A novel super-secondary structure of proteins and the relation between the structure and the amino acid sequence

A.V. Efimov

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

Received 27 October 1983

A novel protein super-secondary structure which is referred to as an $\alpha\alpha$-corner is considered. The $\alpha\alpha$-corner is formed by two consecutive α-helices packed approximately crosswise and connected by two or more peptide units. It is shown that the amino acid sequences coding for the $\alpha\alpha$-corners have a strictly definite order of hydrophobic, hydrophilic and glycine residues. A hypothesis is suggested that the $\alpha\alpha$-corner can be an embryo of protein folding.

Super-secondary structure *α-Helix* *Polypeptide chain conformation* *Amino acid sequence* *Packing of α-helix* *Embryo of folding*

1. INTRODUCTION

One of the most basic features of globular proteins is that some of their structural regions are folded into repetitive units, the so-called super-secondary structures. Authors in [1] were the first to report the super-secondary structure for α/β-proteins. Later the 4-α-helical super-secondary structure of α-proteins [2–4] and the abcd-unit which is the super-secondary structure of β-proteins [5] were revealed and studied.

Here we consider a novel super-secondary structure of proteins which represents a structure formed by two consecutive α-helices connected by a polypeptide chain and packed approximately crosswise. Such locally ordered regions of proteins are denoted as $\alpha\alpha$-corners in contrast to the α-hairpins where α-helices are packed approximately antiparallel.

2. CONFORMATION OF A POLYPEPTIDE CHAIN FOLDED INTO AN $\alpha\alpha$-CORNER

A schematic representation of an $\alpha\alpha$-corner with a short connection consisting of two peptide units is shown in fig.1 (according to the IUPAC-IUB Commission on Biochemical Nomenclature [6] a peptide unit is the -CHR-CO-NH- group of atoms; in the figure such units are shown by virtual bonds joining consecutive C_α-atoms). There are no shorter connections in $\alpha\alpha$-corners because of the

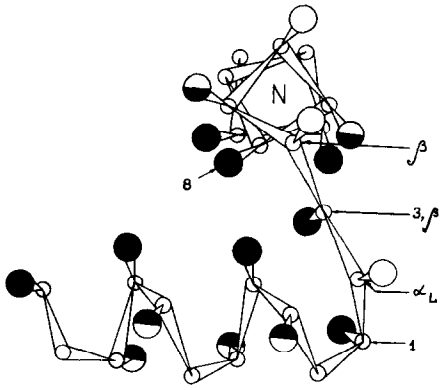

Fig.1. Schematic representation of a right-handed $\alpha\alpha$-corner with a short connection consisting of two peptide units. Hydrophobic side chains are designated by solid circles; hydrophilic side chains and Gly are shown by open circles or not shown at all; side chains which are partly immersed in the hydrophobic core are shown by semi-solid circles, such side chains can be either hydrophobic or hydrophilic; α_L and β are conformations of the corresponding residues. See also the text.

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

overlapping of van-der-Waals radii of the α-helices.

The first α-helix of the $\alpha\alpha$-corner starting from the N-terminal will be referred to here as the A-helix and the second one as the B-helix. The short connection joining two α-helices of the $\alpha\alpha$-corner must have an extended conformation and be oriented approximately perpendicular to the axes of the A- and B-helices. This is possible if the C-terminal residue of the A-helix has an α_L-conformation (or one close to that) and if the N-terminal residue of the B-helix has a β-conformation. Thus, a polypeptide chain of the $\alpha\alpha$-corner with a short connection consisting of two peptide units has an $\ldots\alpha_R\alpha_R\alpha_L\beta\beta\alpha_R\alpha_R\ldots$-conformation, where α_R, α_L and β are residues with ϕ and ψ angles, corresponding to the right-handed α-helix (α_R), left-handed α-helix (α_L) and β-structure (β) regions on the Ramachandran plot. It is clear that this conformation must be considered as the average one. $\alpha\alpha$-Corners with such a conformation of the polypeptide chain will be described here as right-handed $\alpha\alpha$-corners (fig.1). To form a left-handed $\alpha\alpha$-corner from a right-handed one, it is necessary to turn one of the two α-helices by 180° (e.g., the B-helix). It is clear that such a rearrangement leads to a conformation of the connection where strong steric hindrances take place. This may be the reason for the absence of left-handed $\alpha\alpha$-corners in proteins. Details of a stereochemical analysis including statistical data will be described elsewhere [7].

