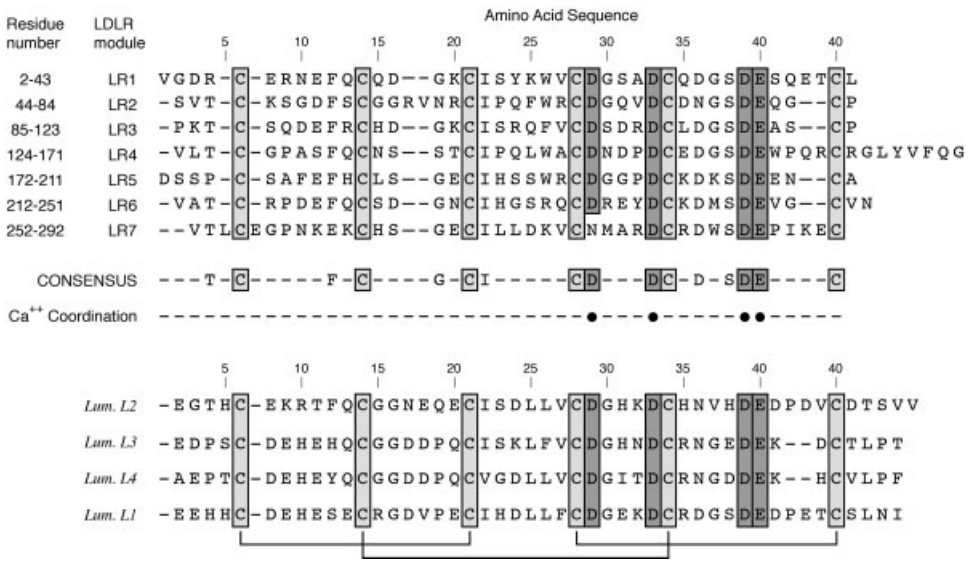

Fig. 7. Sequences of the cysteine-rich LDLR domains and the corresponding segments of linkers L1, L2, L3, and L4.

the calcium-binding site is highly conserved in annelid linker chains it varies in other LDLR-like modules. This patch originates around Leu 25, which forms a dumbbell-shaped protrusion present in the middle of the surface of each linker chain (Fig. 8). Leu 25 is followed by two hydrophobic residues — a Leu or Phe at position 26 and by Val or Leu at position 27.

Structural Role of Linkers

The crystal structure at 5.5 Å resolution[14] shows that the linker complex forms heterotrimers. The arrangement of the linker chains in the whole Hb molecule is diagrammed in Figure 9. The trimer complex of linkers forms a concave surface to which a dodecameric complex of the *abc trimer* and *d* Hb subunits bind. The trimeric core of the linker complex is stabilized by a coiled-coil motif found in the N-terminal end of each linker chain, and by small interactions between adjacent C-terminal domains. The low-resolution structure shows that Leu 25 (Fig. 7) of each linker is part of the interaction between the LDLR-like domain of each linker and the globin *b* subunit (W. E. Royer, personal communication). This interaction appears to be important for the assembly of this complex because Leu 25 is present in all four *Lumbricus* linker sequences (and in all three *b*-subunit LDLR-like domain interactions) but is absent in the corresponding position in the human LDLR domains (Fig. 7).

Calcium Binding

Calcium is required for the folding and maintenance of the structural integrity of the LDLR domain in which a single, unique set of disulfide bonds forms only in the presence of calcium.[46] The exact correspondence of the calcium-binding residues and the disulfide connectivity in the LDLR repeats to those in the linker domains (Fig. 7) suggests that the cysteine-rich segment of the linker chains also requires calcium for folding and establishing the same disulfide connectivity. The presence of calcium at the proposed site in the linkers has recently been confirmed by a new 3.4 Å resolution structure (W. E. Royer, manuscript in preparation). A single Ca^{2+} ion is buried under Asp 29 and Asp 33 on the surface of each LDLR-like domain. This part of each LDLR-like surface together with a hydrophobic patch shown in Figure 8 interacts with the *b* Hb subunits of the dodecameric Hb complex.

