Mass distributions of a macromolecular assembly based on electrospray ionization mass spectrometric masses of the constituent subunits

LEONID HANIN[a], BRIAN GREEN[b], FRANCK ZAL[c] and SERGE VINOGRADOV[d,*]

[a]*Department of Mathematics, Idaho State University, Pocatello, ID 8085, USA*
[b]*Micromass UK Ltd., Altrincham, Cheshire, WA14 5R2, UK*
[c]*Equipe d'Ecophysiologie, Station Biologique, UPMC-CNRS-INSU, 29682 Roscoff, France*
[d]*Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine, Detroit, MI 48201, USA*

**Corresponding author (Fax, 313-577-2765; Email, svinogra@med.wayne.edu)*

Macromolecular assemblies containing multiple protein subunits and having masses in the megadalton (MDa) range are involved in most of the functions of a living cell. Because of variation in the number and masses of subunits, macromolecular assemblies do not have a unique mass, but rather a mass distribution. The giant extracelular erythrocruorins (Ers), ~ 3·5 MDa, comprized of at least 180 polypeptide chains, are one of the best characterized assemblies. Three-dimensional reconstructions from cryoelectron microscopic images show them to be hexagonal bilayer complexes of 12 subassemblies, each comprised of 12 globin chains, anchored to a subassembly of 36 nonglobin linker chains. We have calculated the most probable mass distributions for *Lumbricus* and *Riftia* assemblies and their globin and linker subassemblies, based on the *Lumbricus* Er stoichiometry and using accurate subunit masses obtained by electrospray ionization mass spectrometry. The expected masses of *Lumbricus* and *Riftia* Ers are 3·517 MDa and 3·284 MDa, respectively, with a possible variation of ~ 9% due to the breadth of the mass distributions. The *Lumbricus* Er mass is in astonishingly good agreement with the mean of 23 known masses, 3·524 ± 0·481 MDa.

[Hanin L, Green B, Zal F and Vinogradov S 2003 Mass distributions of a macromolecular assembly based on electrospray ionization mass spectrometric masses of the constituent subunits; *J. Biosci*. **28** 557–568]

1. Introduction

Most of the functions in a living cell are carried out by "molecular machines", large macromolecular assemblies comprised of numerous and different proteins, with masses in the MDa range (Alberts 1998; Nogales and Grigorieff 2001). The variation in the number and masses of the constitueent subunits imply that large protein assemblies do not have a unique mass, but rather a mass distribution. The extracellular erythrocruorins (Ers) of annelids are giant, ~ 3500 kDa, complexes which were among the first protein assemblies investigated by ultracentrifugation (Svedberg 1933) and electron microscopy (EM) (Levin 1963; Roche 1965). Their characteristic EM and scanning transmission electron microscopy (STEM) appearances are those of an hexagonal bilayer (HBL) with diameter ~ 30 nm and height ~ 20 nm (Roche 1965; Terwilliger *et al* 1976; Vinogradov *et al* 1982; Boekema and van Heel 1989; Gotoh and Suzuki 1990; Lamy *et al* 1996). Extensive small angle X-ray scattering (SAXS)

Keywords. Electrospray ionization mass spectrometry; erythrocruorin; lumbricus; mass distribution; riftia

Abbreviations used: cryoEM, cryoelectron microscopy; 3D, three-dimensional; E, mathematical expectation; EM, electron microscopy; Er, erythrocruorin; ESI-MS, electrospray ionization mass spectrometric; HBL, hexagonal bilayer; pdf, probability density function; rv, random variable; SAXS, small angle X-ray scattering; SD, standard deviation; STEM, scanning transmission electron microscopy.

studies have shown the molecular shapes and dimensions of different Ers to be very similar (Stöckel *et al* 1973; Pilz *et al* 1980, 1986, 1988; Wilhelm *et al* 1980; Messerschmidt *et al* 1983; Terwilliger and Terwilliger 1985; Theuer *et al* 1985; Krebs *et al* 1998). More recently, three-dimensional (3D) reconstructions using cryoelectron microscopy (cryoEM) have demonstrated that all the quaternary structures are virtually identical at a resolution of ~ 2 nm (Cejka *et al* 1989, 1991, 1992; De Haas *et al* 1996a, b, c, d, 1997; Taveau *et al* 1999; Jouan *et al* 2001).

Concurrent electrospray ionization mass spectrometric (ESI-MS) studies provided accurate masses for the constituent globin and linker chains and the disulphide-bonded globin dimer/trimer/tetramer subunits (Green *et al* 1995; 1998a, 1999; Weber *et al* 1995; Martin *et al* 1996; Zal *et al* 1996, 1997a, b, 2000). The very recent low resolution crystallographic structure of *Lumbricus* Er (Royer *et al* 2000) established the stoichiometry to be 144 globin chains arranged as 12 dodecamers tethered to 36 linker chains, in agreement with the model proposed earlier (Vinogradov *et al* 1986) and the cryoEM 3D reconstructions (De Haas *et al* 1996a, b, c, d, 1997; Taveau *et al* 1999; Jouan *et al* 2001). In sharp contrast to the evidence for very similar molecular dimensions and quaternary structures, the over 60 known Er masses, obtained by several experimental techniques from the early work of Svedberg to the present, exhibit a surprisingly wide range, i.e. ranging from 2400 to 4470 kDa (Lamy *et al* 1996).

Because the 144 globin chains consist of 4 to 11 different proteins with masses from 16 to 19 kDa, and the 36 linker chains also consist of 2 to 7 different proteins with masses from 24 to 32 kDa (Lamy *et al* 1996), Ers, like all large molecular assemblies, are expected to have a mass distribution rather than a unique mass. There are four levels of mass distribution possible: the dodecamer subassembly of globin chains; the complex of 12 dodecamer subassemblies; the linker subassembly of 36 linker chains which are required for HBL structure formation (Kuchumov *et al* 1999); and finally the overall mass distribution of the complete assembly. We report below the results of calculations of mass distributions for the subassemblies and complete assemblies of the two Ers which differ in the subunit nature of their dodecamer subassemblies: from the earthworm *Lumbricus*; and, from the deep-sea polychaete *Riftia*.

2. Methods

2.1 *Mass distributions for a single dodecamer subassembly*

2.1*a Lumbricus*: Each dodecamer subassembly consists of 3 monomers which are randomly chosen from 3 different globin chains $d1$, $d2$, $d3$ and 3 trimers randomly selected from the observed 4 trimers $t1$-$t4$ (Martin *et al* 1996). We assume that the monomeric and trimeric subunits are incorporated in a dodecamer independently of each other. Let $p(di)$, $i = 1, 2, 3$, and $p(tj)$, $j = 1, 2, 3, 4$, be the probabilities that the corresponding monomer chains and trimers are present in a dodecamer. Let $D = (D1, D2, D3)$ be the vector of random numbers of monomers $d1$, $d2$, $d3$ present in a dodecamer, subject to the condition $D1 + D2 + D3 = 3$. The random vector D takes 10 values that correspond to the following combinations of monomer subunits: $3d1$, $3d2$, $3d3$, $2d1d2$, $2d2d1$, $2d1d3$, $2d3d1$, $2d2d3$, $2d3d2$, $d1d2d3$. Consequently, the random vector D has the multinomial distribution $B[3; p(d1), p(d2), p(d3)]$, i.e.

$$Pr(d) := Pr(D = d) = \frac{3!}{d1!\,d2!\,d3!}\, p(d1)^{d1}\, p(d2)^{d2}\, p(d3)^{d3}, \qquad (1)$$

where $d = (d1, d2, d3)$ is a vector of nonnegative integers such that $d1 + d2 + d3 = 3$.

