fraction α_0 of powder particles, linear absorption coefficient μ, mean chord length $\bar{l}$ of powder particles and scattering angle θ. There are no dramatic effects of microabsorption if the packing fraction of powder particles and the scattering angle are not too small.

The theoretical results agree well with the experimental data after Suortti (1972) especially with respect to the dependence of the microabsorption on the scattering angle. For practical use, the approximate formulae (11), (12) should be sufficient in most cases.

The present method is applicable not only to powders but also to other heterogeneous specimens such as sinter materials. In forthcoming papers, the following related problems will be analysed: (i) microabsorption in quantitative phase analysis; (ii) influence of regularity of packing of powder particles on microabsorption.

The stimulating interest of Drs J. Henke, D. Stephan and N. Mattern is gratefully acknowledged.

References

DEBYE, P. & BUECHE, A. M. (1949). *J. Appl. Phys.* **20**, 518–525.
HARRISON, R. J. & PASKIN, A. (1964). *Acta Cryst.* **17**, 325–331.
OTTO, J. (1984). *Z. Kristallogr.* **167**, 55–64.
SONNTAG, U., STOYAN, D. & HERMANN, H. (1981). *Phys. Status Solidi A*, **68**, 281–288.
STOYAN, D. (1979). *Biom. Z.* **21**, 693–715.
STOYAN, D., KENDALL, W. S. & MECKE, J. (1986). *Stochastic Geometry and its Applications.* Chichester: Wiley & Sons.
SUORTTI, P. (1972). *J. Appl. Cryst.* **5**, 325–331.
TRUCANO, P. & BATTERMAN, B. W. (1970). *J. Appl. Phys.* **41**, 3949–3953.
VALVODA, V. & ČAPCOVÁ, P. (1984). *Phys. Status Solidi A*, **81**, 203–208.
WOLFF, P. M. DE (1956). *Acta Cryst.* **9**, 682–683.

Acta Cryst. (1987). A**43**, 405–417

Deduction and Systematic Classification of Spatial Motifs of the Antiparallel β Structure in Globular Proteins

BY YU. N. CHIRGADZE

Institute of Protein Research, Academy of Sciences of the USSR, 142292 *Pushchino, Moscow Region, USSR*

(*Received* 17 *June* 1986; *accepted* 2 *December* 1986)

Abstract

A systematic classification of complex topological types for polypeptide-chain folding in the antiparallel β form is proposed. Three well known simple topological types, **β**, **m**, **g**, were chosen as the basic ones: single β strand, hairpin of two strands and simple Greek key type of four strands. The new topologically allowed motifs are formed of a combination of the three basic motifs. All spatial motif types possible with this basis were considered for more complicated double Greek key motifs. This was done on the basis of a complete set of 14 basic spatial motifs of simple Greek key topology. Analysis of about 20 globular proteins shows that some spatial motifs appear to be realized as the main part of the chain fold of the molecule. This suggests that chain folds of antiparallel β proteins are necessarily conditioned by simple topological requirements.

1. Introduction

Polypeptide-chain folding in a globular protein molecule is evidently governed by general regularities. In particular, it depends on topological requirements for the main-chain fold and the packing symmetry of the repeated motifs (Ptitsyn & Finkelstein, 1980; McLachlan, 1980; Richardson, 1981). In recent years a number of new three-dimensional protein structures have been determined, and for many known protein structures the data have been considerably improved at higher resolution. This promotes a more successful study of the principles of protein structure. Two general topological motifs of the up-and-down and simple 'Greek key' types were shown to exist in globular β proteins (Richardson, 1977). A chain fold for such proteins or domains was shown to have a highly limited number of topological variants (Ptitsyn, Finkelstein & Falk, 1979; Finkelstein, Ptitsyn & Bendsko, 1979). Some of the possible spatial patterns for the simple 'Greek key' topology were shown to be part of the molecular architecture in β proteins (Efimov, 1982). Therefore, the next step in studying spatial motifs for the antiparallel β structure in proteins became possible. This consists of the deduction and classification of all topologically allowed spatial motifs. A preliminary communication on this topic has recently been published (Chirgadze, 1985). The practical significance of this study is a complete summary of spatial motifs of the antiparallel

0108-7673/87/030405-13$01.50

β form as constructed on the basis of the simple motif types.

2. Outline of deduction and classification

The spatial organization of a globular protein with a regular antiparallel β form presents a structure with one or two pleated sheets. In many cases the three-dimensional arrangement of the two-sheet β form is rather complex and there are two alternatives for its geometrical approximation: one is the cylindrical or 'barrel' model (Richardson, 1981), and the other is the bilayer or 'sandwich' model (Chothia & Janin, 1981, 1982). Each layer is a pleated β sheet with a right-handed twist (Chothia, 1973) and both layers are in face-to-face contact. The surface of the β sheet is often bent and can be described as the surface of a twisted hyperboloid (Novotny, Bruccoleri & Newell, 1984). There will be no topological difference between the two described models if we can distinguish the beginning and end of each sheet. In this case, both models coincide if they are properly deformed. We suggest that, in the majority of cases we have considered in this paper, the bilayer model seems to be more suitable because it can easily represent disruption of the peptide hydrogen bonding between β strands at the ends of the sheets in direct accordance with the data of X-ray structure analysis at high resolution. On the other hand, there is a limited number of cases where the two-sheet structure degenerates to the barrel structure with all the β strands joined to each other by peptide hydrogen bonds.

The extended β strands in the sheet are always bound by hydrogen bonds, and the sheets are bound together by a hydrophobic interaction. The main element of such a structure is a β strand.

We consider a spatial motif of the antiparallel β structure in an *idealized form*. First, the polypeptide chain is assumed to be a homogeneous cord. Such an assumption allows a much easier solution of the deduction problem for spatial motifs. Otherwise we will face a complicated result which seems to be more suitable for an advanced study of the problem. Second, we consider all the sheets as planes with hydrogen-bonded β strands. This condition is very useful for the model graphic representation, but for comparison of such a model with real structures the model should be deformed without changing the topological features of the motif. That is why one can neglect the exact coincidence of the β-strand registers between the sheets. Nevertheless, the third condition for the idealized motif is assumed to be an antiparallel arrangement of the nearest β strands located in different sheets. One can readily see that the last two conditions were introduced just for the convenience of the model representation. A convenient end-on view diagram for β proteins has been proposed by Levitt & Chothia (1976). We will use a simplified version of this diagram which is widely used at present (see, for example, Ptitsyn & Finkelstein, 1980).

We will only consider motifs for one- or two-sheet β structures. Each spatial motif is assumed to belong to a definite *structural class* which is determined by the total number of β strands. An appropriate systematic classification of complex spatial motifs can be proposed on the basis of some simple basic motifs. In globular proteins with the antiparallel β structure, at least three simple topological motifs are encountered very often (Richardson, 1977). These are a single β strand, a two-strand hairpin and a so-called four-strand simple 'Greek key' motif. Thus, the following topological types (and corresponding set of spatial motifs) are obvious choices as *basic ones*:

(1) **β** is a single β strand of the polypeptide chain;

(2) **m** is a motif of the up-and-down topology. Its spatial type consists of two nearest β strands of one sheet;

(3) **g** is a motif of the simple Greek key topology. It consists of four β strands located in one or two sheets.

A schematic plane representation of these motifs is given in Fig. 1. One may recall that the units of the **m** and **g** motifs often occur in symmetrical antique ornaments. The **m** motif is a plane hairpin where two β strands are bound by hydrogen bonds. The motif of simple Greek key topology has two versions, $\mathbf{g}^+$ and $\mathbf{g}^-$. The $\mathbf{g}^-$ motif is derived from the $\mathbf{g}^+$ motif by travelling along the chain in the opposite direction. Correspondingly, in motif $\mathbf{g}^+$ the order of the β strands is 1432, and in motif $\mathbf{g}^-$ it is 4123.

3. Deduction and systematic classification of topological types for the spatial antiparallel β motifs

As described above, an idealized β structure can be represented either by a one-sheet, *i.e.* plane, or a two-sheet structure with antiparallel packing of the strands. There is no other special geometrical condition. However, for simplicity of the model representation, we will assume the β strands to be short straight strings of equal length arranged in sheets along the straight line which is perpendicular to the strands and passes through their centres. It is very important that all the strands belong to only one chain. That is why each spatial motif must satisfy a definite topo-

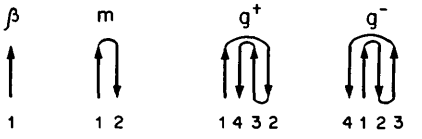


Fig. 1. Two-dimensional schematic representation of the basic topological types **β**, **m** and **g** for the motifs of the antiparallel β structure. Numbers indicate the arrangement of the strands along the chain.

logical requirement. We suggest requirements of the following types.

Structural requirements

(1) The nearest β strands are located in the sheet in an antiparallel manner.

(2) The sheets are parallel to each other, so that the β strands of different sheets are collinear.

(3) The nearest β strands of different sheets are located in an antiparallel manner.

Topological requirements

(1) A chain pathway is prohibited between the β strands in the sheet.

(2) A chain pathway is prohibited between sheets.

(3) Chain crossing and knot forming are prohibited.

All the structural requirements are caused by the geometry of the idealized β structure. The first two topological requirements are conditioned by hydrogen bonding of the β strands in the sheet and strong side-chain interaction in the space between sheets. The third topological condition is explained by the chemical property of the polypeptide chain.

In the complex motif an *overlapping* takes place due to the complexity of the basic **m** and **g** motifs. We have introduced an order of overlapping for the basic motifs constituting the complex motif. The order of overlapping is equal to the number of common overlapped β strands. The basic **β** motif always has an order of overlapping equal to zero. In the protein structure the **β** motif joins neighbouring β strands either in the sheet (hydrogen bonding) or between the sheets (hydrophobic interaction). The order of overlapping of the basic motif is always less than the number of extended strands in the motif, *i.e.* it is equal to 0 for the **β** motif, it can be equal to 0 or 1 for the **m** motif, and it can be equal to 0, 1, 2 or 3 for the **g** motif. Otherwise, a smaller motif is simply a part of a bigger one. When required, the order of overlapping can be marked by subscripts in the motif symbols. Below we will not use this subscript when it is equal to unity for any motif or to zero for the **β** motif. As accepted, we will use superscripts to mark the multiplication of the same motifs with the order of overlapping equal to unity.

