

Combinatorics of giant hexagonal bilayer hemoglobins

Leonid G. Hanin^{a,*}, Serge N. Vinogradov^b

^a *Department of Mathematics, Idaho State University, Pocatello, ID 83209-8085, USA*

^b *Department of Biochemistry and Molecular Biology, School of Medicine, Wayne State University, Detroit, MI 48201, USA*

Received 29 March 1999; accepted 9 September 1999

Abstract

The paper discusses combinatorial and probabilistic models allowing to characterize various aspects of spacial symmetry and structural heterogeneity of the giant hexagonal bilayer hemoglobins (HBL Hb). Linker–dodecamer configurations of HBL are described for two and four linker types (occurring in the two most studied HBL Hb of *Arenicola* and *Lumbricus*, respectively), and the most probable configurations are found. It is shown that, for HBL with marked dodecamers, the number of ‘normal-marked’ pairs of dodecamers in homological position follows a binomial distribution. The group of symmetries of the dodecamer substructure of HBL is identified with the dihedral group D_6 . Under natural symmetry assumptions, the total dipole moment of the dodecamer substructure of HBL is shown to be zero. Biological implications of the mathematical findings are discussed. © 2000 Elsevier Science Inc. All rights reserved.

Keywords: Dipole moment; Group of symmetries; Hemoglobin; Hexagonal bilayer; Linker–dodecamer configuration; Probability distribution

1. Introduction

The giant extracellular hexagonal bilayer (HBL) hemoglobins (Hb) represent a summit of complexity for oxygen-binding heme proteins [1,2]. They occur in most terrestrial, aquatic and marine annelids, represented by about 8000 species, as well as in deep sea annelids and vestimentiferans. These proteins were among the first investigated by ultracentrifugation and electron

* Corresponding author. Tel.: +1-208 236 3293; fax: +1-208 236 2636.

E-mail address: hanin@isu.edu (L.G. Hanin).

microscopy. The most extensively studied HBL Hb is that of the common North American earthworm *Lumbricus terrestris* with the molecular mass of ~ 3.6 MDa.

HBL Hbs have the following biochemical features [2–5]:

- high cooperativity [6, p. 223] of oxygen binding;
- characteristically low iron and heme [6, p. 186] content, about 2/3 of normal;
- acidic isoelectric point [6, p. 48];
- sedimentation coefficient [6, p. 157] of about 60 S (Svedberg units).

Studies of the quaternary structure of various HBL Hbs based on the methods of scanning transmission electron microscopy (STEM) and electrospray ionization mass spectrometry (ESI-MS) revealed that HBL Hbs in different species are remarkably similar and highly symmetric. A ‘bracelet’ model of their structure was proposed to consist of 12 complexes of ~ 200 kDa globin subunits called *dodecamers* tethered to a central scaffolding of non-globin *linker* chains ranging in mass from ~ 24 to ~ 32 kDa [7]. The 12 dodecamers account for $\sim 72\%$ of the total mass, in excellent agreement with the iron and heme content measurements. STEM analysis of negatively stained HBLs gave the height and diameter of the ‘bracelet’ to be ~ 20 and 30 nm, respectively. The essential correctness of the ‘bracelet’ model was demonstrated recently by 3D reconstructions of cryoEM images in vitreous ice [8,9]. In particular, it was shown that the centers of dodecamers form two regular parallel superimposed hexagons rotated on the angle α of about 16° with respect to one another.

Each dodecamer was found to consist of three identical disulfide-bonded trimers (T) and three monomers (M) [10]. A direct mass spectrometric evidence for the existence of dodecamer subunits has been recently obtained in [11]. All 12 globular substructures forming the hexagonally symmetric bilayer were found to have a local 3-fold axis of symmetry, in agreement with the symmetry observed for the dodecamer subunit crystals [12] and expected on the basis of its T_3M_3 structure [10].

An ESI-MS study of *Lumbricus* Hb has identified seven different linker chains: L1a and L1b (27 702 and 27 540 Da), L2 (32 104 Da), L3 (24 912 Da), and L4a, L4b, L4c (24 170, 24 102 and 24 019 Da), respectively [13]. However, the number of linker subunit types in a particular HBL Hb may be less than seven. For, example, Hb of the marine polychaete *Arenicola* contains two types of linkers, while for the terrestrial oligochaete *Lumbricus* the number of linker types in the Hb is four. The role of the linker subunits is predominantly structural; they are required for reassembly of the dodecamer subunits into HBL structures. Furthermore, the total number of linker chains in *Lumbricus* Hb is probably 36, based on the known 6-fold symmetry of the HBL structure and the good agreement between the calculated mass of the latest model of its quaternary structure and the experimental masses obtained by STEM mass mapping and sedimentation equilibrium measurements [13]. Results of the experiments with dissociation and subsequent reassembly of HBL Hbs suggest that binding affinities to dodecamers may differ for various linker types [14].

The present paper addresses the following combinatorial problems related to HBL spacial symmetry and structural heterogeneity, focusing primarily on methodological and computational aspects.

1. *Distribution of linker–dodecamer configurations in HBLs.* The main structural parameters of HBL are the total number of linkers, the number of linker types, and affinities of various types of linkers to dodecamers.

We proceed from the following assumptions:

- (1) The total number of linkers in an HBL is 36.
- (2) The number of types of linkers in an HBL is either 2 (as in *Arenicola*) or 4 (as in *Lumbricus*).
- (3) Each of the 12 constituent dodecamers of an HBL is associated with three linkers.

