Amino acid residue: is it structural or functional?

Alexander P. Golovanov[a], Roman G. Efremov[a,b], Victor A. Jaravine[a], Gérard Vergoten[b], Alexander S. Arseniev[a,*]

[a]*Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Ul. Miklukho-Maklaya, 16/10 Moscow V-437, 117871 GSP-7, Russian Federation*
[b]*Université des Sciences et Technologies de Lille, Centre de Recherches et d'Etudes en Simulations et Modélisation Moléculaires, UFR de Chimie, Bâtiment C8, 59655 Villeneuve d'Ascq Cedex, France*

Abstract A new approach is suggested for delineating the structural and functional amino acid residues in proteins with known three-dimensional structure, basing on the involvement of residues in intramolecular hydrophobic and hydrophilic interactions and additional information about the conservativity of the residues. The approach is applied to the families of homologous neurotoxins and cardiotoxins. The results obtained concerning the role of amino acid residues in both families of toxins accord well with the similarity of their fold, but different mechanisms of action. Current approach can be used for detailed characterization of protein spatial structures, as well as for rational protein engineering.

Key words: Hydrophobic interactions; Protein structure; Neurotoxin; Cardiotoxin

1. Introduction

Once the spatial structure of a protein is known, two major questions arise: how it works, and what should be done to change its activity, stability and/or specificity in the desired direction. The usual way is to introduce mutation(s) to see the resulting changes. Obviously, such approach is time and resources consuming unless there is some initial guess about the role of particular amino acid residues in structure under consideration. In principle, an amino acid residue may be important for holding of the specific spatial structure, or for its specific activity (usually binding with other molecules). Thus some residues form a frame, on which the other residues are disposed in proper alignment for effective binding with the other molecule (or at least, for initial binding with subsequent conformational changes). Some residues, of course, can play structural and functional roles simultaneously. The mutation(s) of amino acid residue(s) changing activity of the molecule can also change its spatial structure (see, for example, [1]), which is not always desirable. That is why the method, which will delineate amino acid residues important for structure holding, and those important for specific activity of the molecule, can be very helpful both for rational protein engineering and for understanding the basic mechanisms of protein action.

It is now accepted that both protein folding and protein–protein recognition are mainly governed by hydrophobic, electrostatic interactions and hydrogen bonds [2–5]. Involvement of amino acid residues in electrostatic interactions and hydro-

gen bonds can be estimated using well-developed techniques, while quantitative evaluation of hydrophobic interactions between amino acid residues remains difficult due to the lack of adequate theory of the hydrophobic effect. 'Structural' role of glycines or prolines can be recognized from simple geometrical reasoning as these residues have, respectively, less or more steric restraints for and angles. To search for the 'structural' residues, recently the method of identifying the hydrophobic cores in the proteins with known spatial structure was described [6]. It is clear however that some residues important for structure holding may be situated outside the hydrophobic core. The methods for prediction of 'active sites' and antigenic determinants of proteins are mainly based on primary structure analysis (for example, [7,8]) and qualitative description of amino acid residues as hydrophobic or hydrophilic. Some methods use amino acid sequences available for protein families to reveal 'functional' amino acid residues (for example, [9]). The information about spatial structure oftenly is not taken into the consideration.

In the present paper we propose a method, which on the basis of precise spatial structure of a protein and usually available information about its conserved residues (within the family of homologous sequences), enables to delineate 'structural' residues important for maintenance of a spatial fold, 'functional' residues, which can be involved in binding with other molecules, but are not important for structure holding, and 'versatile' residues important for structure holding, but which can be also involved in binding with other molecules. The method is based on the analysis of the hydrophobic and hydrophilic contacts between amino acid residues in protein molecules with the help of molecular hydrophobicity potential (MHP). Previously, MHP has been successfully applied to characterize spatial polarity properties of globular [10,11] and membrane proteins [12–14].