There are many $\alpha\alpha$-corners with long connections consisting of three and more peptide units in proteins. Such connections can have conformations of a β-strand, an α-helix and a β-turn or can be formed by a combination of such regions. However irrespective of the length and conformation of the connections, $\alpha\alpha$-corners are practically always right-handed ones in proteins.

3. FEATURES OBSERVED IN AMINO ACID SEQUENCES CODING FOR $\alpha\alpha$-CORNERS

Let us consider the relationship between the amino acid sequence and the structure of an $\alpha\alpha$-corner with a short connection consisting of two peptide units. To avoid strong steric hindrances, the C-terminal residue of the A-helix with an α_L-conformation must be Gly, Ala or a residue with a long and flexible side chain (e.g., Lys, Arg, ...) and, vice versa, this residue must not be Val, Ile, Leu, Phe, Tyr, Trp and Thr. This conclusion is confirmed by experimental data: of the 16 patterns shown in fig.2 [8–21] 11 C-teminal residues of the A-helices are glycines, 3 are lysines, one is arginine, one is histidine and none are massive hydrophobic residues. It is noteworthy that this is characteristic of every α-helix having an α_L-residue terminating it [7,22].

A peculiarity of the first residue of the B-helix is that it has a β-conformation and that its side chain is directed towards the B-helix axis. That is why massive hydrophobic residues cannot occupy the first positions of the B-helices: their side chains cause dehydration of the free NH-groups of the polypeptide backbone which is prohibited. Neither of the 33 patterns shown in fig.2,3 contain Val, Ile, Leu, Phe and Trp residues in the first positions of the B-helices while Tyr occurs twice. More often the first positions are occupied by Ser, Thr, Asn or Asp residues which usually form hydrogen bonds with the free NH-groups of the backbone. This feature is also observed for every α-helix with a β-residue in the first position [7]. The fact that short hydrophilic residues prefer to occupy the first three positions of α-helices was also shown in [33,34].

There are definite features in the order of hydrophobic and hydrophilic residues of amino acid sequences coding for the $\alpha\alpha$-corner structure. To avoid an immersion of polar groups into the hydrophobic environment there should be at least one hydrophobic side chain in each turn of an α-helix. A hydrophobic cluster formed by such side chains on the α-helix surface is denoted here as a 'necessary cluster' (see also [4]). In fig.1 the necessary clusters of the A- and B-helices are shown by solid circles. The main feature is that the last hydrophobic residue of the A-helix necessary cluster and the first hydrophobic residue of the B-helix necessary cluster form a pair in positions 1–8. In other words, there is a 1–8-gap between the hydrophobic necessary clusters of the A- and B-helices. The residue forming a 1–3 pair with the last hydrophobic residue of the A-helix necessary cluster is completely buried in the hydrophobic core and must be hydrophobic. Thus, an amino acid sequence coding for the $\alpha\alpha$-corner with a connection consisting of two peptide units must have:

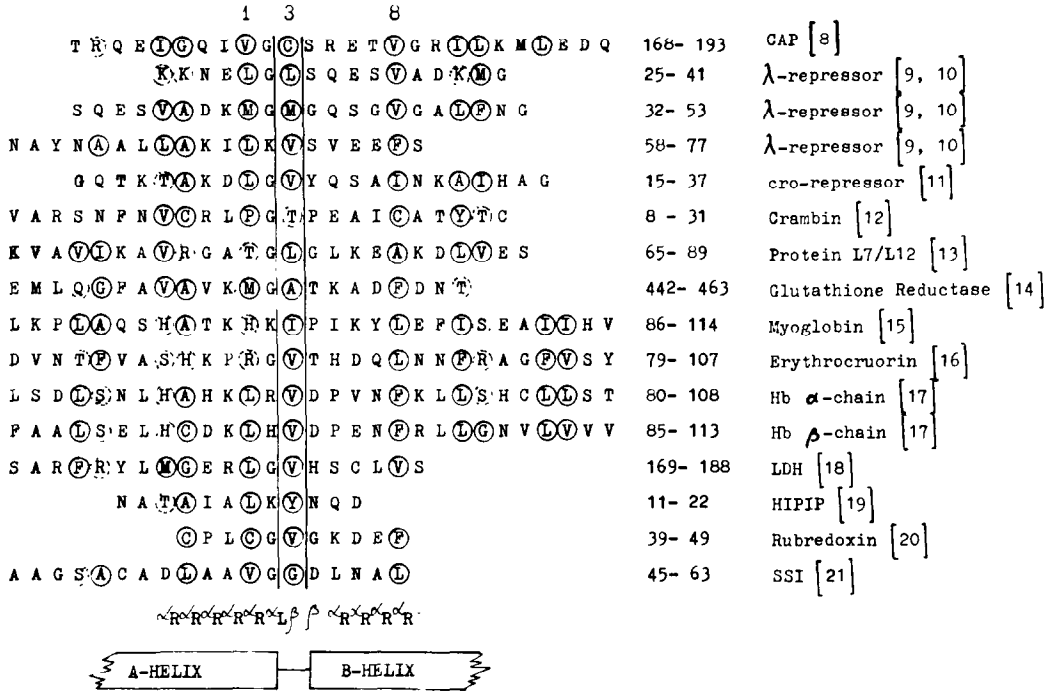

Fig.2. Alignment of the amino acid sequences coding for $\alpha\alpha$-corners with short connections. Regions of the sequences and the proteins containing them are listed in the right part of the figure; α_R, α_L and β are conformations of the residues in the corresponding columns; hydrophobically invariant residues are encircled; vertical lines show the C-terminus of the A-helix and the N-terminus of the B-helix.

(i) hydrophobic residues in the corresponding positions to form necessary clusters of the A- and B-helices;
(ii) a 1–3–8-gap between the necessary clusters;
(iii) residues Gly or Arg and Lys in the last position of the A-helix;
(iv) hydrophilic or small residues (Gly, Ala, Pro) in the first position of the B-helix.

The above is confirmed by the data presented in fig.2 which shows the alignment of 16 amino acid sequences coding for $\alpha\alpha$-corners with short connections. Each column contains structurally similar residues. Columns headed by 1, 3 and 8 represent hydrophobic residues forming the 1–3–8-gaps between the necessary clusters. As seen, these residues are hydrophobically invariant irrespective of whether the proteins are homologous or not. Almost all the positions forming the necessary clusters of the A- and B-helices are hydrophobically invariant as well. Sometimes residues Thr and Ser can be like hydrophobic ones because their side chains form hydrogen bonds with the main chain and no dehydration occurs. Residues Lys, Arg, Gln, Glu can be on the borders of hydrophobic clusters as they have long 'hydrophobic legs'. In our opinion, the features described can be successfully applied to predict the location of the $\alpha\alpha$-corners in protein structures from their amino acid sequences.

Fig.3 shows the alignment of the 17 amino acid sequences coding for $\alpha\alpha$-corners with long connections consisting of three and more peptide units. The features observed are similar to those of the sequences coding for $\alpha\alpha$-corners with short connections. There are hydrophobically invariant residues in the positions forming the necessary clusters of the A- and B-helices whether or not the amino acid sequences are homologous. There is practically always a hydrophobic residue in the position forming a 1–6 pair with the first hydrophobic residue of the B-helix necessary cluster (this 1–6 pair is like the 3–8 pair of the 1–3–8-gaps in $\alpha\alpha$-corners with short connections; see fig.2). It should be noted that at least one