The distributions of linker–dodecamer configurations and the most probable configurations are described in Section 2. It should be mentioned, however, that the methodological approach developed there is equally applicable to the cases when the total number of linkers and the number of their types is different from those assumed above, and the number of linkers per dodecamer is variable.

Analysis of the kinetics of dissociation of *Lumbricus* Hb at neutral pH in the presence of dissociating agents such as urea and guanidinium hydrochloride [15] has demonstrated them to be heterogeneous, consisting generally of three distinct phases with half-times of ~ 2 , 30–50 and 400–500 h⁻¹, respectively. Our hypothesis is that this kinetic heterogeneity is a reflection of a structural heterogeneity within the overall population of HBL structures in the native Hb, and that a possible structural heterogeneity is due to heterogeneity in the linker environment of the dodecamer subunits. If these assumptions are met, then matching similar data for HBLs re-assembled from dodecamers in a controlled medium of linkers with the distributions of linker configurations computed below allows for estimation of relative affinity of various linker types to dodecamers. With this information at hand, the distribution of different linker configurations and the prevailing linker configurations can be predicted for both native and reassembled HBLs.

2. *The distribution of dodecamer pairs ‘normal-marked’ in homological position for HBL containing marked dodecamers.* This distribution is important for planning and interpreting results of the following experiments:

- reassembling hybrid HBL structures with dodecamers from different species;
- study of the fluorescent energy transfer between neighboring dodecamers, one of which is chemically modified with a fluorescent label, during conversion of one state of ligation into another (that is, oxy to deoxy and back).

It is shown in Section 3 that the number of ‘normal-marked’ pairs of dodecamers in homological position follows a binomial distribution $B(6, q)$, where parameter q depends on the probability for a dodecamer to be marked.

3. *Group of symmetries of the dodecamer substructure of HBL.* Processing of cryoEM images of HBLs and reconstruction of native and reassembled HBLs require knowledge of the group of symmetries of the dodecamer substructure of HBL. In Section 4, this group is defined, described, and identified with the classical dihedral group D_6 .

4. *The dipole moment of HBL.* Comparison of the total dipole moment of *Lumbricus* HBL Hb (~ 18000 D) with the dipole moment of a single dodecamer subunit (~ 1200 D) suggests that linker subunits contribute significantly to the dipole moment of HBL [16]. Moreover, it will be shown in Section 5 that under natural symmetry assumptions the total dipole moment of the HBL substructure formed by 12 dodecamers is equal to zero. In particular, if this is the case, the average dipole moment of a linker in *Lumbricus* HBL Hb is ~ 500 D. It appears likely that the large experimental dipole moment of native *Lumbricus* HBL Hb is due to asymmetry in the spacial distribution of linker subunits.

2. Distribution of linker–dodecamer configurations

2.1. HBL with two linker types

Consider HBLs with two types of linkers. These types will be denoted by 1 and 2 and graphically represented by balls carrying these numbers. The four possible combinations of three linkers of the two types associated with a dodecamer are shown and arranged lexicographically in Fig. 1. Let k_1, k_2, k_3, k_4 be the numbers of dodecamers in an HBL attached to which are linker triplets of these four respective types. Vector $k = (k_1, k_2, k_3, k_4)$ with integer components completely characterizes linker–dodecamer binding in HBL and will be called a *configuration*. Denote by \mathcal{K} the set of all configurations. For every $k \in \mathcal{K}$, one has $k_j \geq 0$, $j = 1, 2, 3, 4$,

$$k_1 + k_2 + k_3 + k_4 = 12, \quad (1)$$

and

$$3k_1 + 2k_2 + k_3 = i, \quad (2)$$

where i stands for the total number of linkers of type 1, see Fig. 1. Observe that the relation $k_2 + 2k_3 + 3k_4 = 36 - i$ representing the number of linkers of type 2 follows from (1) and (2).

An HBL is thought of as assembled from dodecamers in a medium containing large equal amounts of linkers of types 1 and 2. Denote by λ the ratio of the affinity to dodecamer of linkers of type 2 to that of linkers of type 1. Then the probabilities for a linker binded to a dodecamer to be of type 1 and 2 are equal to $p = 1/(\lambda + 1)$ and $1 - p = \lambda/(\lambda + 1)$, respectively. Assume that linkers are attached to dodecamers independently of each other. Then the random number I of linkers of type 1 in an HBL thus assembled has the binomial distribution $B(36, p)$:

$$P(I = i) = \binom{36}{i} p^i (1 - p)^{36-i}, \quad p = \frac{1}{\lambda + 1}, \quad 0 \leq i \leq 36. \quad (3)$$

Given the number i of linkers of the first type, we now compute the probability of a configuration $k = (k_1, k_2, k_3, k_4)$ such that $3k_1 + 2k_2 + k_3 = i$. All these configurations have equal probabilities and can be visualized as follows. Think of a dodecamer as a ‘box’ with three ‘nests’ for balls (linkers). Then HBL can be represented by a linear sequence of 12 boxes in which 36 balls are accommodated (i balls of type 1 and $36 - i$ balls of type 2). The total number N_i of such arrangements is equal to the number of choices of i numbers out of 36 (positions of balls of type 1), that is

1	2	3	4
①	①	①	②
①	①	②	②
①	②	②	②

Fig. 1. Linker triplets for two linker types.

$$N_i = \binom{36}{i}.$$

Then the probability of a configuration k is given by $P(k) = n(k)/N_i$, where $n(k)$ is the number of arrangements with k_1, k_2, k_3, k_4 boxes having content of the respective types 1–4 shown in Fig. 1. Observe that all these $n(k)$ arrangements can be obtained from one of them using the following operations:

(1) permutation of boxes;

(2) permutation of balls in the boxes with the content of types 2 and 3.

