Volume 310, number 3, 205–210 FEBS 11589 October 1992


*Hypothesis*

# Reversal of peptide backbone direction may result in the mirroring of protein structure

P. Guptasarma

*Centre for Cellular and Molecular Biology, Hyderabad-500 007, AP, India*

Received 1 June 1992; revised version received 14 August 1992

In linear polypeptides, inversion of amino acid chirality (all-L to all-D) achieves a mirroring of side chain positions and interactions in conformational space. A similar mirroring of side chain positions is independently achieved by a reversal of the direction of the peptide backbone (retro modification). Thus, while an all-D chain could be expected to adopt a perfect 'mirror image' of the three-dimensional structure of its parent all-L protein, the retro-all-L chain could be expected to adopt a topological equivalent of such a mirror image, through the symmetry transformations of side chain interactions. These notions, supported by sequence analyses, modelling studies, and evidence relating to the activity of 'retro–inverso' peptides, are extended towards the proposal, that the backbone reversed chain of a large globular protein might recognize the chiral opposite of the parent protein's substrate(s).

Retroprotein; Peptide backbone direction; Sequence analysis; Structure modelling; Chirality; Protein engineering

## 1. INTRODUCTION

The polypeptide chain of a natural protein may be called a normal-all-L chain; 'normal' defining the direction of peptide bonds along the chain, and 'all-L' defining the chirality of the amino acids used to build it. Using the condition that any transformation of the chain, by switching of amino acid direction or chirality (or both) must be uniformly applied to every residue in the chain, it is possible to think of three deviant forms of the normal-all-L polypeptide, shown in Fig. 1. These are the normal-all-D (inverso), retro-all-L (retro) and retro-all-D (retro-inverso) forms.

Attempts have been made to introduce retro- and inverso modifications of a few different kinds in small peptide hormones [1]. The results of studies conducted on peptides modified throughout the length of the chain, are as follows: normal-all-D [2–6] and retro-all-L [7,8] analogues generally do not possess biological activity. On the other hand, retro-all-D analogues have been found to possess biological activity [1,9–11] when their end groups are suitably modified [1]. The effect of the retro modification thus opposes the effect of chirality inversion, indicating that the two modifications achieve the same transformation with respect to the spatial positions of amino acid side chains. Each modification independently mirrors the side chain across the extended peptide backbone, as is obvious from a consideration of Fig. 1.

The nature of these structural relationships is discussed below in the form of a few 'Gedanken experiments', which consider various aspects of the consequences of these modifications on the structures of globular proteins. Briefly, these experiments lead to the following: (i) a new approach to understanding the topological equivalence between the retro-all-D form and the parent chain; (ii) a reiteration of the enantiomeric relationship between the normal-all-D form and the parent chain; and (iii) a derivation of the structural relationship between the retro-all-L form and the parent chain. This last relationship is shown to involve a topological mirroring of protein structure.

## 2. STRUCTURAL RELATIONSHIPS

### 2.1. *The retro-all-D polypeptide*

Let us imagine that it is possible to convert, instantaneously, every C=O atom pair in the peptide backbone of a compactly folded globular protein into an N–H atom pair, and vice versa. Would such a conversion force a change in the structure of the folded protein? It would seem that it would not, for the following reasons. The interchange would not redistribute any major charge centers. Nor would it be sterically disallowed, since, (i) the peptide bond is essentially planar in nature, (ii) the lengths of the C=O and N–H bonds (C=O ≈ 1.20 Å and N–H ≈ 1.03 Å) are not very different, (iii) the bond angle Cα–C=O in proteins (≈120°) is roughly

*Correspondence address:* P. Guptasarma, Centre for Cellular and Molecular Biology, Hyderabad-500 007, AP, India. Fax: (91) (842) 851 195.