Let us clarify this scheme for some complex motifs. We consider the motifs **mm**, $\boldsymbol{\beta}_0\mathbf{g}^+$ and $\mathbf{g}_0^+\boldsymbol{\beta}$ composed of the overlapping motifs (Fig. 2). The first motif consists of three extended strands, and the second and third ones of five strands. In the first case the order of overlapping is equal to unity, in the others to zero. The example of the last two motifs shows clearly that $\boldsymbol{\beta}\mathbf{g}^+ \neq \mathbf{g}^+\boldsymbol{\beta}$, *i.e.* the permutation leads to a different motif.

There are two important general features of the spatial motifs. First, all the two-sheet spatial motifs can be divided into two subgroups of *enantiomorphous motifs*. Suppose we place a given motif in the coordinate system with the origin at any point located outside the space occupied by the motif trajectory. For example, the origin can be located at the zero point of the first β strand. Then the motif trajectory can be obtained by moving the end of the position vector along the chain. Similarly, a centrosymmetric trajectory can be obtained relative to the inversion centre placed at the origin. It can be shown that this new motif is a mirror-image equivalent, *i.e.* both motifs form an enantiomorphous pair.

Second, all the spatial motifs can be divided also into two other subgroups of *direct and opposite motifs* relative to the direction of travel along the chain. This phenomenon is taken into account by introducing the opposite basic motifs $\boldsymbol{\beta}^-$, $\mathbf{m}^-$ and $\mathbf{g}^-$. In fact, we are already using such basic motifs since the direct **β** and **m** motifs are identical with their opposite analogues in the approximation of polypeptide chains as a homogeneous cord.

The generation of the new topologically permitted motifs is created by a combination of the three basic motifs **β**, **m** and **g**. Thus, we can compile some *topological types* (Table 1). To avoid ambiguity in designation, it is necessary to define in complex motifs the basic motifs of highest possible rank. The class G_k of the β structure contains all motifs for the structure with k β strands. The motifs for the structural class G_{k+1} are generated by an addition of the motif **β** without overlapping, and the motifs **m** or **g** with overlapping. A symbolic description of the topological type for the complex motif corresponds to the location of the basic motifs along the polypeptide chain always beginning at the N terminal.

Let us consider Table 1. For the third structural class there are some spatial motifs of three topological types $\mathbf{m}^2$, **mβ** and **βm**. In the case of a one-sheet structure, all these spatial motifs degenerate into the motif $\mathbf{m}^2$. For the fourth structural class, a new basic motif of topological type **g** appears in addition to the several motifs composed of **m** and **β** types. For higher structural classes, the number of topological types rapidly increases. From the structural class $G_k = 5$ onwards, an important new quality appears in the form of the overlapped Greek key motifs $\mathbf{g}_n\mathbf{g}$, where n is the order of overlapping. It is interesting to note

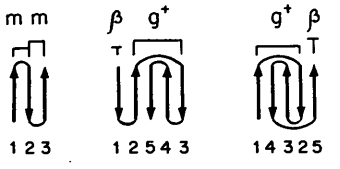


Fig. 2. Examples of complex topological motifs with overlapping of the basic motifs.

Table 1. *General classification of the topological types of the spatial motifs of the antiparallel β structure*

Class number of the β structure	Topological types																						
	1	2				3			4				5	6			7						
k	$\mathbf{m}^{k-1}$	$\mathbf{m}^{k-2}\boldsymbol{\beta}$	$\boldsymbol{\beta}\mathbf{m}^{k-2}$	$\boldsymbol{\beta}\mathbf{m}^{k-3}\boldsymbol{\beta}$	$\mathbf{m}^{k-4}\mathbf{g}$	$\mathbf{g}\mathbf{m}^{k-4}$	$\mathbf{m}^{i}\mathbf{g}\mathbf{m}^{j}$	$\mathbf{m}^{k-5}\boldsymbol{\beta}\mathbf{g}$	$\boldsymbol{\beta}\mathbf{g}\mathbf{m}^{k-5}$	$\mathbf{m}^{k-5}\mathbf{g}\boldsymbol{\beta}$	$\mathbf{g}\boldsymbol{\beta}\mathbf{m}^{k-5}$	$\mathbf{g}_{8-k}\mathbf{g}$	$\mathbf{m}^{i}\mathbf{g}_{9-k}\mathbf{g}$	$\mathbf{g}_{9-k}\mathbf{g}\mathbf{m}^{j}$	$\mathbf{m}^{i}\mathbf{g}_{10-k}\mathbf{m}^{j}$	$\boldsymbol{\beta}^{i}\mathbf{g}_{9-k}\mathbf{g}$	$\mathbf{g}_{9-k}\mathbf{g}\boldsymbol{\beta}^{j}$	$\boldsymbol{\beta}^{i}\mathbf{g}_{10-k}\mathbf{g}\boldsymbol{\beta}^{j}$
2	$\mathbf{m}$																		
3	$\mathbf{m}^2$	$\mathbf{m}\boldsymbol{\beta}$	$\boldsymbol{\beta}\mathbf{m}$																
4	$\mathbf{m}^3$	$\mathbf{m}^2\boldsymbol{\beta}$	$\boldsymbol{\beta}\mathbf{m}^2$	$\boldsymbol{\beta}\mathbf{m}\boldsymbol{\beta}$	$\mathbf{g}$	$\mathbf{g}$													
5	$\mathbf{m}^4$	$\mathbf{m}^3\boldsymbol{\beta}$	$\boldsymbol{\beta}\mathbf{m}^3$	$\boldsymbol{\beta}\mathbf{m}^2\boldsymbol{\beta}$	$\mathbf{m}\mathbf{g}$	$\mathbf{g}\mathbf{m}$		$\boldsymbol{\beta}\mathbf{g}$	$\boldsymbol{\beta}\mathbf{g}$	$\mathbf{g}\boldsymbol{\beta}$	$\mathbf{g}\boldsymbol{\beta}$	$\mathbf{g}_2\mathbf{g}$					$\boldsymbol{\beta}\mathbf{g}_2\mathbf{g}$	$\mathbf{g}_2\mathbf{g}\boldsymbol{\beta}$	
6	$\mathbf{m}^5$	$\mathbf{m}^4\boldsymbol{\beta}$	$\boldsymbol{\beta}\mathbf{m}^4$	$\boldsymbol{\beta}\mathbf{m}^3\boldsymbol{\beta}$	$\mathbf{m}^2\mathbf{g}$	$\mathbf{g}\mathbf{m}^2$	$\mathbf{m}\mathbf{g}\mathbf{m}$	$\mathbf{m}\boldsymbol{\beta}\mathbf{g}$	$\boldsymbol{\beta}\mathbf{g}\mathbf{m}$	$\mathbf{m}\mathbf{g}\boldsymbol{\beta}$	$\mathbf{g}\boldsymbol{\beta}\mathbf{m}$	$\mathbf{g}_2\mathbf{g}$	$\mathbf{m}\mathbf{g}_3\mathbf{g}$	$\mathbf{g}_3\mathbf{g}\mathbf{m}$			$\boldsymbol{\beta}\mathbf{g}_3\mathbf{g}$	$\mathbf{g}_3\mathbf{g}\boldsymbol{\beta}$	
7	$\mathbf{m}^6$	$\mathbf{m}^5\boldsymbol{\beta}$	$\boldsymbol{\beta}\mathbf{m}^5$	$\boldsymbol{\beta}\mathbf{m}^4\boldsymbol{\beta}$	$\mathbf{m}^3\mathbf{g}$	$\mathbf{g}\mathbf{m}^3$	$\mathbf{m}\mathbf{g}\mathbf{m}^2$	$\mathbf{m}^2\boldsymbol{\beta}\mathbf{g}$	$\boldsymbol{\beta}\mathbf{g}\mathbf{m}^2$	$\mathbf{m}^2\mathbf{g}\boldsymbol{\beta}$	$\mathbf{g}\boldsymbol{\beta}\mathbf{m}^2$	$\mathbf{g}\mathbf{g}$	$\mathbf{m}\mathbf{g}_2\mathbf{g}$	$\mathbf{g}_2\mathbf{g}\mathbf{m}$	$\mathbf{m}\mathbf{g}_3\mathbf{g}\mathbf{m}$		$\boldsymbol{\beta}\mathbf{g}_2\mathbf{g}$	$\mathbf{g}_2\mathbf{g}\boldsymbol{\beta}$	$\boldsymbol{\beta}\mathbf{g}_3\mathbf{g}\boldsymbol{\beta}$
8	$\mathbf{m}^7$	$\mathbf{m}^6\boldsymbol{\beta}$	$\boldsymbol{\beta}\mathbf{m}^6$	$\boldsymbol{\beta}\mathbf{m}^5\boldsymbol{\beta}$	$\mathbf{m}^4\mathbf{g}$	$\mathbf{g}\mathbf{m}^4$	$\mathbf{m}^2\mathbf{g}\mathbf{m}$	$\mathbf{m}^3\boldsymbol{\beta}\mathbf{g}$	$\boldsymbol{\beta}\mathbf{g}\mathbf{m}^3$	$\mathbf{m}^3\mathbf{g}\boldsymbol{\beta}$	$\mathbf{g}\boldsymbol{\beta}\mathbf{m}^3$	$\mathbf{g}_0\mathbf{g}$	$\mathbf{m}\mathbf{g}\mathbf{g}$	$\mathbf{g}\mathbf{g}\mathbf{m}$	$\mathbf{m}\mathbf{g}_2\mathbf{g}\mathbf{m}$		$\boldsymbol{\beta}\mathbf{g}\mathbf{g}$	$\mathbf{g}\mathbf{g}\boldsymbol{\beta}$	$\boldsymbol{\beta}\mathbf{g}_2\mathbf{g}\boldsymbol{\beta}$

Notes: The basic motifs are **β**, **m** and **g** (see text). The upper index designates a multiplication and the lower index an order of overlapping. An order of overlapping is not indicated if it is equal to unity. The **β** motif always has zero order of overlapping.

that each topological type can be described by the general formula applied for any structural class. The table presented is not exhaustive. It shows only the regularities of the new motif generation and advantages of the general classification. For example, some topological types with permutation of **β** and **m** motifs are not presented in the table. This table will be infinite when higher structural classes are considered. In this case, new families will appear having a new general formula. For example, a new family of heterooligomers will be generated with the general formula

$$M = M_a^i M_b^j \ldots M_d^n,$$

where $M_a, M_b, \ldots, M_d$ are some spatial motifs; $i, j, \ldots, n$ are the multiplications of these motifs.

