

===== TRANSLATIONS OF PUBLISHED ARTICLES =====

UDC: 519

## The Study of Interhelical Angles in the Structural Motifs Formed by Two Helices

Tikhonov D.A.<sup>\*1</sup>, Kulikova L.I.<sup>†1</sup>, Efimov A.V.<sup>2</sup>

<sup>1</sup>*Institute of Mathematical Problems of Biology RAS – the Branch of Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Pushchino, Russia*

<sup>2</sup>*Institute of Protein Research, Russian Academy of Sciences, Pushchino, Russia*

**Abstract.** In this paper a statistical analysis of distributions of inter-helical angles in pairs of consecutive and connected  $\alpha$ -helices in spatial structures of proteins are presented. A number of rules for selection of the helical pairs from a set of protein structures obtained from Protein Data Bank (PDB) are developed. The set of helical pairs has been analyzed for the purpose of classification and finding out the features of protein structural organization. All the pairs of connected helices are divided into three subsets according to the criterion of crossing of projections of the helices on parallel planes, which pass through the axes of the helices. It is shown that the distribution of all the types of helical pairs, whose projections do not cross each other, covers almost the entire range of inter-helical angles. The distribution has a single minimum which is close to right angle. Most pairs in this set constitute helical pairs consisting of  $\alpha$ - and  $3_{10}$ -helices, and most pairs with the crossing projections of helices are helical pairs formed by two  $\alpha$ -helices. It is also shown that a great amount of the pairs of connected  $\alpha$ -helices has acute angle ( $20^\circ \leq \varphi \leq 60^\circ$ ) between the axes of the helices. The distribution of all the types of helical pairs depending on the length of the inter-helical connections is also analyzed. It is shown that the structures with short connections occur most often in all the subsets.

**Key words:** structural motifs of proteins, point model, helical pairs, angle between the axes of the helices.

### INTRODUCTION

In earlier works [1, 2] some simple rules of a polypeptide chain arrangement were revealed. As a result structural trees describing a diversity of protein structures were constructed. This made possible automatic recognition and prediction of various structural motives of protein molecules.

At present, attention of researchers is drawn to structural motives having unique spatial arrangements [1]. This interest is caused by the uniqueness of their structures and their capacity to serve as nuclei in the course of protein folding. However irrespective of the mechanism by which the protein folding proceeds, structural motives can be used as starting

---

<sup>\*</sup>[dmitry.tikhonov@gmail.com](mailto:dmitry.tikhonov@gmail.com)

<sup>†</sup>[likulikova@mail.ru](mailto:likulikova@mail.ru)

structures in the course of identification of possible arrangements of a polypeptide chain in modelling the protein structure [1, 3].

The simplest structural motives are those having unique spatial arrangements of polypeptide chains [1–4]. In previous works [5, 6], the problem of recognition of the structural motifs of  $\alpha$ - $\alpha$ -corners proteins was solved using the analytical description of the main chain of the protein globule and the spectral method of repeats recognition [7–9]. They dealt with  $\alpha$ - $\alpha$ -corners with a short connection. It is known that  $\alpha$ - $\alpha$ -corner is a frequent structural motif in proteins [3, 10–15]. This super secondary structure is formed by two neighboring  $\alpha$ -helices connected by connections and packed orthogonally. In proteins  $\alpha$ - $\alpha$ -corners occur in the form of a left-handed superhelix. Their sequences are arranged in a special way in a chain of hydrophobic, hydrophilic and glycine residues. The structures found were verified with the use of a conformational pattern characterized by limit values of the angles on the Ramachandra plot. Using the molecular dynamics method it was shown that  $\alpha$ - $\alpha$ -corners with a short connection are stable in water surroundings [5, 6]. Earlier stability of  $\alpha$ - $\alpha$ -corners was indirectly proved *in vitro* [16]. Our hypothesis about autonomous stability of structural motives was independently checked *in silico* in molecular dynamics computational experiments.

It is known from literature that  $\alpha$ -helices are packed densely. Structural motives formed of two neighboring  $\alpha$ -helices connected by one or more nonhelical amino acid residues are described in papers [2–4]. They are densely packed spatial structures. It is known that  $\alpha$ -helices pack in one of three characteristic arrangements, aligned parallel or antiparallel, orthogonal, or slanted. Some examples of such packings in  $\alpha$ - $\alpha$ -corners,  $\alpha$ - $\alpha$ -hairpins, L-shaped and V-shaped structures were studied in [2]. In view of the above it seems reasonable to form a set of all possible motives and investigate them.

In our earlier work [17] we studied the structures formed of two helices of any type:  $\alpha$ -helices,  $3_{10}$ -helices and  $\pi$ -helices. There we showed that 72.16 % of the total number of helices are  $\alpha$ -helices; 27.73 % are  $3_{10}$ -helices; and 0.1 % are  $\pi$ -helices. The analysis was made for the structural motives from the Protein Data Bank [18] formed by two neighboring helices of any type coupled by connections of different lengths and conformations. A statistical analysis of the distribution of interhelical distances in the helical pairs from the PDB was performed.

Here we investigate the structures found in our previous work from the viewpoint of distribution of the angles between the helical axes. The fundamental difference between our work and other investigations of interhelical angles in protein structures [19–23] is that we deal with the angles between the axes of the helices connected by a connection. The connection is a nonhelical fragment of an amino acid sequence whose length is one or several amino acids.

It should be noted that these angles may yield little information to biologists and biochemists since on their basis they cannot differentiate between the supersecondary structures available such as L-shaped structures,  $\alpha$ - $\alpha$ -corners etc. Investigation of interhelical angles has formal character. We analyze all the characteristics of helical pairs – the interhelical distance, connection length, square and perimeter of the overlapping polygon of crossing projections, etc. Thus, we stress, that the angles  $\varphi$  between the helical axes which we calculate are not the interhelical torsion angles  $\Omega$ . The torsion angles  $\Omega$  will be analyzed in the paper to follow.

It should be noted that we dealt with the helical pairs selected from the Protein Data Bank without regard for the protein classification. This classification is made on the basis of some peculiarities of the secondary and tertiary structure of proteins. It is known that in terms of availability and combination of  $\alpha$ -helices and  $\beta$ -structures all the protein structures are divided into several classes [24]:  $\alpha$ -,  $\beta$ -, ( $\alpha/\beta$ )-, ( $\alpha + \beta$ )-proteins, etc. Thus,  $\alpha$ -proteins are those consisting of  $\alpha$ -helices only. Popular examples of this class of proteins are hemoglobin and myoglobine. Proteins consisting of  $\beta$ -structures only are classified as  $\beta$ -proteins exemplified

by immunoglobulin. The class of  $(\alpha + \beta)$ -proteins includes both the structures within one polypeptide chain (lysozyme molecules) and the class of  $\alpha/\beta$ -proteins contains alternating  $\alpha$ - and  $\beta$ -structures (lactic dehydrogenase). In the future we are planning to investigate the distribution of interhelical angles and other characteristics of helical pairs for each class of the proteins.

## METHODS

In order to investigate the structural motives of proteins we elaborated special rules for recognition and selection of helical pairs. The main of them are:

1. Analysis of the secondary structure of the polypeptide chains is made by a method developed by the authors of the Dictionary for Secondary Structure of Protein [25]. Helices of three different types have been analyzed: the first type is  $\alpha$ -helix or H in DSSP notation, the second type (G-helix) involves  $3_{10}$ -helices, the third type (I-helix) is  $\pi$ -helix.

2. A candidate for the structure sought-for is a protein site containing two helices and a protein strand between the helices called a connection. All such structures can also be described in terms of the helices they contain.