Calcium binding, together with the disulfide connectivity, dictates a relatively rigid surface architecture. Removal of calcium or reduction of linker chain disulfide bonds would be expected to cause substantial conformational changes as it does in the LDLR repeats and to eventually disrupt the assembly of the entire Hb complex. Thus the use of dithionite to reduce ferric hemes may compromise the structural integrity of the Hb because dithionite can also reduce disulfide bonds.[47]

The binding of a single Ca^{2+} ion within each of the linker chains results in a total of 36 linker Ca^{2+} ions in the Hb complex. Calcium is also required for the assembly of the hexagonal bilayer structure of the Hb and as an allosteric factor in oxygenation. Although the allosteric effect of calcium is shown by the isolated *abc* trimeric Hb subunit of the Hb complex,[48] the identification of additional specific assembly sites has not yet been made.

The human LDLR has seven slightly different ligand-binding repeats, different combinations of which permit the receptor to recognize and bind the quite different apoB and apoE lipoprotein ligands.[49] This suggests that the combination of the different linkers (L1–L4) may optimize interactions with different globin interfaces in the assembly of the Hb. Although some of the linkers may be interchangeable[13] this does not necessarily mean that interchanging linkers will lead to the same energetic stability.

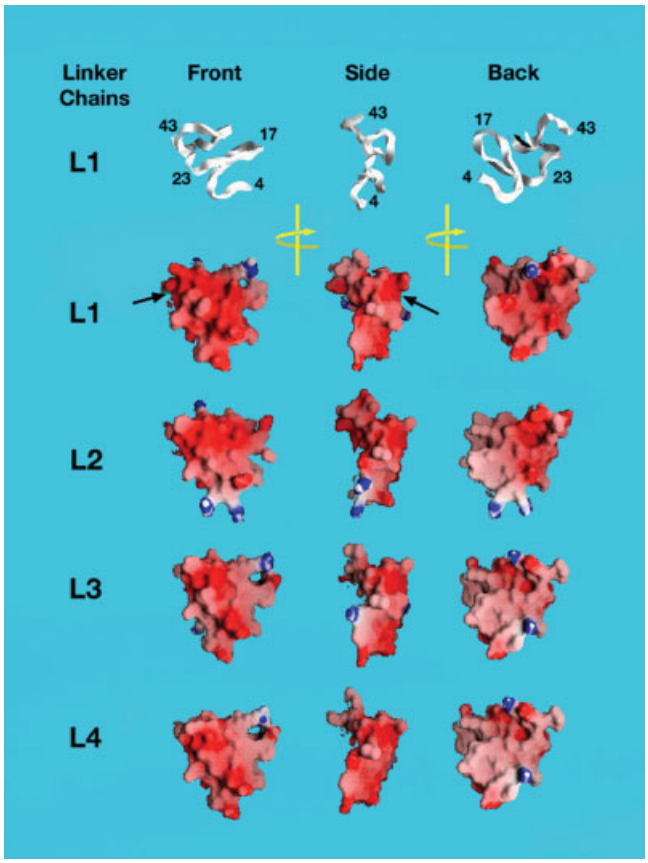

Fig. 8. A ribbon diagram of the LDLR-like domain from linker L1 is shown in three orientations in the top row. The corresponding views of the electrostatic potentials of the surfaces of the four linker chains are shown underneath. Each representation shows the front, side, and back views of the model structure with each view being related to its neighbor by a 90° rotation around the vertical axis. Numbers on the ribbon diagram give the positions of the residue as shown in Figure 7. Negative, neutral, and positive potentials of each surface are shown in red, white, and blue, respectively. The acidic and basic potentials are contoured at −12 kT and 10 kT, respectively. The black arrows point to the calcium-binding site found in the human LDLR which is probably also present in the LDLR-like domains in each of the four linker chains of *Lumbricus*.