In fact, repeated motifs with the homooligomer formula $M = M_a^n$ occur very often in molecules of globular proteins (Richardson, 1981).

The *spatial motifs* of any given topological type can be obtained by a combination of the basic spatial motifs of the **β**, **m** and **g** topologies. As a rule, the spatial motifs are degenerate. There are a few types of degeneracy which are conditioned by the following factors:

(1) The new basic motif can be placed in two positions, *i.e.* at either side of the original spatial motif.

(2) The basic motif **g** exists in two forms $\mathbf{g}^+$ and $\mathbf{g}^-$, and their spatial motifs are highly degenerate.

(3) The mutual disposition of the sheets can be changed.

As a result, a number of spatial motifs can correspond to one topological type.

4. Deduction of spatial motifs of the simple Greek key topology

Each of the basic topological types **β** and **m** has only one spatial motif (Fig. 1). The basic topological type **g** is more complicated and gives rise to a number of spatial motifs. Let us deduce all these motifs. Their schematic perspective views and butt-end projections are presented in Fig. 3.

In the *one-sheet* structure all four β strands can be placed in one sheet. In this case only two plane spatial

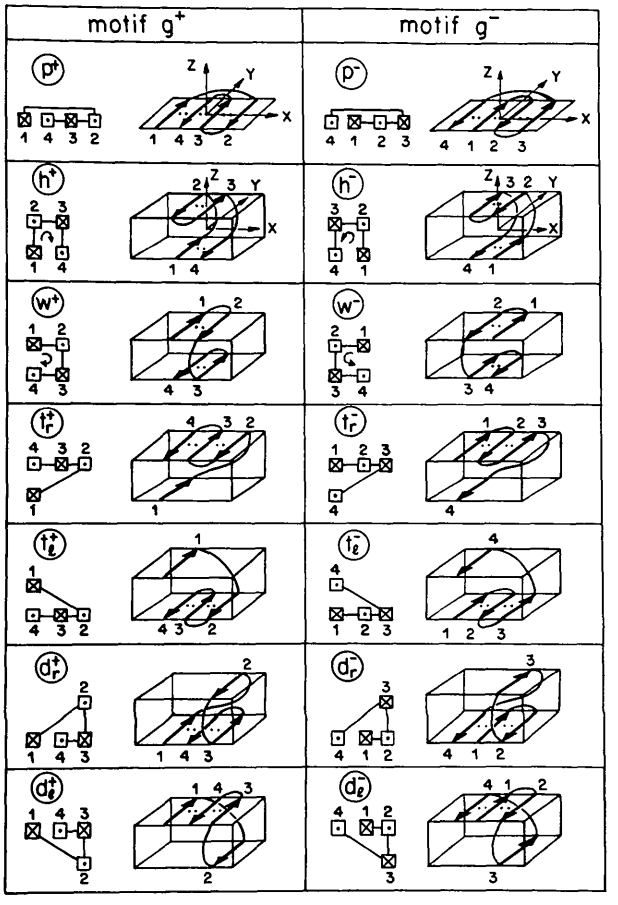


Fig. 3. A complete set of the basic spatial motifs derived for two topological types of the simple 'Greek key' topology. The perspective views and butt-end projections are presented. The superscript index + or − is related to the direct or opposite type of the **g** motif, the subscript index is related to the mirror-image equivalents.

Table 2. *Basic spatial motifs of topological type* **g** *for the antiparallel* β *structure, structural class* G_4

No.	Physical type	Sheet number	β-strand number in the sheets	Enantiomorphous type *r* type	*l* type
1	**p**	1	4	$\mathbf{p}^+, \mathbf{p}^-$	—
2	**h**	2	2 and 2	$\mathbf{h}^+$	$\mathbf{h}^-$
3	**w**	2	2 and 2	$\mathbf{w}^+$	$\mathbf{w}^-$
4	**t**	2	1 and 3	$\mathbf{t}_r^+, \mathbf{t}_r^-$	$\mathbf{t}_l^+, \mathbf{t}_l^-$
5	**d**	2	1 and 3	$\mathbf{d}_r^+, \mathbf{d}_r^-$	$\mathbf{d}_l^+, \mathbf{d}_l^-$

motifs $\mathbf{p}^+$ and $\mathbf{p}^-$ are allowed in accordance with the $\mathbf{g}^+$ and $\mathbf{g}^-$ topology (see Fig. 3).

In the *two-sheet* structure all possible spatial motifs can be obtained by different combinations of the β strands, provided the structural and topological requirements are satisfied. First, we consider the motifs with two β strands placed in each sheet. For the $\mathbf{g}^+$ motifs (1432) the following pair combinations of the antiparallel β strands are possible: 14, 12, 43 and 32. They give two different hairpin motifs $\mathbf{h}^+$ (14 and 23) and $\mathbf{w}^+$ (12 and 34). In a similar manner, the spatial motifs $\mathbf{h}^-$ and $\mathbf{w}^-$ are deduced for the topological type $\mathbf{g}^-$. Second, we consider the motifs with three β strands placed in one of the sheets. For the $\mathbf{g}^+$ motif (1432) only the combinations 143 and 432 of the antiparallel β strands are possible. Now we take into account the permutation of the sheets. As a result, we take two 'right', $\mathbf{t}_r^+$ and $\mathbf{d}_r^+$, and two 'left', $\mathbf{t}_l^+$ and $\mathbf{d}_l^+$, mirror-image spatial motifs (see Fig. 3). Similarly, for the $\mathbf{g}^-$ topological type we have four corresponding spatial motifs. It can be noted that two *abcd* structures (Efimov, 1982) built from the N and C terminals of the polypeptide chain coincide with the spatial motifs $\mathbf{d}_r^-$ and $\mathbf{d}_r^+$, both beginning at the N terminal of the chain.

Thus, for each of the topological types **g** we obtained the motifs of five *physical spatial types*: **p**, a plane structure; **h** and **w**, hairpins; **t** and **d**, different two-sheet structures with three β strands in one of the sheets. As shown earlier, all the two-sheet spatial motifs obtained should be divided into two subgroups of the mirror-image motifs. It is easy to see that hairpin motifs $\mathbf{h}^+$ and $\mathbf{w}^+$ have their mirror-image equivalent motifs which exactly coincide with the $\mathbf{h}^-$ and $\mathbf{w}^-$ spatial motifs. Now, we can present all deduced motifs in a general form (Fig. 4). Finally, we have 14 spatial motifs of the **g** type (Table 2).

Therefore, we have a total of 16 basic spatial motifs of the topological types **β**, **m** and **g**. Any other possible spatial motif of the antiparallel β structure can be obtained from combinations of these.

5. Deduction of the spatial motifs of the double Greek key $\mathbf{g}_n\mathbf{g}$ topology

The complex spatial motifs built from different basic spatial ones can easily be obtained. On the other hand, the spatial motifs of the topological type $\mathbf{g}_n\mathbf{g}$ in the structural classes G_5–G_8 are much more complicated. The relationship between the topological, physical and spatial types for such motifs is shown in Table 3. Allowed spatial motifs are deduced by the treatment of all 196 combined pairs of 14 spatial **g** motifs, and we will consider them in detail below. The main result in the form of a selection-rule matrix for all combination pairs of all physical types **p**, **h**, **w**, **t** and **d** is presented in Table 4. The value of unity in the table denotes the existence of one or several spatial motifs with given physical type. It is interesting to note that the spatial motifs of the diagonal terms are pseudosymmetrical. There is a mirror plane for the motifs of structural classes G_5 and G_7, an inversion centre for class G_6, and a dyad axis for class G_8. All allowed non-diagonal spatial motifs are asymmetrical.