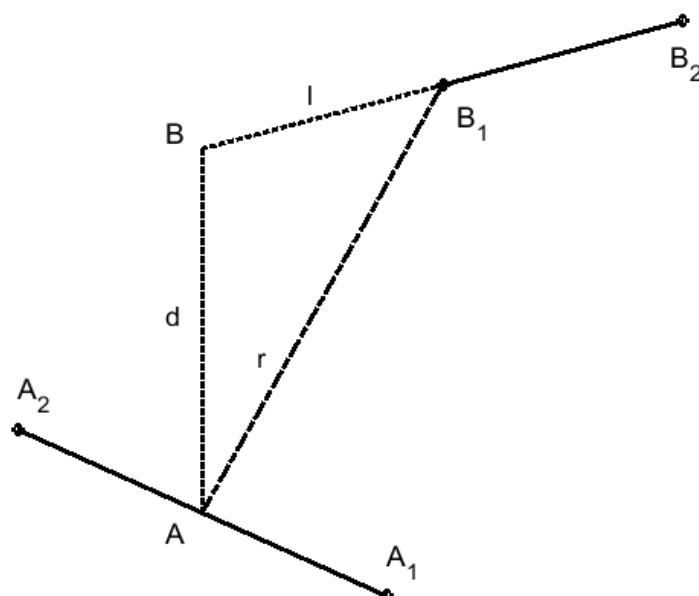
3. For each helix of the structure, we find the axis of the cylinder around which it is wound. The axis of the cylinder is determined by the least square method which implies minimizing the deviation of the helical parameters from those of an ideal helix. The quality of the axes assessment is characterized by the value of the root mean-square deviation. Only the helices (and accordingly, the structures) for which the accuracy of the assessments satisfies a certain predetermined criterion are selected.

4. Two helical axes completely determine the three dimensional arrangement of two cylinders of the helical pair. It is known that one can place two parallel planes onto two noncrossing right lines so that the first line would belong to the first plane, while the other one – to the second plane. The axis lying on one plane can be projected onto the other one. Thus, the three dimensional arrangement of the cylinders is fully described by the distance between the parallel planes and the axes projections onto the plane.

5. All the helical pairs selected are subdivided into three subsets according to criterion of crossing helix projections on the parallel planes passing through the axes of the helices:

- subset  $\{A\}$  involves the helical pairs not having crossing projections;
- subset  $\{B\}$  involves the helical pairs having crossing projections except for the helical pairs where the overlapping polygon contains the cross point of the helical axes projection;
- subset  $\{C\}$  involves the helical pairs for which the overlapping polygon contains the cross point of the helical axes projection.

6. From known coordinates of the points  $A_1, A_2, B_1$  и  $B_2$ , which are the initial and finite points of the axes of two helices [12] we calculate coordinates of the vectors  $\overrightarrow{A_2A_1}, \overrightarrow{B_1B_2}$  and find a cosine of the angle between these vectors  $\cos(\overrightarrow{A_2A_1}, \overrightarrow{B_1B_2})$ . The point model of the helical pair is shown in fig. 1 where the axes of the pair are shown. The interval  $[A_1, A_2]$  is the axis of the cylinder of the first helix,  $[B_1, B_2]$  is the axis of the cylinder of the second helix. The figure also demonstrates all possible distances  $d, r$  and  $l$  between the helices. From the viewpoint of mutual arrangement of the helices, three distances naturally come up. The first one is the interplane distance  $d$ . As is known, one can uniquely place two parallel planes onto two noncrossing right lines so that the shortest distance between the planes be the same as the distance between the lines. The distance between the right lines on which the cylinder axes lie will be called the interplane distance  $d$  of a helical pair. The second characteristic of the helical pair is the shortest distance  $r$  between the cylinder axes. Obviously, the minimum value of  $r$  will always be less than the interplane distance  $d$ , therefore we can introduce the value of the leg such that  $l = \sqrt{r^2 - d^2}$ . The leg  $l$  is the third distance which describes the relative arrangement of helical cylinders in a helical pair.



**Fig. 1.** Geometry of a helical pair – four points forming two legs in space.

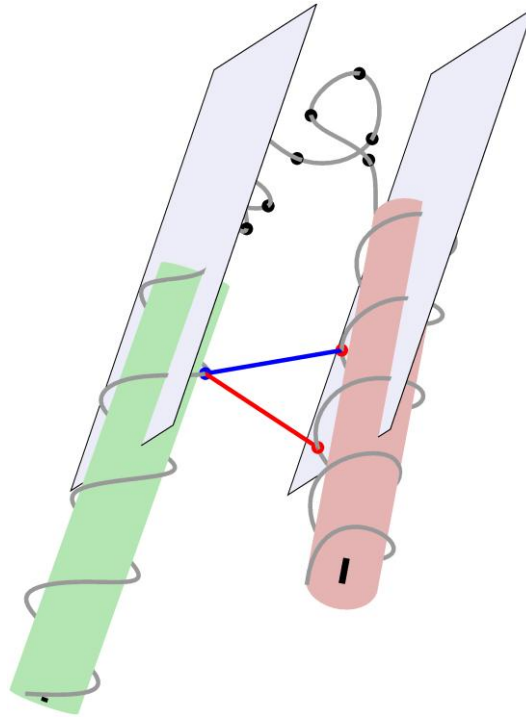
Using this algorithm we formed a base of structural motives and developed structure processing software which enables one to get information for each structure, such as interhelical axes  $\varphi$ , for example.

## RESULTS AND DISCUSSION

In our previous work [12] we analyzed the distribution of interhelical distances in the helical pairs connected by connections. Here we investigate the structures found in [12] from the viewpoint of distribution of the angles between the helical axes.

To solve this problem we suggested a point model for the description of the structures sought-for which consist of two helices connected by connections of different lengths. Scanning protein chains we select the structures consisting of two helices and irregular segments of different lengths between them.

Each helix is presented as a cylinder whose axis is found by the least square method. The quality of the axes assessment is characterized by the value of the root mean-square deviation. Only the structures which include helices maximally resembling an ideal one and for which the accuracy of the assessments satisfies a certain predetermined criterion are selected [26, 27]. In figure 2 we give a diagram of a structure formed by two helices connected by a connection. This is an example of a helical pair which is a fragment of an amino acid chain from the Protein Data Bank (PDB ID 3VKH, chain B, fragment: 2846–2891). The cylinders approximating the helix and the planes passing through the axes of the cylinders are shown. The curve is approximated by the positions of  $C_\alpha$ -atoms of the protein chain. This helical pair consists of two  $\alpha$ -helices (two helices of H type), the connection length is 12. The segments (red and blue) indicate the least distance (6.67 Å and 7.43 Å, accordingly) between the helices. This parameter of the studied structures will be described in detail in the work to follow.



**Fig. 2.** An example of a helical pair. Fragment of a protein chain from the Protein Data Bank (PDB ID 3VKH, chain B, fragment: 2846–2891). The cylinders approximating the helix, their axes and the planes passing through the cylinder axes are shown. The curve is approximated by the positions of  $C_{\alpha}$ -atoms of the protein chain, the atoms on the curves are indicated by points. Red and blue segments are the least distances (6.67 Å and 7.43 Å, accordingly) between the helices.

The data on the number of protein molecules processed and the number of helical pairs found are presented in Table 1.

**Table 1.** Data – the number of protein molecules processed and the number of helical pairs found

Number of processed protein structures from PDB	Number of processed aminoacid residues	Number of processed aminoacid chains	Number of helices of H type	Number of helices of G type	Number of helices of I type
100397	66546491	384666	1952658	750605	2908

The table suggests that the helices of H type prevail.

After all the helices in an amino acid chain had been recognized helical pairs were composed from neighboring helices. The rules for selecting the helical structures are described in detail in [12]. All such structures can also be described in terms of the types of helices they involve. There are six types of the structures: HH, GG, II, HG, HI, GI. If the name of a type consists of two similar letters, the helical pair of this type is formed by two helices of the same type. For example, the helical pair of HH-type consists of two H-helices. If a helical pair consists of helices of different types its name contains the letters of the types from which it consists. For example, HG is a helical pair containing one helix of H-type and the other helix of G-type. The data are symmetrized, i.e. if a helical pair consists of helices of different types we do not differentiate between the pairs where the sequence orders of the helices are different. For example, the helical pairs HG and GH belong to HG-type.