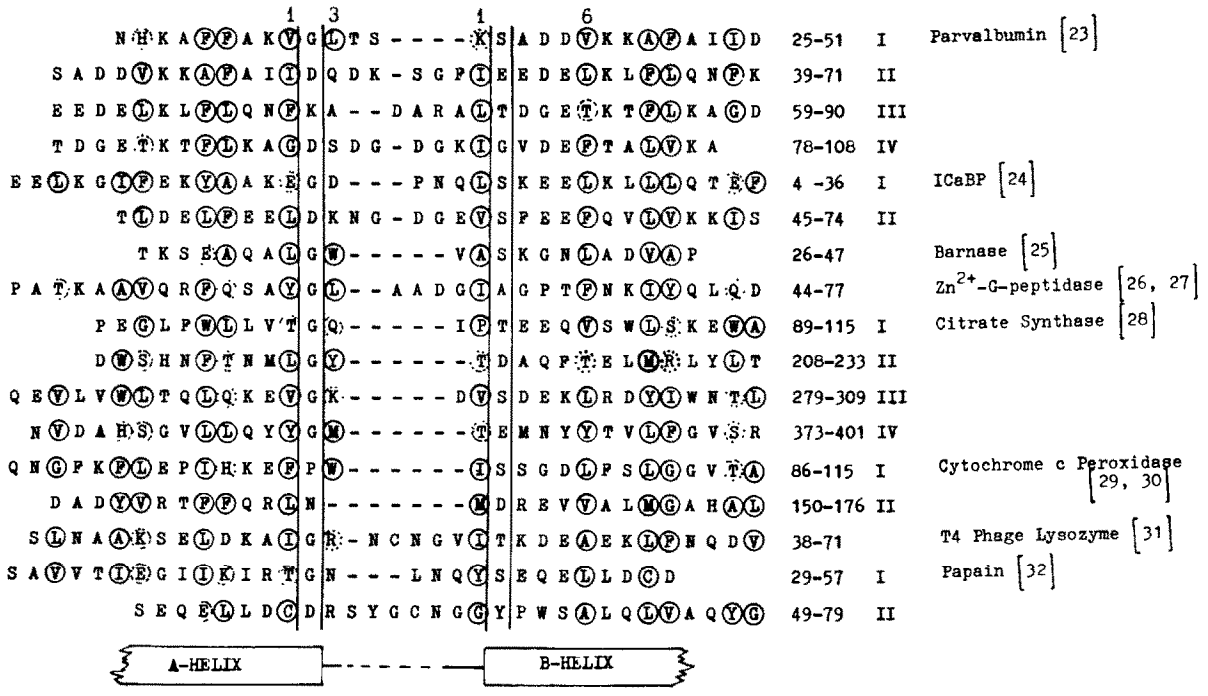

Fig.3. Alignment of the amino acid sequences coding for $\alpha\alpha$-corners with long connections. Designations are the same as in fig.2.

residue, either Gly or Arg and Lys is found at the end of the A-helix or in the connection region.

4. PACKING OF THE REMAINING α-HELICES OF A PROTEIN MOLECULE RELATIVE TO THE $\alpha\alpha$-CORNER

α-Helices joined to the A-helix in a protein molecule will be designated here as A_1, A_2, ... in the order of their receding from the A-helix in the chain. α-Helices joined to the B-helix will be correspondingly designated as B_1, B_2, ... The packing of the A_1, A_2, ... and B_1, B_2, ... helices relative to the $\alpha\alpha$-corner in proteins has the following features. Three consecutive α-helices packed to form two $\alpha\alpha$-corners are folded into a left-handed superhelix (see structures labeled I, V, VII, X, XII, XIII in fig.4). If only two of the three consecutive α-helices form an $\alpha\alpha$-corner, the third α-helix is packed in such a way that the B_1- and B-helices form a BB_1-hairpin (see structure II in the figure), and the A_1- and A-helices form an AA_1-hairpin (structure III). Possible pathways of growing the structures with a step-wise addition of other α-helices are shown in fig.4. The packing of a subsequent α-helix depends on the packing of α-helices in the structures shown in the left part of the scheme. In other words, there exists an intramolecular relationship between the arrangement of secondary structure elements in different parts of a protein molecule (see also [4,5]).

The peculiarity is that there is a limited number of pathways to obtain all the protein structures containing $\alpha\alpha$-corners. Thus, the scheme can be considered as a key for searching similarities between these protein structures. Common structural features rather than the homology of the amino acid sequences is the basis of such similarities.

The analysis of the features described here has allowed us to propose a hypothesis that the $\alpha\alpha$-corner can be an embryo of protein folding (an analogous hypothesis has been suggested for β-proteins in which an abcd-unit can be an embryo of folding [5]).