The number of permutations of 12 boxes with k_j boxes having content of type j , $j = 1, 2, 3, 4$, is given by

$$\frac{12!}{k_1!k_2!k_3!k_4!}.$$

Further, the number of permutations of balls in a box with the content of types 2 or 3 is obviously equal to 3. Hence,

$$n(k) = \frac{12!}{k_1!k_2!k_3!k_4!} 3^{k_2+k_3}.$$

Therefore, for the conditional probability $P_i(k) := \Pr(k|I = i)$ of a configuration k given the number $i = 3k_1 + 2k_2 + k_3$ of linkers of type 1 we have

$$P_i(k) = \frac{12! 3^{k_2+k_3}}{\binom{36}{i} k_1!k_2!k_3!k_4!}. \quad (4)$$

Proceeding from the formula $P(k) = \Pr(k|I = i)\Pr(I = i)$, we compound (3) and (4) to obtain the distribution of configurations for an HBL with two linker types and a given ratio λ of affinities to dodecamer:

$$P(k) = \frac{12! 3^{k_2+k_3} \lambda^{k_2+2k_3+3k_4}}{(\lambda + 1)^{36} k_1!k_2!k_3!k_4!}, \quad k \in \mathcal{K}. \quad (5)$$

Configurations $k(\lambda)$ with the maximal probability $P(\lambda)$, that were computed using a MAPLE based program for $\lambda = 1, 2, 3, 5$, are shown in Table 1.

Practical use of formula (5), however, is hampered by the fact that the number of configurations is large. To compute it, observe that according to (1), the number of elements in the set \mathcal{K} is equal to the number of ways to represent 12 as a sum of four nonnegative integers with the order of summands taken into account. More generally, we will compute the number $N(n, r)$ of similar

Table 1

The most probable linker–dodecamer configurations for different values of the ratio λ of linker affinities to dodecamer

λ	$k(\lambda)$	$P(\lambda)$
1	(1, 5, 5, 1)	0.0286
2	(0, 2, 6, 4)	0.0407
3	(0, 1, 5, 6)	0.0587
5	(0, 0, 4, 8)	0.0905

representations of any natural number n as a sum of r non-negative integers. Thinking of n as a string of n units we see that every representation of n as a sum of r terms is uniquely determined by selection of places for $r - 1$ delimiters breaking the string of units into r blocks. The latter can be done in

$$N(n, r) = \binom{n+r-1}{r-1}$$

ways. In particular, the total number of HBL configurations is equal to

$$N(12, 4) = \binom{15}{3} = 455.$$

Given the affinity ratio λ , one can consider HBLs with *the most probable* number $m = m(\lambda)$ of linkers of type 1. This model leads to a smaller number of configurations and serves as an alternative to the model of random assembly of HBL described above.

Since the distribution of the random number I of linkers of the first type in HBL is $B(36, p)$ with $p = 1/(\lambda + 1)$, the most probable value m of I is subject to the inequalities

$$(n+1)p - 1 \leq m \leq (n+1)p$$

as easily follows by comparing the probabilities (3) for the values of i neighboring with m . Setting $l := np = n/(\lambda + 1)$ we rewrite these inequalities in the form

$$l - \frac{\lambda}{\lambda + 1} \leq m \leq l + \frac{1}{\lambda + 1}.$$

If l is an integer then this implies that $m = l$ in which case

$$\frac{n-m}{m} = \lambda$$

as is quite natural to expect. In particular, for $n = 36$ and $\lambda = 1, 2, 3, 5$, we have $m = 18, 12, 9, 6$, respectively. These are the cases for which the distributions (4) of linker–dodecamer configurations with $i = m$ are described in what follows.

(1) $\lambda = 1$. In this case HBL contains $m = 18$ linkers of the first type and 18 linkers of the second type. Linker–dodecamer configurations are described by the system of equations (1) and (2) with $i = 18$. This system has 25 non-negative integer solutions arranged in a lexicographic order in Table 2. It follows that in this case the set of configurations consists of nine symmetric pairs (2 and 5, 3 and 9, 4 and 14, 7 and 10, 12 and 16, 13 and 19, 18 and 20, 22 and 23) and of seven configurations (1, 6, 11, 17, 21, 24, 25) which are self-symmetric.

The results of computation of the probabilities $P_m(k)$ given by Eq. (3) for $m = 18$ are shown in Table 2. They suggest that configuration (1, 5, 5, 1) has the highest probability of about 0.2165. This configuration also serves as a maximizer for the distribution (5) of *all* configurations with $\lambda = 1$, as follows from the comparison of distributions (4) with $i = 18$ and (5) with $\lambda = 1$, see also Table 1. Configuration (2, 4, 4, 2) has the second largest probability of about 0.150, while the third largest probability of about 0.120 occurs for a pair of symmetric configurations (1, 6, 3, 2) and (2, 3, 6, 1). The total probability of these four dominating configurations is about 0.61.