Having calculated the overlapping polygon of the helical axes projection [28] on the parallel planes passing through the helical axes we subdivide all the helical pairs obtained into three subsets:  $\{A\}$ ,  $\{B\}$  and  $\{C\}$  [12]:

- Subset  $\{A\}$  involves the helical pairs not having crossing projections;
- Subset  $\{B\}$  involves the helical pairs having crossing projections except for the helical pairs where the overlapping polygon [17] contains the cross point of the helical axes projection;
- Subset  $\{C\}$  involves the helical pairs for which the overlapping polygon contains the cross point of the helical axes projection.

In Table 2 we present the result of such subdivision. The number of helical pairs in the subset  $\{A\}$  is equal to the number of elements in the subset  $\{B\}$ , each contributing 44 % to the total number of helical pairs, the rest 12 % are accounted for by the elements from  $\{C\}$ . This means that most of helical pairs in amino acid chains (56 %) have crossing projections.

As for the subdivision by the types of the helices, it should be noted that though H-helices prevail (72 %), helical pairs of HH-type account for as little as 54.7 % of the total number of helical pairs. Helical pairs involving G helices account for 45 %, less than 0.2 % are the pairs involving I helices.

In the subset  $\{A\}$  where the helical pairs do not have crossing projections most of the pairs (45 %) belong to HG-type. In the subset  $\{B\}$ , HH-pairs are predominant (59 %). In the subset  $\{C\}$ , helical pairs of HH type are vastly predominant (87.5 %).

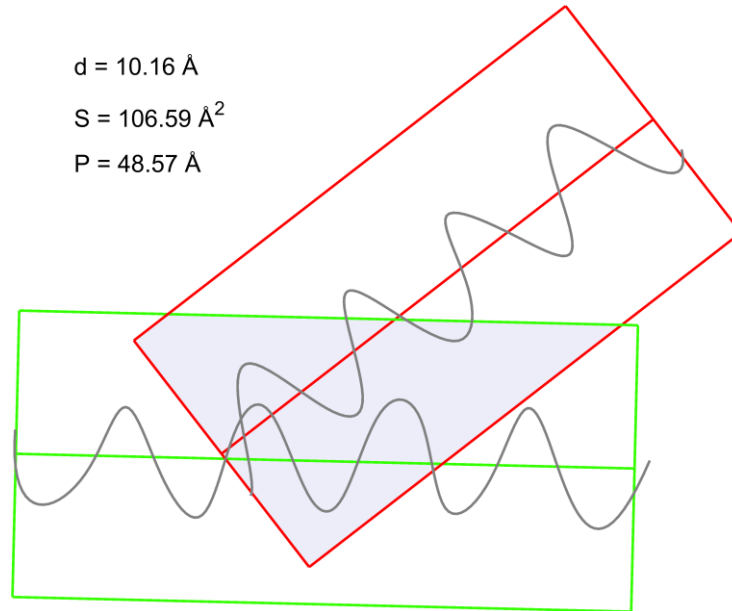
**Table 2.** Number of different-type helical pairs in the subsets of helical pairs

Subsets of helical pairs	Types of helical pairs						Number of elements in the subset
	HH	HG	GG	HI	GI	II	
$\{A\}$	402912	441055	125766	1588	643	0	971964
$\{B\}$	570830	349024	45513	1677	244	1	967289
$\{C\}$	234000	31719	1598	26	9	0	267352
Total number of helical pairs per type	1207742	821798	172877	3291	896	1	<b>2206605</b>

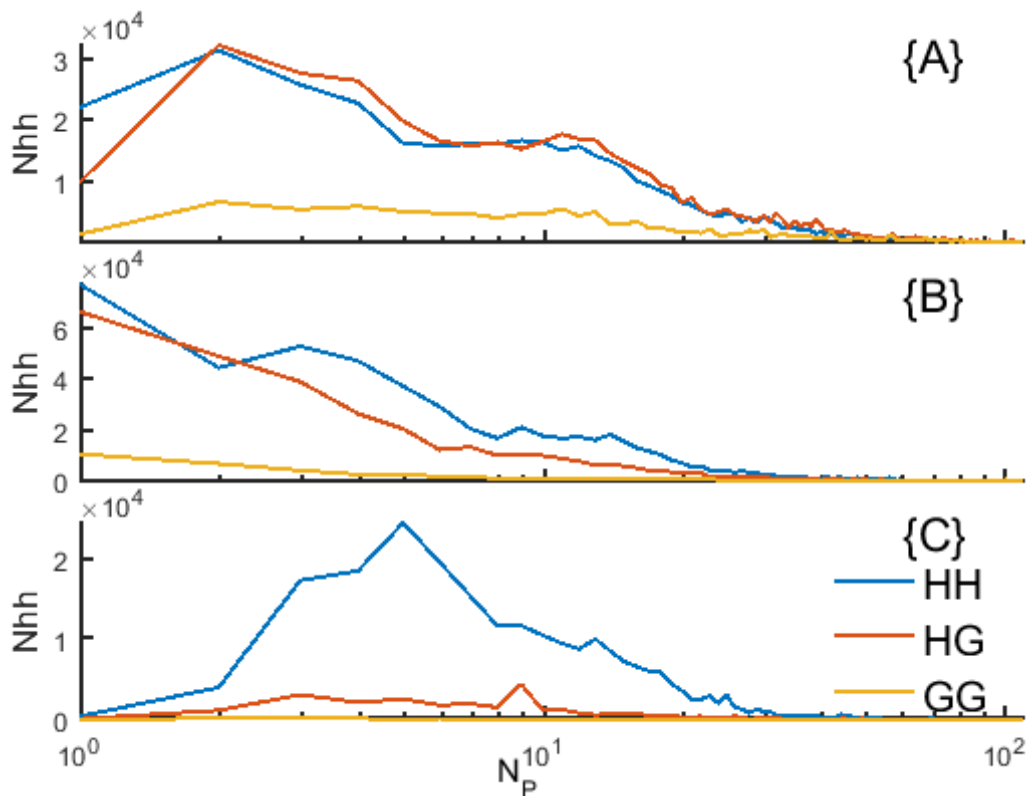
Figure 3 demonstrates crossing projections for the helical pair shown in figure 2 (PDB ID 3VKH, chain B, fragment: 2846–2891). Red and green rectangles are projections of the helices of this pair and the axes of relevant helices. It is evident from the figure that the projections of the helices of this pair overlap while the axes of the helices do not. The overlapping polygon is marked in gray color; its square  $S$  and perimeter  $P$  are given. In this case both the helices are  $\alpha$ -helices. Hence this helical pair of HH-type belongs to the subset  $\{B\}$ . The value of the interplane distance  $d$  is also presented.

The total number of the helical pairs selected is 2206605. A considerable body of the data obtained provides sufficient reliability of the results.

So then, a helical pair is a fragment of a protein chain consisting of two neighboring helices connected by one or more amino acids whose secondary structure is non-helical (connection). We investigate the structures formed by two helices of any type (including  $\alpha$ -helices) coupled by a connection of any conformation and any nonzero length  $N_p$ . We present the results of calculations of the distribution of the helical pairs of all the types belonging to different subsets depending on the connection length.



**Fig. 3.** Overlapping of projections of cylinders of helices of a helical pair. Overlapping polygon of the helices projections for the helical pair shown in fig. 2 (PDB ID 3VKH, chain B, fragment: 2846–2891). Red and green rectangles are projections of the helices and helical axes. The overlapping polygon is marked in gray color. The values of the square  $S$ , perimeter  $P$  and the interplane distance  $d$  are given.



**Fig. 4.** Distribution of the selected structures belonging to different types and different subsets depending on the number of amino acids in the connection. Along the X-axis – the connection length; along the Y-axis – actual number of the selected structures with a relevant connection length. The blue line represents HH-structural motives of proteins formed by two  $\alpha$ -helices and G-helices; the yellow line GG – motives formed by G-helices. {A}, {B} and {C} – subsets of helical pairs selected according to the criterion of crossing axes projections.