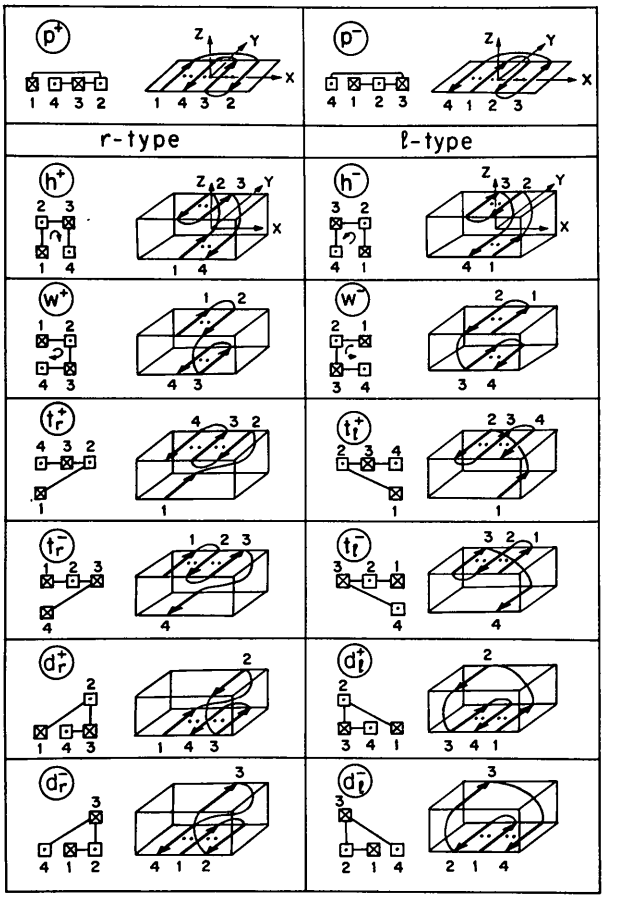
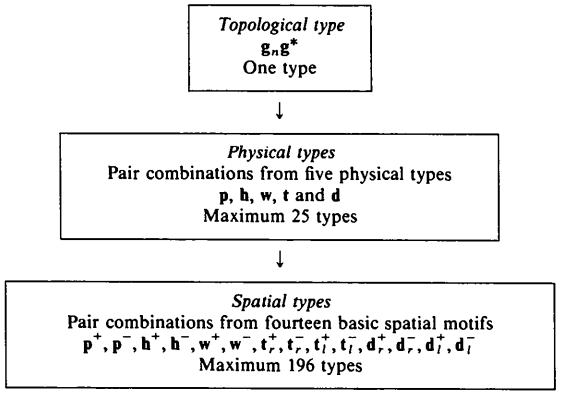


Fig. 4. General representation of the basic spatial motifs of topological type **g**. The structural class of the antiparallel β structure is G_4. The group of two-sheet motifs is divided into two subgroups of enantiomorphous, *i.e.* mirror-image-equivalent, spatial motifs.

Table 3. *Hierarchy for the motif types of the antiparallel β structure with double Greek key topology*



* Here n is the order of overlapping of the g motifs; it can be equal to 0, 1, 2 or 3.

(*a*) *Spatial motifs of the* $\mathbf{g_3g}$ *type, structural class* G_5

The selection rules for spatial motifs of this type are presented in matrix form in Table 5, and perspective views of them are given in Fig. 5. First of all, we reveal the selection-rule matrix to be asymmetric in contrast to the corresponding matrix of the general physical types (see Table 4). This is a direct consequence of the permutations of two basic spatial motifs. There are only 11 possible spatial motifs, mainly because of the high value of the order of overlapping of the basic motifs. Three motifs, $\mathbf{p^+p^-}$, $\mathbf{t_r^+t_r^-}$ and $\mathbf{t_l^+t_l^-}$ are pseudosymmetrical with a mirror plane perpendicular to the sheet planes and intersecting the third β strand. The two-sheet motifs are divided into two subgroups of the *r* and *l* enantiomorphous types. All two-sheet spatial motifs of the $\mathbf{g_3g}$ type are listed in Table 6. Spatial motif pairs **pt–tp** and **hd–dh** are direct and opposite motif pairs respectively.

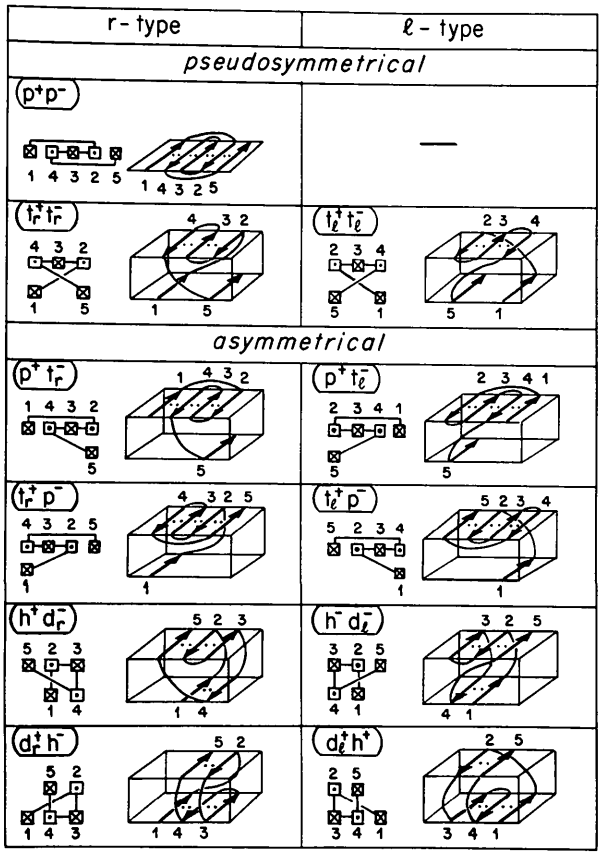


Fig. 5. Allowed spatial motifs of topological type $\mathbf{g_3g}$ for the structural class G_5.

(*b*) *Spatial motifs of the* $\mathbf{g_2g}$ *type, structural class* G_6

There are only ten possible spatial motifs of this topological type, and only two-sheet motifs are allowed. The selection-rule matrix is presented in Table 7, and all the spatial motifs are given in Fig. 6. There are six pseudosymmetrical motifs with the inversion centre located between the sheets. These motifs have very compact chain folds. All the motifs are listed in Table 8. Spatial motif pairs **pw–wp** are direct and opposite motif pairs respectively.

(*c*) *Spatial motifs of the* **gg** *type, structural class* G_7

There are 55 possible spatial motifs of this topological type. The selection-rule matrix is given in Table

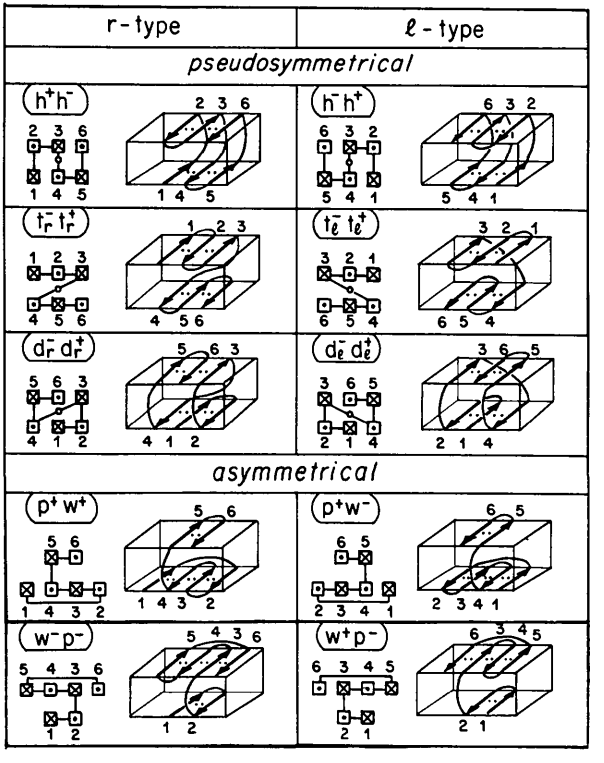


Fig. 6. Allowed spatial motifs of topological type $\mathbf{g_2g}$ for the structural class G_6.

Table 4. *Selection-rule matrices for general physical types of spatial motifs with topological type* $\mathbf{g}_n\mathbf{g}$

Type $\mathbf{g}_3\mathbf{g}$, Class G_5

First motif \ Second motif	p	h	w	t	d
p	1_m	0	0	1	0
h	0	0	0	0	1
w	0	0	0	0	0
t	1	0	0	1_m	0
d	0	1	0	0	0

Type $\mathbf{g}_2\mathbf{g}$, Class G_6

First motif \ Second motif	p	h	w	t	d
p	0	0	1	0	0
h	0	1_i	0	0	0
w	1	0	0	0	0
t	0	0	0	1_i	0
d	0	0	0	0	1_i

Type $\mathbf{gg}$, Class G_7

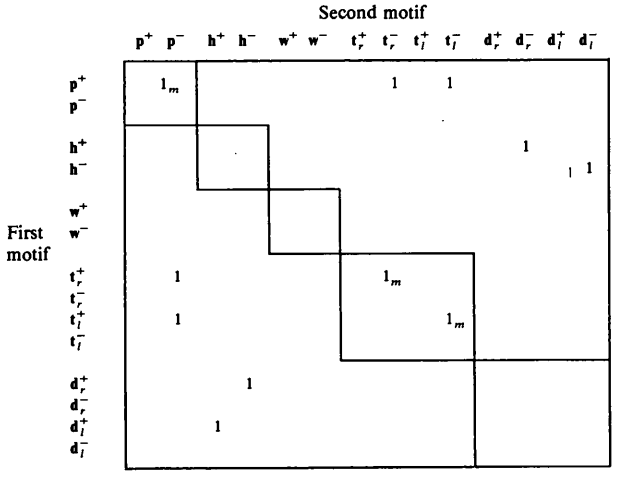
First motif \ Second motif	p	h	w	t	d
p	1_m	1	1	1	1
h	1	0	0	0	1
w	1	0	0	0	1
t	1	0	0	0	1
d	1	1	1	1	1_m

Type $\mathbf{g}_0\mathbf{g}$, Class G_8

First motif \ Second motif	p	h	w	t	d
p	1_2	1	1	1	1
h	1	1_2	1	1	1
w	1	1	1_2	1	1
t	1	1	1	1_2	1
d	1	1	1	1	1_2

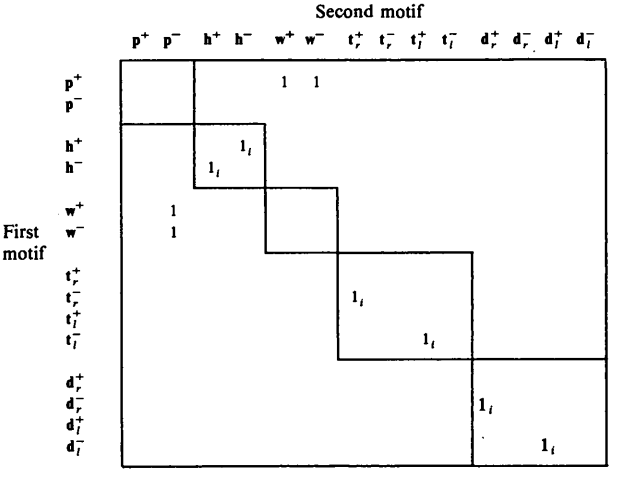
Note: Description of the general physical types **p**, **h**, **w**, **t** and **d** for the **g** topology is given in the text. Non-zero terms mean allowed motifs. Symbol subscripts designate the pseudosymmetry elements: *m* - mirror plane, *i* - inversion centre, 2 - dyad axis.