As the check and illustration of the proposed method, spatial structures of proteins belonging to two homologous families of short postsynaptic neurotoxins and cardiotoxins (cytotoxins) are analyzed. It is well known, that neurotoxins and cardiotoxins share similar folds (three extended loops, formed by β-structures), but reveal different mechanisms of action. Neurotoxins bind to the nicotinic acetylcholine receptor with very high affinities and block postsynaptic neurotransmission. Cardiotoxins exert a variety of actions on different cells causing cytotoxicity, depolarization of membranes of excitable cells, muscle contraction and hemolysis, suggesting that these toxins act by perturbing of cell membranes (see review [15]). Still the mode of action of cardiotoxins is poorly understood at the molecular level. That is why it is interesting to delineate resi-

---

*Corresponding author. Fax: (7) (095) 335-7103.

dues in both types of toxins responsible for maintenance of the common fold, those, governing the specificity of each family, and to identify residues defining moderate differences in toxin action within the families.

2. Materials and methods

The atomic coordinates of neurotoxin II from *Naja naja oxiana* (NTII) were taken from [16,17]. Atomic coordinates of toxin from *Naja nigricollis* (1NEA) [18], α-neurotoxin from the *Dendroaspis polylepis polylepis* (1NTX) [19], erabutoxin b from *Laticauda semifasciata* in crystal (3EBX) [20] and in solution (1FRA) [21], cardiotoxin V4 from *Naja mossambica mossambica* (1CDT) [22], cardiotoxin CTXIIB from *Naja mossambica mossambica* (2CCX) [23], cardiotoxin γ from *Naja nigricollis* (1TGX) and cardiotoxin V from *Naja naja atra* (1CVO) [24] were taken from the Brookhaven Protein Data Bank (PDB) [25].

The values of MHP_{ij}, created by the side chain atoms of amino acid residue i (source residue) in geometrical center of the side chain of residue j (target residue, $i \neq j$) were calculated according to the formula:

$$MHP_{ij} = \sum_{k=1}^{N} f_k \cdot \exp(- r_{kj})$$

where f_k is the hydrophobicity constant of atom k belonging to the side chain of residue i, and r_{kj} is the distance (in Å) between the geometrical center of the side chain of residue j and atom k. N is the number of side chain atoms in residue i. In the current study we were mainly interested in interactions, specific for an amino acid residue, so we did not consider contribution from the backbone atoms to MHP. The atomic hydrophobicity constants f_k, evaluated basing on the octanol–water partition coefficients of numerous chemical compounds, were taken from [26]. In the case when several models of a toxin were presented in the PDB entry, MHP contact maps for all models were calculated, and then averaged map was further considered.

The contact between the two amino acid residues was considered as hydrophobic if both MHP_{ij} 0.0001 and MHP_{ji} ≥0.0001 (in other words, this contact corresponds to positive peaks symmetric across the diagonal on the MHP map). The contact was considered to be hydrophilic if both MHP_{ij} −0.0001 and MHP_{ji} −0.0001. The contact was named 'unfavourable' if MHP_{ij} and MHP_{ji} differed in sign, but their absolute values were greater than or equal to 0.0001.

The contribution of residue i in structure holding was calculated as:

$$C_{phob}^{i} = \sum_{j} MHP_{ij} \times MHP_{ji}, \quad \text{if } MHP_{ij}, MHP_{ji} \geq 0.0001$$

$$C_{phil}^{i} = \sum_{j} MHP_{ij} \times MHP_{ji}, \quad \text{if } MHP_{ij}, MHP_{ji} \leq - 0.0001$$

$$C_{unfav}^{i} = \sum_{j} MHP_{ij} \times MHP_{ji}, \quad \text{if } \begin{cases} (MHP_{ij} \times MHP_{ji}) < 0 \\ |MHP_{ij}| \geq 0.0001 \\ |MHP_{ji}| \geq 0.0001 \end{cases}$$

for hydrophobic, hydrophilic and unfavourable contributions, respectively.

The residue i was considered to be involved in structure holding (residue type 'S') via hydrophobic or hydrophilic interactions if C_{phob}^{i} 2.0 × 10⁻⁴ or C_{phil}^{i} 2.5 × 10⁻⁵, respectively (the cutoffs were chosen arbitrary to keep balance between the number of 'structural' and 'functional' residues). The residue was considered to be 'functional' (residue type 'F') if it was conserved in a family of proteins with a similar function, but it was not of type 'S'. The residue was considered to be 'versatile' (residue type 'V') if it belonged to a partially solvent-exposed hydrophobic cluster (i.e. its C_{phob} 2.0 × 10⁻⁴, the number of hydrophobic contacts for the residue was 60% of the maximal number of hydrophobic contacts per residue for this protein, and accessible surface area of such residue was greater than 20 Å²). Sequence alignment was adopted from [15] with minor changes.