Figure 4 demonstrates histograms of the distribution of helical pairs of different types belonging to the subsets {A}, {B} or {C}, depending on the length of the irregular fragment

between the helices. Maxima in the distribution of the helical pairs belonging to the subsets  $\{A\}$  and  $\{B\}$  correspond to small values of the connection (from one to three amino acids). Distribution of the helical pairs belonging to the subset  $\{A\}$  has much more long-ranging character than the distribution of the pairs belonging to  $\{B\}$ . It is evident from figure 4 that the distribution of the helical pairs of all the types belonging to the subset  $\{A\}$  has a maximum for the connection length equal to 2. At the same time the mean values of  $N_P$  calculated from these distributions differ considerably. Thus for the pairs from the subset  $\{A\}$ , the mean value of the connection is 15–20 amino acids, while for the structures from the subset  $\{B\}$ , this value is equal to 9–11 amino acids (see Table 3). A great number of helical pairs from the subset  $\{B\}$  belongs to HH and HG types and has a connection of the length of 1. An example of such helical pairs is hairpins which are very often met in proteins. Helical pairs of GG type are few and their number is maximal for the connection length equal to 1. In the subset  $\{C\}$ , the absolute majority are helical pairs of HH type. It is shown that among the structures selected according to the above-formulated criterion, the structures with the connection length equal to 5 prevail. At the same time the mean value of the connection length is approximately equal to the mean value  $N_P$  of the helical pairs from the subset  $\{B\}$ . The helical pairs of GG type are none or very few in this subset. The number of HG-type helical pairs belonging to subset  $\{C\}$  is maximal for  $N_P = 9$ .

It should be noted that no helical pairs with the connection length equal to one amino acid ( $N_P = 1$ ) were detected among the structures belonging to the subset with crossing axes projections (subset  $\{C\}$ ).

It is also demonstrated that in all the subsets, the structural motives with a short connection prevail.

Taking into account that the mean value mainly depends on the long-ranging character of the distribution, we may conclude that the long-ranging distributions of the pairs are approximately similar for subsets  $\{B\}$  and  $\{C\}$  (see Table 3). Table 3 lists statistical estimates of the connection lengths of different-type helical pairs belonging to different subsets.

**Table 3.** Statistical estimates of the distributions of connection lengths  $N_P$  for helical pairs of different types belonging to different subsets

Statistical estimates	connection length $N_P$ for helical pairs of different types		
	HH	HG	GG
mode $\{A\}$	2	2	2
median $\{A\}$	11	12	14
mean $\{A\}$	15.34	17.67	24.26
rms deviation $\{A\}$	17.14	20.52	28.93
mode $\{B\}$	1	1	1
median $\{B\}$	6	4	4
mean $\{B\}$	10.59	9.01	11.09
rms deviation $\{B\}$	11.78	13.41	18.72
mode $\{C\}$	5	9	3
median $\{C\}$	9	9	9
mean $\{C\}$	11.97	12.65	18.34
rms deviation $\{C\}$	10.36	13.22	24.25



Thus, we have investigated the distribution of helical pairs depending on the connection length which is of relevance since the length of the irregular fragment between the helices, surely, plays a certain role in the spatial arrangement of helical pairs.

And now let us pass on to investigation of the structures selected from the viewpoint of the distribution of the angles between the helical axes, in particular in terms of the connection length and the distance between the helical axes. By an interhelical angle  $\varphi$  or an angle between the helical axes we mean an angle between the vectors lying on the helical axes where the origin of the first vector is the end of the first helix and the end of the first vector is the origin of this helix, while the second vector originates at the origin of the second helix and the end of the second helix is the end of the second vector.

Figure 5 shows histograms of the distribution of helical pairs of different types belonging to different subsets depending on the angles between the helices. Table 4 lists statistical estimates of the distribution of interhelical angles  $\varphi$  in helical pairs of different types belonging to different subsets.

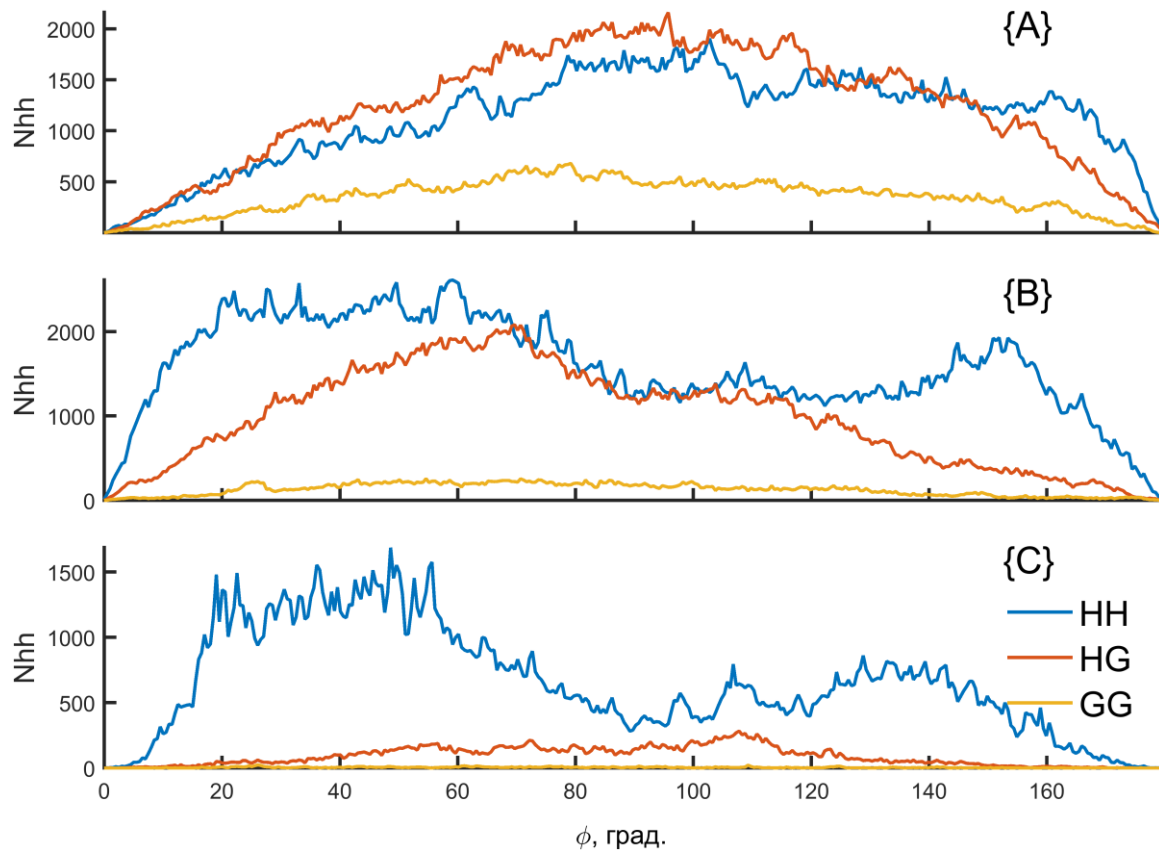
Figure 5 clearly demonstrates a significant difference between the distributions of the structures from different subsets depending on the angle between the helical axes. The top figure illustrates the distribution of the pairs from subset  $\{A\}$  depending on the angle. The distribution is rather wide and covers practically the whole range of the angles. It has one maximum which is slightly different for helical pairs of different types. For helical pairs of HH type it is slightly displaced towards the region of obtuse angles which is confirmed by the mean value of the angles equal to  $100.17^\circ$  (see Table 4). For the helical pairs of other types, we can observe asymmetry rather than displacement. The maximum corresponds to approximately the right angle, the mean values confirm this result ( $93.99^\circ$  and  $91.45^\circ$  for HG and GG types, respectively).