(2) $\lambda = 2$. This case deals with $m = 12$ linkers of type 1 and 24 linkers of type 2. Such configurations are described by the system of equations (1) and (2) with $i = 12$. This system has 19

Table 2

Distribution of HBL configurations for two linker types with $\lambda = 1$ ($m = 18$)

No.	k	$P_m(k)$
1	(0, 6, 6, 0)	0.0541
2	(0, 7, 4, 1)	0.0773
3	(0, 8, 2, 2)	0.0193
4	(0, 9, 0, 3)	0.0005
5	(1, 4, 7, 0)	0.0773
6	(1, 5, 5, 1)	0.2165
7	(1, 6, 3, 2)	0.1203
8	(1, 7, 1, 3)	0.0115
9	(2, 2, 8, 0)	0.0193
10	(2, 3, 6, 1)	0.1203
11	(2, 4, 4, 2)	0.1504
12	(2, 5, 2, 3)	0.0401
13	(2, 6, 0, 4)	0.0011
14	(3, 0, 9, 0)	0.0005
15	(3, 1, 7, 1)	0.0115
16	(3, 2, 5, 2)	0.0401
17	(3, 3, 3, 3)	0.0297
18	(3, 4, 1, 4)	0.0037
19	(4, 0, 6, 2)	0.0011
20	(4, 1, 4, 3)	0.0037
21	(4, 2, 2, 4)	0.0019
22	(4, 3, 0, 5)	0.0001
23	(5, 0, 3, 4)	0.0001
24	(5, 1, 1, 5)	0.0000
25	(6, 0, 0, 6)	0.0000

solutions given in Table 3. The most probable configuration with the probability of about 0.291 is (3, 6, 3, 0). It is worth noting that it differs significantly from the corresponding most probable configuration (0, 2, 6, 4) for the distribution (5) with $\lambda = 2$, see Table 1, for which the number of linkers of type 1 is equal to 10. Configurations with the next three largest probabilities are (4, 4, 4, 0), (4, 5, 2, 1) and (2, 8, 2, 0), see Table 3; they carry the probabilities of approximately 0.182, 0.145, and 0.140, respectively. The total probability of the four dominating configurations is about 0.76.

(3) $\lambda = 3$. In this case there are $m = 9$ linkers of type 1 and 27 linkers of type 2. All their configurations are described by the system of equations (1) and (2) with $i = 9$. This system has 12 solutions presented in Table 4. The most probable configuration with the probability of about 0.386 is (5, 5, 2, 0). Observe that the most probable configuration of the distribution (5) with $\lambda = 3$ is (0, 1, 5, 6), see Table 1, for which $i = 7$. The next two most probable configurations are (4, 7, 1, 0) and (6, 3, 3, 0) with the probabilities of approximately 0.276 and 0.143, respectively, see Table 4. The total probability of the three most probable configurations is about 0.81.

(4) $\lambda = 5$. This case deals with $m = 6$ linkers of type 1 and 30 linkers of type 2 whose configurations are described by the system of equations (1) and (2) with $i = 6$. This system has seven solutions shown in Table 5. The most probable configuration with the probability of about 0.494 is (7, 4, 1, 0), to be compared with the most probable configuration (0, 0, 4, 8) for the distribution (5) with $\lambda = 5$, for which $i = 4$, see Table 1. The second highest probability of about 0.346 occurs

Table 3

Distribution of HBL configurations for two linker types with $\lambda = 2$ ($m = 12$)

No.	k	$P_m(k)$
1	(0, 12, 0, 0)	0.0004
2	(1, 10, 1, 0)	0.0187
3	(2, 8, 2, 0)	0.1401
4	(2, 9, 0, 1)	0.0104
5	(3, 6, 3, 0)	0.2906
6	(3, 7, 1, 1)	0.0830
7	(4, 4, 4, 0)	0.1816
8	(4, 5, 2, 1)	0.1453
9	(4, 6, 0, 2)	0.0081
10	(5, 2, 5, 0)	0.0291
11	(5, 3, 3, 1)	0.0646
12	(5, 4, 1, 2)	0.0161
13	(6, 0, 6, 0)	0.0005
14	(6, 1, 4, 1)	0.0054
15	(6, 2, 2, 2)	0.0054
16	(6, 3, 0, 3)	0.0004
17	(7, 0, 3, 2)	0.0002
18	(7, 1, 1, 3)	0.0001
19	(8, 0, 0, 4)	0.0000

Table 4

Distribution of HBL configurations for two linker types with $\lambda = 3$ ($m = 9$)

No.	k	$P_m(k)$
1	(3, 9, 0, 0)	0.0460
2	(4, 7, 1, 0)	0.2760
3	(5, 5, 2, 0)	0.3864
4	(5, 6, 0, 1)	0.0429
5	(6, 3, 3, 0)	0.1431
6	(6, 4, 1, 1)	0.0716
7	(7, 1, 4, 0)	0.0102
8	(7, 2, 2, 1)	0.0204
9	(7, 3, 0, 2)	0.0023
10	(8, 0, 3, 1)	0.0006
11	(8, 1, 1, 2)	0.0006
12	(9, 0, 0, 3)	0.0000

for the configuration (6, 6, 0, 0), see Table 5. The total probability of the two dominating configurations is about 0.84.

2.2. HBL with four linker types

We now consider HBLs containing linkers of four types. These linker types will be represented graphically by ‘balls’ numbered 1, 2, 3, 4. In these terms, a linker configuration can be identified with a disposition of 36 balls of four types in 12 boxes with three balls going into each box. All 20 possibilities for the content of a box (that is, for a triplet of linkers attached to a dodecamer) are shown in Fig. 2.