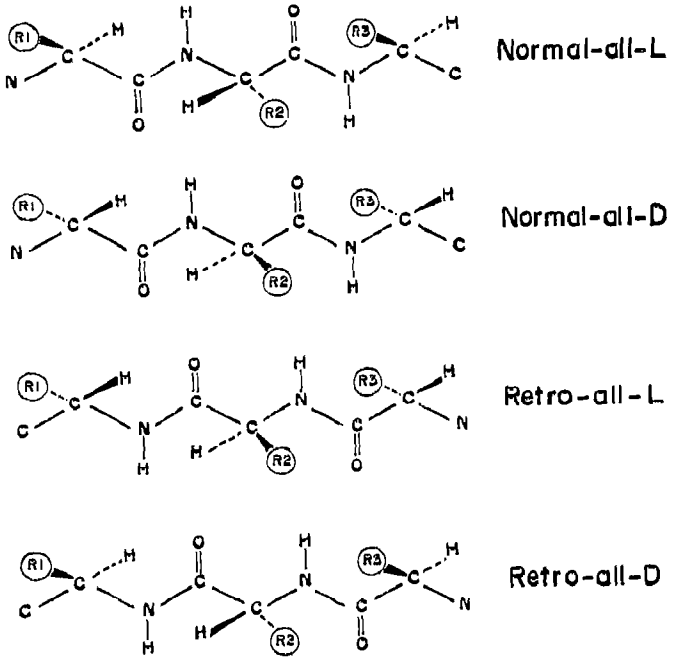


Fig. 1. The variants of the normal-all-L polypeptide obtained by inverting amino acid chirality (normal-all-D), reversing peptide backbone direction (retro-all-L), or both (retro-all-D). R1, R2 and R3 represent the side chains attached to α carbon atoms.

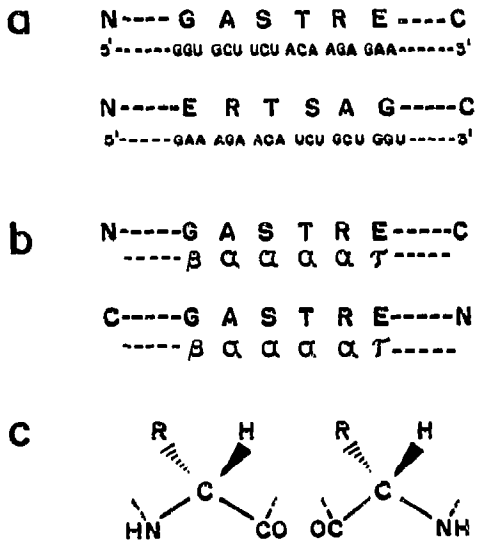


Fig. 2. (a) Corresponding segments of a protein (above) and its retro-all-L isomer (below) running in mutually reversed directions; nucleotide sequences are shown below the amino acids. (b) Schematic representation of hypothetical secondary structure prediction profiles (β, beta-structure, α, alpha-helix, τ, turn) of the two sequences. On lateral inversion of the representation of the retro isomer in (a), the predicted profiles are found to overlap. (c) Exchange of the C=O and N–H atom pairs of a residue in a folded protein, without modification of side chain position, amounts to an inversion of chirality. Note: the dashed lines in all three figures indicate the extension of the polypeptide chain on both sides of the chain.

equal to the bond angle Cα–N–H (≈115°), (iv) the van der Waals radii of the atoms involved (C = 1.65, N = 1.55, O = 1.50 and H = 1.20–1.45, Å) are comparable, and (v) the angle between the C–Cα and Cα–N bonds (≈110°) is not affected by the transformation. The transformation could, therefore, be expected not to disturb the protein's topology. Let us consider what it amounts to.

There is an obvious reversal of peptide backbone direction owing to the interconversion of the C=O and N–H atom pairs. Since the side chains are not moved from their positions, the reversal of backbone direction is effectively mediated through the exchanging of positions between the occupants of two of the four corners of the tetrahedron surrounding each α carbon atom in the chain (Fig. 2c). This amounts to an inversion of amino acid chirality, as well. The protein obtained is thus a retro-all-D molecule, which is topologically equivalent to the parent normal-all-L protein. Since the C–N interchange replaces the torsion angle C–N–Cα–C with N–C–Cα–N, the dihedral angle $\phi$ characterizing any residue in the retro-all-D protein would approximately equal the dihedral angle $\psi$ associated with the same residue in the parent protein.

### 2.2. *The normal-all-D polypeptide*

In the case of short peptides, it is easy to see that the mirroring of side chain positions in conformational space due to the L-to-D conversion would result in symmetry transformations (mirroring) of side chain interactions as well. For a globular protein refolding from the denatured state, such a mirroring of interactions may be expected to make the molecule fold into a three-dimensional structure which would constitute a perfect mirror image of the native structure of the normal-all-L parent protein. The rotations about the N–Cα and Cα–C bonds necessary for such mirroring are sterically and energetically allowed; Ramachandran et al. [12] have suggested that the preferred conformational space of D-amino acid residues inclined to participate in the formation of regular secondary structures ($-\phi,-\psi$), is obtained by reflecting the conformational space of L-amino acid residues across the origin of the Ramachandran diagram. A right-handed alpha-helix in the parent protein would, therefore, be represented by a left-handed helix in the normal-all-D protein. Polypeptide chains made of D-amino acids might therefore adopt structures that would mirror their parent molecules so completely that they could recognize the chiral opposites of the substrates of the original peptides.