**Table 4.** Statistical estimates of the distribution of interhelical angles  $\varphi$  for helical pairs of different types belonging to different subsets

Statistical estimates	interhelical angles $\varphi$ for helical pairs of different types		
	HH	HG	GG
mode $\{A\}$	82.78	104.97	67.54
median $\{A\}$	100.54	94.03	89.01
mean $\{A\}$	100.17	93.99	91.45
rms deviation $\{A\}$	42.4	39.5	39.81
mode $\{B\}$	58.83	66.72	58.64
median $\{B\}$	71.65	72.41	75.69
mean $\{B\}$	80.14	77.04	79.46
rms deviation $\{B\}$	47.62	37.05	38.43
mode $\{C\}$	44.29	96.17	109.48
median $\{C\}$	60.62	86.08	71.98
mean $\{C\}$	73.56	84.49	75.56
rms deviation $\{C\}$	43.5	31.98	39.08

Central and bottom figures illustrate distributions of the helical pairs of all the types belonging to the subsets  $\{B\}$  and  $\{C\}$ . The distributions of the helical pairs from the subsets  $\{B\}$  and  $\{C\}$  differ from the distribution of the structures belonging to the subset  $\{A\}$ . These

helical pairs are characterized by distributions with two maxima: one in the region of acute angles and the other – in the region of obtuse angles.



**Fig. 5.** Distribution of helical pairs of different types belonging to different subsets depending on the angle  $\phi$  between the helical axes.

**Table 5.** Statistical estimates of the distributions of interhelical angles  $\phi$  in the region of acute and obtuse angles for helical pairs of different types belonging to different subsets

Statistical estimates	Interhelical angles $\phi$ in the region of acute angles for helical pairs of different types			Interhelical angles $\phi$ in the region of obtuse angles for helical pairs of different types		
	HH	HG	GG	HH	HG	GG
mode {B}	58.83	66.72	58.64	151.72	93.29	132.26
median {B}	46.77	56.46	55.81	133.12	114.84	117.47
mean {B}	<b>46.68</b>	<b>54.49</b>	<b>54.73</b>	<b>131.77</b>	<b>118.83</b>	<b>120.51</b>
rms deviation {B}	23	20.9	21.3	23.6	20.46	21.75
mode {C}	44.29	78.16	22.46	133.69	96.17	109.48
median {C}	45.04	60.56	52.79	129.12	109.31	117.15
mean {C}	<b>45.92</b>	<b>59.35</b>	<b>51.8</b>	<b>127.82</b>	<b>112.76</b>	<b>120.22</b>
rms deviation {C}	19.84	19.02	21.66	19.72	15.99	21.43

The maximum in the region of acute angles is much more pronounced. The bay in the distribution corresponds to the vicinity of the right angle between the axes of a helical pair. This is especially evident for the helical pairs belonging to the subset {C} where the bay of the distribution is one fourth of the maximum. The figure demonstrates the presence of two

peaks for the structures formed by two  $\alpha$ -helices: one peak corresponds to the acute angle ( $20^\circ$ – $60^\circ$ ) and second peak corresponds to the obtuse angle ( $130^\circ$ – $150^\circ$ ). The mean values of the angles: in the region of acute angles –  $45.92^\circ$  and in the region of obtuse angles –  $127.82^\circ$ , this is confirmed by the data of Table 5.

For the structures formed by one  $\alpha$ -helix and one helix of a different type, there is a scatter from  $50^\circ$  to  $120^\circ$ . For the helical pairs from the subset  $\{B\}$ , the bay though being observable, is less pronounced. The peak of the distribution is displaced towards the region of obtuse angles and a bay in the vicinity of the right angle is lacking.

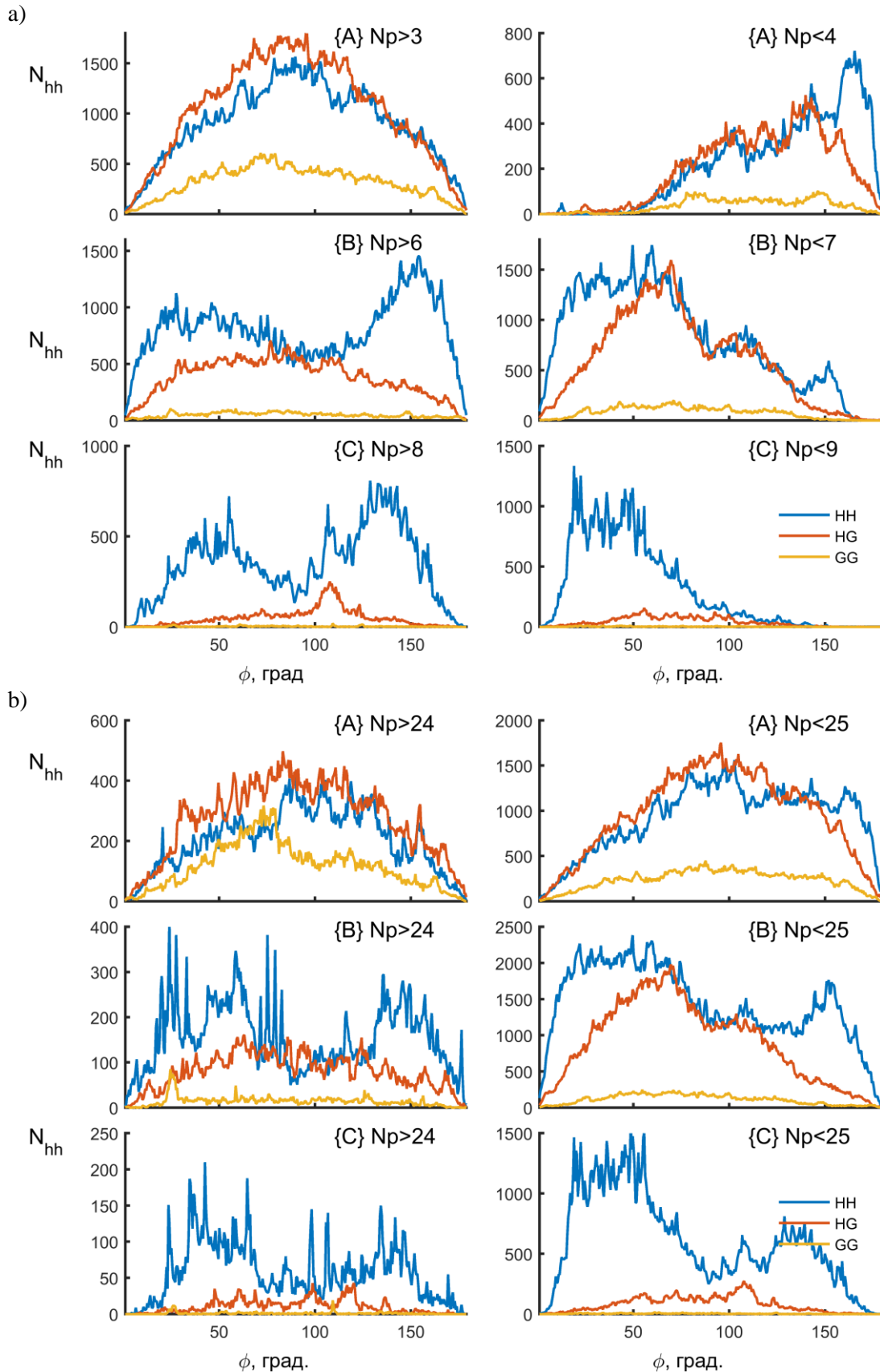
Figure 6 demonstrates the histograms of the distribution of helical pairs of different types belonging to different subsets depending on the angles between the helical axes with regard for the connection length. In Figure 6,a, the upper histograms illustrate the distribution of the helical pairs of different types belonging to the subset  $\{A\}$  depending on the angle between the helical axes for the connection length  $N_p < 4$  и  $N_p > 3$ .