All the calculations of MHP, accessible surface areas (according to [27]), as well as automatic assignment of residues to one of three types (S, F or V) were done using the home-build program EXPO.

3. Results and discussion

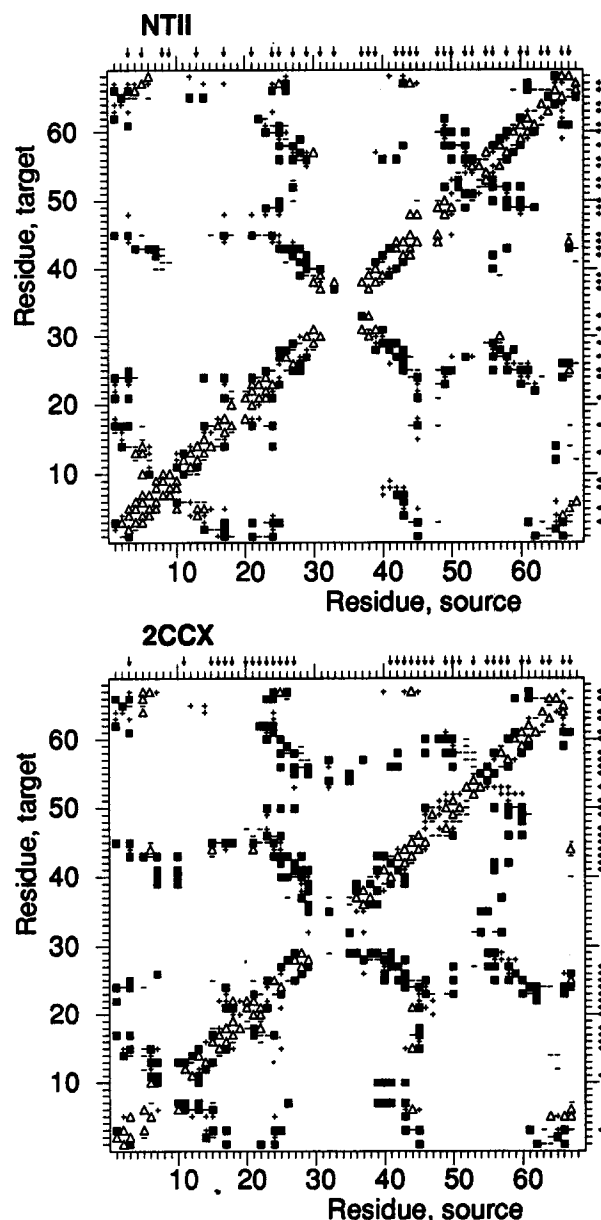Preliminary analysis of 9 spatial structures of neurotoxins

Fig. 1. MHP contact maps for NTII and 2CCX. Hydrophobic and hydrophilic contacts between side chains of amino acid residues are shown by black squares and triangles, respectively. Unfavourable contacts are labelled with '+' and '−' for positive and negative MHP, respectively. Conservative residues are marked with arrows.