### 2.3. *The retro-all-L polypeptide*

Reversal of peptide bond direction achieves a mirroring of side chain positions in conformational space when it is not accompanied by inversion of amino acid chirality (Fig. 1). The resulting symmetry transformations of the side chain interactions could therefore cause

the retro-all-L chain to fold, as well, into a mirror image of the parent normal-all-L protein. This mirror image would, however, be only a topological equivalent of the structure adopted by the normal-all-D protein, since every C=O atom pair in the latter would be represented by an N–H atom pair in the former.

The dihedral angle $\phi$, corresponding to any residue in the retro-all-L protein, would therefore represent the dihedral angles $\psi$, $-\psi$, and $-\phi$ respectively, in the normal-all-D, normal-all-L and retro-all-D proteins. It is noted parenthetically that a change of sign in these transformations indicates a mutual mirroring of side chain positions, while a change of dihedral angle indicates the replacement of the torsion angle C–N–Cα–C, by the torsion angle N–C–Cα–N, owing to the reversal of backbone direction. The relationship of the parent normal-all-L molecule to its three deviant forms is outlined below.

normal-all-D: $\phi$ corresponds to $-\phi$ of parent molecule (and $\psi$ to $-\psi$) **(I)**

retro-all-L: $-\phi$ corresponds to $\psi$ of parent molecule (and $-\psi$ to $\phi$)? **(II)**

retro-all-D: $\phi$ corresponds to $\psi$ of parent molecule (and $\psi$ to $\phi$) **(III)**

It may be noted that the transformations described by expressions I and II, add up to that described by expression III, as expected.

## 3. MODELLING THE RETRO-ALL-L STRUCTURE

The following computational experiment was done to test the validity of the proposed mirror imaging principle. The pair of mutually retro-running hexapeptides, $H_2N$–GASTRE–COOH (comprising residues 56–61 of the hypervariable L2 loop in the light chain of the phosphorylcholine binding antibody, McPC 603 [13]), and its retro-peptide, $H_2N$–ERTSAG–COOH, were constructed in an extended conformation on the Desktop Molecular Modeller (DTMM, Version 1.0, (C) Oxford University Press, 1989) software, and put through an energy minimization routine using the derivative method developed by Vinter et al. [14], on an IBM PC/AT 386 equipped with an 80387 coprocessor and a 4 MB RAM, running at a clock speed of 25 MHz. Identical side chain conformations were used for corresponding amino acids in the two chains. Energy minimization was carried out for 1,000 cycles of iteration during which simultaneous minimization of bond length energy, bond angle energy, torsion angle energy and van der Waals interaction energy was carried out.

Fig. 3 shows the starting and final conformations of the two molecules, while Fig. 4 shows a comparative view of their peptide backbones. The structures in Fig. 3 tend to mirror each other topologically in a manner that could allow them to recognize chirally opposite substrates. At some α carbons, the corresponding dihedral angles $\phi$ and $\psi$, can already be seen to be tending towards equal and opposite values at the end of 1,000 iterations. Emphasis here is laid not on the accuracy or feasibility of structure prediction by energy minimization methods, but on the fact that the same considerations during energy minimization (applied for the same number of iterations, to two starting structures made with identical side chain conformations, and no bias in backbone conformation) resulted in such structures.