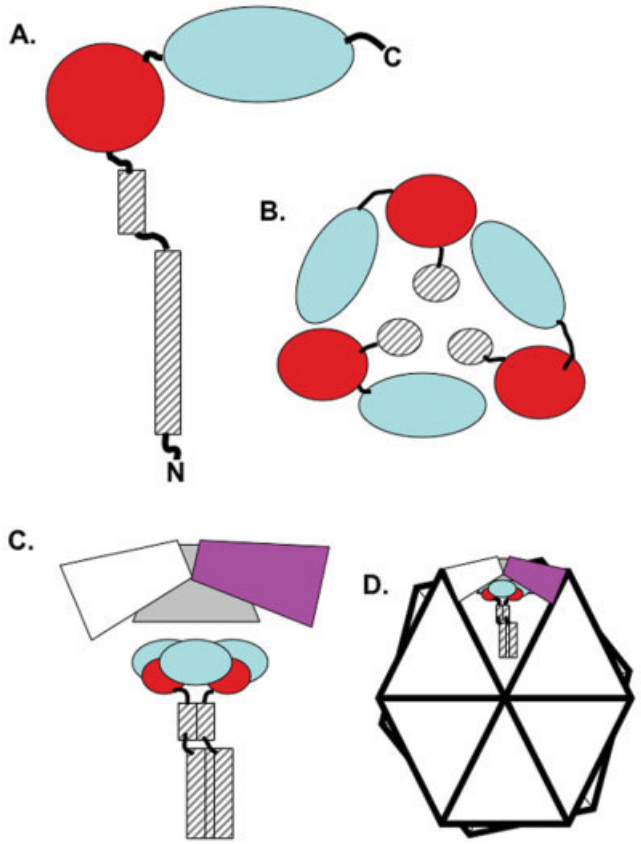

Fig. 9. A diagrammatic representation of the domain structure of *Lumbricus* linker chains and their orientation in the whole molecule as determined by X-ray crystallography.[14] (**A**) A single linker subunit has two N-terminal helical regions (grey stripes) connected by a short nonhelical segment. Each linker also has a middle LDLR-like domain (red) and a C-terminal domain (blue). (**B**) Three different linker subunits form a heterotrimeric structure, largely stabilized by a coiled-coil domain and by interactions between the LDLR-like domain and its neighboring C-terminal domain. The orientation of the LDLR-like domains in the trimeric complex corresponds to the side views shown in Figure 8. (**C**) The hemoglobin dodecamer is shown as white, grey, and purple trapezoids with each trapezoid representing one of the three *abcd* heterotetramers. The dodecamer binds to the top face of the linker heterotrimer with the molecular threefold of the dodecamer being aligned with the quasi-threefold axis of the linker trimer. (**D**) Twelve of the subcomplexes (as shown in (C) form a hexagonal bilayer structure with the bottom layer being offset from the top layer by a 15° rotation.

Posttranslational Adducts

Comparison of the deduced molecular weights of linker polypeptides and the masses determined by mass spectrometry (Table IV) shows that only linkers L1 and L2 could contain carbohydrate.

Linker L1

Martin and colleagues[50] found two components for L1 with masses 1865.8 Da (L1a) and 1704.2 Da (L1b) higher than that of the L1 polypeptide (Table IV). They suggest that these masses can be accommodated by assuming that L1a has two *N*-acetyl hexosamines and nine hexoses that would give 1865.7 Da. Linker L1b has a mass 161.5 Da lower which suggests one less hexose. The deglycoslylated product has a mass within 0.9 Da of that of the L1 polypeptide confirming the carbohydrate nature of the adduct.

Linker L2

The sequence-derived molar mass of 30,429.3 Da for L2 is ~ 1600 to 1800 Da lower than obtained by mass spectrometry (Table IV). The nature of the additional mass is under investigation but has not yet been identified. Although the absence of change in molar mass after treatment with either N- or O-deglycanases[50] suggests that the adduct may not be carbohydrate, this possibility has not been completely excluded.

Linker L3

The close correspondence of the sequence and mass spectrometry–derived masses clearly indicate the absence of any carbohydrate or other posttranslational modification of linker L3.