and cardiotoxins revealed that small number of conserved ion pairs and hydrogen bonds, involving side chains of amino acid residues, cannot explain similarity of their folding. To identify the other sources of folding similarities, MHP contact maps were calculated (see section 2) for 5 structures of neurotoxins and 4 structures of cardiotoxins. The representative MHP contact maps for neurotoxin NTII and cardiotoxin CTXIIB are shown in Fig. 1. The maps delineate hydrophobic and hydrophilic as well as unfavourable contacts (see section 2) in the molecules. It can be seen that each side chain can form simultaneously both hydrophobic and hydrophilic contacts with the other amino acid residues. This illustrates that MHP approach provides much more detailed picture of hydrophobic and hy-
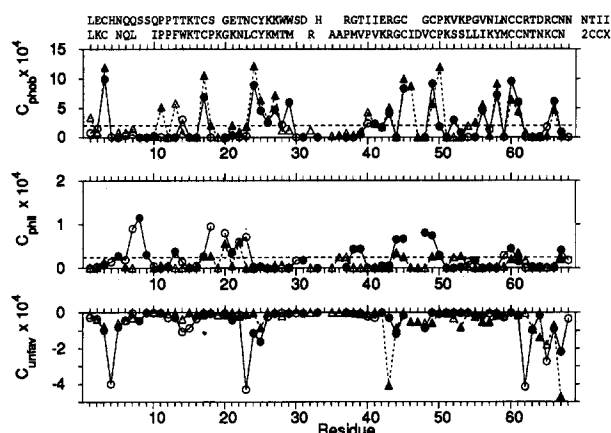
Fig. 2. Contributions of amino acid residues of NTII and 2CCX to the overall hydrophobic (C_{phob}), hydrophilic (C_{phil}) and unfavourable (C_{unfav}) intramolecular interactions. Circles and triangles correspond to NTII and 2CCX residues, respectively. Filled symbols correspond to conserved residues. Amino acid residues above the cutoff values (horizontal broken lines in C_{phob} and C_{phil} plots) are considered as 'structural'. Amino acid sequences of NTII and 2CCX are shown above the plots.

drophilic interactions then widely used various hydrophobicity scales of amino acid residues (see [28] for review). For both toxin families the hydrophilic contacts usually appear between sequential residues and close to the ends of antiparallel β-

structure (near the turns). For sequentially distant residues hydrophilic contacts are not so frequent as the hydrophobic ones (see Fig. 1). The hydrophobic contacts between side chains are frequent both for sequential residues and the residues forming β-strands, and also for hydrophobic core of the toxins, from which three loops are extended (see Fig. 4). MHP maps are very similar for all toxins (Fig. 1 shows maps only for NTII and 2CCX).

The contributions (C_{phil}, C_{phob} and C_{unfav}) of amino acid residues to the structure holding via hydrophobic and hydrophilic interactions are shown in Fig. 2. It could be seen that side chains contribute to hydrophobic contacts approximately ten times more than to the hydrophilic ones. Most of the residues revealing the largest contributions to the hydrophobic interactions are conserved in toxin sequences (filled circles and triangles on Fig. 2). In addition, there is a good correlation between contributions of conservative residues to structure holding within each family of neurotoxins and cardiotoxins (data not shown). The values of C_{unfav} (Fig. 2) reflect participation of the residues in unfavourable hydrophobic–hydrophilic contacts. Usually, the strongest unfavourable contacts in short neurotoxins and cardiotoxins are caused by the variable amino acid residue(s). For example, 1NEA has less unfavourable contacts then NTII. Comparison of their contact maps reveals that such contacts in NTII are caused by the presence of arginins instead of Thr-62 and Lys-65 in 1NEA sequence (numbering as in Fig. 3). Also, His-6 in 1NTX, 3EBX and 1FRA, and His-26 in 3EBX
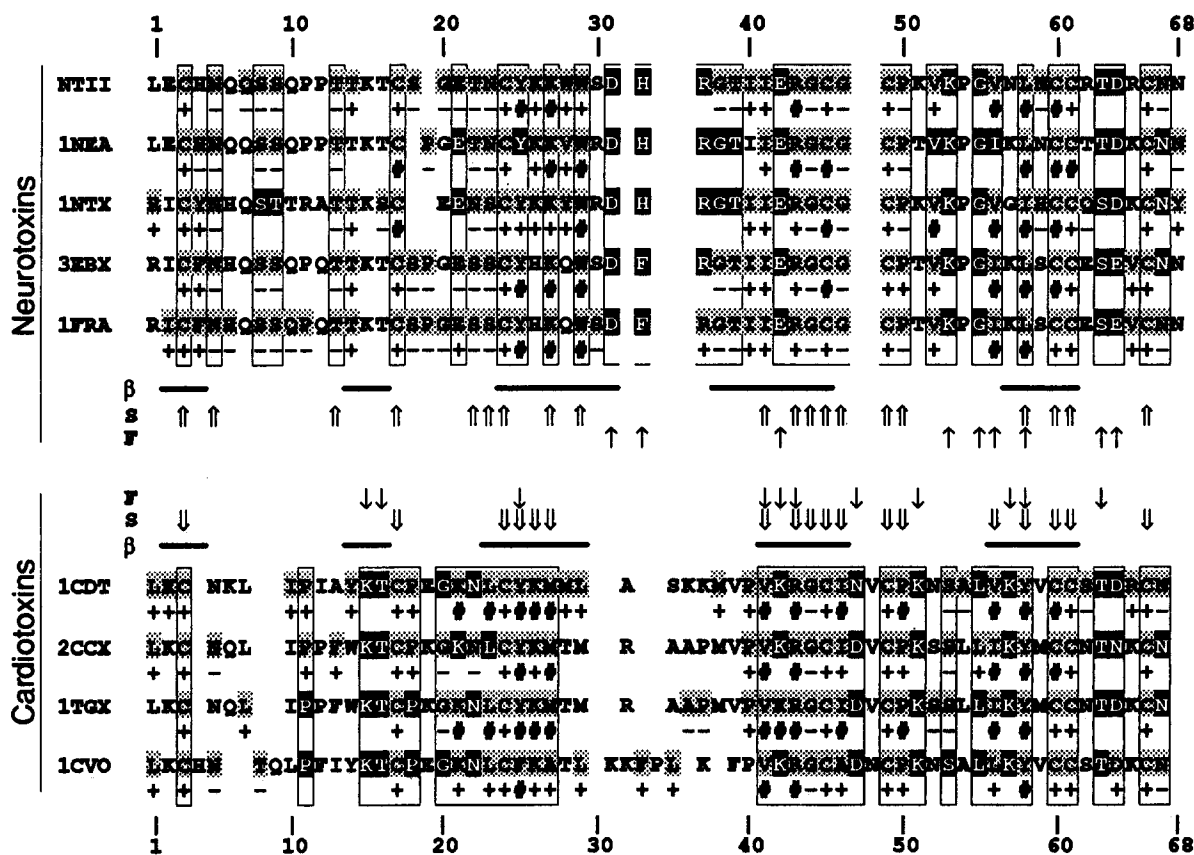