Table 5

Distribution of HBL configurations for two linker types with $\lambda = 5$ ($m = 6$)

No.	k	$P_m(k)$
1	(6, 6, 0, 0)	0.3458
2	(7, 4, 1, 0)	0.4940
3	(8, 2, 2, 0)	0.1235
4	(8, 3, 0, 1)	0.0274
5	(9, 0, 3, 0)	0.0030
6	(9, 1, 1, 1)	0.0061
7	(10, 0, 0, 2)	0.0000

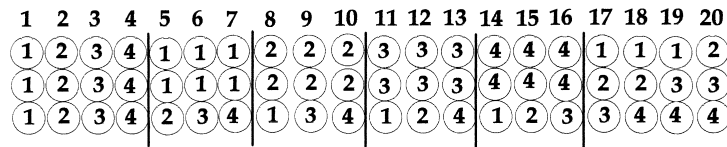


Fig. 2. Linker triplets for four linker types.

As in Section 2.1, we are assuming that HBL is assembled in a medium containing large equal amounts of linkers of the four types, and that all linkers are binded to dodecamers independently of each other. Let p_j be the probability for a linker binded to a dodecamer to be of type j , $j = 1, 2, 3, 4$, then $p_1 + p_2 + p_3 + p_4 = 1$. The ratios of these probabilities represent relative affinities of linkers of the four types to dodecamer.

Denote by I the four-dimensional integer-valued random vector whose components are the number of linkers of the four respective types in a randomly assembled HBL. It follows from our assumptions that the random vector I follows the multinomial distribution $B(36, p_1, p_2, p_3, p_4)$:

$$Pr(I = i) = \frac{36!}{i_1! i_2! i_3! i_4!} p_1^{i_1} p_2^{i_2} p_3^{i_3} p_4^{i_4}, \quad i = (i_1, i_2, i_3, i_4) \in \mathbb{Z}_+^4, \quad i_1 + i_2 + i_3 + i_4 = 36. \quad (6)$$

The structure of HBL is determined by the configuration vector $k = (k_1, k_2, \dots, k_{20})$ showing the number of dodecamers attached to which are triplets of linkers of the 20 respective types given in Fig. 2. Indeed, $k_j \geq 0$, $j = 1, 2, \dots, 20$, and

$$k_1 + k_2 + \dots + k_{20} = 12. \quad (7)$$

Hence the total number of configurations is

$$N(12, 20) = \binom{31}{19} = 141\,120\,525. \quad (8)$$

Counting the number of linkers of the types 1–4 yields the following equations:

$$3k_1 + 2k_5 + 2k_6 + 2k_7 + k_8 + k_{11} + k_{14} + k_{17} + k_{18} + k_{19} = i_1, \quad (9.1)$$

$$3k_2 + k_5 + 2k_8 + 2k_9 + 2k_{10} + k_{12} + k_{15} + k_{17} + k_{18} + k_{20} = i_2, \quad (9.2)$$

$$3k_3 + k_6 + k_9 + 2k_{11} + 2k_{12} + 2k_{13} + k_{16} + k_{17} + k_{19} + k_{20} = i_3, \quad (9.3)$$

$$3k_4 + k_7 + k_{10} + k_{13} + 2k_{14} + 2k_{15} + 2k_{16} + k_{18} + k_{19} + k_{20} = i_4. \quad (9.4)$$

Observe that Eq. (7) follows by adding Eqs. (9.1)–(9.4).

Using the method of Section 2.1, we will compute the probability $P(k)$ of a configuration k . Given the vector $i = (i_1, i_2, i_3, i_4)$ representing the number of linkers of the four types, the probability of a configuration k satisfying (9.1)–(9.4) equals

$$P(k|I = i) = \frac{n(k)}{N_i},$$

where

$$N_i = \binom{36}{i} = \frac{36!}{i_1!i_2!i_3!i_4!}$$

is the total number of arrangements of i_1, i_2, i_3, i_4 balls of four respective types, $i_1 + i_2 + i_3 + i_4 = 36$, in 12 boxes with three balls going into each box, and $n(k)$ is the number of arrangements with the content of boxes described by the vector k . All such arrangements can be obtained from one of them by, first, permuting boxes and, second, permuting balls within a box. The number of permutations of 12 boxes is

$$\frac{12!}{k_1!k_2! \cdots k_{20}}.$$

The number of permutations of balls within a box is 3 for a box with the content of types 5–16, and is 6 for boxes with the content of types 17–20, see Fig. 2. Therefore,

$$Pr(k|I = i) = \frac{12!}{\binom{36}{i} k_1!k_2! \cdots k_{20}} 3^{k_5+k_6+\cdots+k_{16}} 6^{k_{17}+k_{18}+k_{19}+k_{20}},$$

compare with (4). Compounding this with (6) we obtain the distribution of linker–dodecamer configurations to be

$$P(k) = \frac{12!}{k_1!k_2! \cdots k_{20}} 3^{k_5+k_6+\cdots+k_{16}} 6^{k_{17}+k_{18}+k_{19}+k_{20}} p_1^{i_1} p_2^{i_2} p_3^{i_3} p_4^{i_4}, \quad (10)$$

where i_1, i_2, i_3, i_4 are related to k through (9.1)–(9.4). Owing to (7), formula (10) can be rewritten in the form

$$P(k) = \frac{12!3^{12}2^{k_{17}+k_{18}+k_{19}+k_{20}}}{3^{k_1+k_2+k_3+k_4}k_1!k_2! \cdots k_{20}} p_1^{i_1} p_2^{i_2} p_3^{i_3} p_4^{i_4}. \quad (11)$$

In what follows, we assume for simplicity that $p_1 = p_2 = p_3 = p_4 = 1/4$, that is, linkers of all four types have equal affinities to dodecamer. In this case, (11) reduces to

$$P(k) = \frac{12!3^{12}2^{k_{17}+k_{18}+k_{19}+k_{20}}}{4^{36}3^{k_1+k_2+k_3+k_4}k_1!k_2! \cdots k_{20}}. \quad (12)$$