TABLE IV. Comparison of Masses of Linker Chains Obtained from Sequence and Mass Spectrometry Data

Linker	Number of Cysteines	Sequence Mass[a] (Da)	Mass Spectrometry				Estimated Additonal Mass
				Ref. 50	Ref. 21	Ref. 45	
L1	10	25,836.6	L1a	27,702.4 ± 2	27,728 ± 15	27,684	1,865.8
			L1b	27,540.8	—		1,704.2
			L1	25,837.5 ± 3 (deglycosylated)	—		0
L2	10	30,330.2[b]		32,104.3 ± 5	32,251 ± 20	32,085	~1685–1851[e]
L3	8	24,913.4[d]		24,912.9 ± 2	24,919 ± 10	24,942	0
							Observed Difference
L4	8	24,248.0	L4a	24,169.9 ± 2	—	24,120	78.1
			L4b	24,102.3	—		140.7
			L4c	24,019.0	—		229

[a]Masses corrected for the hydrogens removed upon disulfide formation
[b]The value with 2 Ala replacing 2 Gly is 30,358.3 Da (see text). An NH_2-terminal acetyl group brings the mass to 30,400.3 Da.
[c]Difference calculated from mass with an acetyl group as in footnote b. The estimated adduct may not be carbohydrate because deglycosylation did not change the mass.[50]
[d]This mass is with Phe in position 114 in the L3 sequence. The close correspondence between the calculated and observed masses supports this assignment. The position appears to be heterogeneous; see text.

Linker L4

Martin and colleagues[50] found components L4a, L4b, and L4c that differ from the polypeptide mass by an amount which indicates the *loss* of 1, 2, or 3 units of average mass 74.9 Da. The difference between this value and that of an Ala residue (71.1 Da) might be due to an unknown amino acid difference, measurement inaccuracies or unresolved microheterogeneity. A possible explanation for this discrepancy is found by examining the NH_2-terminal amino acid sequence including the signal sequence: MRGPFIGVVVVVLAAVACLLQDA/A/AEED—. The slashes indicate the position of the cleavage of the signal peptide as demonstrated by NH_2-terminal sequencing of L4 that gave AAEED together with a minor component that began with AEE.[23] If the signal sequence were terminated one residue earlier, after D, then the mature peptide would begin with three NH_2-terminal alanyl residues rather than two.

Origin of linkers

The origin of the linkers remains elusive. What is the origin of the polypeptides into which the LDLR-like segment was inserted? Did they arise from globins? Residues 106 to 180 in L1 are 20% to 23% identical to residues 70 to 153 in globin *c* (helices E, F, G, and H) but this low degree of correspondence renders any possible globin relationship uncertain.[20] Although linkers have been reported to bind heme, it appears that they do so only under conditions where the globin chains lose heme.[36] The NH_2-terminal glycine-rich and the carboxyl-terminal Asp-Asp-His repeats in L2, discussed above, might have been inserted by a process of exon swapping as has evidently happened with cysteine-rich sequences that are homologous with the binding repeats of the LDL receptor.

Super Oxide Dismutase Activity and Metal Binding

Both Zn and Cu have been found in *Lumbricus* Hb.[51] This has led to the finding that *Lumbricus* Hb has some superoxide dismutase (SOD) activity[52]. The linker sequences do not show any obvious motifs that might correspond to an active site for SOD. However, BLASTp searches were used as queries against a nonredundant assembly of protein sequence databases to reveal sequence matches independently in the NH_2- (before the cysteine-rich segment) and C-terminal sequences (after the cysteine-rich segment) of all *Lumbricus* linker chains. These searches showed that the C-terminal DDH repeats of the L2 chain also occur in pernin, a self-associating hemolymph protein of a bivalve mollusk.[53] Although the sequence of this protein is homologous with Cu/Zn SODs, it lacks SOD activity and does not bind Cu or Zn but does bind Fe, possibly within its DDH region. We speculate that the similar C-terminal DDH repeats of L2 might constitute a set of metal binding sites.