Fig. 3. Sequence alignment of neurotoxins and cardiotoxins. The conserved residues within each family are closed into boxes. 'Structural' residues are gray-shaded, the signs below reflect involvement of the residue in hydrophobic (+) or hydrophilic (−) interactions. 'Versatile' residues are marked (#). 'Functional' residues are shown with inverted colours. Black bars indicate β-structural regions for both families. Arrows in the middle show the 'functional' residues, common for each family, while the 'structural' residues are labelled with double arrows.
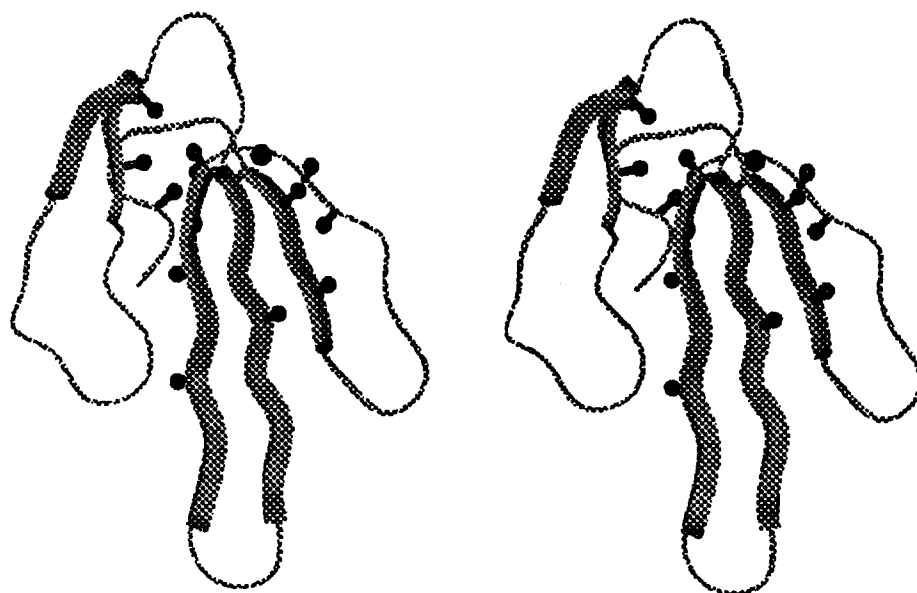
Fig. 4. The MHP-based 'structural' framework of toxins: disposition of amino acid residues determining the common spatial fold for cardiotoxins and neurotoxins. β-structural regions are shown with ribbons.

and 1FRA participate in unfavourable hydrophobic–hydrophilic contacts and can destabilize spatial structures. Thus simple analysis leads to identification of residues which probably influence the stability of the structure. It should be also mentioned that some weak unfavourable contacts might be necessary for creating the energy barriers essential for specificity of protein folding.