Similarly, let $T = (T1, T2, T3, T4)$ be the vector of random numbers of trimers $t1$–$t4$ contained in a dodecamer. Obviously, Tj, $1 \le j \le 4$, are nonnegative integer-valued random variables such that $\sum_{j=1}^{4} Tj = 3$. The random vector T takes 20 values that correspond to the following combinations of trimer subunits: $3t1$, $3t2$, $3t3$, $3t4$, $2t1t2$, $2t2t1$, $2t1t3$, $2t3t1$, $2t1t4$, $2t4t1$, $2t2t3$, $2t3t2$, $2t2t4$, $2t4t2$, $2t3t4$, $2t4t3$, $t1t2t3$, $t1t2t4$, $t1t3t4$, $t2t3t4$. Consequently, the random vector T follows the multinomial distribution $B[3; p(t1), p(t2), p(t3), p(t4)]$, i.e.

$$p(t) := Pr(T = t) = 3! \prod_{j=1}^{4} \frac{p(tj)^{tj}}{tj!}, \qquad (2)$$

where $t = \{t\,j : 1 \le j \le 4\}$ is a vector of nonnegative integers such that $\sum_{j=1}^{4} tj = 3$. Thus, the number of different *Lumbricus* dodecamer subassemblies is equal to $10 \times 20 = 200$. Each of them is determined by the two integer vectors d and t described above, and its probability is equal to the product of probabilities (1) and (2). The mass of a randomly assembled dodecamer is given by:

$$m = \sum_{i=1}^{3} m(di) D_i + \sum_{j=1}^{4} m(tj) T_j + h, \qquad (3)$$

where $h = 12 \times 616{\cdot}5 = 7398$ Da is the total heme mass, and the observed electrospray ionization mass spectrometric masses of monomers and trimers are $m(d1) = 15993$ Da, $m(d2) = 15978$ Da, $m(d3) = 15962$ Da and $m(t1) = 52923$ Da, $m(t2) = 52760$ Da, $m(t3) = 52599$ Da, $m(t4) = 52435$ Da, with probabilities derived from the observe intensities of the peaks, $p(d1) = 0{\cdot}55$, $p(d2) = 0{\cdot}28$, $p(d3)$

= 0·17, $p(t1) = 0·24$, $p(t2) = 0·40$, $p(t3) = 0·08$, $p(t4) = 0·28$ (Martin *et al* 1996). In particular, if $D = d$ and $T = t$, the mass of the dodecamer is,

$$m(d, t) = \sum_{i=1}^{3} m(di)d_i + \sum_{j=1}^{4} m(tj)t_j + h. \quad (4)$$

From (3), the expected mass of a dodecamer is:

$$m_L = 3\left[\sum_{i=1}^{3} m(di)\,p(di) + \sum_{j=1}^{4} m(tj)\,p(tj)\right] + h, \quad (5)$$

and the standard deviation s_L of the mass distribution (4) is given by:

$$s_L^2 = \sum_{d,t} m^2(d, t)\,p(d)\,p(t) - m_L^2. \quad (6)$$

2.1*b Riftia*: Each dodecamer subassembly is assumed to consist of a random number D of dimers d and random numbers A and B of monomers a and b with $A, B, D \geq 0$ and $2D + A + B = 12$. The electrospray ionization mass spectrometric masses of the dimer d and monomers a, b (including heme) are $m_d = 32954$ Da, $m_a = 16750$ Da and $m_b = 17422$ Da and the monomer probabilities derived from the peak intensities are $p_a = 0·75$ and $p_b = 0·25$ (Zal *et al* 1996). We proceed using the following two-stage model of random assembly of a dodecamer. In the first stage, each dimer is included in a dodecamer independently of other dimers with some probability p_d. This implies that the distribution of the number of dimers in a dodecamer is binomial $B(6, p_d)$. The expected number of dimers in a dodecamer is therefore $6\,p_d$, and the expected number of dimers in the entire HBL is $72\,p_d$. Since the observed mean number of dimers in HBL is 48, we set $p_d = 2/3$. Suppose the number $D = k$, $0 \leq k \leq 6$, of dimers is selected. Then the remaining $12 - 2k$ monomers a and b are chosen randomly according to their frequencies. Hence, given that $D = k$, the number A of monomers a follows the binomial distribution $B(12 - 2k, pa)$, and similarly the number B of monomers b follows the binomial distribution $B(12 - 2k, pb)$. Given that $D = k$, $0 \leq k \leq 6$, the number of possible combinations of monomers a and b is $12 - 2k + 1$ and the total number of possible dodecamer compositions is,

$$N = \sum_{k=0}^{6} (12 - 2k + 1) = 49. \quad (7)$$

Each dodecamer is characterized by the number k of dimers and the number i of monomers a, where $k, i \geq 0$ and $2k + I \leq 12$, with the number of monomers b being $12 - 2k - i$. The probability of each dodecame composition is given by:

$$p(k, i) := Pr(D = k, A = i) = Pr(D = k)Pr(A = i|D = k)$$

$$= \binom{6}{k}\binom{12-2k}{i} p_d^k (1 - p_d)^{6-k}\, p_a^i\, p_b^{12-2k-i}. \quad (8)$$

The mass of randomly assembled dodecamer with D dimers, A monomers a and B monomers b is $m_dD + m_aA + m_bB$. The expected dodecamer mass is,

$$m_R = m_dED + m_aEA + m_b(12 - 2ED - EA)$$

$$= 12m_b + (m_d - 2m_b)ED + (m_a - m_b)EA, \quad (9)$$

where E indicates expectation; thus, $ED = 6\,p_d = 4$ and

$$EA = \sum_{k=0}^{6} E(A \mid D = k)Pr(D = k)$$

$$= \sum_{k=0}^{6} p_a(12 - 2k)Pr(D = k)$$

$$= 2\,p_a(6 - ED) = 3. \quad (10)$$

Hence the expected dodecamer subassembly mass is,

$$m_R = 12m_b + 4(m_d - 2\,m_b) + 3(m_a - m_b)$$

$$= 3m_a + m_b + 4m_d, \quad (11)$$

with the standard deviation,

$$s_R^2 = \sum_{2k+i\leq 12} [(m_d - 2m_b)(k - 4) + (m_a - m_b)(i - 3)]^2\, p(k, i), \quad (12)$$

2.2 *Mass distribution for 12 dodecamer subassemblies*

We assume that the 12 dodecamer subassemblies are included in the HBL structure independently of each other from the pool of all possible dodecamers described above (200 for *Lumbricus* and 49 for *Riftia*) according to their probabilities. Let $x1, \ldots, x12$ be the individual masses of the dodecamer subassemblies and let $x = x1 + \ldots + x12$ be the total mass of 12 subassemblies. It follows from our assumptions that the random variables $x1$–$x12$ are independent and identically distributed. Denote by m and s their expected value and standard deviation, respectively. The random variable x has a discrete distribution which may contain up to $200^{12} \approx 4·1 \times 10^{27}$ points for *Lumbricus* and $49^{12} \approx 1·9 \times 10^{20}$ points for *Riftia*. To make the distribution of x more tractable, we approximate it, on the basis of the central limit theorem (Taylor and Karlin 1998), by the normal distribution $N(12m, \sqrt{12s})$ with the probability density function

$$f_D(x) = \frac{1}{2\sqrt{6ps}} \exp\left[-\frac{(x - 12m)^2}{24s^2}\right]. \quad (13)$$

2.3 *Mass distribution for 36 linker chains*

Both *Lumbricus* and *Riftia* Ers have 4 types of linkers, $L1$–$L4$. Let $K = (K1, K2, K3, K4)$ be a random vector describing the number of linkers. $K1$–$K4$ are nonnegative integer-valued random variables such that $K1 + K2 + K3 + K4 = 36$. Consequently, the distribution of the random vector K is multinomial $B(36; p1, p2, p3, p4)$, that is,

$$Pr(K = k) = \frac{36!}{k1!k2!k3!k4!}\, p1^{k1}\, p2^{k2}\, p3^{k3}\, p4^{k4}, \quad (14)$$

where $k = (k1, k2, k3, k4)$ is a vector of non-negative integers such that $k1 + k2 + k3 + k4 = 36$. The expected total mass of linkers h and its standard deviation will be $36n$ and $6s$, respectively, where $n = m1p1 + m2p2 + m3p3 + m4p4$ is the expected mass of a single linker with the corresponding standard deviation, $s^2 = (m1 - n)^2 p1 + (m2 - n)^2 p2 + (m3 - n)^2 p3 + (m4 - \nu)^2 p4$. The random variable h has a discrete distribution which contains

$$\binom{36+4-1}{4-1} = 9139 \text{ points.}$$

Again, by the central limit theorem (Taylor and Karlin 1998), h can be approximated by the normal distribution $N(36n, 6s)$ with the probability density function

$$f_L(x) = \frac{1}{2\sqrt{6ps}} \exp\left[-\frac{(x - 36n)^2}{72s^2}\right]. \quad (15)$$

The electrospray ionization mass spectrometric masses of *Lumbricus* linkers are: $m_L1 = 27607$ Da, $m_L2 = 32104$ Da, $m_L3 = 24913$ Da, $m_L4 = 24090$ Da and their observed probabilities based on the intensities of the peaks are $p_L1 = 0{\cdot}24$, $p_L2 = 0{\cdot}16$, $p_L3 = 0{\cdot}4$, $p_L4 = 0{\cdot}2$, respectively (Martin *et al* 1996). The corresponding data for *Riftia* are: $m_R1 = 23505$ Da, $m_R2 = 23851$ Da, $m_R3 = 26342$ Da, $m_R4 = 27426$ Da, and $p_R1 = 0{\cdot}35$, $p_R2 = 0{\cdot}32$, $p_R3 = 0{\cdot}16$, $p_R4 = 0{\cdot}17$ (Zal *et al* 1996).