CONCLUSIONS

A complete understanding of the assembly of the extracellular hemoglobin of the earthworm requires the determination of the amino acid sequences of all eight of the constituent polypeptides. The present study completes this task. Each of the four linkers has a highly conserved, negatively charged, cysteine-rich segment that is homologous with the ligand-binding domains of the human LDL receptor. This segment has three disulfide bonds with exactly the same connectivity as found in the LDL receptor. This connectivity is made possible by a calcium ion whose binding makes possible the correct folding of the segment and the right disulfide connectivity as has been shown for the LDL receptor. These segments interact with positively charged residues of the globin chains in the process of assembly. The functions of other parts of the linker chains remain to be determined. Although superoxide dismutase activity has been reported for this hemoglobin the location of the presumed Zn/Cu site within the molecule has not been identified. However, the imperfect Asp-Asp-His repeats in L2 form an attractive sequence for

further study as a possible metal binding site. Although four different linkers have been identified, it appears likely that the very similar linkers L3 and L4 are alternative forms. Neither L3 nor L4 has any posttranslational adduct. In contrast, both L1 and L2 have 1.6 to 1.8 kDa adducts. Although the adduct of L1 has been identified as carbohydrate the nature of the L2 adduct remains uncertain.

ACKNOWLEDGMENTS

The experimental work on cDNA described here was performed primarily by W. -Y. Kao and that on disulfide connectivity by Jun Qin. We thank William E. Royer for valuable discussions, Maria Person for MALDI measurements, and Mehdi Moini, Steven Halls, and Klaus Linse for analyses at the Proteomics and Mass Spectrometry Facility of the Institute of Cellular and Molecular Biology of the University of Texas.