The immense number (8) of configurations calls for a simplified model of HBL formation. Following the idea that was employed in Section 2.1, we assume the vector i of the numbers of linker types to take the value which is the *most probable* for given affinities of linkers of the four

types to dodecamer. In the case of equal affinities, this obviously implies that $i_1 = i_2 = i_3 = i_4 = 9$. Then the distribution of configurations is given by (12), where the set \mathcal{K} of all admissible configurations is described by the system of equations:

$$3k_1 + 2k_5 + 2k_6 + 2k_7 + k_8 + k_{11} + k_{14} + k_{17} + k_{18} + k_{19} = 9, \quad (13.1)$$

$$3k_2 + k_5 + 2k_8 + 2k_9 + 2k_{10} + k_{12} + k_{15} + k_{17} + k_{18} + k_{20} = 9, \quad (13.2)$$

$$3k_3 + k_6 + k_9 + 2k_{11} + 2k_{12} + 2k_{13} + k_{16} + k_{17} + k_{19} + k_{20} = 9, \quad (13.3)$$

$$3k_4 + k_7 + k_{10} + k_{13} + 2k_{14} + 2k_{15} + 2k_{16} + k_{18} + k_{19} + k_{20} = 9. \quad (13.4)$$

The number of configurations in \mathcal{K} , that is, the number of solutions of the system of equations (13.1)–(13.4), that was computed using MAPLE turns out to be 183 712.

Next, we are going to find the number of the most probable configurations in the set \mathcal{K} , that will be referred to as *extremal configurations*, and their total probability. Note that (12) can be written as

$$P(k) = C\varphi(k),$$

where

$$C = \frac{12!3^{12}}{4^{36}} \simeq 5.3 \times 10^{-8},$$

and

$$\varphi(k) = \frac{2^{k_{17}+k_{18}+k_{19}+k_{20}}}{3^{k_1+k_2+k_3+k_4} k_1! k_2! \dots k_{20}!}.$$

To maximize $\varphi(k)$ over $k \in \mathcal{K}$, observe that the sequence $2^n/n!$, $n \in \mathbb{Z}_+$, takes on the maximum value 2 for $n = 1, 2$. Therefore,

$$\varphi(k) \leq \frac{2^{k_{17}}}{k_{17}!} \frac{2^{k_{18}}}{k_{18}!} \frac{2^{k_{19}}}{k_{19}!} \frac{2^{k_{20}}}{k_{20}!} \leq 2^4 = 16,$$

and $\varphi(k) = 16$ if and only if

$$k_1 = k_2 = k_3 = k_4 = 0; \quad k_5, k_6, \dots, k_{16} = 0 \text{ or } 1; \quad k_{17}, k_{18}, k_{19}, k_{20} = 1 \text{ or } 2.$$

Using this criterion and the system of equations (13.1)–(13.4), a MAPLE based program computed the number of extremal configurations to be 240. Therefore, the total probability of extremal configurations is $240 \times 16C \simeq 2 \times 10^{-4}$. This suggests that in HBLs with four linker types the probabilities of linker configurations are spread more uniformly than in HBLs with two linker types.

3. The distribution of the number of ‘normal-marked’ pairs of dodecamers in homological position

We assume that 12 dodecamers of an HBL are marked with the same probability p , $0 < p < 1$, independently of each other. Two dodecamers in different hexagonal layers are said to be in

homological position if, after the rotation of one of the layers through the angle $\alpha \simeq 16^\circ$ about the axis through the centers of the two layers, the centers of these dodecamers become ‘superimposed,’ that is, belong to a line perpendicular to the planes of the layers. Obviously, HBL contains six pairs of dodecamers in homological position. These pairs are of two types: heterogeneous (i.e. consisting of one normal and one marked dodecamer) and homogeneous (i.e. consisting of two normal or two marked dodecamers).

Since dodecamers are marked independently, the probability for a pair of dodecamers in homological position to be heterogeneous is $q := 2p(1 - p)$. The same assumption implies that the types of six pairs of dodecamers in homological position are independent. Therefore, the number X of heterogeneous pairs of dodecamers in homological position follows the binomial distribution $B(6, q)$:

$$Pr(X = n) = \binom{6}{n} q^n (1 - q)^{6-n} = \binom{6}{n} (2p)^n (1 - p)^n (1 - 2p + 2p^2)^{6-n}, \quad 0 \leq n \leq 6.$$

In particular, if HBL is assembled in a medium containing large equal amounts of normal and marked dodecamers then $p = 1/2$. In this case, $q = 1/2$, and hence X has symmetrical binomial distribution $B(6, 1/2)$:

$$Pr(X = n) = \binom{6}{n} / 64, \quad 0 \leq n \leq 6.$$

4. Group of symmetries of the dodecamer substructure of HBL

The dodecamer substructure of HBL is represented by the set S of 12 points in the space thought of as centers of dodecamers. These points form two congruent parallel superimposed regular hexagons H_1 and H_2 , one of which is rotated on the angle $\alpha \simeq 16^\circ$ with respect to the other.

A *symmetry* of a finite set $A \subset \mathbb{R}^3$ is defined as an isometry $U : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ that induces a permutation of the set A . The collection $G(A)$ of all symmetries of the set A is a group with respect to the composition of isometries, termed the *group of symmetries* of the set A . The unit element e of this group is the identical isometry I of \mathbb{R}^3 .