2.4 *Mass distribution of the HBL structures*

The total HBL mass is equal to $x + h$, where x and h are defined by equations (13) and (15), respectively, and are considered to be independent random variables. The expected total HBL mass is $12x + 36h$ and its standard deviation is $2\sqrt{3(s^2 + 3s^2)}$. The distribution of the total HBL mass is approximately normal, $N[12x + 36h, 2\sqrt{3(s^2 + 3s^2)}]$ and can be represented by the probability density function

$$f_L(x) = \frac{1}{4\sqrt{3p(s^2 + 3s^2)}} \exp\left[-\frac{(x - 12m - 36n)^2}{24(s^2 + 3s^2)}\right]. \quad (16)$$

All computations were carried out using MAPLE (Adams 1998).

3. Results

Figure 1a shows the mass distributions for the total mass of the 36 linker subassemblies from *Riftia* and *Lumbricus* Ers calculated from the probability density function, eq. (14). Figure 1b shows plots of the mass distribution for the total masses of the 12 dodecamer subassemblies of *Riftia* and *Lumbricus* Ers determined by the probability density function given by eq. (12). Figure 1c shows the total mass distributions of the two complete assemblies calculated from the probability density function given by eq. (16). Note that if an absolutely continuous random variable is measured in units L, then its probability density function has units $1/L$; therefore, the units for probability density functions of the mass distributions are MDa^{-1}. Table 1 shows the expected masses together with their standard deviations, for the dodecamer subassemblies, the complex of 12 dodecamer subassemblies, the 36-linker subassemblies and the complete assemblies of *Lumbricus* and *Riftia* Ers. The two assemblies differ in the subunit compositions of their globin dodecamer subassemblies, monomers and disulphide-bonded trimers of globin chains in *Lumbricus* (Martin *et al* 1996) and monomers and disulphide-bonded dimers in *Riftia* (Zal *et al* 1996). In the latter case, a dodecamer subassembly can, in principle, vary from 12 monomers to 6 dimers and include all the intermediate combinations with a correspondingly broad mass distribution.

4. Discussion

The 3D reconstructions obtained by cryoEM at resolutions of ~ 2·0 nm provide overwhelming evidence in favour of a common quaternary structure for seven different Ers (De Haas *et al* 1996a, b, c, d, 1997; Taveau *et al* 1999; Jouan *et al* 2001): 144 globin chains within 12 dodecamer subassemblies, arranged in two hexagonal layers, with each globin dodecamer subassembly anchored to 3 linker chains (Royer *et al* 2000) forming a central subassembly of 36 linker chains.

It is obvious that the mass of a protein assembly–consisting of 12 dodecamers of 4 to 11 different globin chains, which vary in mass from 16 to 19 kDa, combined with 36 linker chains which consist of 2 to 7 linker chains and which vary in mass from 24 to 32 kDa (Lamy *et al* 1996) – can only be represented by a mass distribution.

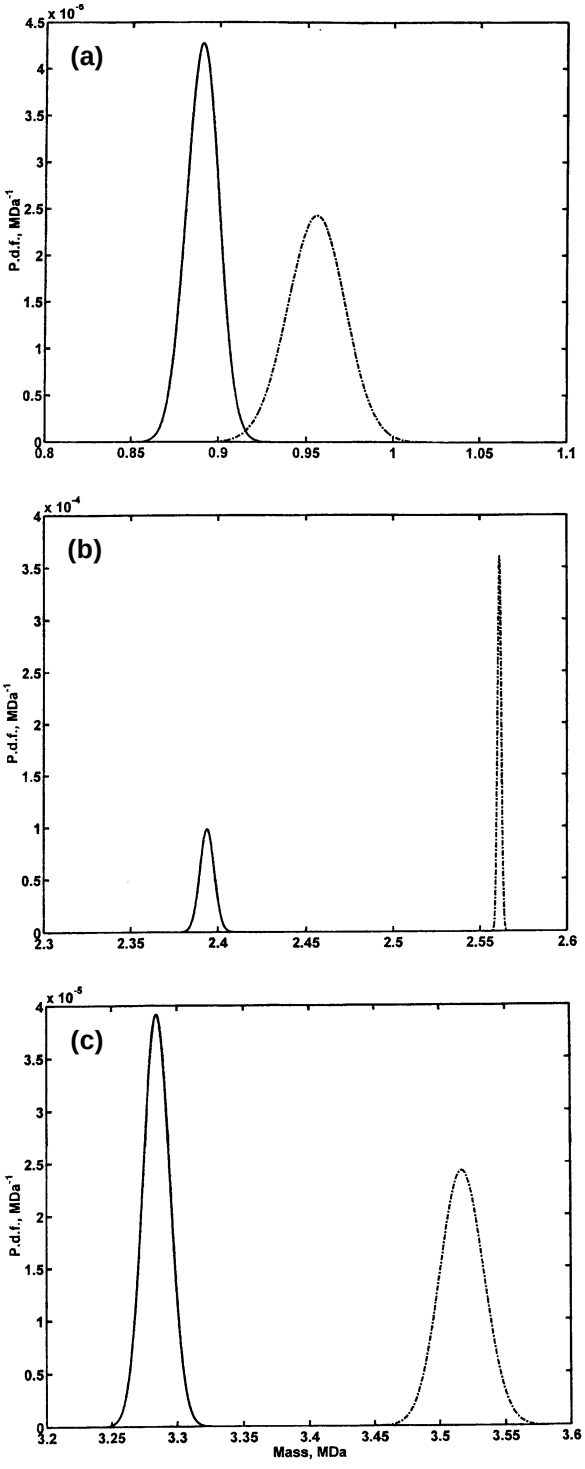

Figure 1. Plots of the normal approximations to the distributions of the total mass of globin and linker subassemblies and complete hexagonal bilayer structures of *Riftia* (left) and *Lumbricus* (right) erythrocruorins: 36 linker chains (**a**), 12 dodecamers (**b**) and complete hexagonal bilayer assembly (**c**). Note that if an absolutely continuous random variable is measured in units L, then its probability density function has units $1/L$; therefore, the units for the probability density functions of the mass distributions are MDa^{-1}.

The combinatorics of a multisubunit complex such as *Lumbricus* Er are very complex (Hanin and Vinogradov 2000). There are four levels of mass distribution possible: (i) the dodecamer of globin chains; (ii) the HBL complex of 12 dodecamer subassemblies; (iii) the linker subassembly of 36 linker chains which are required for HBL structure formation (Kuchumov *et al* 1999); and (iv) finally the overall mass distribution of the complete assembly of 12 dodecamers and 36 linker chains. Although the two Ers are presumed to have the same globin to linker stoichiometry, they differ in the subunit nature of their dodecamer subassemblies i.e. in *Lumbricus* Er they are comprised of 3 monomer globin chains and 3 disulphide-bonded trimers (Martin *et al* 1996); while in *Riftia*, the globin subassemblies consist of monomer and disulphide-bonded dimer subunit (Zal *et al* 1996). Table 1 shows the expected masses, their standard deviations (SD) and the range of possible masses for randomly assembled subassemblies corresponding to each of the foregoing four levels calculated for *Lumbricus* and *Riftia* Ers. The calculated mass distributions for the 36 linker complex, the 12 dodecamer subassembly, and the complete assemblies are illusrated in figure 1. The percent variation in the mass of each subassembly, defined as range × 100/expected mass, differs substantially from *Lumbricus* to *Riftia*. The *Lumbricus* dodecamer and linker subassemblies have very disparate variations, 0·73 and 30·2, respectively compared to 5·7 and 15·9 for *Riftia*. Nevertheless, the percent variation for the complete assembly is similar for both Ers, 8·7 and 8·4, respectively.