We can see that among the helical pairs not having crossing projections, the pairs with the connection length  $N_p > 3$  prevail over the pairs with a short connection (one, two or three amino acids). The pairs of HG type are fewer than the pairs formed by two  $\alpha$ -helices. All the distributions have similar character but for HH and HG pairs, the maximum of the distribution is observed at the angle  $\varphi \approx 90^\circ$ , while for GG pairs, it is slightly displaced towards the region of acute angles ( $\varphi \approx 75^\circ$ ). The distribution of the helical pairs with the connection length  $N_p < 4$  is quite different from the distribution of the pairs with the connection  $N_p > 3$ . The pairs with a short connection and an acute angle between the helical axes ( $0^\circ \leq \varphi \leq 60^\circ$ ) are very few if any. The structures with this connection appear for  $\varphi \approx 60^\circ$  and their number grows with an increase in the interhelical angle. For HG pairs, maximum of the distribution is observed for the angles  $120^\circ \leq \varphi \leq 130^\circ$ , for HH pairs – for  $\varphi \approx 170^\circ$ . We can state with assurance that this maximum of the distribution of HH helical pairs with a short connection for  $\varphi \approx 170^\circ$  is provided by L-structures.

The distribution of GG-type pairs has uniform character without any peaks for all the angles  $60^\circ \leq \varphi \leq 180^\circ$ .

The central histograms of Figure 6,a demonstrate the distribution of helical pairs of different types belonging to the subset  $\{B\}$  depending on the angle between the helical axes for the connection lengths  $N_p > 6$  and  $N_p < 7$ . We can see that the helical pairs belonging to the subset  $\{B\}$  (the pairs have crossing projections but do not have crossing axes projections) are very many. Especially abundant are HH-type pairs. For the connection length  $N_p > 6$ , the distributions of the pairs of different types are different. Helical pairs of HH type demonstrate a distribution with two maxima: one in the region of acute angles ( $30^\circ \leq \varphi \leq 50^\circ$ ) and the other – in the region of obtuse angles ( $140^\circ \leq \varphi \leq 150^\circ$ ). The maximum in the region of obtuse angles is much more pronounced. The bay of the distribution corresponds to the vicinity of the right angle between the axes of a helical pair. The distribution of HG-type structures is rather wide. It covers nearly the whole range of the angles and has one maximum in the vicinity of the right angle ( $\varphi \approx 90^\circ$ ). Maximum of the distribution of HG-type pairs corresponds to the region where there is a bay of the distribution of HH-type pairs. The pairs of GG type belonging to the subset  $\{B\}$  are very few for the connection length  $N_p > 6$  and  $N_p < 7$ . They are distributed uniformly over the whole range of the angles.

The distribution of different-type helical pairs belonging to the subset  $\{B\}$  whose connection length is  $N_p < 7$  has different character. It is evident from the figure that most of these pairs belong to HH type. Most of HH pairs are localized in the region of acute angles ( $10^\circ < \varphi < 90^\circ$ ), their number is nearly the same over the whole range of these angles and maximum of the distribution corresponds to the angles ( $50^\circ \leq \varphi \leq 60^\circ$ ).



**Fig. 6.** Distribution of helical pairs of different types belonging to different subsets depending on the angle  $\phi$  between the helical axes and the connection length.

In the region of obtuse angles two more local maxima are observed: one – in the region of ( $100^\circ < \phi < 120^\circ$ ), and the other – for  $\phi \approx 160^\circ$ . Most of the helical pairs of HG type whose

connection length is  $N_p < 7$  have an acute angle between the helical axes. Such pairs are very few for very small angles and their number grows as the angle between the axes increases in the region of acute angles. In the region of obtuse angles ( $100^\circ < \varphi < 120^\circ$ ) there is also a peak of the distribution but it is less pronounced. This peak coincides in its value with the peak of the distribution of HH-type helical pairs. The lower histograms of Figure 6,a illustrate the distribution of different-type helical pairs from the subset  $\{C\}$  depending on the angle between the helical axes for the connection length  $N_p > 8$  (left figure) and  $N_p < 9$  (right figure). We can see that the helical pairs with the connection length equal to 1 to 8 amino acids are much more abundant than the helical pairs with a longer connection. Most of such pairs (right figure) are distributed in the region of acute angles ( $10^\circ < \varphi < 90^\circ$ ) very few pairs occur in the vicinity of  $100^\circ$ . Helical pairs from the subset  $\{C\}$  with this connection length and an obtuse angle between the helical axes ( $\varphi > 100^\circ$ ) are very few if any. Maximum of the distribution corresponds to the angles ( $25^\circ < \varphi < 60^\circ$ ). Helical pairs of HG type are very few, they are mainly localized in the region ( $40^\circ < \varphi < 100^\circ$ ). Pairs of GG type are lacking at all. It should be noted that the right figures for the subsets  $\{B\}$  and  $\{C\}$  illustrating the distribution of the helical pairs with a short connection demonstrate the distribution of the pairs of  $\alpha$ -proteins.

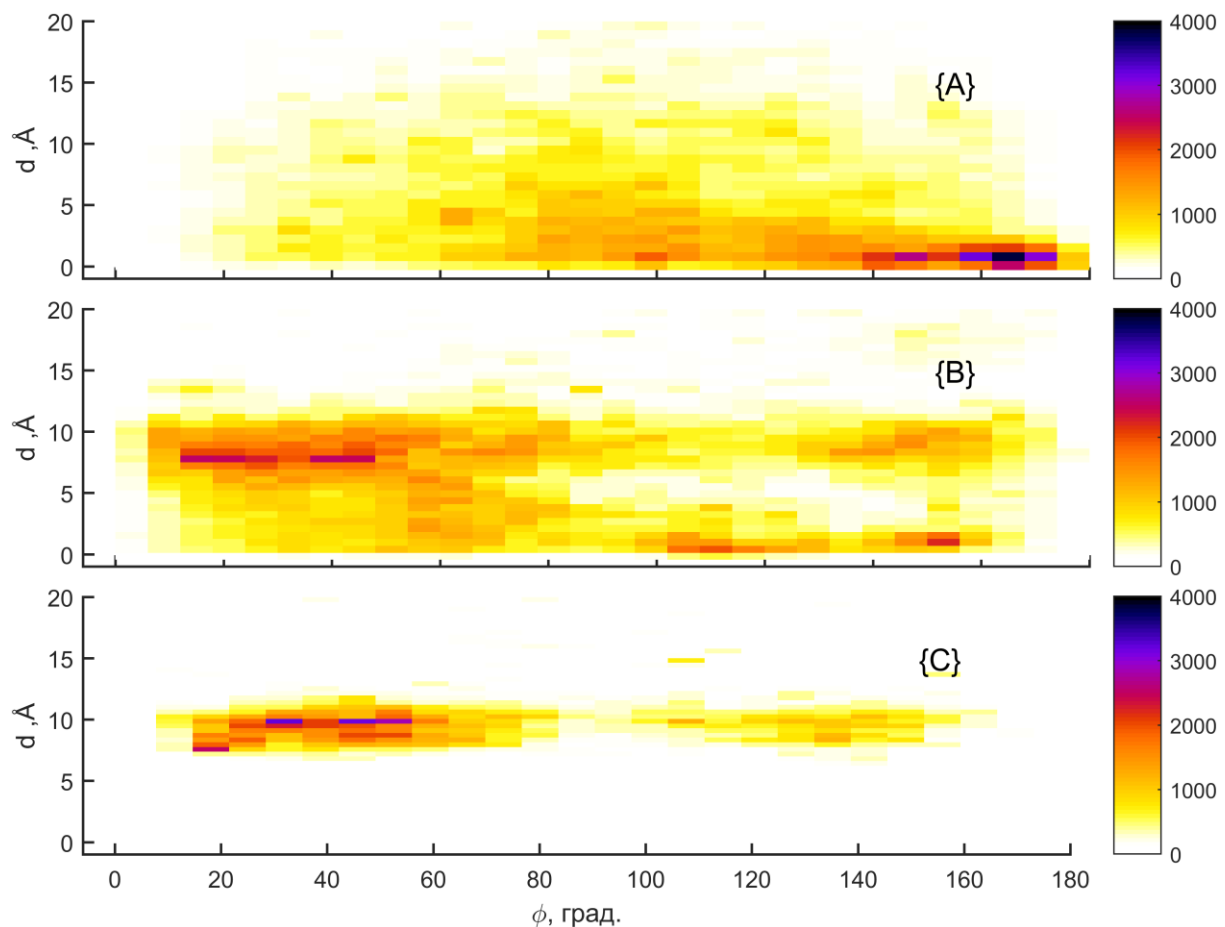
Distribution of the helical pairs with crossing projections and crossing axes projections with the connection length  $N_p > 8$  (left figure) reminds the distribution of the helical pairs from the subset  $\{C\}$  without regard for the connection length (see Fig. 5). But in this case maximum is observed in the region of obtuse angles ( $130^\circ < \varphi < 145^\circ$ ). There is a kind of a peak in the vicinity of  $110^\circ$ . For the same value of the angle, the structures of HG type demonstrate maximum of the distribution.