REFERENCES

1. Kao W-Y, Fushitani K, Riggs AF. Structures of non-globin chains required for assembly of the gigantic extracellular hemoglobin of the earthworm. FASEB J 1996;10:A1387.
2. Kao W.-Y, Fushitani K, Riggs CK, Riggs AF. The linker chains of the gigantic hemoglobin of the earthworm: sequences of linkers and connectivity of disulfide bonds Biophys J 2000;78:166A.
3. Svedberg, T, Eriksson, I-B. Molecular weights of the blood pigments of *Arenicola* and of *Lumbricus*. Nature 1932;130:434–435.
4. Levin Ö. Electron microscope observations on some 60 s erythrocruorins and their split products. J Mol Biol 1963;6:95–101.
5. Roche J. Electron-microscope studies on high molecular weight erythrocruorins (invertebrate haemoglobins) and chlorocruorins of annelids. Studies Comp Biochem 1965;23:62–80.
6. Chew MY, Scutt PB, Oliver IT, Lugg JWH. Erythrocruorin of *Marphysa sanguinea*: isolation and some physical, physicochemical and other properties. Biochem J 1965;94:378–383.
7. Waxman L. The hemoglobin of *Arenicola cristata*. J Biol Chem 1971;246:7318–7327.
8. Garlick RL, Riggs AF. Purification and structure of the polypeptide chains of earthworm hemoglobin. Arch Biochem Biophys 1981;208:563–575.
9. Shlom JM, Vinogradov, SN. A study of the subunit structure of the extracellular hemoglobin of *Lumbricus terrestris*. J Biol Chem 1973;248:7904–7912.
10. Vinogradov SN. The structure of invertebrate extracellular hemoglobins (erythrocruorins and chlorocruorins). Comp Biochem Physiol 1985;82B:1–15.
11. Vinogradov SN, Lugo SD, Mainwaring MG, Kapp OH, Crewe AV. Bracelet protein: a quaternary structure proposed for the giant extracellular hemoglobin of *Lumbricus terrestris*. Proc Natl Acad Sci U S A 1986;83:8034–8038.
12. Zhu H, Ownby DW, Riggs CK, Nolasco NJ, Stoops JK, Riggs AF. Assembly of the gigantic hemoglobin of the earthworm *Lumbricus terrestris*. J Biol Chem 1996;271:30007–30021.
13. Kuchumov AR, Taveau JC, Lamy JN, Wall JS, Weber RE, Vinogradov SN. The role of linkers in the reassembly of 3.6 MDa hexagonal bilayer hemoglobin from *Lumbricus terrestris*. J Mol Biol 1999;289:1361–1374.
14. (a) Royer WE, Strand K, van Heel M, Hendrickson WA. Structural hierarchy in erythrocruorin, the giant respiratory assemblage of annelids. Proc Natl Acad Sci U S A 2000;97:7107–7111. (b) Daniel E, Lustig A, David MM, Tsfadia Y. Towards a resolution of the long-standing controversy regarding the molecular mass of extracellular erythrocruorin of the earthworm *Lumbricus terrestris*. Biochim Biophys Acta 2003;1649:1–15.
15. Suzuki T, Ohta T, Yuasa HJ, Takagi T. The giant extracellular hemoglobin from the polychaete *Neanthes diversicolor*. The cDNA-derived amino acid sequence of linker chain L2 and the exon/intron boundary conserved in linker genes. Biochim Biophys Acta 1994;1217:291–296.
16. Suzuki T, Takagi T, Gotoh T. Primary structure of two linker chains of the extracellular hemoglobin from the polychaete *Tylorrhynchus heterochaetus*. J Biol Chem 1990;265:12168–12177.
17. Suzuki T, Takagi T, Ohta S. Primary structure of a linker subunit of the tube worm 3000-kDa hemoglobin. J Biol Chem 1990;265:1551–1555.
18. Pallavicini A, Negrisolo E, Barbato R, Dewilde S, Ghiretti-Magaldi A, Moens L, Lanfranchi G. The primary structure of globin and linker chains from the chlorocruorin of the polychaete *Sabella spallanzanii*. J Biol Chem 2001;276:26384–26390.
19. Suzuki T, Vinogradov SN. Globin and Linker sequences of the giant extracellular hemoglobin from the leech *Macrobdella decora*. J Protein Chem 2003;22:231–242.
20. Suzuki T, Riggs AF. Linker chain L1 of earthworm hemoglobin. J Biol Chem 1993;268:13548–13555.
21. Ownby DW, Zhu H, Schneider K, Beavis RC, Chait BT, Riggs AF. The extracellular hemoglobin of the earthworm, *Lumbricus terrestris*. Determination of subunit stoichiometry. J Biol Chem 1993;268:13539–13547.
22. Jhiang SM, Riggs AF. The structure of the gene encoding chain c of the hemoglobin of the earthworm, *Lumbricus terrestris*. J Biol Chem 1989;264:19003–19008.
23. Fushitani K, Higashiyama K, Asao M, Hosokawa K. Characterization of the constituent polypeptides of the extracellular hemoglobin from *Lumbricus terrestris*: heterogeneity and discovery of a new linker chain L4. Biochim Biophys Acta 1996;1292:273–280.
24. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
25. Combet C, Blanchet C, Geourjon C, Deléage G. NPS@: network protein sequence analysis. TIBS 2000;25:147–150.
26. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 1997;25:3389–3402.
27. Beavis RC, Chait BT. Matrix assisted laser desorption ionization mass-spectrometry of proteins. Methods Enzymol 1996;270:519–551.
28. Qin J, Steenvoorden RJJM, Chait BT. A practical ion trap mass spectrometer for the analysis of peptides by matrix-assisted laser desorption/ionization. Anal Chem 1996;68:1784–1791.
29. Fass D, Blacklow S, Kim PS, Berger JM. Molecular basis of familial hypercholesterolaemia from structure of LDL receptor module. Nature 1997;388:691–693.
30. Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallgr 1991;A47:110–119.
31. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283–291.
32. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.
33. Fushitani K, Matsuura MSA, Riggs AF. The amino acid sequences of chains a, b and c that form the trimer subunit of the extracellular hemoglobin from *Lumbricus terrestris*. J Biol Chem 1988;263:6502–6517.
34. Xie Q, Donahue RA Jr, Schneider K, Mirza UA, Haller I, Chait BT, Riggs AF. Structure of chain d of the gigantic hemoglobin of the earthworm. Biochim Biophys Acta 1997;1337:241–247.
35. Maier CS, Arbogast B, Hahn U, Deinzer ML, Kuchumov AR, Vinogradov SN, Walz DA. A mass spectrometric study of the heterogeneity of the monomer subunit of *Lumbricus terrestris* hemoglobin. J Am Soc Mass Spec 1997;8:352–364.
36. Zhu H, Hargrove M, Xie Q, Nozaki Y, Linse K, Smith SS, Olson JS, Riggs AF. Stoichiometry of subunits and heme content of hemoglobin from the earthworm *Lumbricus terrestris*. J Biol Chem 1996;271:29999–30006.
37. DeBaere I, Liu L, Moens L, van Beeumen J, Gielens C, Richelle J, Trotman C, Finch J, Gerstein M, Perutz M. Polar zipper sequence in the high-affinity hemoglobin of *Ascaris suum*: amino acid sequence and structural interpretation. Proc Natl Acad Sci U S A 1992;89:4638–4642.
38. Qin J, Chait BT. Preferential fragmentation of protonated gas-phase peptide ions adjacent to acidic amino acid residues. J Am Chem Soc 1995;117:5411–5412.