Let O_1 and O_2 be the centers of the two hexagons, and let O be the midpoint of the segment O_1O_2 (the ‘center’ of the set S). Let a be the counter-clockwise rotation of \mathbb{R}^3 through the angle 60° about the axis O_1O_2 . Clearly, $a \in G(S)$ and $a^6 = e$. Observe that symmetries e, a, a^2, a^3, a^4, a^5 of S are all distinct and represent rotations of \mathbb{R}^3 about the axis O_1O_2 through the angles that are multiples of 60° .

Shown in Fig. 3 are projections P_1 (marked white) and P_2 (marked black) of the two hexagons on the ‘central’ plane of S , that is, the plane through the center O of the set S which is parallel to the planes of the hexagons. Let b be the rotation of \mathbb{R}^3 on the angle 180° about the line through a pair of intersection points of the two hexagons P_1 and P_2 that are symmetric about the center O (line l on Fig. 3). This rotation induces a permutation of S that interchanges the two hexagons. Therefore, $b \in G(S)$. Indeed, we have $b^2 = e$. Note that symmetries $b, ab, a^2b, a^3b, a^4b, a^5b$ are all different and also distinct from a^i , $0 \leq i \leq 5$.

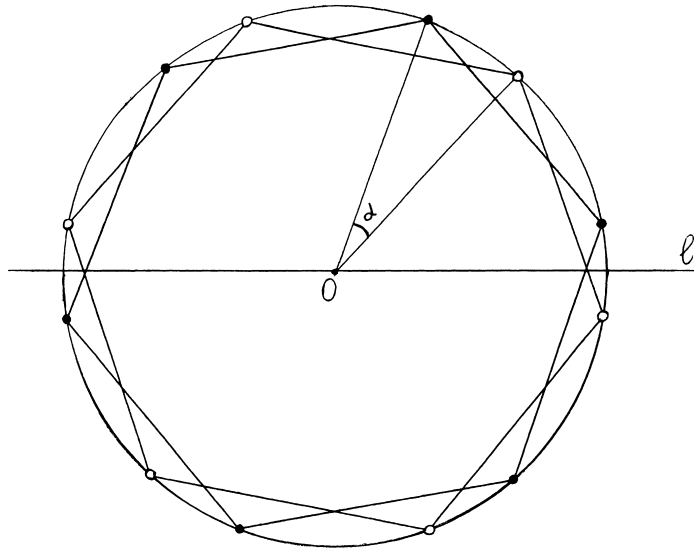


Fig. 3. Projections of the hexagonal layers of HBL on the central plane.

Denote by G the subgroup of $G(S)$ generated by a and b . Since $a^{-1} = a^5$ and $b^{-1} = b$, every element of G is an intermittent product of the element b and elements a^i , $0 \leq i \leq 5$. From Fig. 3 we find by inspection that $aba = b$ and hence $bab = a^5$. Therefore, every such product is either a^i or ba^i , $0 \leq i \leq 5$. This shows that

$$G = \{e, a, a^2, a^3, a^4, a^5, b, ab, a^2b, a^3b, a^4b, a^5b\}, \quad (14)$$

which coincides with the classical dihedral group D_6 with two generators a and b satisfying the relations $a^6 = e$, $b^2 = e$, $aba = b$.

We claim that $G(S) = G$. To prove this, we have to show that every isometry U of \mathbb{R}^3 that induces a permutation of S coincides with one of those listed in (14).

Observe that point O is equidistant from all points of the set S ; moreover, this is the only point in \mathbb{R}^3 with this property. Since $U(S) = S$, the point $U(O)$ is equidistant from all points of S . This implies that $U(O) = O$, and hence U is a linear map with respect to the point O . The image $U(H_1)$ of the hexagon H_1 belongs to a plane that contains six points of the set S . Therefore, $U(H_1) = H_1$ or $U(H_1) = H_2$.

Consider the case $U(H_1) = H_1$. Then also $U(H_2) = H_2$. Since O_1 is the only point in the plane of the hexagon H_1 that is equidistant from all points of H_1 , we have $U(O_1) = O_1$. Therefore, the line O_1O_2 is mapped by U into itself, and every point of this line is a fixed point of U . Let A_0 be a vertex of H_1 , then $U(A_0) = A_i$ for some i , $0 \leq i \leq 5$. Consequently, isometry $V := a^{-i}U$ fixes both the line O_1O_2 and the point A_0 . Hence, all points of the plane A that is generated by O_1O_2 and A_0 are fixed points of V .

We will show that $V = I$. If this were not the case, then we would have $V(A) \neq A$ for all points A in \mathbb{R}^3 that do not belong to the plane A . Let A be any such point. For $L \in A$, in view of $V(L) = L$, the point L is equidistant from A and $V(A)$. Therefore, L belongs to the plane that is perpendicular to the segment $[A, V(A)]$ and passes through its midpoint. Thus, this plane coincides

with A , so that the points A and $V(A)$ are symmetric about A . This means that V is a reflexion about the plane A . However, the hexagon H_2 is not symmetric about A , because it is rotated with respect to H_1 on an angle which is not a multiple of 60° . This contradicts $V(H_2) = H_2$. Therefore, $V = I$, and hence $U = a^i$.

Consider now the case $U(H_1) = H_2$. For the isometry $W := bU$ we have $W(H_1) = H_1$. Therefore, as shown above, $W = a^i$ for some i , $0 \leq i \leq 5$. This implies that $U = bW = ba^i$, as required.