Fig. 3 shows the alignment of neurotoxin and cardiotoxin sequences, the conserved residues and classification of residues according to the role they can play in particular protein. Residue is considered to be 'structural' (type 'S') if its contribution to intramolecular hydrophobic or hydrophilic interactions is high (see section 2). Residue is considered to be 'functional' (type 'F') if it is conserved, but not of type 'S'. Partially solvent-exposed hydrophobic residues belonging to the hydrophobic clusters are important for structure holding, but as it is energetically unfavourable for such residues to be exposed to the polar environment, they will tend to be shielded from the solvent. Thus such kind of residues may possess the potency for binding with hydrophobic surfaces and that is why they are considered as a special type of 'versatile' residues (type 'V'). Quasi-conserved residues (conserved in a subfamily of homologous molecules, for example Lys-15 for neurotoxins) which are not of type 'S', can also be 'functional' (type 'F') in the subfamilies of proteins, and this can explain small differences in toxin action within the family.

As one can see in Fig. 3, the most of 'S' residues are conserved for neurotoxin and cardiotoxin families and also for both families. Residues of type 'F' are almost conservative in each family, but they are quite different in neurotoxins and cardiotoxins (Fig. 3). This fact accords well with the similarity of the folding motif for two families of toxins and strong differences in their functioning. There is also a good correlation with experimentally observed participation of residues 25, 27, 29, 31, 33, 37, 38, 42 and 53 of neurotoxins in binding with acetylcholine receptor ([1,29], see also references in [15]), and involvement of residues 15, 21, 25 and 43 in cardiotoxin's func-

tioning ([30] and references therein). Recent NMR experiments also suggest the interaction of the side chains of hydrophobic residues of toxin γ in the segments 5–14, 25–35, 39–46 and 54–58 with the detergent micelles [31]. This also accords well with the predicted in the current study 'versatile' and 'functional' residues for cardiotoxins. It should be noted that there is no available experimental data about the functional importance of some amino acid residues, predicted as 'functional' in the current study (see Fig. 3). From Fig. 3 one can also see some discrepancies in the roles of residues (for example, for erabutoxin b in crystal, 3EBX, and in solution, 1FRA). This can be explained by the high sensitivity of MHP to spatial structure and sharp criteria used to define the residue types (see section 2). The 'structural' residues, common for neurotoxins and cardiotoxins, are indicated in Fig. 4. We propose that these residues are responsible for the general folding motif, typical for the both families of toxins.

Current approach resembles in some details the other techniques, based on the analysis of distance maps [32], with one main advantage. It can discriminate between hydrophobic and hydrophilic intramolecular contacts. As it was shown recently [33], such structural features as side-chain to side-chain contacts and solvent accessibility are not necessary conserved among the proteins with similar spatial structures, and more general features of a protein structure, such as requirements for burial of hydrophobic residues and solvent exposure for the polar ones, should be taken into account. The proposed approach gives simple formalism for describing such 'general features' as hydrophobic and hydrophilic intramolecular contacts. At least in the case of neurotoxins and cardiotoxins the 'general features' responsible for the given folding motif are conserved. Further analysis is needed to check the conservation of such contact patterns inside other families of proteins with similar folding.

In summary, the new approach was proposed for efficient identification of 'structural' and 'functional' amino acid residues in proteins with known spatial structure. The approach

can be used for structure analysis and rational protein engineering.

References

[1] Pillet, L., Trémeau, O., Ducancel, F., Drevet, P., Zinn-Justin, S., Pinkasfeld, S., Boulain, J.-C. and Ménez, A. (1993) J. Biol. Chem. 268, 909–916.

[2] Ponnuswamy, P.K. (1993) Prog. Biophys. Mol. Biol. 59, 57–103.