39. Qin J, Chait BT. Identification and characterization of posttranslational modifications of proteins by MALDI ion trap mass spectrometry. Anal Chem 1997;69:4002–4009.
40. Bieri S, Djordjevic JT, Daly NL, Smith R, Kron PA. Disulfide bridges of a cysteine-rich repeat of the LDL receptor ligand-binding domain. Biochemistry 1995;34:13059–13065.
41. Südhof TC, Goldstein JL, Brown MS, Russell DW. The LDL receptor gene: a mosaic of exons shared with different proteins. Science 1985;228:815–822.
42. North CL, Blacklow SC. Solution structure of the sixth LDL-A module of the LDL receptor. Biochemistry 2000;39:2564–2571.
43. Mahley RW. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. Science 1988;240:622–630.
44. Wilson C, Wardell MR, Weisgraber KH, Mahley RW, Agard DA. Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E. Science 1991;252:1817–1822.
45. Lamy J, Kuchumov A, Taveau J-C, Vinogradov SN, Lamy JN. Reassembly of *Lumbricus terrestris* hemoglobin: a study by matrix-assisted laser desorption/ionization mass spectrometry and 3D reconstruction from frozen-hydrated specimens. J Mol Biol 2000; 298:633–647.
46. Atkins AR, Brereton IM, Kroon PA, Lee HT, Smith R. Calcium is essential for the structural integrity of the cysteine-rich, ligand-binding repeat of the low-density lipoprotein receptor. Biochemistry 1998;37:1662–1670.
47. Wang P-F, Veine, DM, Ahn, SH, Williams, CH. A stable mixed disulfide between thioredoxin reductase and its substrate, thioredoxin: preparation and characterization. Biochemistry1996;35: 4812–4819.
48. Fushitani K, Riggs AF. The extracellular hemoglobin of the earthworm, *Lumbricus terresris*. Oxygenation properties of isolated chains, trimer and a reassociated product. J Biol Chem 1991;266:10275–10281.
49. Brown MS, Herz J, Goldstein JL. Calcium cages, acid baths and recycling receptors. Nature 1997;388:629–630.
50. Martin PD, Kuchumov AR, Green BN, Oliver RWA, Braswell EH, Wall JS, Vinogradov SN. Mass spectrometric composition and molecular mass of *Lumbricus terrestris* hemoglobin: a refined model of its quaternary structure. J Mol Biol 1996;255: 154–169.
51. Standley PR, Mainwaring MG, Gotoh T, Vinogradov SN. The calcium, copper and zinc content of some annelid extracellular hemoglobins. Biochem J 1998;249:915–916.
52. Liochev SI, Kuchumov AR, Vinogradov SN, Fridovitch I. Superoxide dismutase activity in the giant hemoglobin of the earthworm, *Lumbricus terrestris*. Arch Biochem Biophys 1996;330:281–284.
53. Scotti PD, Dearing SC, Greenwood DR, Newcomb RD. Pernin: a novel, self-aggregating hemolymph protein from the New Zealand green-lipped mussel, *Perna canaliculus* (Bivalvia: Mytilidae). Comp Biochem Physiol Part B 2001;128:767–779.