Table 5. *A selection-rule matrix for the spatial motifs of topological type* $\mathbf{g}_3\mathbf{g}$, *structural class* G_5



Note: Non-marked terms are not allowed, subscript *m* designates a mirror plane.

Table 7. *A selection-rule matrix for spatial motifs of topological type* $\mathbf{g}_2\mathbf{g}$, *structural class* G_6



Note: Non-marked terms are not allowed, subscript *i* designates an inversion centre.

Table 6. *Allowed spatial motifs of the* $\mathbf{g}_3\mathbf{g}$ *type, structural class* G_5

No.	Physical type	Pseudosymmetry	Enantiomorphous type *r* type	*l* type
1	**pp**	*m*	$\mathbf{p}^+\mathbf{p}^-$	—
2	**tt**	*m*	$\mathbf{t}_r^+\mathbf{t}_r^-$	$\mathbf{t}_l^+\mathbf{t}_l^-$
3	**pt**	1	$\mathbf{p}^+\mathbf{t}_r^-$	$\mathbf{p}^+\mathbf{t}_l^-$
4	**tp**	1	$\mathbf{t}_r^+\mathbf{p}^-$	$\mathbf{t}_l^+\mathbf{p}^-$
5	**hd**	1	$\mathbf{h}^+\mathbf{d}_r^-$	$\mathbf{h}^-\mathbf{d}_l^-$
6	**dh**	1	$\mathbf{d}_r^+\mathbf{h}^-$	$\mathbf{d}_l^+\mathbf{h}^+$

Table 8. *Allowed spatial motifs of the* $\mathbf{g}_2\mathbf{g}$ *type, structural class* G_6

No.	Physical type	Pseudosymmetry	Enantiomorphous type *r* type	*l* type
1	**hh**	$\bar{1}$	$\mathbf{h}^+\mathbf{h}^-$	$\mathbf{h}^-\mathbf{h}^+$
2	**tt**	$\bar{1}$	$\mathbf{t}_r^-\mathbf{t}_r^+$	$\mathbf{t}_l^-\mathbf{t}_l^+$
3	**dd**	$\bar{1}$	$\mathbf{d}_r^-\mathbf{d}_r^+$	$\mathbf{d}_l^-\mathbf{d}_l^+$
4	**pw**	1	$\mathbf{p}^+\mathbf{w}^+$	$\mathbf{p}^+\mathbf{w}^-$
5	**wp**	1	$\mathbf{w}^-\mathbf{p}^-$	$\mathbf{w}^+\mathbf{p}^-$

9. There are only three pseudosymmetrical motifs, $\mathbf{p}^+\mathbf{p}^-$, $\mathbf{d}_r^+\mathbf{d}_r^-$ and $\mathbf{d}_l^+\mathbf{d}_l^-$, with a mirror plane perpendicular to the sheet planes and intersecting the fourth β strand (Fig. 7). Some spatial motifs of the general physical **pt** and **tp** types are doubly degenerate. This type of degeneracy is related to changing the disposition of one of the sheets. For example, the second version of the spatial motif $\mathbf{p}^+\mathbf{t}_r^+$ can be obtained by rotating the $\mathbf{t}_r^+$ motif around the fourth β strand of the $\mathbf{p}^+$ motif by 180°. All possible spatial motifs of the **gg** type are listed in Table 10. There are direct-opposite pairs in motifs **ph–hp**, **pw–wp** *etc.*

(*d*) *Spatial motifs of the* $\mathbf{g}_0\mathbf{g}$ *type, structural class* G_8

An essential feature of these motifs is the zero order of overlapping of the basic motifs. This means that both simple **g** motifs have no common β strands. The binding contact occurs either through the intrasheet hydrogen bonds or through the intersheet hydrophobic interactions. As a result, the limitations on motif formation are removed. This explains the high degree of degeneracy for some motifs and the large total number of all possible spatial motifs. Therefore, we have considered only the spatial motifs of the homogeneous physical types **pp**, **hh**, **ww**, **tt** and **dd**.

Table 9. *A selection-rule matrix for spatial motifs of topological type* **gg**, *structural class* G_7

First motif \ Second motif	p^+	p^-	h^+	h^-	w^+	w^-	t_r^+	t_r^-	t_l^+	t_l^-	d_r^+	d_r^-	d_l^+	d_l^-
p^+							2		2					
p^-	1_m		1	1	1	1	2	1	2	1	1		1	
h^+	1												1	
h^-	1										1			
w^+	1										1			
w^-	1												1	
t_r^+	1												1	
t_r^-	2	2									1	1		
t_l^+	1										1			
t_l^-	2	2											1	1
d_r^+							1							
d_r^-	1		1			1	1			1	1_m			
d_l^+									1					
d_l^-	1			1	1			1	1					1_m

Note: Non-marked terms are not allowed, subscript m designates a mirror plane.

Table 10. *Allowed spatial motifs of the* **gg** *type, structural class* G_7

No.	Physical type	Pseudo-symmetry	Enantiomorphous type *r* type			*l* type		
1	pp	*m*	p^-p^+	–	–	–	–	–
2	dd	*m*	$d_r^-d_r^+$	–	–	$d_l^-d_l^+$	–	–
3	ph	1	p^-h^+	–	–	p^-h^-	–	–
4	hp	1	h^-p^+	–	–	h^+p^+	–	–
5	pw	1	p^-w^+	–	–	p^-w^-	–	–
6	wp	1	w^-p^+	–	–	w^+p^+	–	–
7	pt	1	$p^+t_r^+$	$p^-t_r^+$	$p^-t_r^-$	$p^+t_l^+$	$p^-t_l^+$	$p^-t_l^-$
8	tp	1	$t_r^-p^-$	$t_r^-p^+$	$t_r^+p^+$	$t_l^-p^-$	$t_l^-p^+$	$t_l^+p^+$
9	pd	1	$p^-d_r^+$	–	–	$p^-d_l^+$	–	–
10	dp	1	$d_r^-p^+$	–	–	$d_l^-p^+$	–	–
11	hd	1	$h^-d_r^+$	–	–	$h^+d_l^+$	–	–
12	dh	1	$d_r^-h^+$	–	–	$d_l^-h^-$	–	–
13	wd	1	$w^+d_r^+$	–	–	$w^-d_l^+$	–	–
14	dw	1	$d_r^-w^-$	–	–	$d_l^-w^+$	–	–
15	td	1	$t_r^+d_r^+$	$t_r^-d_r^+$	$t_r^-d_r^-$	$t_l^+d_r^+$	$t_l^-d_l^+$	$t_l^-d_l^-$
16	dt	1	$d_r^-t_r^-$	$d_r^-t_r^+$	$d_r^+t_r^+$	$d_r^-t_l^-$	$d_l^-t_l^+$	$d_l^+t_l^+$

Note: Some spatial motifs of the **pt** and **tp** types are doubly degenerate.

The selection rules for such motifs are presented in the block-diagonal matrix of Table 11.

We should take into account the different location and orientation of the second motif at the formation of the complex motif of the **gg** type. The sheet permutation must also be performed for the **pp** motif.

We will use the following system for designation of the disposition of the second motif relative to the first one. The main coordinate system is associated with the first motif, and its origin is placed at the geometrical centre of the first motif, as shown in Fig. 4. The transformation of vector $\mathbf{A}_1$ by the rotation and translation is described by the formula $\mathbf{A}_2 = \mathbf{A}_1\mathbf{R} + \mathbf{T}$, where $\mathbf{R}$ is the rotation matrix and $\mathbf{T}(a, b, c)$ is the translation vector.

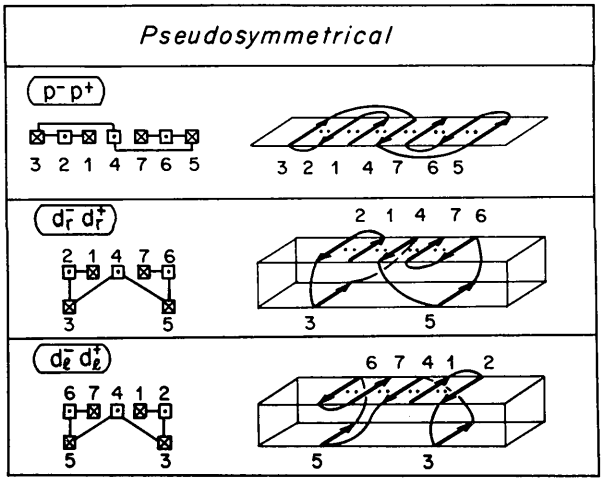


Fig. 7. Allowed pseudosymmetrical spatial motifs of topological type **gg** for the structural class G_7. The two two-sheet motifs are enantiomorphous.

Table 11. *A selection-rule matrix for spatial motifs of topological type* $\mathbf{g_0g}$, *structural class* G_8

First motif \ Second motif	p^+	p^-	h^+	h^-	w^+	w^-	t_r^+	t_r^-	t_l^+	t_l^-	d_r^+	d_r^-	d_l^+	d_l^-
p^+	6_2	6												
p^-	6	6_2												
h^+			4_2	0										
h^-			0	4_2										
w^+					4_2	0								
w^-					0	4_2								
t_r^+							2_2	2	2	2				
t_r^-							2	2_2	2	2				
t_l^+							2	2	2_2	2				
t_l^-							2	2	2	2_2				
d_r^+											1_2	2	0	2
d_r^-											2	1_2	2	0
d_l^+											0	2	1_2	2
d_l^-											2	0	2	1_2

Notes: Only block-diagonal terms were considered, zero terms are not allowed. There are four pseudosymmetrical motifs for all the physical types **pp**, **hh**, **ww** and **dd**, but type **tt** has eight motifs. Subscript 2 designates a dyad axis.