[3] Matthews, B.W. (1993) Curr. Opin. Struct. Biol. 3, 589–593.

[4] Waldburger, C.D., Schildbach, J.F. and Sauer, R.T. (1995) Nature Struct. Biol. 2, 122–128.

[5] Matthews, B.W. (1995) Nature Struct. Biol. 2, 85–86.

[6] Swindells, M.B. (1995) Protein Sci. 4, 93–102.

[7] Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Mornon, J.P. (1990) Biochimie 72, 555–574.

[8] Kolaskar, A.S. and Tongaonkar, P.C. (1990) FEBS Lett. 276, 172–174.

[9] Casari, G., Sander, C. and Valencia, A. (1995) Struct. Biol. 2, 171–178.

[10] Efremov, R.G. and Alix, A.J.P. (1993) J. Biomol. Struct. Dyn. 11, 483–507.

[11] Efremov R.G., Golovanov, A.P., Vergoten, G., Alix, A.J.P., Tsetlin V.I. and Arseniev, A.S. (1995) J. Biomol. Struct. Dyn. 12, 971–991.

[12] Brasseur, R. (1991) J. Biol. Chem. 266, 16120–16127.

[13] Efremov, R.G., Gulyaev, D.I., Vergoten, G. and Modyanov, N.N. (1992) J. Protein Chem. 11, 665–675.

[14] Sansom, M.S.P. and Kerr, I.D. (1993) Protein Eng. 6, 65–74.

[15] Rees, B. and Bilwes, A. (1993) Chem. Res. Toxicol. 6, 385–406.

[16] Golovanov, A.P., Lomize, A.L., Arseniev, A.S., Utkin, Y.N. and Tsetlin, V.I. (1993) Eur. J. Biochem. 213, 1213–1223.

[17] Golovanov, A.P. and Arseniev, A.S. (1995) Bioorgan. Khim. (Russian) 21, 571–579.

[18] Zinn-Justin, S., Roumestand, C., Gilquin, B., Bontems, F., Ménez, A. and Toma, F. (1992) Biochemistry 31, 11335–11347.

[19] Brown, L.R. and Wüthrich, K. (1992) J. Mol. Biol. 227, 1118–1135.

[20] Smith, J.L., Corfield, P.W., Hendrikson, W.A. and Low, B.W. (1988) Acta Crystallogr. sect. A. 44, 357–368.

[21] Hatanaka, H., Oka, M., Kohda, D., Tate, S., Suda, A., Tamiya, N. and Inagaki, F. (1994) J. Mol. Biol. 240, 155–166.

[22] Rees, B., Bilwes, A., Samama, J.P. and Moras, D. (1990) J. Mol. Biol. 214, 281–297.

[23] O'Connell, J., Bougis, P. and Wüthrich, K. (1993) Eur. J. Biochem. 213, 891–900.

[24] Singhal, A.K., Chien, K.-Y., Wu, W. and Rule, G. (1993) Biochemistry 32, 8036–8044.

[25] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.

[26] Viswanadhan, V.N., Ghose, A.K., Revankar, G.R. and Robins, R.K. (1989) J. Chem. Inf. Comput. Sci., 29, 163–172.

[27] Lee, B. and Richards, F.M. (1971) J. Mol. Biol., 55, 379–400.

[28] van de Waterbeemd, H., Karajiannis, H. and El Tayar, N. (1994) Amino Acids 7, 129–145.

[29] Tsetlin, V.I., Karlsson, E., Utkin, Yu.N., Pluzhnikov, K.A., Arseniev, A.S., Surin, A.M., Kondakov, V.V., Bystrov, V.F., Ivanov, V.T. and Ovchinnikov, Yu.A. (1982) Toxicon 20, 83–93.

[30] Gilquin, B., Roumestand, C., Zinn-Justin, S., Ménez, A. and Toma, F. (1993) Biopolymers 33, 1659–1675.

[31] Dauplais, M., Neumann, J.-M., Pinkasfeld, S., Ménez, A. and Roumestand, C. (1995) Eur. J. Biochem. 230, 213–220.

[32] Kikuchi, T. (1992) J. Protein Chem. 11, 305–320.

[33] Russell, R.B. and Barton, G.J. (1994) J. Mol. Biol. 244, 332–350.