Figure 6,b shows the histograms of the distribution of different-type helical pairs from the subsets  $\{A\}$ ,  $\{B\}$  and  $\{C\}$  with the connection length  $N_p > 24$  and  $N_p < 25$ . The upper histograms illustrate the distribution of the helical pairs from the subset  $\{A\}$  with these connection lengths. It can be seen that the distributions of the helical pairs of all the types (HH, HG and GG) belonging to the subset  $\{A\}$  with the connection length  $N_p < 25$  resemble the distribution of the pairs from the same subset without regard for the connection length (see Fig. 5). It should be noted that the overwhelming majority of the pairs from this subset have a connection whose length is less than 25 amino acids. The structures with a shorter connection are much fewer, their distribution has different character. If the pairs of HG type also demonstrate a wide scatter over the whole range of the angle values with a maximum for  $\varphi \approx 90^\circ$ , the distribution of HH pairs demonstrates a certain bay in the vicinity of  $\varphi \approx 75^\circ$ , the maximum of the distribution is displaced towards obtuse angles ( $90^\circ < \varphi < 125^\circ$ ). There is a local maximum in the region of acute angles for  $\varphi \approx 60^\circ$ . The distribution of the helical pairs of GG type also changes its character in the vicinity of  $\varphi \approx 75^\circ$  where there is a bay for HH pairs, maximum of the distribution, a bay in the vicinity of  $100^\circ$  and a local maximum for  $\varphi \approx 120^\circ$ .

The central histograms of Figure 6,b illustrate the distributions of different-type helical pairs from the subset  $\{B\}$  with the connection length  $N_p > 24$  and  $N_p < 25$ . It is evident that the helical pairs with the connection length  $N_p < 25$  prevail. The distributions of the helical pairs of all the types from the subsets  $\{B\}$  and  $\{C\}$  with the connection length  $N_p < 25$  also retain the character of the distribution of the pairs from this subset without regard for the connection length (see Fig. 5). The pairs with a longer connection are much fewer and the character of their distribution changes. The distribution of HG structures from the subset  $\{B\}$  with a long connection covers nearly the whole range of the angles and demonstrates one maximum in the vicinity of the right angle. The GG helical pairs from the subsets  $\{B\}$  and  $\{C\}$  are very few, they are uniformly distributed over the whole range of the angles and

demonstrate a local peak in the region of acute angles ( $\varphi \approx 25^\circ$ ). The distribution of HH pairs demonstrates a lot of peaks over the whole range of definition, though the general tendency (bays, two maxima in the region of acute and obtuse angles) retains.

Figure 7 illustrates the distribution of the helical pairs consisting of two  $\alpha$ -helices (HH-type helical pair), belonging to different subsets depending on two parameters: angle  $\varphi$  between the helical axes and interplane distance  $d$ . The upper diagram demonstrates the distribution of the helical pairs of HH type belonging to the subset  $\{A\}$ . It is seen that the helical pairs from the subset  $\{A\}$  have interhelical angles in the range of ( $20^\circ < \varphi < 180^\circ$ ) and the interplanar distance from 0 to 15 Å. There is a fairly large number of pairs whose helices are set at right angle to each other and the interplane distance varies in the range of 0–5 Å. There are very many helical pairs not having crossing axes projections whose interhelical distance is 2–3 Å and the angle between the helical axes falls on the interval ( $145^\circ < \varphi < 170^\circ$ ). Maximum number of the pairs is observed for  $\varphi \approx 165^\circ$ .



**Fig. 7.** Distribution of the helical pairs of HH type belonging to different subsets depending on the angle between the helical axes and the interplane distance  $d$ . Along the  $Ox$  axis – angle  $\varphi$  between the helical axes, along the  $Oy$  axis – interplanar distance  $d$ .  $\{A\}$ ,  $\{B\}$  and  $\{C\}$  are subsets of the helical pairs selected according to the criterion of crossing helix projections. On the right-hand side there is a scale for correspondence between the color (from white to black) and the number of helical pairs.

The central diagram illustrates the distribution of the helical pairs of HH type belonging to the subset  $\{B\}$ . These structures demonstrate a different distribution. Most of the pairs have the interplane distance of 5–10 Å and an acute angle between the helical axes, exhibiting two maxima for the angles  $\varphi$ : ( $15^\circ < \varphi < 25^\circ$ ), ( $35^\circ < \varphi < 50^\circ$ ) and the interplane distance  $d \approx 8$  Å. It should be noted that there are two more local maxima in the region of obtuse angles,  $\varphi \approx 110^\circ$  and  $\varphi \approx 150^\circ$ , the interplane distance being approximately equal to 2 Å.

The lower diagram shows the distribution of the helical pairs consisting of two  $\alpha$ -helices belonging to the subset  $\{C\}$  depending on the angle  $\varphi$  between the helical axes and the interplane distance  $d$ . It is seen that nearly all the helical pairs of this subset are localized in the region of acute ( $10^\circ < \varphi < 80^\circ$ ) and obtuse ( $110^\circ < \varphi < 150^\circ$ ) angles and have the interplane distance equal to 8–10 Å. The helical pairs with an acute angle  $\varphi$  between the helical axes are predominant. The distribution demonstrates two peaks for  $\varphi$  in the range of ( $30^\circ < \varphi < 35^\circ$ ) and ( $40^\circ < \varphi < 50^\circ$ ); the interplane distance in both the cases is equal to 10 Å



**Fig. 8.** Examples of compact structures formed by two helices, in this particular case by two  $\alpha$ -helices: **a)** structure with a short connection, protein 4DL6 (coordinates 275–291), **b)** structure with a long connection, 4EAG ( $C_\alpha$ : 211–249).

Figure 8 illustrates some examples of compact structures selected according to the suggested criteria: a structure with a short connection (in this particular case –  $\alpha$ - $\alpha$ - corner) and a structure with a long connection where  $\beta$ -strands and irregular fragments of different lengths occur between two  $\alpha$ -helices.

## CONCLUSION

Using a point model of helical pairs we selected a set of protein molecule structures to be investigated. The set was selected from the Protein Data Bank with the use of special rules and subdivided into three subsets according to the criterion of crossing helix projections on the parallel planes passing through the axes of the helices.

We analyzed how the helical pairs of all the types belonging to different subsets are distributed depending on the connection length. It is shown that in all the subsets the structure motives with a short connection prevail. The distribution of the helical pairs of all the types from the subset  $\{A\}$  demonstrates one maximum for the connection length equal to 2 amino acid residues. A great many helical pairs from the subset  $\{B\}$  belong to HH and HG types and have a connection length equal to 1 amino acid residue.

It is shown that the vast majority of all the pairs from the subset  $\{C\}$  are the helical pairs of HH type. Among the helical pairs of all the types from the subset  $\{C\}$  there are no pairs whose connection consists of one amino acid residue ( $N_P = 1$ ). It is also demonstrated that among the structures selected according to the above-formulated criterion, the structures with the connection length equal to 5 amino acid residues ( $N_P = 5$ ) prevail.