In all our cases, the second motif rotates only by 180° around the x, y, z axes, *i.e.*

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad R_y = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

$$R_z = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We will designate these rotations by a triple index which consists of only diagonal terms of these rotation matrices: $1\bar{1}\bar{1}$, $\bar{1}1\bar{1}$ and $\bar{1}\bar{1}1$.

Similarly, we will designate the translation vector of the second motif along the x, y and z axes by an additional index $a00$, $0b0$ and $00c$ if the vector co-

Table 12. *Allowed pseudosymmetrical spatial motifs of the* $\mathbf{g_0g}$ *type, structural class* G_8

No.	Physical type	Pseudo-symmetry	Enantiomorphous type *r* type	*l* type
1	pp	2	$(\mathbf{p}^+\mathbf{p}^+)_{\bar{1}1\bar{1},00c}$	$(\mathbf{p}^+\mathbf{p}^+)_{\bar{1}1\bar{1},00\bar{c}}$
2	pp	2	$(\mathbf{p}^-\mathbf{p}^-)_{\bar{1}1\bar{1},00c}$	$(\mathbf{p}^-\mathbf{p}^-)_{\bar{1}1\bar{1},00\bar{c}}$
3	hh	2	$(\mathbf{h}^+\mathbf{h}^+)_{\bar{1}1\bar{1},a00}$	$(\mathbf{h}^-\mathbf{h}^-)_{\bar{1}1\bar{1},\bar{a}00}$
4	hh	2	$(\mathbf{h}^+\mathbf{h}^+)_{\bar{1}1\bar{1},\bar{a}00}$	$(\mathbf{h}^-\mathbf{h}^-)_{\bar{1}1\bar{1},a00}$
5	ww	2	$(\mathbf{w}^+\mathbf{w}^+)_{\bar{1}1\bar{1},a00}$	$(\mathbf{w}^-\mathbf{w}^-)_{\bar{1}1\bar{1},\bar{a}00}$
6	ww	2	$(\mathbf{w}^+\mathbf{w}^+)_{\bar{1}1\bar{1},\bar{a}00}$	$(\mathbf{w}^-\mathbf{w}^-)_{\bar{1}1\bar{1},a00}$
7	tt	2	$(\mathbf{t}_r^+\mathbf{t}_r^+)_{\bar{1}1\bar{1},a00}$	$(\mathbf{t}_l^+\mathbf{t}_l^+)_{\bar{1}1\bar{1},a00}$
8	tt	2	$(\mathbf{t}_r^+\mathbf{t}_r^-)_{\bar{1}1\bar{1},\bar{a}00}$	$(\mathbf{t}_l^+\mathbf{t}_l^-)_{\bar{1}1\bar{1},a00}$
9	tt	2	$(\mathbf{t}_r^-\mathbf{t}_r^-)_{\bar{1}1\bar{1},a00}$	$(\mathbf{t}_l^-\mathbf{t}_l^-)_{\bar{1}1\bar{1},a00}$
10	tt	2	$(\mathbf{t}_r^-\mathbf{t}_r^+)_{\bar{1}1\bar{1},\bar{a}00}$	$(\mathbf{t}_l^-\mathbf{t}_l^+)_{\bar{1}1\bar{1},a00}$
11	dd	2	$(\mathbf{d}_r^+\mathbf{d}_r^+)_{\bar{1}1\bar{1},\bar{a}00}$	$(\mathbf{d}_l^+\mathbf{d}_l^+)_{\bar{1}1\bar{1},a00}$
12	dd	2	$(\mathbf{d}_r^-\mathbf{d}_r^-)_{\bar{1}1\bar{1},\bar{a}00}$	$(\mathbf{d}_l^-\mathbf{d}_l^-)_{\bar{1}1\bar{1},a00}$

ordinates are positive, or by the index $\bar{a}00$, $0\bar{b}0$ and $00\bar{c}$ if they are negative.

Then, for example, the spatial motif formula $(\mathbf{p}^+\mathbf{p}^+)_{\bar{1}1\bar{1},00\bar{c}}$ means that the second motif $\mathbf{p}^+$ is rotated around the y axis relative to the first by 180° and is displaced in the z direction by the value $-c$.

The total number of spatial motifs considered is 92. Among them are 24 pseudosymmetrical motifs with a dyad axis (Table 12). Their three-dimensional

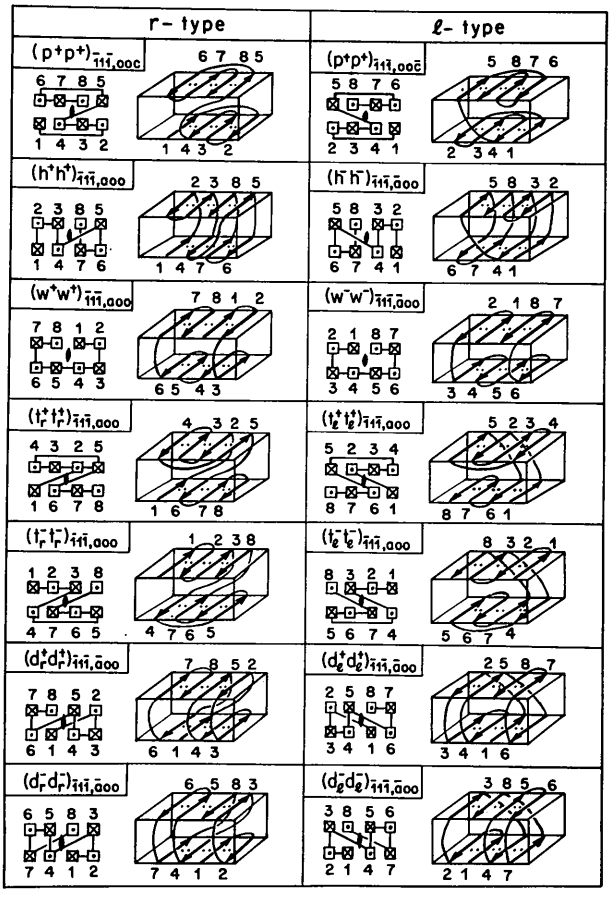


Fig. 8. Some allowed pseudosymmetrical two-sheet spatial motifs of topological type $\mathbf{g_0g}$ for the structural class G_8.

patterns can easily be built with the use of the subscript index information. Some examples of these motifs are given in Fig. 8. The dyad axis is located between the sheets along the line parallel to the β strands. All these motifs are very compact. They have different connectivity of the two sheets. The motifs of the **dd** type display maximal connectivity: their sheets are bound by five side-sheet linkers. Such structural patterns appear to be much more stable than the others.

6. Elaboration of the deduced spatial motifs in globular proteins

In globular protein molecules, an antiparallel β structure differs somewhat from the idealized model considered. First, the real structure is a pleated sheet; this feature does not change any result obtained. Second, the real structure has twisted β sheets. In some cases a bilayer structure has unaligned orthogonal β-sheet packing (Chothia & Janin, 1982). These features do not violate the topological requirements but before starting to classify the spatial patterns of some real protein structure one must unbend it, so that it approaches as closely as possible the idealized model. In some cases, the β strands can be of different length, but this does not seriously increase the discrepancy between the unbent real and idealized motifs. In the most complicated cases, it is suggested that a convenient computational approach be applied in order to derive an adequate idealized model. Novotny, Bruccoleri & Newell (1984) give an idea of such an approach. These authors propose a method for approximation of the surface of a β barrel structure by a twisted hyperboloid. This surface can be unwrapped, then all sheet regions are detected and an idealized model is built. However, in all cases which we analysed such a procedure was unnecessary.

The number of β protein structures known at present seems to be insufficient to observe all the revealed spatial motifs. We have considered about 20 protein structures with an antiparallel β form. All these proteins contain the spatial motifs as the main part of their structure. Examples are presented in Table 13. All motifs are related to the structural classes with five to eight β strands. Good examples of the plane one-sheet motifs are domain 3 of glutathione reductase (Schulz, Schirmer, Sachenheimer & Pai, 1978; Thieme, Pai, Schirmer & Schulz, 1981) and the subtilisin inhibitor from *Streptomyces* (Mitsui, Sato, Watanabe & Itaka, 1979). γ-Chymotrypsin has a two-domain structure (Cohen, Silverton & Davies, 1981). Each domain has a twisted barrel structure which can be described as a motif of topological type $\mathbf{g_2g}$, structural class G_6. The spatial motif of this structure is $\mathbf{t}_l^+\mathbf{t}_l^- \simeq \mathbf{h}^-\mathbf{h}^+ \simeq \mathbf{d}_r^-\mathbf{d}_r^+$. The degeneracy occurs owing to the high degree of deformation of the two-sheet spatial motif into the structure of the closed barrel

Table 13. *Examples of spatial motifs of the antiparallel β structure in globular proteins*