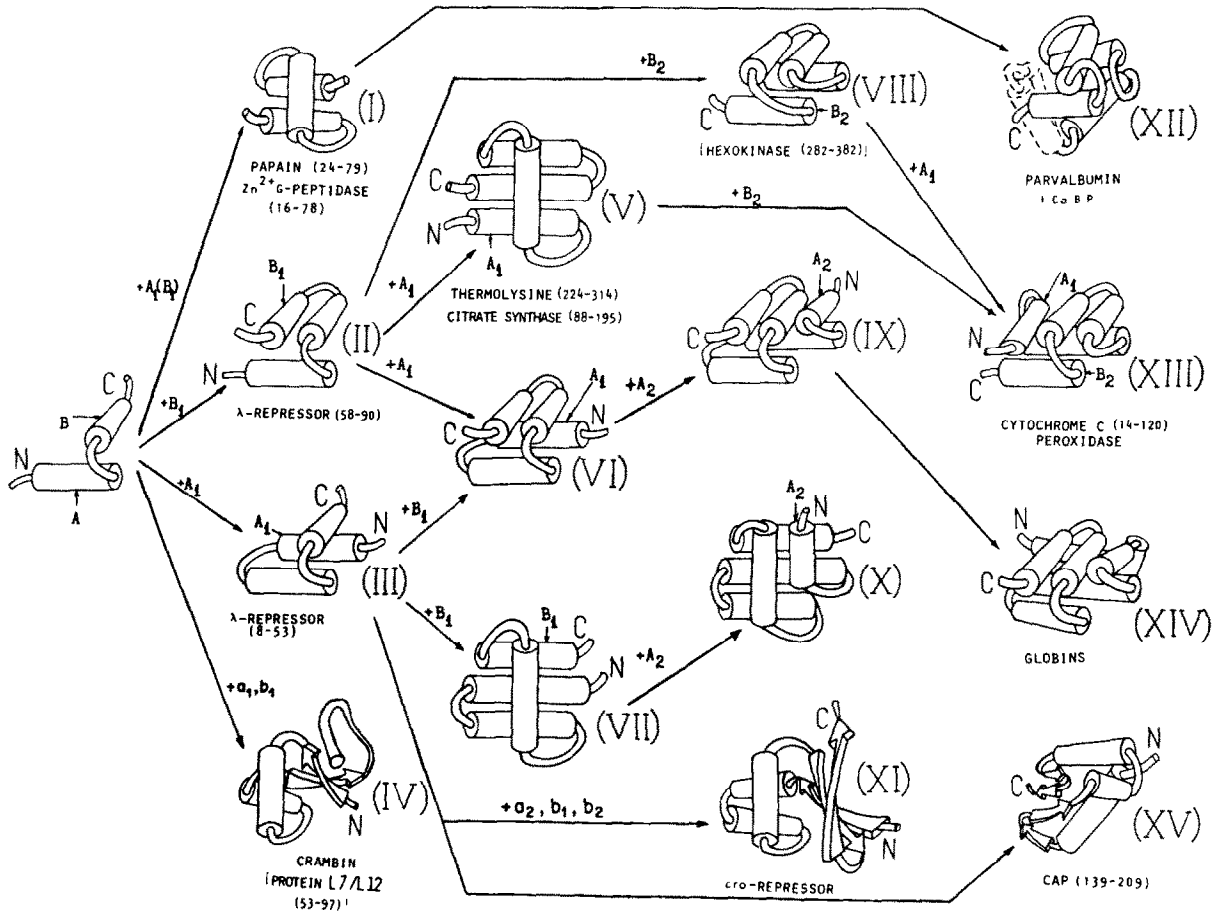

Fig.4. A scheme of growing the structures using a stepwise addition of α-helices when an $\alpha\alpha$-corner is taken as the initiating complex. α-Helices are shown by cylinders; connections are simplified. Proteins or their regions which have corresponding structures are denoted under them. See also the text.

5. DISCUSSION

A comparative analysis of the known super-secondary structures reveals some common features in their arrangement. The super-secondary structures are formed by secondary structure segments adjacent along the polypeptide chain. As a rule, super-secondary structures contain two layers of α-helices and (or) β-strands. A super-secondary structure has a unique handedness; e.g., Rossmann's fold is right-handed and only right-handed $\alpha\alpha$-corners occur in the proteins. In each super-secondary structure the spatial arrangement of α-helices and (or) β-strands is practically the same for all the proteins in which it is found, no matter whether they are homologous or not. The analysis shows that there is at least one super-secondary structure in a domain of most known proteins. As has been shown in [5] and also in fig.4 of this paper, a super-secondary structure can determine the packing of the remaining secondary structure elements in a protein molecule or a domain. Apparently, there is a definite relation between the amino acid sequence of a region folded into a super-secondary structure and its structure independent of the remaining sequence of a molecule (that is shown for the $\alpha\alpha$-corners and for the 4-α-helical super-secondary structure [4]). All this has led us to a hypothesis that a super-secondary structure can fold independently of the remaining part of a protein molecule and can be an embryo of folding.

REFERENCES

[1] Rao, S.T. and Rossmann, M.G. (1973) J. Mol. Biol. 76, 241–256.