Table 2 compares the expected masses of *Lumbricus* and *Riftia* Ers (table 1) with the calculated means of the 23 published masses of *Lumbricus* Er, the 28 masses of polychaete Ers, and of all the 67 known Er masses. The latter are provided in table 3 in the supporting information section and are plotted in figure 2 as a function of the year of publication for oligochaete and leech Ers (top panel) and polychaete Ers (bottom panel). The crosses mark the values reported prior to 1950. Despite the scatter in the 23 published masses for *Lumbricus* Er over a very broad range, from 2730 to 4470 kDa, which encompasses the published masses of the other Ers (Lamy *et al* 1996) as can be seen in figure 2, the expected mass, 3517 ± 16 kDa is in astonishingly good agreement with the calculated mean of known masses, 3524 ± 481 kDa (N = 23). It is also within 1% of the overall mean of all oligochaete Ers including other earthworms, 3481 ± 434 kDa (N = 37). Furthermore, there was no significant asymmetry in the scatter of the published masses, since the recalculated means, omitting values outside ± 1 SD of the original mean, were 3577 ± 271 kDa (N = 16) and 3500 ± 279 kDa (N = 26), respectively.

The expected mass of *Riftia* Er, 3284 ± 10 kDa (table 2) is also in good agreement with the calculated mean of

Table 1. Expected masses of globin and linker subassemblies and the complete HBL structure of *Lumbricus* and *Riftia* Ers.

Result	*Lumbricus*	*Riftia*
Expected dodecamer mass ± SD, Da	213,436 ± 319	199,488 ± 1,173
Range, Da	212,624–214,176	197,724–209,064
Percent variation[a]	0·73	5·7
Expected mass of 12 dodecamers ± SD, Da	2,561,208 ± 1,107	2,393,856 ± 4,063
Range, Da	2,551,068–2,569,752	2,372,688–2,508,768
Percent variation[a]	0·73	5·7
Expected mass of 36 linkers ± SD, Da	955,639 ± 16,416	890,504 ± 9,341
Range, Da	867,240–1,155,744	846,180–987,336
Percent variation[a]	30·2	15·9
Total mass of hexagonal bilayer assembly, Da	3,516,847 ± 16,453	3,284,360 ± 10,187
Range, Da	3,418,308–3,725,496	3,218,868–3,496,104
Percent variation[a]	8·7	8·4

[a]Range × 100/expected value.

all the known polychaete Er masses, 3209 ± 389 kDa (N = 30). The latter is again unaffected by excluding values outside 1 SD of the original mean, 3236 ± 234 kDa (N = 19) (table 2). In general, most polychaete Er masses are lower than the masses for the oligochaete and achaete Ers, as reflected in the lower means, 3209 ± 389 kDa (N = 30) versus 3481 ± 434 kDa (N = 37) and as clearly illustrated in figure 2. However, the difference, 272 kDa, is well within the SD of the two mean values.

The results shown in tables 1–3 and figure 2, show that the inherent spread in the molecular masses of Ers is much smaller than the spread of experimental masses, about 8–9% (range × 100/expected mass) compared to ~ 50%, respectively. Several different factors are responsible for this 5-fold greater variability in experimental masses. These include experimental errors such as the handling of protein samples, errors proper to the biophysical method used to determine the mass and errors due to the biological variability of Ers.

It is interesting to note that Svedberg (1933) and Svedberg and Hedenius (1934) had observed the sedimentation constants of polychaete Ers to be lower than those of oligochaete Ers, 57·1 S versus 60·8 S. However, the mean of all known polychaete Er masses determined after 1980, 3398 ± 350 kDa (N = 14) (table 2) is in agreement with the calculated mean, 3387 ± 45 kDa (N = 8). The generally lower masses obtained for polychaete Ers, particularly prior to 1980 (see figure 2) are probably due to improper handling of the protein samples. Polychaete Ers have a greater proclivity towards dissociation as a result of proteolytic degradation either during preparation or storage than do oligochaete and leech Ers and the use of protease inhibitors in buffers used in Er preparation is relatively recent (Vinogradov and Sharma 1994).

An illustration of errors pertaining to a biophysical method of mass determination is the large effect of the specific volume of a protein on the mass determined by ultracentrifugation, whether by sedimentation equilibrium or calculated from the S and D values using the Svedberg equation (Svedberg and Pedersen 1940). Generally, the specific volume was calculated from the amino acid composition of Ers, providing values ranging from 0·73 to 0·75 (table 3). However, our experimentally determined value for *Lumbricus terrestris* Er, 0·714 (Martin *et al* 1996), indicates that the choice of the specific volume can well result in an error of 10% or more.

In addition to the above uncertainties, it should be emphasized that biological variability is inherent in the Er preparations, even of the same species, obtained by different investigators in different parts of the world.

Table 2. Comparison of the mean Er masses calculated from ESI-MS data with the means of published experimental masses.

Species	Mean calculated mass, kDa	Mean experimental mass, kDa[a]
Lumbricus terrestris	3517 ± 16[b]	3524 ± 481 (N = 23) 3577 ± 271 (N = 16)[c]
All oligochaete and achaete Ers		3481 ± 434 (N = 37) 3500 ± 279 (N = 26)[c]
Riftia pachyptila	3284 ± 10[b]	–
All polychaete Ers		3209 ± 389 (N = 30) 3236 ± 234 (N = 19)[c]
All polychaete Ers since 1980		3398 ± 350[a] (N = 14)
All annelid Ers	3387 ± 45[a] (N = 12)	3366 ± 428 (N = 67) 3·314 ± 0·235 (N = 47)[c]

[a]From table 3 in supporting information section.
[b]From table 1.
[c]Mean of data within ± 1 SD of original mean.

Table 3. Experimental and calculated masses of annelid Ers.

Species	Experimental mass, kDa	Method[a]	Calculated mass, kDa[b]	Reference
Oligochaetes				
Lumbricus terrestris	2730	SE		Svedberg and Eriksson-Quensel (1933)
	2950	SE		Svedberg and Eriksson-Quensel (1934)
	3140[c]	SD		Svedberg and Pedersen (1940)
	3450	LS		Rossi Fanelli *et al* (1970)
	2500	GF		Wiechelman and Parkhurst (1973)
	3100	LS		Harrington *et al* (1973)
	3230 ± 180[d]	SD		Shlom and Vinogradov (1973)
	2940 ± 100	LS		Harrington and Herskovits (1975)
	3860 ± 90	SE		Wood *et al* (1976)
	3680 ± 170	SE		Vinogradov *et al* (1977)
	3950 ± 150	SAXS		Pilz *et al* (1980)
	4120[e]	SD		Vinogradov *et al* (1980)
	3800 ± 300[f]	EM		Vinogradov and Kolodziej (1988)
	3340 ± 510	GF		Barnikol *et al* (1989)
	3410 ± 390	SE	3517	Martin *et al* (1996)
	3560 ± 130	STEM	3460[h]	Martin *et al* (1996)
	3850[g]	SD		
	3840	SE		David and Daniel (1974)
	3590	GEMMA		Kaufman *et al* (1998)
	4470	SE		Tsfadia and Daniel (1999)
	4100 ± 100	MALLS		Zhu *et al* (1996)
	3755 ± 80	MALLS		F Zal (unpublished observations)
Lumbricus rubellus	3680	SE		Ellerton *et al* (1987a)
Eisenia fetida	4010[j]	SD		Ochiai and Enoki (1981)
	3820 ± 0·05	SE		Frossard (1982)
Glossoscolex paulistus	3100[k]	S		Costa *et al* (1988)
	3200	SAXS		El Idrissi Slitine *et al* (1990)
Maoridrilus montanus	3200	GF		Ellerton *et al* (1987b)
Pheretima communissima	3070[l]	SD		Ochiai and Enoki (1979)
Limnodrilus	3010[m]	SD		Yamagishi *et al* (1966)
Tubifex tubifex	3010	SE	3368[n]	Scheler and Schneiderat (1959)
	3630	LS		Russell and Osborn (1968)
	3090 ± 150	SAXS		Theuer *et al* (1985)
Achaetes				
Haemopis sanguisuga	3710	SE	3361[o]	Wood *et al* (1976)
Macrobdella decora	3560 ± 160	STEM	3355	Kapp *et al* (1990)
	3540 ± 80	SE		Weber *et al* (1995)
	3804 ± 70	MALLS		
Polychaetes				
Riftia pachyptila	3503 ± 13	MALLS	3284	Zal *et al* (1996)
	3396 ± 540	STEM		
Spirographis spallanzanii	2800	SE, SD	3360[p]	Antonini *et al* (1962)
Potamilla leptochaeta	2900	SE, GF		Himmel and Squires (1981)
Eudistylia vancouverii	3100	SE	3393[q]	Terwilliger *et al* (1975a,b)
	3480 ± 230	STEM		Qabar *et al* (1991)
	3261	MALLS		F Zal (unpublished observations)
Arenicola marina	2850	SE	3385[r]	Svedberg and Eriksson-Quensel (1933)
	3000	SE		Roche (1965)
	3300 ± 360[s]	SD		Pionetti and Pouyet (1980)
	3400 ± 40	SE		Wilhelm *et al* (1980)
	3000	SAXS		
	3850 ± 150	SAXS		Wilhelm *et al* (1980)
	3650 ± 20	MALLS		Zal *et al* (1997b)

(**Table 3.** *Contd*.)