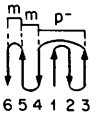
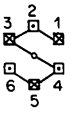
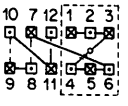
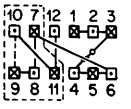
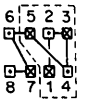
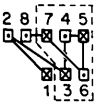
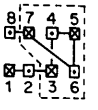
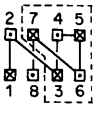
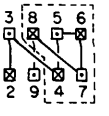
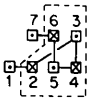
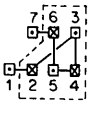
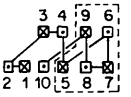
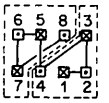
No.	Protein, reference	Schematic diagram of chain tracing	Structural class	Motif type Topological	Spatial
1	Glutathione reductase (Thieme *et al.*, 1981)		G_5	mg	mp^+
2	Subtilisin inhibitor, *Streptomyces* (Mitsui *et al.*, 1979)		G_6	gm^2	p^-m^2
3	γ-Chymotrypsin, N- and C-terminal domains (Cohen *et al.*, 1981)		G_6	g_2g	$t_l^-t_l^+ \approx h^-h^+ \approx d_r^-d_r^+$
4	Concanavalin A, N-terminal domain (Reeke *et al.*, 1975)		G_6	g_2g	$t_r^-t_r^+$
5	Concanavalin A, C-terminal domain (Reeke *et al.*, 1975)		G_5	g_3g	$h^+d_r^-$
6	Prealbumin (Blake *et al.*, 1978)		G_5	g_3g	$h^+d_r^-$
7	Plastocyanin (Guss & Freeman, 1983)		G_5	g_3g	$h^+d_r^-$
8	Superoxide dismutase, Cu, Zn (Tainer *et al.*, 1982)		G_5	g_3g	$h^+d_r^-$
9	CAP protein, N-terminal domain (McKay & Steitz, 1981)		G_5	g_3g	$h^+d_r^-$
10	Virus coat proteins, SBMV, STNV and TBSV (*S* domain) (Rossmann *et al.*, 1983)		G_5	g_3g	$h^+d_r^-$
11	Fab fragment of immunoglobulin G New, C_L and C_H1 domains (Saul *et al.*, 1978)		G_5	g_3g	$d_r^+h^-$
12	Actinoxantin (Pletnev *et al.*, 1981)		G_5	g_3g	$d_r^+h^-$

Table 13 (*cont.*)

No.	Protein, reference	Schematic diagram of chain tracing	Structural class	Motif type Topological	Spatial
13	Virus coat protein, TBSV, *P* domain (Harrison *et al.*, 1978)		G_5	$\mathbf{g_3g}$	$\mathbf{d_r^+h^-}$
14	γ-Crystallin, N- and C-terminal domains (Wistow *et al.*, 1983)		G_8	$\mathbf{g_0g}$	$\mathbf{d_r^-d_r^-}$

type. Other serine proteases have a similar structure. In all other cases the β structure appears to be described by a more or less twisted two-sheet motif. This is clearly demonstrated on peptide hydrogen bonds: they are not formed, as a rule, at the ends of the sheets.

The concanavalin A molecule (Reeke, Becker & Edelman, 1975) consists of two spatial motifs, $\mathbf{t_r^-t_r^+}$ and $\mathbf{h^+d_r^-}$ of the structural classes G_6 and G_5. In this protein we observe at one end of the bilayer a short extra hairpin between strands 9 and 10 which cannot be included in the bilayer structure. The next five examples of Table 13 display the same spatial motif $\mathbf{h_3^+d_r^-}$, topological type $\mathbf{g_3g}$, structural class G_5. This motif is met in prealbumin (Blake, Geison, Swan, Rerat & Rerat, 1974; Blake, Geison, Oatley, Rerat & Rerat, 1978), plastocyanin (Guss & Freeman, 1983), Cu, Zn superoxide dismutase (Tainer, Getzoff, Beem, Richardson & Richardson, 1982), N-terminal domain of CAP proteins (McKay & Steiz, 1981) and virus coat proteins (Rossmann, Abad-Zapatero & Murthy, 1983). Incrementing the basic spatial patterns $\mathbf{h_3^+d_r^-}$ by one or two additional β strands in different sheets

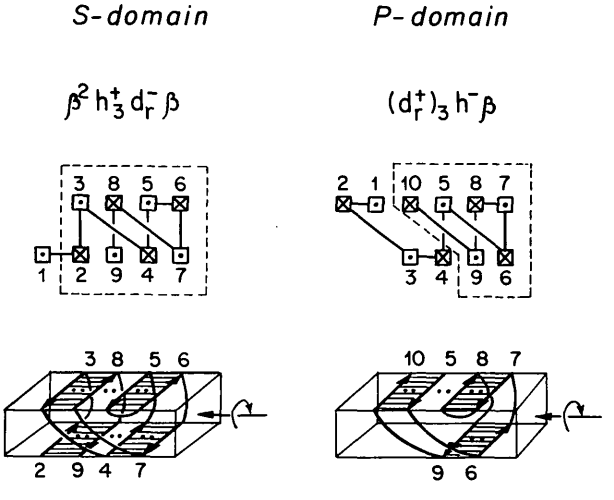


Fig. 9. Superhelical hairpin spatial patterns in *S* and *P* domains of the tomato bushy stunt virus coat protein, beginning with the spatial motifs $\mathbf{h_3^+d_r^-}$ and $(\mathbf{d_r^+})_3\mathbf{h^-}$.

gives rise to a complete spatial structure of the domain or the whole protein molecule.

The next three proteins listed in Table 13 (Nos. 11, 12, 13) give an example of the spatial motif $(\mathbf{d_r^+})_3\mathbf{h^-}$, topological type $\mathbf{g_3g}$, structural class G_5. This pattern is a basic part of the whole three-dimensional structure in the constant domains of the Fab fragment of immunoglobulin G New (Saul, Amzei & Poljak, 1978) and actinoxantin (Pletnev, Kuzin, Trakhanov, Khokhlov & Ovchinnikov, 1981). In the *P* domain of the tomato bushy stunt virus (TBSV) coat protein (Harrison, Olson, Schutt & Winkler, 1978), this spatial motif represents about half of the whole structure. Growth of the other half of the *P* domain leads to the formation of a specific two-sheet pattern which can also be described as a large superhelical right-handed hairpin beginning with the motif $\mathbf{h^-}$ and having one and a half turns. It is interesting to note that a very similar structural pattern is observed in the *S* domain of the TBSV coat protein. However, in this case it is a superhelical right-handed hairpin beginning with the motif $\mathbf{h^+}$ and having two complete turns (Fig. 9). Such a structure with one turn is also observed in the N-terminal domain of the CAP protein. It should be noted that for the first time this spatial pattern has been revealed in the structure of the satellite tobacco necrosis virus protein (Liljas, Unge, Jones, Fridborg, Lövgren, Skoglund & Strandberg, 1982).

The last example in Table 13 presents a highly pseudosymmetrical spatial motif $\mathbf{d_r^-d_r^-}$, topological type $\mathbf{g_0g}$, structural class G_8. The structure has pseudosymmetrical dyad axes (see Fig. 8). This spatial motif is observed very clearly in the N- and C-terminal domains of the calf eye lens homologous proteins, γ-crystallins (Wistow, Turnell, Summers, Sligsby, Moss, Miller, Lindley & Blundell, 1983; Chirgadze, Sergeev, Fomenkova & Oreshin, 1981; Chirgadze, Sergeev, Fomenkova, Oreshin & Nikonov, 1981). The perspective view, butt-end projection and topological diagram of main-chain folding are presented in Fig. 10.

We see that for the two-sheet antiparallel β structures the spatial motifs $\mathbf{h^+d_r^-}$ and $\mathbf{d_r^+h^-}$, topological

type $\mathbf{g}_3\mathbf{g}$, are realized in most cases. This result coincides with the result of Efimov (1982). Thus, both these motifs initiate right-handed superhelical hairpins. In all cases these motifs are located at the end of the two-sheet structure.

Of all the 14 spatial motifs of the **g** topology only the motifs $\mathbf{w}^+$, $\mathbf{w}^-$ and $\mathbf{d}_l^+$, $\mathbf{d}_l^-$ have not been observed. The reason for this is not clear; nevertheless, the right-handedness of the twisted β sheets in the proteins seems to be significant.

The spatial motifs of the whole β structure of the protein molecule have usually been described by two-dimensional topological diagrams. The deduced spatial motifs can easily be applied to such diagrams. In addition, we propose to designate the peptide hydrogen bonds between the β strands by two dots, if they are known. A general symbol formula is composed of different spatial motifs, and some examples are given in Fig. 11. As we can see, the formulae of the protein spatial motifs very often include an overlapped motif of the $\mathbf{g}_n\mathbf{g}$ type. This can be explained in the following way. The simple **g**-type spatial motifs, for example $\mathbf{h}^+$ or $\mathbf{d}_r^-$, can be considered as cooperative systems. In such a case, a complex motif of the $\mathbf{g}_n\mathbf{g}$ type with a high order of overlapping, for example $\mathbf{h}_3^+\mathbf{d}_r^-$ or $(\mathbf{d}_r)_3^+\mathbf{h}^-$, will be much more cooperative. This is one of the reasons why these motifs occur more frequently than others. As far as the motif of the $\mathbf{g}_0\mathbf{g}$ type is concerned, the formula of the observed spatial motif can also be represented as the equivalent expression

$$(\mathbf{d}_r^-)_0\mathbf{d}_r^- = \mathbf{d}_r^-\mathbf{h}_3^+\mathbf{d}_r^-.$$

Therefore, this motif also satisfies the same explanation.

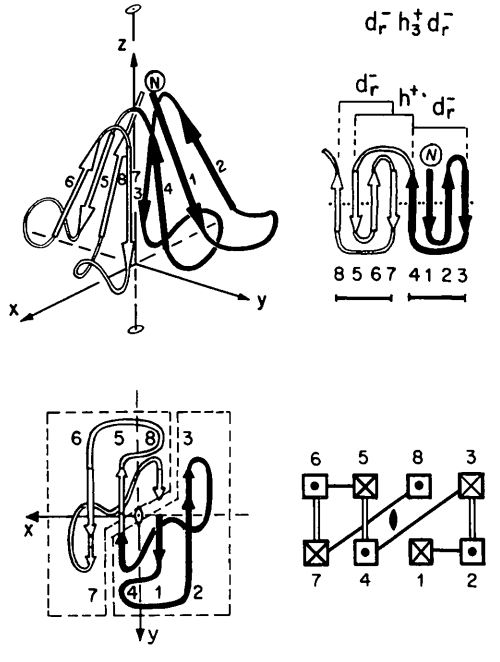


Fig. 10. Perspective view, butt-end projection and topological diagram of the N- and C-terminal domains of the calf eye lens protein γ-crystallin IIIb (Chirgadze *et al.*, 1981).