[2] Argos, P., Rossmann, M.G. and Johnson, J.E. (1977) Biochem. Biophys. Res. Commun. 75, 83–86.

[3] Weber, P.C. and Salemme, F.R. (1980) Nature 287, 82–84.

[4] Efimov, A.V. (1982) Mol. Biol. (USSR) 16, 271–281.

[5] Efimov, A.V. (1982) Mol. Biol. (USSR) 16, 799–806.

[6] IUPAC-IUB Commission on Biochemical Nomenclature (1970) J. Mol. Biol. 52, 1–17.

[7] Efimov, A.V. (1983) Mol. Biol. (USSR), in press.

[8] McKay, D.B., Weber, I.T. and Steitz, T.A. (1982) J. Biol. Chem. 257, 9518–9524.

[9] Pabo, C.O. and Lewis, M. (1982) Nature 298, 443–447.

[10] Samer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M. and Pabo, C.O. (1982) Nature 298, 447–451.

[11] Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Matthews, B.W. (1981) Nature 290, 754–758.

[12] Hendrickson, W.A. and Teeter, M.M. (1981) Nature 290, 107–113.

[13] Leijonmarck, M., Eriksson, S. and Liljas, A. (1980) Nature 286, 824–826.

[14] Tieme, R., Pai, E.F., Schirmer, R.H. and Schulz, G.E. (1981) J. Mol. Biol. 152, 763–782.

[15] Takano, T. (1977) J. Mol. Biol. 110, 537–568.

[16] Steigemann, W. and Weber, E. (1979) J. Mol. Biol. 127, 309–338.

[17] Ladner, R.C., Heidner, E.J. and Perutz, M.F. (1977) J. Mol. Biol. 114, 385–414.

[18] Adams, M.J., Ford, G.C., Liljas, A. and Rossmann, M.G. (1973) Biochem. Biophys. Res. Commun. 53, 46–51.

[19] Carter, C.W. jr, Kraut, J., Freer, S.T., Xuong, N., Alden, R.A. and Bartsch, R.G. (1974) J. Biol. Chem. 249, 4212–4225.

[20] Watenpaugh, K.D., Sieker, L.C. and Jensen, L.H. (1979) J. Mol. Biol. 131, 509–522.

[21] Mitsui, Y., Satow, Y., Watanabe, Y., Hirono, S. and Iitaka, Y. (1979) Nature 277, 447–452.

[22] Schellman, C. (1980) in: Protein Folding (Jaenicke, R. ed) pp.53–61, Elsevier, Amsterdam, New York.

[23] Kretsinger, R.H. and Nockolds, C.E. (1973) J. Biol. Chem. 248, 3313–3326.

[24] Szebenyi, D.M.E., Obendorf, S.K. and Moffat, K. (1981) Nature 294, 327–332.

[25] Manguen, Y., Hartley, R.W., Dodson, E.J., Dodson, G.G., Bricogne, G., Chothia, C. and Jack, A. (1982) Nature 297, 162–164.

[26] Dideberg, O., Charlier, P., Dive, G., Joris, B., Frere, J.M. and Chuysen, J.M. (1982) Nature 299, 469–470.

[27] Joris, B., Van Beeumen, J., Casagrande, F., Gerday, C., Frere, J.-M. and Chuysen, J.-M. (1983) Eur. J. Biochem. 130, 53–69.

[28] Remington, S., Wiegand, G. and Huber, R. (1982) J. Mol. Biol. 158, 111–152.

[29] Poulos, T.L., Freer, S.T., Alden, R.A., Edwards, S.L., Skogland, U., Takio, K., Eriksson, B., Xuong, N.H., Yonetani, T. and Kraut, J. (1980) J. Biol. Chem. 255, 575–580.

[30] Takio, K., Titani, K., Eriksson, L.H. and Yonetani, T. (1980) Arch. Biochem. Biophys. 203, 615–629.

[31] Remington, S.J., Ten Eyck, L.F. and Matthews, B.W. (1977) Biochem. Biophys. Res. Commun. 75, 265–270.

[32] Drenth, J., Jansonius, J.N., Koekoek, R. and Wolthers, B.G. (1971) Advan. Protein Chem. 25, 79–116.

[33] Finkelstein, A.V. (1976) Mol. Biol. (USSR) 10, 507–513.

[34] Finkelstein, A.V. (1976) Mol. Biol. (USSR) 10, 879–886.