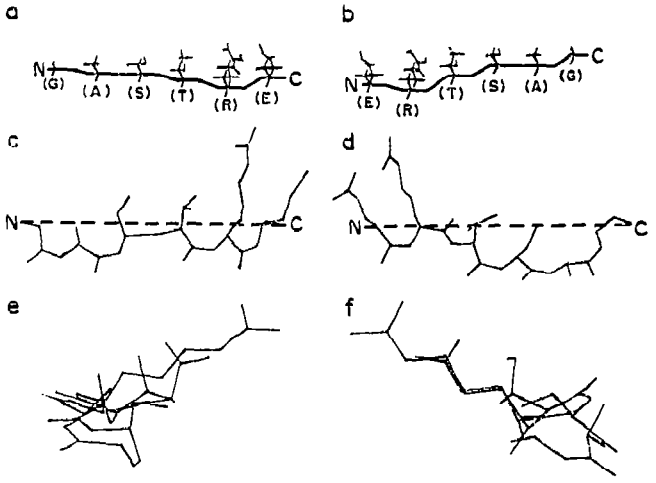


Fig. 3. (a and b) The two chains, $H_2N$–GASTRE–COOH and $H_2N$–ERTSAG–COOH, as constructed on DTMM, prior to minimization. The energies of the molecules are 27,984 and 20,588.3 kJ/mol, respectively. (c and d) The corresponding structures obtained after 1,000 iterations. The energies of the molecules at this stage are 3,729.28 and 3,901.49 kJ/mol, respectively. The dotted line joins the carbon of the C-terminal carboxyl group to the nitrogen of the N-terminal amino group. Of the atoms at the termini, only N and C are shown here. (e) View of (c), looking down from N to C. (f) View of (d) looking down from C to N. The amino acid G (Gly) is on top in both cases.

## 4. ANALYZING RETRO-ALL-L SEQUENCES

What about large proteins with significant amounts of secondary structure? Would regions of the retroprotein, corresponding to helix- or sheet-forming regions in the parent protein, be equally disposed to forming regular structures? The transformations described in expressions I and II show that residues participating in right-handed alpha-helices in the normal-all-L chain are likely to form left-handed helices in both the normal-all-D and retro-all-L chains, since the $\phi$,$\psi$ pairs obtained by both transformations fall in the 'left-handed helix' region of the Ramachandran diagram. Few would contest the suggestion that a right-handed helix in a normal-all-L protein would be replaced by a left-handed helix in its normal-all-D counterpart. Considering that most

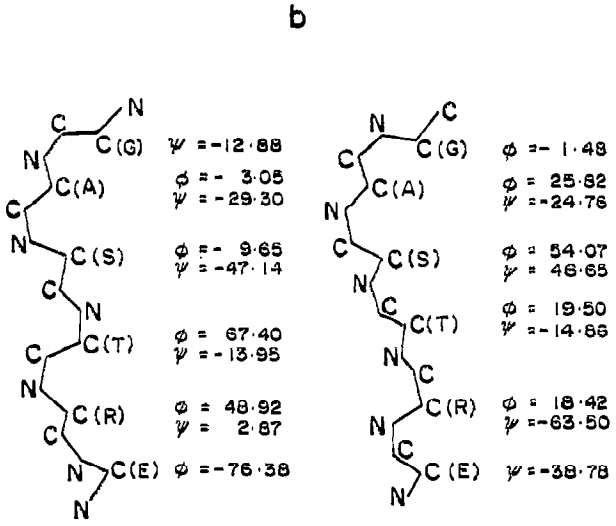


Fig. 4. (a) Peptide backbone of the molecule shown in Fig. 3c, rotated clockwise by about 90°. (b) Peptide backbone of the molecule shown in Fig. 3d, rotated anti-clockwise by about 90° and then rotated by 180° about the C–N axis.

known all-L proteins possess only right-handed helices, however, the plausibility of the second contention, namely that a similar replacement would occur in the retro-ali-L protein as well, needs to be examined. Can an all-L protein contain left-handed α-helices? Or, for that matter, can an all-D protein contain right-handed helices, as is suggested by our earlier consideration that the retro-all-D protein would be equivalent to the normal-all-L protein?

That the left-handed helical conformation is not disallowed in all-L chains per se is evident from the fact that left-handed helical residues are not unknown among proteins crystallized to date [15,16]. While glycine adopts right- and left-handed helical conformations with roughly equal frequencies, a large number of non-glycine residues, too, are found in the left-handed helical conformation [16], indicating that L-amino acids are not sterically disallowed from adopting such a conformation, in an otherwise all-L context. Weaver et al. [17] found that glycines and non-glycines occurring in left-handed helical conformation in loop regions, or at the termini of helices, have very similar backbone conformational energies (within 0.5 kcal/mol of each other). This indicates that, energetically speaking, left-handed conformations of this kind are not prohibited. But then, how does the existence of the occasional left-handed alpha-helical residue in natural all-L proteins relate to the larger question of the formation of a left-handed alpha-helix? Obviously, the formation of such a structure would require that a stretch of contiguously placed residues in a chain be similarly disposed to adopting a left-handed helical conformation. Would such a stretch be allowed to form a left-handed helix?