5. Dipole moment of the dodecamer substructure of HBL

Recall the definition of the dipole moment. Let q_1, q_2, \dots, q_n be point charges located at respective points A_1, A_2, \dots, A_n in the space. The dipole moment \mathbf{M} of this system of charges with respect to a point O in the space is a vector defined by

$$\mathbf{M} := \sum_{i=1}^n q_i \mathbf{OA}_i,$$

where \mathbf{OA}_i is the position vector of the point A_i with respect to the point O , $i = 1, 2, \dots, n$. It follows easily from this definition that if the system of charges is electrically neutral, that is,

$$\sum_{i=1}^n q_i = 0,$$

then the dipole moment is independent of the choice of point O . In particular, this is the case when the system consists of several elementary dipoles with charges $q_i, -q_i$ located at points A_i, B_i , $i = 1, 2, \dots, n$. For such system of charges,

$$\mathbf{M} = \sum_{i=1}^n q_i \mathbf{A}_i \mathbf{B}_i.$$

In other words, the dipole moment of a system of charges composed of elementary dipoles is equal to the sum of their dipole moments:

$$\mathbf{M} = \sum_{i=1}^n \mathbf{M}_i, \quad \text{where } \mathbf{M}_i = q_i \mathbf{A}_i \mathbf{B}_i, \quad i = 1, 2, \dots, n.$$

When computing the dipole moment of the system of 12 dodecamers forming an HBL, we proceed from the following assumptions:

1. Each dodecamer is an elementary dipole with the same numerical value of the dipole moment.
2. The two layers of HBL are parallel and hexagonally symmetric.
3. Up to a rotation of one of the layers on some angle about the axis that passes through its center and is perpendicular to its plane, the layers are symmetric about the central plane parallel to the planes of the layers.

It follows from assumption 1 that the dipole moment of the system of 12 dodecamers is equal to the sum of their dipole moments. By Assumption 2, vectors of the dipole moments of dodecamers in one layer can be obtained from one of them by rotation through the angles which are multiples of 60° about the axis through the center of the layer that is perpendicular to its plane. Therefore,

the resulting vectors of the dipole moments of both layers are orthogonal to these layers. Finally, according to assumption 3, these two vectors are opposite. Thus, under assumptions 1–3, the total dipole moment of the dodecamer substructure of HBL is equal to zero.

Acknowledgements

The authors are grateful to Cathy Kriloff and Robert Fisher for their fruitful discussions of symmetry groups. The authors would also like to thank Boris Hanin for his help in producing figures.

References

- [1] S.N. Vinogradov, D.A. Walz, B. Pohajdak, L. Moens, O.H. Kapp, T. Suzuki, C.N.A. Trotman, *Comput. Biochem. Physiol.* 106B (1993) 1.
- [2] J.N. Lamy, B.N. Green, A. Toulmond, J.S. Wall, R.E. Weber, S.N. Vinogradov, *Chem. Rev.* 96 (1996) 3113.
- [3] S.N. Vinogradov, O.H. Kapp, M. Ohtsuki, in: J. Harris (Ed.), *Electron Microscopy of Proteins*, vol. 3, Academic Press, New York, 1982, p. 135.
- [4] S.N. Vinogradov, in: J. Lamy, J.P. Truchot, R. Gilles (Eds.), *Respiratory Pigments in Animals*, Springer, Berlin, 1985, p. 9.
- [5] T. Gotoh, T. Suzuki, *Zool. Sci.* 7 (1990) 1.
- [6] C.K. Mathews, K.E. vanHolde, *Biochemistry*, Benjamin/Cummings, Redwood City, 1990.
- [7] S.N. Vinogradov, S.D. Lugo, M.G. Mainwaring, O.H. Kapp, A.V. Crewe, *Proc. Natl. Acad. Sci. USA* 83 (1986) 8034.
- [8] M. Schatz, E.V. Orlova, P. Dube, J. Jäger, M. Van Heel, *J. Struct. Biol.* 114 (1995) 28.
- [9] F. De Haas, A.R. Kuchumov, J.-C. Taveau, N. Boisset, S.N. Vinogradov, J.N. Lamy, *Biochemistry* 36 (1997) 7330.
- [10] S.N. Vinogradov, P.K. Sharma, A.N. Qabar, J.S. Wall, J.A. Westrick, J.H. Simmons, S.J. Gill, *J. Biol. Chem.* 266 (1991) 13091.
- [11] B.N. Green, R.S. Bordoli, L.G. Hanin, F.H. Lallier, A. Toulmond, S.N. Vinogradov, *J. Biol. Chem.* 274 (1999) 28206.
- [12] P.D. Martin, K.L. Eisele, M.A. Doyle, A.R. Kuchumov, D.A. Walz, E.G. Arutyunyan, S.N. Vinogradov, B.F.P. Edwards, *J. Mol. Biol.* 255 (1996) 170.
- [13] P.D. Martin, A.R. Kuchumov, B.N. Green, R.W.A. Oliver, E.H. Braswell, J.S. Wall, S.N. Vinogradov, *J. Mol. Biol.* 255 (1996) 154.
- [14] A.R. Kuchumov, J.-C. Taveau, J.N. Lamy, J.S. Wall, R.E. Weber, S.N. Vinogradov, *J. Mol. Biol.* 289 (1999) 1361.
- [15] P.K. Sharma, A.R. Kuchumov, G. Chottard, P.D. Martin, J.S. Wall, S.N. Vinogradov, *J. Biol. Chem.* 271 (1996) 8754.
- [16] S. Takashima, A.R. Kuchumov, S.N. Vinogradov, *Biophys. Chem.* 77 (1999) 27.