We also analyzed how the helical pairs of all the types belonging to different subsets are distributed depending on the angles between the helical axes. It was shown that the distribution of the helical pairs of all the types from the subset  $\{A\}$  covers nearly the whole range of the angles and demonstrates one maximum in the region of ( $90^\circ \leq \varphi \leq 100^\circ$ ). The helical pairs of HG type prevail over the pairs not having crossing projections. The pairs from the subsets  $\{B\}$  and  $\{C\}$  demonstrate the distributions with two maxima – one in the region of



acute angles and the other – in the region of obtuse angles. The maximum in the region of acute angles is much more pronounced. It is shown that a considerable portion of HH-type helical pairs (which are predominant in the subset  $\{C\}$ ) have an acute angle ( $20^\circ \leq \varphi \leq 50^\circ$ ) between the helical axes. For all these structures, the interplane distance is equal to 10 Å.

The study was made with the support from the RFBR (project № 18-07-01031).

## REFERENCES

1. Efimov A.V. Standard structures in proteins. *Prog. Biophys. Molec. Biol.* 1993;60:201–239. doi: [10.1016/0079-6107\(93\)90015-C](https://doi.org/10.1016/0079-6107(93)90015-C)
2. Gordeev A.B., Kargatov A.M., Efimov A.V. PCBOST: Protein classification based on structural trees. *Biochemical and Biophysical Research Communications*. 2010;397:470–471. doi: [10.1016/j.bbrc.2010.05.136](https://doi.org/10.1016/j.bbrc.2010.05.136)
3. Efimov A.V. Super-secondary structures and modeling of protein folds. In: *Methods in Molecular Biology*. Ed. Kister A.E. Clifton: Humana Press; 2013. V. 932. P. 177–189.
4. Brazhnikov E.V., Efimov A.V. Structure of  $\alpha$ - $\alpha$ -hairpins with short connections in globular proteins. *Molecular Biology*. 2001;35(1):89–97. doi: [10.1023/A:1004859003221](https://doi.org/10.1023/A:1004859003221)
5. Rudnev V.R., Pankratov A.N., Kulikova L.I., Dedus F.F., Tikhonov D.A., Efimov A.V. Recognition and Stability Analysis of Structural Motifs of  $\alpha$ - $\alpha$ -corner Type in Globular Proteins. *Mathematical Biology and Bioinformatics*. 2013;8(2):398–406 (in Russ.). doi: [10.17537/2013.8.398](https://doi.org/10.17537/2013.8.398)
6. Rudnev V.R., Pankratov A.N., Kulikova L.I., Dedus F.F., Tikhonov D.A., Efimov A.V. Conformational Analysis of Structural Motifs of  $\alpha$ - $\alpha$ -Corner in the Computational Experiment of Molecular Dynamics. *Mathematical Biology and Bioinformatics*. 2014;9(2):575–584 (in Russ.). doi: [10.17537/2014.9.575](https://doi.org/10.17537/2014.9.575)
7. Dedus F.F., Makhortykh S.A., Ustinin M.N., Dedus A.F. *Obobshchennyi spektral'no – analiticheskii metod obrabotki informatsionnykh massivov. Zadachi analiza izobrazhenii i raspoznavaniia obrazov* (Generalized Spectral-Analytic Method in Information Processing. Problems of Pattern Recognition and Image Analysis). Moscow; 1999. 357 p. (in Russ.).
8. Dedus F.F., Makhortykh S.A., Ustinin M.N. Application of the Generalized Spectral-Analytic Method in Information Problems. *Pattern Recognition and Image Analysis*. 2002;12(4):429–437.
9. Pankratov A.N., Tetuev R.K., Pyatkov M.I., Toigildin V.P., Popova N.N. Spectral Analytical Method of Recognition of Inexact Repeats in Character Sequences. *Proceedings of the Institute for System Programming of the RAS*. 2015;27(6):335–344 (in Russ.). doi: [10.15514/ISPRAS-2015-27\(6\)-21](https://doi.org/10.15514/ISPRAS-2015-27(6)-21)
10. Finkelstein A.V. *Fizika belkovykh molekul* (Physics of Protein Molecules). Moscow–Izhevsk; 2014. 424 p. (in Russ.).
11. Ptitsyn O.B., Finkelstein A.V. In: *Itogi nauki i tekhniki* (Results in Science and Technology). Ed. Vol'kenshtein M.D. Moscow; 1979. V. 15. P. 6–41. (Series “Molecular Biology”) (in Russ.).
12. Shul'ts G.E., Shirmer R.Kh. *Printsipy strukturnoi organizatsii belkov*. Moscow; 1982. 354 p. (Translation of: Schulz G.E., Schirmer R.H. Principles of Protein Structure Springer-Verlag New York, 1979. (Series “Springer Advanced Texts in Chemistry”)).
13. Miller S., Janin J., Lesk A.M., Chothia C. Interior and surface of monomeric proteins. *Molecular Biology*. 1987;196:641–656. doi: [10.1016/0022-2836\(87\)90038-6](https://doi.org/10.1016/0022-2836(87)90038-6)
14. Creighton T.E. *Proteins*. N.Y.: W.H. Freeman & Co; 1991.
15. Stepanov V.M. *Molekuliarnaia biologiya. Struktura i funktsii belkov* (Molecular biology. The structure and function of proteins). Moscow; 1996. 336 p. (in Russ.).



16. Tsai F.C., Sherman J.C. Circular dichroism analysis of a synthetic peptide corresponding to the  $\alpha$ - $\alpha$ -corner motif of hemoglobin. *Biochemical and Biophysical Research Communications*. 1993;196(1):435–439. doi: [10.1006/bbrc.1993.2268](https://doi.org/10.1006/bbrc.1993.2268)
17. Tikhonov D.A., Kulikova L.I., Efimov A.V. Statistical Analysis of the Internal Distances of Helical Pairs in Protein Molecules. *Mathematical Biology and Bioinformatics*. 2016;11(2):170–190 (in Russ.). doi: [10.17537/2016.11.170](https://doi.org/10.17537/2016.11.170)
18. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. The Protein Data Bank. *Nucleic Acids Research*. 2000;28:235–242. doi: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235)
19. Crick F.H.C. The Packing of  $\alpha$ -Helices: Simple Coiled-Coils. *Acta Crystallographica*. 1953;6:689–697. doi: [10.1107/S0365110X53001964](https://doi.org/10.1107/S0365110X53001964)
20. Lee H.S., Choi J., Yoon S. QHELIX: A Computational Tool for the Improved Measurement of Inter-Helical Angles in Proteins. *Protein*. 2007;26:556–561. doi: [10.1007/s10930-007-9097-9](https://doi.org/10.1007/s10930-007-9097-9)
21. Walther D., Eisenhaber F., Argos P. Principles of Helix-Helix Packing in Proteins: The Helical Lattice Superposition Model. *Molecular Biology*. 1996;255:536–553. doi: [10.1006/jmbi.1996.0044](https://doi.org/10.1006/jmbi.1996.0044)
22. Chothia C., Levitt M., Richardson D. Structure of proteins: Packing of  $\alpha$ -helices and pleated sheets. *Proc. Natl. Acad. Sci.* 1977;74:4130–4134. doi: [10.1073/pnas.74.10.4130](https://doi.org/10.1073/pnas.74.10.4130)
23. Chothia C., Levitt M., Richardson D. Helix to Helix Packing in Proteins. *Molecular Biology*. 1981;145:215–250. doi: [10.1016/0022-2836\(81\)90341-7](https://doi.org/10.1016/0022-2836(81)90341-7)
24. Levitt M., Chothia C. Structural patterns in globular proteins. *Nature*. 1976;261:552–558. doi: [10.1038/261552a0](https://doi.org/10.1038/261552a0)
25. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22(12):2577–2637. doi: [10.1002/bip.360221211](https://doi.org/10.1002/bip.360221211)
26. Kabsch W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica*. 1976;32:922–923. doi: [10.1107/S0567739476001873](https://doi.org/10.1107/S0567739476001873)
27. Kabsch W. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica*. 1978;34:827–828. doi: [10.1107/S0567739478001680](https://doi.org/10.1107/S0567739478001680)
28. Legland D. MatGeom: Matlab geometry toolbox for 2D/3D geometric computing. <http://github.com/dlegland/matGeom> (accessed 11 September 2018).

Received 29.01.2019.

Published 08.02.2019.