Table 3. *Contd.*

Species	Experimental mass, kDa	Method[a]	Calculated mass, kDa[b]	Reference
Arenicola cristata	2850[t]	SD		Waxman (1971)
Abarenicola pacifica	2600	GF		Garlick and Terwilliger (1977)
Abarenicola affinis	3290	SE		Chung and Ellerton (1982)
Marphysa sanguinea	2400[u]	SD	3370[n]	Chew *et al* (1965)
	3609	MALLS		F Zal (unpublished observations)
Eunice aphroditois	3440	SE		Bannister *et al* (1976)
Cirriformia grandis	3000	SE		Swaney and Klotz (1981)
Pista pacifica	3400	SE		Terwilliger *et al* (1975a,b)
Thelepus crispus	3300	GF		Garlick and Terwilliger (1974)
Alvinella caudata	3517 ± 14	MALLS	3401	Zal *et al* (2000)
Paralvinella grasslei	3822 ± 28	MALLS	3426	Zal *et al* (2000)
Alvinella pompejana	3833 ± 14	MALLS	3412	Zal *et al* (2000)
Euzonus mucronata	3400	SE		Terwilliger *et al* (1977)
Ophelia bicornis	2700[v]	SD		Mezzasalma *et al* (1985)
Perinereis cultrifera	2700[w]	SD		Chiancone *et al* (1977)
Tylorrhynchus heterochaetus	3370	SE	3422[x]	Suzuki and Gotoh (1986)

[a]Abbreviations: GF, gel filtration; LS, light scattering; MALLS, multiple angle laser light scattering; GEMMA, gas-phase electrophoretic mobility molecular analysis; SAXS, small angle x-ray scattering; SD, calculated from the Svedberg equation and diffusion coefficient; SE, sedimentation equilibrium; STEM, scanning transmission electron microscopic mass mapping.
[b]Calculated on the basis of *Lumbricus* erythrocruorin stoichiometry, 144 globin and 36 linker chains (Royer *et al* 2000), using subunit masses obtained by ESI-MS.
[c]Calculated using $S°_{20,w} = 60{\cdot}9$, $D_{20,w} = 1{\cdot}81 \times 10^{-7}\text{cm}^2.\text{s}^{-1}$.
[d]Calculated using $S°_{20,w} = 58{\cdot}9$, $D_{20,w} = 1{\cdot}66 \times 10^{-7}\text{cm}^2.\text{s}^{-1}$ and $n = 0{\cdot}733$.
[e]Calculated using $S°_{20,w} = 58{\cdot}9$ (Shlom and Vinogradov 1973), $D_{20,w} = 1{\cdot}3 \times 10^{-7}\text{cm}^2.\text{s}^{-1}$ (Gros 1978) and $n = 0{\cdot}733$.
[f]Determined by counting number of particles with ferritin as the mass standard.
[g]Calculated using $S°_{20,w} = 58{\cdot}9$ (Shlom and Vinogradov 1973), $D_{20,w} = 1{\cdot}3 \times 10^{-7}\text{cm}^2.\text{s}^{-1}$ (Gros 1978) and experimental $n = 0{\cdot}714$ from Martin *et al* (1996).
[h]Mass calculated without carbohydrate.
[i]Value recalculated using earlier data (David and Daniel 1974) and $n = 0{\cdot}738$.
[j]Calculated using $S°_{20,w} = 62{\cdot}2$, $D_{20,w} = 1{\cdot}41 \times 10^{-7}\text{cm}^2.\text{s}^{-1}$ and $n = 0{\cdot}733$.
[k]Estimated from $S°_{20,w} = 58$ and $M_1/M_2 = (S_1/S_2)^{3/2}$.
[l]Calculated using $S°_{20,w} = 59$, $D_{20,w} = 1{\cdot}9 \times 10^{-7}\text{cm}^2.\text{s}^{-1}$ and $n = 0{\cdot}745$.
[m]Calculated using $S°_{20,w} = 60{\cdot}8$, $D_{20,w} = 1{\cdot}8 \times 10^{-7}\text{cm}^2.\text{s}^{-1}$ and $n = 0{\cdot}733$.
[n]Calculated from B N Green (unpublished data).
[o]Calculated from ESI-MS data of Green *et al* (1999).
[p]Calculated from sequence data (Pallavicini *et al* 2001).
[q]Calculated from ESI-MS data of Green *et al* (1998a).
[r]Calculated from ESI-MS data of Zal *et al* (1997b).
[s]Calculated using $S°_{20,w} = 54 \pm 0{\cdot}5$, $D_{20,w} = 1{\cdot}66 \pm 0{\cdot}15 \times 10^{-7}\text{cm}^2.\text{s}^{-1}$.
[t]Calculated using $S°_{20,w} = 57{\cdot}8$, $D_{20,w} = 2{\cdot}06 \times 10^{-7}\text{cm}^2.\text{s}^{-1}$ and $n = 0{\cdot}726$.
[u]Calculated using $S°_{20,w} = 58{\cdot}2$, $D_{20,w} = 2{\cdot}06 \times 10^{-7}\text{cm}^2.\text{s}^{-1}$ and $n = 0{\cdot}73$.
[v]Calculated using $S°_{20,w} = 55{\cdot}1$, $D_{20,w} = 1{\cdot}84 \times 10^{-7}\text{cm}^2.\text{s}^{-1}$ and $n = 0{\cdot}73$.
[w]Calculated using $S°_{20,w} = 55$, $f/f_o = 1{\cdot}2$ and $n = 0{\cdot}74$.
[x]Calculated from ESI-Ms data of Green *et al* (1995).

Since Er preparations are obtained generally from the pooled blood of many worms, it is not known whether the Ers from different *Lumbricus* populations collected over the world would have the same subunit composition. Several instances of highly polymorphic single chain Ers are known. ESI-MS has demonstrated that the monomeric Ers from the larvae of the midge *Chironomus* and the intracellular Ers of the marine polychaete *Glycera* are complex mixtures of more than 20 components (Green *et al* 1998b, c). Thus, it is conceivable that some of the variation in the published molecular masses of the giant Ers could be due to variation in the species.

The HBL Ers are known to contain about 40 to 80 Ca^{2+} and 2 to 4 Cu^{2+} and Zn^{2+} (Standley *et al* 1988), with about half of the Ca^{2+} required for the formation and maintenance of the HBL structure (Kuchumov *et al*