7. Discussion

The proposed approach allows one to deduce any possible spatial motif of the idealized antiparallel β structure. It consists of the combination of all the basic spatial motifs. The only limitation is conditioned by a few very reasonable structural and topological requirements. We can see that at least part of the deduced spatial motifs is realized in known protein structures. Therefore, a general regularity can be deduced for polypeptide-chain folds in globular proteins. That is why the skeletons of many different proteins are built of the same spatial motifs.

The spatial motifs obtained could give a sound basis for engineering newly synthesized or artificial protein molecules. One can also easily deduce any

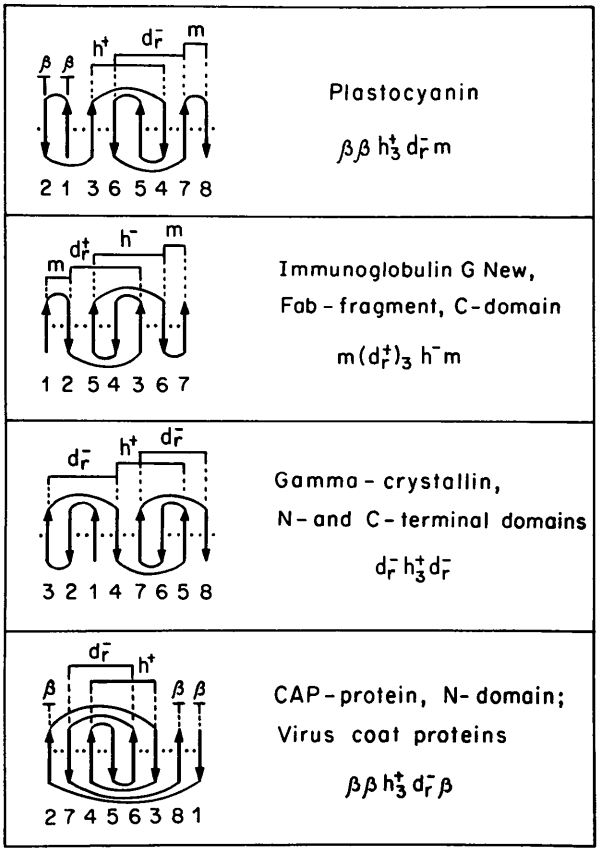


Fig. 11. Topological diagrams and general symbol formulae of the spatial structure for some β proteins. The basic motifs are listed in the formula beginning at the N-terminal end of the polypeptide chain. The subscript designates the order of overlapping of the β strands. The β motif always has zero order of overlapping. The order of overlapping is not designated if it is equal to unity.

topological type and then obtain all possible spatial motifs. They can have very different features. For example, they can be *complete* or *incomplete* depending on the number of β strands in each of the two sheets. The complete motifs are suitable for monomeric proteins and the incomplete ones for the subunit associates. The complete pseudosymmetrical motifs, for example the motifs of the even structural classes G_6 and G_8, are highly suitable for monomeric proteins or domains. In fact, such motifs occur in the domains of serine proteases, concanavalin and γ-crystallins. Incomplete motifs fit very well some protein dimers which have an expanded cooperative two-sheet β structure. Prealbumin dimers are a very good example of such a structure (Blake, Geison, Swan, Rerat & Rerat, 1974).

The variety of the geometrical features of spatial motifs can be used for creating some specific functions. For example, the spatial motifs with linkers on both sides of the sheets seem to be very *stable*. The motifs with only one-side linkers seem to have a potential *mobility* which can be revealed by the opening of the two sheets.

The spatial motifs can be *polar* or *non-polar* depending on whether the chain ends are placed at long or short distances. The spatial motifs $\mathbf{t}_r^-\mathbf{t}_r^+$ and $\mathbf{d}_r^-\mathbf{d}_r^+$, structural class G_6, are examples of such motifs (see Fig. 6). The polar motifs can be easily joined together into homo- or heterooligomeric associates with the united two-sheet β structure. These motifs can belong to different domains of a single polypeptide chain. Coding of such a protein system in one complex gene seems to proceed without difficulties. However, such giant protein structures have not yet been observed. On the contrary, non-polar motifs can be realized as separate domains without the united two-sheet β structure.

It is an interesting fact that the whole family of spatial motifs is divided into two enantiomorphous subgroups. At present, a choice of mirror type cannot be made in advance. However, one can imagine the possibility of creating artificial polypeptides from the D-amino acid. Then the spatial motifs of the corresponding mirror type would seem to be realized.

Finally, it may be hoped that the structures of some oligomeric protein associates will contribute to the understanding of the new principle of ultrastructural organization, at least for proteins with the antiparallel β structure.

The author expresses his gratitude to Professor Natalya Andreeva, Professor O. B. Ptitsyn, Dr A. V. Finkelstein, Dr A. V. Efimov and Dr S. V. Nikonov for many valuable discussions.

References

BLAKE, C. C. F., GEISON, M. J., OATLEY, S. J., RERAT, B. & RERAT, C. (1978). *J. Mol. Biol.* **121**, 339-356.

BLAKE, C. C. F., GEISON, M. K., SWAN, I. D. A., RERAT, C. & RERAT, B. (1974). *J. Mol. Biol.* **88**, 1-12.

CHIRGADZE, YU. N. (1985). *Dokl. Akad. Nauk SSSR*, **283**, 1500-1503. Engl. transl: *Dokl. Biophys. Proc. Acad. Sci. USSR*, pp. 183-186.

CHIRGADZE, YU. N., SERGEEV, YU. V., FOMENKOVA, N. P. & ORESHIN, V. D. (1981). *FEBS Lett.* **131**, 81-84.

CHIRGADZE, YU. N., SERGEEV, YU. V., FOMENKOVA, N. P., ORESHIN, V. D. & NIKONOV, S. V. (1981). *Dokl. Akad. Nauk SSSR*, **259**, 1502-1505. Engl. transl: *Dokl. Biochem. Proc. Acad. Sci. USSR*, pp. 316-319.

CHOTHIA, C. (1973). *J. Mol. Biol.* **75**, 295-302.

CHOTHIA, C. & JANIN, J. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 4146-4150.

CHOTHIA, C. & JANIN, J. (1982). *Biochemistry*, **21**, 3955-3965.

COHEN, G. H., SILVERTON, E. W. & DAVIES, D. R. (1981). *J. Mol. Biol.* **148**, 449-479.

EFIMOV, A. V. (1982). *Mol. Biol. (USSR)*, **16**, 799-806. Engl. transl: *Mol. Biol. (USSR)*, pp. 635-641.

FINKELSTEIN, A. V., PTITSYN, O. B. & BENDSKO, P. (1979). *Biofizika (USSR)*, **24**, 21-26. Engl. transl: *Biophysics (USSR)*, pp. 18-25.

GUSS, J. M. & FREEMAN, H. C. (1983). *J. Mol. Biol.* **169**, 521-563.

HARRISON, S. G., OLSON, A. J., SCHUTT, C. E. & WINKLER, F. K. (1978). *Nature (London)*, **276**, 368-373.

LEVITT, M. & CHOTHIA, C. (1976). *Nature (London)*, **261**, 552-558.

LILJAS, L., UNGE, T., JONES, T. A., FRIDBORG, K., LÖVGREN, S., SKOGLUND, U. & STRANDBERG, B. (1982). *J. Mol. Biol.* **159**, 93-108.

MCKAY, D. B. & STEITZ, T. A. (1981). *Nature (London)*, **290**, 744-749.

MCLACHLAN, A. D. (1980). *Protein Folding*, edited by R. JAENICKE, pp. 79-96. Amsterdam: Elsevier/North-Holland Biomedical Press.

MITSUI, Y., SATO, Y., WATANABE, Y. & ITAKA, Y. (1979). *J. Mol. Biol.* **131**, 697-724.

NOVOTNY, J., BRUCCOLERI, R. E. & NEWELL, J. (1984). *J. Mol. Biol.* **177**, 567-573.

PLETNEV, V. Z., KUZIN, A. P., TRAKHANOV, S. D., KHOKHLOV, A. S. & OVCHINNIKOV, YU. A. (1981). *Kristallografiya (USSR)*, **26**, 1046-1052. Engl. transl: *Sov. Phys. Crystallogr.* pp. 596-599.

PTITSYN, O. B. & FINKELSTEIN, A. V. (1980). *Q. Rev. Biophys.* **13**, 339-386.

PTITSYN, O. B., FINKELSTEIN, A. V. & FALK, P. (1979). *FEBS Lett.* **101**, 1-5.

REEKE, G. N., BECKER, J. W. & EDELMAN, G. M. (1975). *J. Biol. Chem.* **250**, 1525-1547.

RICHARDSON, J. S. (1977). *Nature (London)*, **268**, 495-500.

RICHARDSON, J. S. (1981). *Adv. Protein Chem.* **34**, 167-339.

ROSSMANN, M. G., ABAD-ZAPATERO, C. & MURTHY, M. R. M. (1983). *J. Mol. Biol.* **165**, 711-736.

SAUL, F. A., AMZEI, L. M. & POLJAK, R. J. (1978). *J. Biol. Chem.* **253**, 585-597.

SCHULZ, G. E., SCHIRMER, R. H., SACHENHEIMER, W. & PAI, E. F. (1978). *Nature (London)*, **273**, 120-124.

TAINER, J. A., GETZOFF, E. D., BEEM, K. M., RICHARDSON, J. S. & RICHARDSON, D. C. (1982). *J. Mol. Biol.* **160**, 181-217.

THIEME, R., PAI, E. F., SCHIRMER, R. H. & SCHULZ, G. E. (1981). *J. Mol. Biol.* **152**, 763-782.

WISTOW, G., TURNELL, B., SUMMERS, L., SLINGSBY, C., MOSS, D., MILLER, L., LINDLEY, P. & BLUNDELL, T. (1983). *J. Mol. Biol.* **170**, 175-202.