Ramachandran et al. [12] point out that both the right- and left-handed alpha-helical forms lie outside the 'normally allowed' regions of the Ramachandran diagram but within the 'outer limits' of the sterically allowed regions. In L-amino acid chains, the perfect right-handed alpha-helical form ($\phi = -50$, $\psi = -50$) is somewhat more stable than the perfect left-handed alpha-helical form ($\phi = 50$, $\psi = 50$), while the converse is true for D-amino acid chains. Helices in proteins, however, rarely conform to such standards; the average helix in a natural protein is a substantially distorted version of the perfect right-handed alpha-helix [15]. Thus, while it is true that right-handed helices of L-amino acids are likely to tolerate much more distortion than left-handed helices [12], since they occupy a larger region of the Ramachandran map, some sequences of L-amino acids could find themselves having to choose between: (i) adopting a very distorted (unstable) right-handed alpha-helical structure; and (ii) adopting a not-so-distorted (slightly more stable) left-handed alpha-helical structure. In such a situation, some sequences might very well adopt continuous left-handed helices.

Perhaps the reason that we do not see such structures in natural proteins is a consequence of the fact that sequences inclined to form such structures have not been chosen by evolution, or have otherwise gone undetected. I suggest that the retro-all-L sequences of natural proteins constitute such sequences, and that, therefore, they have a very fair chance of adopting left-handed alpha-helices.

How does the extended (beta) conformation respond to the $\phi = -\psi$, $\psi = -\phi$ transformation achieved by the retro-all-L modification? This question is most easily addressed by taking the coordinates of any point from within the region defining the beta-structure in the Ramachandran diagram, and putting it through the above transformation to see where it lies after the transformation. Such an exercise shows that, for every such point, the coordinates of the transformed point continue to lie within the same region of the map; for instance (−150,100) becomes (−100,150). Thus, the retro-all-L modification would appear to conserve extended beta-structures.

How do available routines for protein sequence analysis respond to the suggestion that secondary structural elements are conserved through backbone reversal? Before such a question is addressed, of course, it is necessary to realize the two predictable consequences of asking such a question: (i) no available program could be expected to predict the occurrence of a left-handed alpha-helix; and (ii) some of these programs use an empirically derived set of structure forming probabilities, to examine the probability of formation of a particular secondary structure by an amino acid, in the context of its immediate neighbourhood. Since the neighbourhood of any residue in a normal-all-L protein happens to be equivalent to that of its corresponding resi-

due in the retro-all-L protein, the outputs of such programs cannot be influenced substantially by the direction of the chain. Thus, the predictions for a natural protein and its retro-protein, represented graphically as plots of calculated amino acid characteristics (Y-axis) vs. amino acid sequence (X-axis), may be expected to mirror each other across the Y-axis (or be superimposable upon lateral inversion of either profile). As is described below, this is indeed what happens.

A number of protein sequences were reversed and analyzed on the sequence analysis software PC/Gene (Version 6.50, (C) Amos Bairoch, University of Switzerland, Geneva). The software was used to carry out secondary structural (and other) analyses, using the routines ANTIGEN (antigenicity of fragments), SOAP (hydropathy profile), GGBSM, GARNIER, NOVOTNY, BETATURN (secondary structure prediction), RADIALOC (prediction of radial locations in globular proteins) and FLEXPRO (prediction of chain flexibility). In some secondary structure prediction routines, i.e. those which output very different predictions for the normal-all-L and retro-all-L sequences when default settings of user-defined parameters (e.g. percentages of secondary structures etc.) are used, these parameters were adjusted to predict a close match to the known secondary structure of a protein before the routine was applied to analyzing its retro-protein.

As expected, the predictions for mutually reversed chains mirrored each other (Fig. 2b illustrates this point schematically). Thus, the regions predicted to form helices, extended sheets, and even beta-turns, comprise the same set of residues in two proteins with mutually reversed chain directions. For instance, the residues 4–7 and 40–43 in the 46 residue plant toxin, crambin (sequence obtained from [18]), which have the highest predicted potential for beta-turn formation, overlap (three residues out of four) with regions 41–44 and 5–8, which have the highest corresponding potential in the retro-crambin sequence obtained by reversing the crambin sequence. The shift occurs presumably to accommodate the residue, proline, in the second position of the turn. Similarly, the retro-sequence of the protein, felix (a de novo designed four helix bundle) turns out to have the same groups of residues forming predicted helices and beta-turn-containing loops, as its parent protein (sequence from [19]).