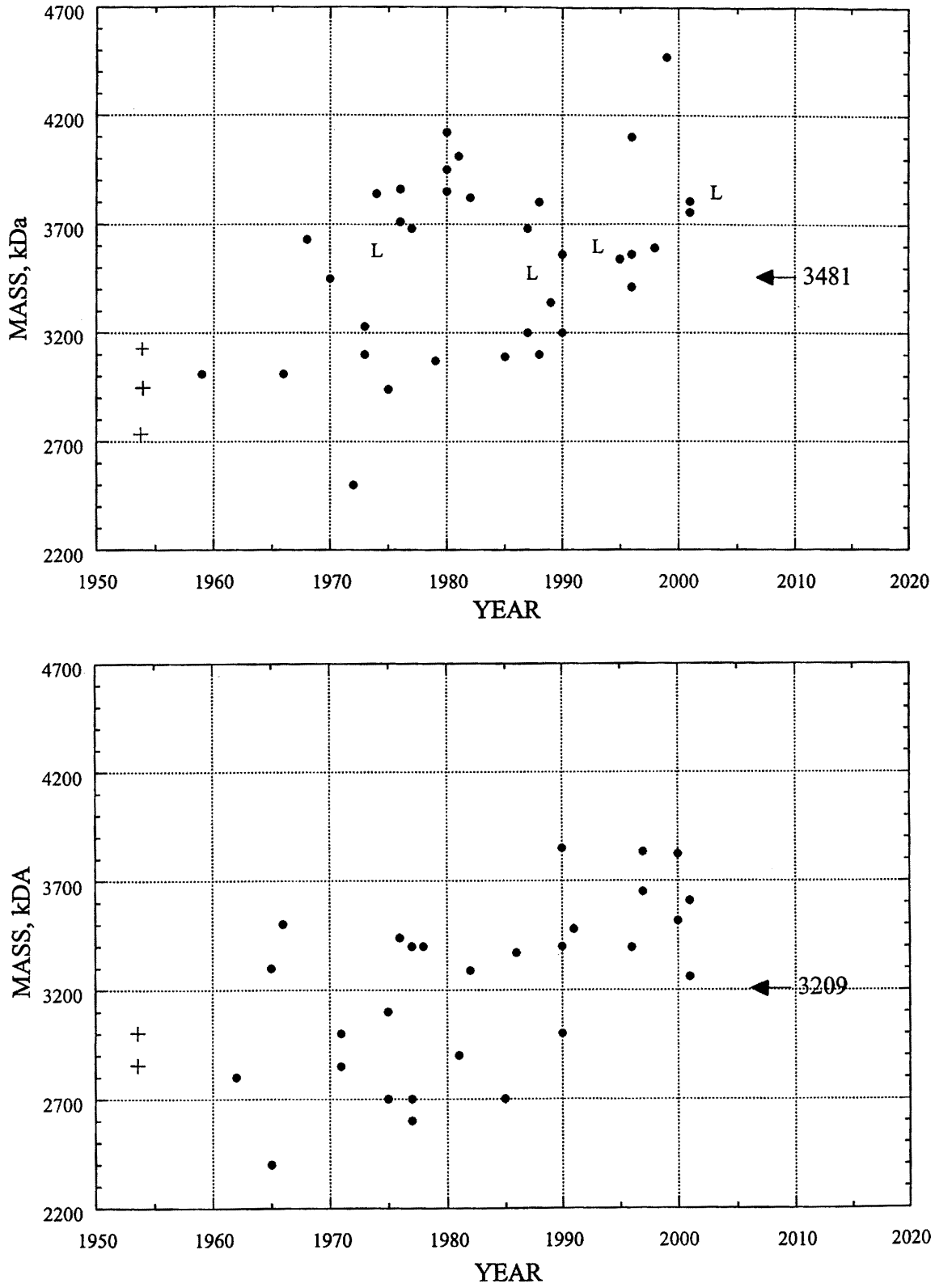

Figure 2. Plot of available molecular masses of oligochaete and leech Ers (top) and of polychaete Ers (bottom). The masses for leech Ers are marked L. The crosses indicate the values obtained before 1950 (Svedberg 1933; Svedberg and Ericksson-Quensel 1933, 1934).

2000). Clearly, the resulting additional mass of ~ 1700 to 3500 Da is a negligible contribution to the total Er mass.

References

Adams S G 1998 *Maple talk* (Upper Saddle River, NJ: Prentice Hall)

Alberts B 1998 The cell as a collection of protein machines; *Cell* **92** 291–294

Antonini E, Rossi-Fanelli A and Caputo A 1962 Studies on chlorocruorin. II. Some physicochemical properties of *Spirographis* chlorocruorin; *Arch. Biochem. Biophys.* **97** 343–350

Bannister J, Bannister W, Anastasi A and Wood E J 1976 Isolation, characterization and oxygen equilibrium of an extracellular haemoglobin from *Eunice aphroditois; Biochem. J.* **159** 35–42

Barnikol W, Birkhard O and Pötzchke H 1989 Das Erythrocruorin des Regenwurms (*Lumbricus terrestris*) als Eichsubstanz in der Gelchromatographie; *J. Chromtaogr.* **497** 231–235

Boekema E J and van Heel M 1989 Molecular shape of *Lumbricus terrestris* erythrocruorin studied by electron microscopy and image analysis; *Biochim. Biophys. Acta* **957** 370–379

Cejka Z, Magaldi A, Tognon G, Zanotti G and Baumeister W 1989 The molecular architecture of extracellular hemoglobin of *Eophila tellinii*; *J. Ultrastruct. Mol. Res.* **102** 71–81

Cejka Z, Santini C, Tognon G and Ghiretti-Magaldi A 1991 The molecular architecture of the extracellular hemoglobin of *Ophelia bicornis*: analysis of two-dimensional crystalline arrays; *J. Struct. Biol.* **107** 259–267

Cejka Z, Kleinz J, Santini C, Hegerl R and Ghiretti-Magaldi A 1992 The molecular architecture of the extracellular hemoglobin of *Ophelia bicornis*: analysis of individual molecules; *J. Struct. Biol.* **109** 52–60

Chew M Y, Scutt P, Oliver I and Lugg J 1965 Erythrocruorin of *Marphysa sanguinea*. Isolation and some physical, physicochemical and other properties; *Biochem. J.* **94** 378–383

Chiancone E, Ascoli F, Giardina B, Vecchini P, Antonini E, Musmeci M, Cina R, Zagra M, D'Amelio V and De Leo G 1977 Physicochemical and functional properties of *Perinereis cultrifera* (Grube) erythrocruorin; *Biochim. Biophys. Acta* **494** 1–8

Chung M and Ellerton H D 1982 The physico-chemical and functional properties of *Abarenicola affinis affinis* extracellular haemoglobin (erythrocruorin); *Biochim. Biophys. Acta* **702** 6–16

Costa M, Bonafé C, Meirelles N and Galembeck F 1988 Sedimentation coefficient and minimum molecular weight of extracellular hemoglobin of *Glossoscolex paulistus* (Oligochaeta); *Braz. J. Med. Res.* **21** 115–118

David M and Daniel E 1974 Subunit structure of earthworm erythrocruorin; *J. Mol. Biol.* **87** 89–101

De Haas F, Taveau J-C, Boisset N, Lambert O, Vinogradov S N and Lamy J N 1996a Three-dimensional reconstruction of the chlorocruorin of the polychaete annelid *Eudistylia vancouverii; J. Mol. Biol.* **255** 140–153

De Haas F, Boisset N, Taveau J-C, Lambert O, Vinogradov S N and Lamy J N 1996b Three-dimensional reconstruction of *Macrobdella decora* (Leech) hemoglobin by cryoelectron microscopy; *Biophys. J.* **70** 1973–1984

De Haas F, Zal F, Lallier F, Toulmond A and Lamy J N 1996c Three-dimensional reconstruction of the hexagonal bilayer hemoglobin of the hydrothermal vent tube worm *Riftia pachyptila* by cryoelectron microscopy; *Proteins Struct. Funct. Genet.* **26** 241–256

De Haas F, Zal F, You V, Lallier F, Toulmond A and Lamy J N 1996d Three-dimensional reconstruction by cryoelectron microscopy of the giant hemoglobin of the polychaete worm *Alvinella pompejana*; *J. Mol. Biol.* **264** 111–120

De Haas F, Kuchumov A, Taveau J-C, Boisset N, Vinogradov S N and Lamy J N 1997 Three-dimensional reconstruction of native and reassembled *Lumbricus terrestris* extracellular hemoglobin. Localization of the monomeric globin chains; *Biochemistry* **36** 7330–7338

El Idrissi Slitine F, Torriani I and Vachette P 1990 Small angle x-ray scattering study of two annelid extracellular hemoglobins; in *Invertebrate dioxygen carriers* (eds) G Préaux and R Lontie (Leuven: Leuven University Press) pp 225–228

Ellerton H D, Chen C and Lim A 1987a Subunit structure of the extracellular hemoglobin (erythrocruorin) from the earthworm *Lumbricus rubellus*; *Comp. Biochem. Physiol.* **87B** 1011–1016

Ellerton H D, Bearman C and Loong P 1987b Erythrocruorin from the New Zealand earthworm *Maoridrilus montanus*: a multi-subunit annelid extracellular hemoglobin; *Comp. Biochem. Physiol.* **87B** 1017–1023

Frossard P 1982 The erythrocruorin of *Eisenia fetida*. I. Properties and subunit structure; *Biochim. Biophys. Acta* **704** 524–534

Garlick R L and Terwilliger R C 1974 Coelomic cell hemoglobin of the terebellid polychaete *Thelepus crispus* Johnson; *Comp. Biochem. Physiol.* **47A** 543–554

Ghiretti-Magaldi A, Zanotti G, Tognon G and Mezzasalma V 1985 The molecular architecture of the extracellular haemoglobin of *Ophelia bicornis*; *Biochim. Biophys. Acta* **829** 144–149

Gotoh T and Suzuki T 1990 Molecular assembly and evolution of multi-subunit extracellular annelid hemoglobins; *Zool. Sci.* **7** 1–16

Green B N, Suzuki T, Gotoh T, Kuchumov A and Vinogradov S N 1995 Electrospray ionization mass spectrometric determination of the complete polypeptide chain composition of *Tylorrhynchus heterochaetus* hemoglobin; *J. Biol. Chem.* **270** 18209–18211

Green B N, Kuchumov A, Walz D A, Moens L and Vinogradov S N 1998a A hierarchy of disulfide-bonded subunits: the quaternary structure of *Eudistylia* chlorocruorin; *Biochemistry* **37** 6598–6605

Green B N, Kuchumov A R, Schmidt E R, Bergtrom G A, Walz D A and Vinogradov S N 1998b An electrospray ionization mass spectrometric study of the extracellular hemoglobins from *Chironomus thummi thummi*; *Biochim. Biophys. Acta* **1383** 143–150

Green B N, Loo J A, Walz D A and Vinogradov S N 1998c Electrospray ionization mass spectrometric study of the multiple intracellular monomeric and polymeric hemoglobins of *Glycera dibranchiata*; *J. Protein Chem.* **17** 85–97

Green B N, Kuchumov A, Klemm D and Vinogradov S N 1999 An electrospray ionization mass spectrometric study of the giant extracellular bilayer hemoglobin of the leech *Haemopis grandis* provides a complete enumeration of its subunits; *Int. J. Mass Spectrom. Ion Phys.* **188** 105–112

Gros G 1978 Concentration dependence of the self-diffusion of human and *Lumbricus terrestris* hemoglobins; *Biophys. J.* **22** 453–468

Hanin L G and Vinogradov S N 2000 Combinatorics of giant hexagonal bilayer hemoglobins; *Math. Biosci.* **163** 59–73

Harrington J and Herskovits T 1975 The effect of salts on the subunit structure and dissociation of *Lumbricus terrestris* hemoglobin; *Biochemistry* **14** 4972–4976

Harrington J, Pandolfelli E and Herskovits T 1973 Solution studies on heme proteins: circular dichroism and optical rotation of *Lumbricus terrestris* and *Glycera dibranchiata* hemoglobins; *Biochim. Biophys. Acta* **328** 61–73

Himmel M and Squires P 1981 High-performance size exclusion chromatography of sea worm chlorocruorin and other large proteins, viruses and polysaccharides on a TSK G5000 PW preparative column; *J. Chromatogr.* **210** 443–452

Jouan L, Taveau J-C, Marco S and Lamy J N 2001 Occurrence of two architectural types of extracellular hemoglobin in annelids: comparison of 3D reconstruction volumes of *Arenicola marina* and *Lumbricus terrestris* hemoglobins; *J. Mol. Biol.* **305** 757–771

Kapp O H, Qabar A, Bonner M, Stern M, Walz D A, Schmuck M, Pilz I, Wall J and Vinogradov S N 1990 Quaternary structure of the giant extracellular hemoglobin of the leech *Macrobdella decora*; *J. Mol. Biol.* **213** 141–158

Kaufman S L, Kuchumov A R, Kazakevich M and Vinogradov S N 1998 Analysis of a 3·6 MDa hexagonal bilayer hemoglobin from *Lumbricus terrestris* using a gas-phase electrophoretic mobility molecular analyzer; *Anal. Biochem.* **259** 195–202

Krebs A, Lamy J N, Vinogradov S N and Zipper P 1998 *Lumbricus terrestris* hemoglobin: a comparison of EM data and SAXS data; *Biopolymers* **45** 289–298

Kuchumov A R, Taveau J-C, Lamy J N, Wall J, Weber R and Vinogradov S N 1999 The role of linkers in the reassembly of the 3·6 MDa hexagonal bilayer hemoglobin from *Lumbricus terrestris*; *J. Mol. Biol.* **289** 1361–1374

Kuchumov A R, Loo J A and Vinogradov S N 2000 Subunit distribution of Ca-binding sites in *Lumbricus terrestris* hemoglobin; *J. Protein Chem.* **19** 139

Lamy J N, Green B N, Toulmond A, Wall J S, Weber R E and Vinogradov S N 1996 Giant hexagonal hemoglobins; *Chem. Rev.* **96** 3113–3124

Levin O 1963 Electron microscope observations on some 60S erythrocruorins and their split products; *J. Mol. Biol.* **6** 95–101

Martin P, Kuchumov A, Green B N, Oliver R, Braswell E, Wall J and Vinogradov S N 1996 Mass spectrometric composition and molecular mass of *Lumbricus terrestris* hemoglobin: a refined model of its quaternary structure; *J. Mol. Biol.* **255** 154–169

Messerschmidt U, Wilhelm P, Pilz I and Vinogradov S N 1983 Molecular shape and conformation of the extracellular hemoglobin of *Nephtys incisa; Biochim. Biophys. Acta* **742** 366–373

Mezzasalma V, Di Stefano L, Piazzese S, Zagra M, Salvato B, Togonon G and Ghiretti-Magaldi A 1985 Physicochemical and structural properties of the extracellular hemoglobin of *Ophelia bicornis; Biochim. Biophys. Acta* **829** 135–143

Nogales E and Grigorieff N 2001 Molecular machines: putting the pieces together; *J. Cell Biol.* **152** 1–14

Ochiai T and Enoki Y 1979 Earthworm (*Pheretima communissima* and *Pheretima hilgendorfi*) hemoglobins: giant assemblies and subunits; *Biochim. Biophys. Acta* **579** 442–451

Ochiai T and Enoki Y 1981 The molecular architecture and the subunits of earthworm (*Eisenia foetida*) hemoglobin; *Comp. Biochem. Physiol.* **68B** 275–279

Ownby D W, Zhu H, Schneider K, Beavis R C, Chait B T and Riggs A F 1993 The extracellular hemoglobin of the earthworm, *Lumbricus terrestris*. Determination of subunit stoichiometry; *J. Biol. Chem.* **268** 13539–13547

Pallavicini A, Negrisolo E, Barbato R, Dewilde S, Ghiretti-Magaldi A, Moens L and Lanfranchi G 2001 The primary structure of globin and linker chains from the chlorocruorin of the polychaete *Sabella spallanzanii*; *J. Biol. Chem.* **276** 26384–26390

Pilz I, Schwartz E and Vinogradov S N 1980 Small angle x-ray scattering studies of *Lumbricus terrestris* hemoglobin; *Int. J. Biol. Macromol.* **2** 279–283

Pilz I, Theuer M, Wilhelm P, Schwarz E and Vinogradov S N 1986 Structure of annelid hemoglobins – small angle x-ray studies; in *Invertebrate oxygen carriers* (ed.) B Linzen (Berlin: Springer-Verlag) pp 57–60

Pilz I, Schwartz E, Suzuki T and Gotoh T 1988 Small angle x-ray scattering studies of *Tylorrhynchus heterochaetus* hemoglobin; *Int. J. Biol. Macromol.* **10** 356–360

Pionetti J-M and Pouyet J 1980 Molecular architecture of annelid erythrocruorins. Extracellular hemoglobin of *Arenicola marina* (Polychaeta); *Eur. J. Biochem.* **105** 131–138

Qabar A, Stern M, Walz D, Chiu J-T, Timkovich R, Wall J, Kapp O and Vinogradov S N 1991 Hierarchy of globin complexes. The quaternary structure of the extracellular chlorocruorin of *Eudistylia vancouverii*; *J. Mol. Biol.* **222** 1109–1129

Roche J 1965 Electron microscope studies on high molecular weight erythrocruorins (invertebrate haemoglobins) and chlorocruorins of annelids; in *Studies in comparative biochemistry* (ed.) K A Munday (Oxford: Pergamon Press) pp 62–80

Rossi Fanelli M, Chiancone E, Vecchini P and Antonini E 1970 Studies on eryhtrocruorin. I. Physicochemical properties of earthworm erythrocruorin; *Arch. Biochem. Biophys.* **141** 278–283