Although these outcomes appear to support the contentions made in this paper, their significance may be called into question, since the methods used for these predictions are somewhat insensitive to chain direction. However, considering that these programs use a knowledge-based approach to predict the secondary structures of all-L chains with a modest degree of success [20], the prediction of beta-structures, in line with the expectations outlined earlier, may be taken to support the notion that a beta-structure remains a beta-structure even upon retro-all-L modification. Alternatively, these results probably only bring to light a serious lacuna in current secondary structure prediction methods. Certainly, the prediction of helical stretches in corresponding segments of the two proteins arises from the identification and propagation of helix-nucleating regions in a direction-independent manner.

## 5. PERSPECTIVES

Retroproteins are likely to adopt three-dimensional structures which are topological mirror images of the native structures of the parent proteins. They could thus recognize and modify the chiral opposite(s) of the parent protein's substrate(s). While such a function would, in all likelihood, be performed just as well by normal-all-D molecules, these are not easily synthesized. Retroproteins, on the other hand, are very easily made in living systems (Fig. 2a) through modern recombinant DNA technology. Such molecules could find applications in the introduction of new metabolic pathways into living systems, e.g. enzymes designed to recognize and use L-glucose could enhance the nutritional adaptability of microorganisms. Others could be used to produce unnatural isomers of biological molecules inside living systems, or distinguish between chirally unselected products of chemical reactions.

*Acknowledgements:* For insightful discussions and help with the preparation and editing of this manuscript, I am deeply indebted to Prof. D. Balasubramanian, CCMB, Hyderabad. I am also indebted to Manni Guptasarma and Dipankar Manna, CCMB, for helpful discussions, and to Sushil Chandani for comments on the manuscript. I thank Dr. Matti Saraste, Editor, FEBS Letters, for bringing independently to my notice, an article by Kent and colleagues ([1992] Science 256, 1445), which appeared during the refereeing of this manuscript. This article shows that the normal-all-D HIV protease does actually fold into a structural mirror image of its parent protein to recognize the chiral opposite of its substrate, as proposed in this paper.

## REFERENCES

[1] Goodman, M. and Chorev, M. (1979) Acc. Chem. Res. 12, 1.
[2] Schroeder, E., Lubke, K. and Hempel, R. (1965) Experientia 21, 70.
[3] Stewart, J.M. and Wooley, D.W. (1965) Nature 206, 619.
[4] Flouret, G. and du Vigneaud, V. (1965) J. Am. Chem. Soc. 87, 3775.
[5] Vogler, K., Studer, R.O., Lergier, W. and Lanz, P. (1965) Helv. Chim. Acta 48, 1407.
[6] Morley, J., Tracy, H.J. and Gregory, R.A. (1965) Nature 207, 1356.
[7] Bodanszky, M., Ondetti, M.A. and Sheehan, J.T. (1963) Ann. NY Acad. Sci. 104, 24.
[8] Vogler, K., Lanz, P. and Lergier, W. (1962) Helv. Chim. Acta 45, 561.
[9] Chorev, M., Willson, G.C. and Goodman, M. (1977) J. Am. Chem. Soc. 99, 8075.
[10] Shemyakin, M.M., Ovchinnikov, Yu.A. and Ivanov, V.T. (1969) Angew. Chem., Int. Ed. Engl. 8, 492.
[11] Goissis, G., Nonailhetas, V.L.A. and Paiva, A.C.M. (1976) J. Med. Chem. 19, 1287.

[12] Ramachandran, G.N., Ramakrishnan, C. and Sasisekharan, V. (1963) J. Mol. Biol. 7, 95.
[13] Satow, Y., Cohen, G.H., Padlan, E.A. and Davies, D.R. (1986) J. Mol. Biol. 190, 593.
[14] Vinter, J.G., Davis, A. and Saunders, M.R. (1987) J. Comp. Aided Mol. Design 1, 31.
[15] Presta, L.G. and Rose, G.D. (1988) Science 240, 1632.
[16] Nicholson, H., Solderlind, E., Tronud, D.E. and Matthews, B.W. (1989) J. Mol. Biol. 210, 181.
[17] Weaver, L.H., Tronud, D.E., Nicholson, H. and Matthews, B.W. (1990) Curr. Sci. 51, 833.
[18] Teeter, M.M., Mazer, J.A. and L'Italien, J.J. (1981) Biochemistry 20, 5427.
[19] Hecht, M.H., Richardson, J.S., Richardson, D.C. and Ogden, R.C. (1990) Science 249, 884.
[20] Garnier, J. (1990) Biochimie 72, 513.