Royer W E, Strand K, van Heel M and Hendrickson W 2000 Structural hierarchy in erythrocruorin, the giant respiratory assemblage of annelids; *Proc. Natl. Acad. Sci. USA* **97** 7107–7111

Russell J and Osborn J 1968 New tetramer of *Tubifex* haemoglobin; *Nature* (*London*) **220** 1125–1127

Scheler W and Schneiderat L 1959 Über die Praparation und die Bestimmung des mollekular-gemichtes wie der Konzentrazion des Hämoglobins von *Tubifex tubifex*; *Acta Biol. Med. Germ.* **3** 588–597

Shlom J M and Vinogradov S N 1973 A study of the subunit structure of the extracellular hemoglobin of *Lumbricus terrestris; J. Biol. Chem.* **248** 7904–7912

Standley P M, Mainwaring M G, Gotoh T and Vinogradov S N 1988 The calcium, copper and zinc content of some annelid extracellular haemoglobins; *Biochem. J.* **249** 915–917

Stöckel P, Mayer A and Keller R 1973 x-ray small angle scattering investigation of a giant respiratory protein: hemoglobin of *Tubifex tubifex; Eur. J. Biochem.* **37** 193–200

Suzuki T and Gotoh T 1986 The complete amino acid sequence of giant multisubunit hemoglobin from the polychaete *Tylorrhynchus heterochaetus; J. Biol. Chem.* **261** 9257–9267

Svedberg T and Pedersen K 1940 *The ultracentrifuge* (Oxford: Clarendon Press)

Svedberg T 1933 Sedimentation constants, molecular weights and isoelectric points of the respiratory proteins; *J. Biol. Chem.* **103** 311–325

Svedberg T and Eriksson-Quensel I-B 1933 The molecular weight of erythrocruorin. I; *J. Am. Chem. Soc.* **55** 2834–2841

Svedberg T and Eriksson-Quensel I-B 1934 The molecular weight of erythrocruorin. II; *J. Am. Chem. Soc.* **56** 1700–1706

Svedberg T and Hedenius A 1934 The sedimentation constant of respiratory proteins; *Biol. Bull.* **66** 191–223

Swaney J and Klotz I 1971 Properties of erythrocruorin from *Cirriformia grandis; Arch. Biochem. Biophys.* **147** 475–486

Taveau J-C, Boisset N, Vinogradov S N and Lamy J N 1999 Three-dimensional reconstruction of *Lumbricus terrestris* hemoglobin at 22 Å resolution: intramolecular localization of the globin and linker chains; *J. Mol. Biol.* **289** 1343–1359

Taylor H and Karlin T 1998 *An introduction to stochastic modeling*, 3rd edition (San Diego: Academic Press)

Terwilliger R C and Terwilliger N B 1985 Respiratory proteins of hydrothermal vent animals; *Bull. Biol. Soc. Wash.* **6** 273–287

Terwilliger R C, Terwilliger N B and Roxby R 1975a Quaternary structure of *Pista pacifica* vascular hemoglobin; *Comp. Biochem. Physiol.* **50B** 225–232

Terwilliger R C, Garlick R, Terwilliger N B and Blair D 1975b Molecular weight of *Eudistylia vancouveerii* chlorocruorin and its subunits; *Biochim. Biophys. Acta* **400** 302–309

Terwilliger R C, Terwilliger N B and Schabtach E 1976 Comparison of chlorocruorin and annelid hemoglobin quaternary structures; *Comp. Biochem. Physiol.* **55A** 51–55

Terwilliger R C, Terwilliger N B and Schabtach E 1980 The structure of hemoglobin from an unusual deep sea worm (Vestimentifera); *Comp. Biochem. Physiol.* **65B** 531–535

Terwilliger R C, Terwilliger N B, Schabtach E and Dangott L 1977 Erythrocruorins of *Euzonus mucronata* Treadwell: evidence for a dimeric annelid extracellular hemoglobin; *Comp. Biochem. Physiol.* **57A** 143–149

Theuer M, Pilz I, Schwarz E, Wilhelm P, Mainwaring M G and Vinogradov S N 1985 Molecular size and shape of the native and reassociated forms of the extracellular haemoglobin of *Tubifex tubifex*; *Int. J. Biol. Macromol.* **7** 25–29

Tsfadia Y and Daniel E 1999 A reevaluation of the molecular mass of earthworm extracellular hemoglobin from meniscus depletion sedimentation equilibrium; *Biochim. Biophys. Acta* **1433** 217–228

Vinogradov S N and Kolodziej P 1988 The molecular mass of the extracellular hemoglobin of *Lumbricus terrestris* determined by electron microscopy; *Comp. Biochem. Physiol.* **91B** 577–580

Vinogradov S N and Sharma P K 1994 Preparation and characterization of invertebrate globin complexes; *Methods Enzymol.* **231** 112–124

Vinogradov S N, Kapp O and Ohtsuki M 1982 The extracellular haemoglobins and chlorocruorins of annelids; in *Electron microscopy of proteins* (ed.) J Harris (London: Academic Press) vol. 3, pp 135–163

Vinogradov S N, Lugo S D, Mainwaring M G, Kapp O H and Crewe A V 1986 Bracelet protein: a quaternary structure proposed for the giant extracellular hemoglobin of *umbricus terrestris*; *Proc. Natl. Acad. Sci. USA* **83** 8034–8038

Vinogradov S N, Shlom J M, Hall B, Kapp O and Mizukami H 1977 The dissociation of *Lumbricus terrestris* hemoglobin: a model of its subunit structure; *Biochim. Biophys. Acta* **492** 136–155

Vinogradov S N, Shlom J M, Kapp O and Frossard P 1980 The dissociation of annelid extracellular hemoglobins and their quaternary structure; *Comp. Biochem. Physiol.* **67B** 1–16

Waxman L 1971 The hemoglobin of *Arenicola cristata; J. Biol. Chem.* **246** 7318–7327

Weber R E, Malte H, Braswell E, Oliver R, Green B N, Sharma P K, Kuchumov A R and Vinogradov S N 1995 Mass spectrometric composition, molecular mass and oxygen binding of *Macrobdella decora* hemoglobin and its tetramer and monomer subunits; *J. Mol. Biol.* **251** 703–720

Wiechelman K and Parkhurst L 1973 Kinetics of ligand binding in the hemoglobin of *Lumbricus terrestris; Biochemistry* **11** 4515–4520

Wilhelm P, Pilz I and Vinogradov S N 1980 A comparison of *Arenicola marina* and *Lumbricus terrestris* hemoglobin by small angle x-ray scattering; *Int. J. Biol. Macromol.* **2** 383–385

Wood E J, Mosby L and Robinson M 1976 Characterization of the extracellular haemoglobin of *Haemopis sanguisuga; Biochem. J.* **153** 589–596

Yamagishi M, Kajita A, Shukuya R and Kaziro K 1966 Quaternary structure of *Limnodrilus* haemoglobin; *J. Mol. Biol.* **21** 467–472

Zal F, Lallier F, Green B N, Vinogradov S N and Toulmond A 1996 The multi-hemoglobin system of the hydrothermal vent tube worm *Riftia pachyptila*. II. Complete polypeptide chain composition investigated by maximum entropy analysis of mass spectra; *J. Biol. Chem.* **271** 8875–8881

Zal F, Green B N, Lallier F and Toulmond A 1997a Investigation by electrospray ionization mass spectrometry of the extracellular hemoglobin from the polychaete annelid *Alvinella pompejana*: an unusual hexagonal bilayer hemoglobin; *Biochemistry* **36** 11777–11786

Zal F, Green B N, Lallier F, Vinogradov S N and Toulmond A 1997b Quaternary structure of the extracellular haemoglobin of the lugworm *Arenicola marina*: a multi-angle-laser-light-scattering and electrospray-ionisation mass-spectrometry analysis; *Eur. J. Biochem.* **243** 85–92

Zal F, Green B N, Martineu P, Lallier F, Toulmond A, Vinogradov S N and Childress J J 2000 Polypeptide chain composition diversity of hexagonal bilayer haemoglobins within a single family of annelids, the Alvinellidae; *Eur. J. Biochem.* **267** 5227–5236

Zhu H, Ownby D, Riggs C, Nolasco N, Stoops J and Riggs A 1996 Assembly of the gigantic hemoglobin of the earthworm *Lumbricus terrestris*; *J. Biol. Chem.* **271** 30007–30021

MS received 1 February 2003; accepted 22 April 2003

Corresponding editor: DIPANKAR CHATTERJI