ELSEVIER

FEBS Letters 340 (1994) 202–206

FEBS 13723

FEBS LETTERS

Predicted topology of the N-terminal domain of the hydrophilic subunit of the mannose transporter of *Escherichia coli*

Zora Marković-Housley[a,*], Jochen Balbach[c], Beat Stolz[a], Jean-Claude Génovésio-Taverne[b]

[a]*Institut für Biochemie der Universität Bern, Freiestrasse 3, CH-3012 Bern, Switzerland*
[b]*Abteilung Strukturbiologie, Biozentrum der Universität Basel, CH-4056 Basel, Switzerland*
[c]*Organisch Chemische Institut, Technische Universität München, 85747 Garching, Germany*

Received 14 December 1993

Abstract

A folding topology for the homodimeric N-terminal domain (IIA, 2 × 14 kDa) of the hydrophilic subunit ($IIAB^{man}$) of the mannose transporter of *E coli* is proposed. The prediction is based on (i) tertiary structure prediction methods, and (ii) functional properties of site-directed mutants in correlation with NMR-derived α/β secondary structure data. The 3D structure profile suggested that the overall fold of IIA is similar to that of the unrelated protein, flavodoxin, which is an open-stranded parallel β-sheet with a strand order of 5 4 3 1 2. The 3D model of IIA, constructed using the known atomic structure of flavodoxin, is consistent with the results from site-directed mutagenesis. Recently NMR results confirmed the open parallel β-sheet with a strand order of 4 3 1 2 (residues 1–120) of our model whereas β-strand 5 (residues 127–130) was shown to be antiparallel to β-strand 4. The correctly predicted fold includes 90% of the monomeric subunit sequence and contains all functional sites of the IIA domain.

Key words: Mannose transporter; α/β protein; Protein topology; 3D structure prediction

1. Introduction

Numerous examples of proteins with similar 3D structures but little or no sequence similarity (< 15%) [1,2] have been identified over the last decade as more and more 3D structures have become available. This has led to the proposal that the number of folding motifs is limited [3] and that conservation at the level of structure as opposed to amino acid sequence is more likely. Thus, the environmental properties of each residue in the structure (for example, solvent accessibility and local secondary structure) appear to be more conserved than the identity of the residues themselves. In view of this, several groups (e.g. Bowie, Eisenberg and colleagues [4,5] and Johnson, Blundell and colleagues [6,7]) developed highly automated methods for tertiary structure prediction. These methods align test amino acid sequences with 3D profiles of known structures, represented by a chain of environmental properties at each residue site. Using this approach it was demonstrated that several proteins, though lacking sequence similarity, nonetheless possess highly similar tertiary structures [4]. However, while these methods can usually identify the relationship between unrelated proteins of highly similar tertiary structures (e.g. actin and heat-shock protein), the detection of proteins with similar overall topologies but varying tertiary structures (e.g. insertions and deletions within a similar domain, or existence of an extra domain) is more difficult (e.g. periplasmic arabinose and ribose-binding proteins) [4].

The mannose transporter protein complex in *Escherichia coli* consists of a hydrophilic subunit, $IIAB^{man}$, and two transmembrane subunits, IIC^{man} and IID^{man}. The hydrophilic subunit ($IIAB^{man}$) is a homodimer of two 35 kDa subunits which, when complexed with IIC^{man} and IID^{man}, catalyses the phosphate transfer from phospho carrier protein (HPr) to the sugar substrate [8]. Each monomer consists of two structurally and functionally distinct domains, N-terminal domain IIA (14 kDa, residues 1–136) and C-terminal domain IIB (20 kDa, residues 156–323) which are linked by an Ala–Pro-rich flexible hinge. Stable isolated domains can be produced by limited trypsin digestion followed by column chromatography [8]. The 14 kDa N-terminal domain (IIA) is a homodimer that carries the sites for dimerisation and for the interaction with HPr. During catalysis the phosphoryl group released by HPr is accepted by His^{10} in IIA, further transferred to the IIB domain and hence to sugar.

In view of our interest in the structure–function relationship of this complex membrane protein, various studies were recently initiated with the entire $IIAB^{man}$ subunit and with the isolated domains. The N-terminal

*Corresponding author. Fax: (41) (31) 631 4499.
E-mail: zhousley@ibc.unibe.ch

Abbreviations: IIA, amino-terminal domain of hydrophilic subunit of mannose transporter; HPr, histidine containing phosphoryl carrier protein; NMR, nuclear magnetic resonance.

SSDI 0014-5793(94)00107-7

domain IIA has been particularly well characterised. Structural studies using two- and three-dimensional NMR techniques have recently provided evidence for the presence of α/β secondary structure elements [9]. Single crystals diffracting to 2 Å resolution have also been characterised by X-ray diffraction [10]. In addition, functional characterisation of several site-specific mutants were reported [11]. These include a Ser-to-Cys mutant as possible target for heavy atom attachment, and a Trp-to-Phe mutant designed for fluorescence studies. The results of these studies in correlation with the general features of α/β proteins [12,13] have provided sufficient information to propose a possible folding topology of the IIA domain. To corroborate our proposal, predictive methods based on the comparison of the IIA primary structure [14] with 3D structure profiles were employed. This method suggested a structural relationship between IIA and clostridial flavodoxin. An approximate 3D model of the IIA domain, based on the known structure of flavodoxin, was constructed. This model is used for interpreting current results and for planning new experiments.

Fig. 1. Comparison of the secondary structure of IIA determined by NMR (solid underlined) and predicted by the Chou-Fasman method (dashed overlined). Bold and thin lines (dashed and solid) denote β-strands and α-helices, respectively. The preliminary sequential order of β-strands and α-helices determined by NMR is indicated by the numbered β and α symbols. Boxed letters indicate residues involved in the activity of IIA, as inferred from mutant studies.

2. Materials and methods

The biochemical studies of wild-type and mutant $IIAB^{man}$ proteins have been reported elsewhere [11]. The programs PROFILE [15] and QSLAVE [7] were used to align amino acid sequences to tertiary profiles constructed from protein coordinates deposited in the PDB databank [16]. The programs used for the secondary structure prediction were part of the Genetic Computer Group (GCG) software package [17]. Modeling was performed on an ESV graphic workstation with the program, FRODO [18]. Replacement of the amino acid side chains was done manually with the program, INSIGHTII [19]. The resulting model was energy minimised with the program, XPLOR [20] on a CONVEX C120 computer.

3. Results and discussion

3.1. Hypothesis based on experimental data

Two independent experimental observations led to the intuitive prediction of the overall folding topology of the IIA domain. Firstly, studies of the $IIAB^{man}$ mutants suggest that the sequence-remote residues near Trp^{69} and Ser^{72} in the IIA domain are most probably spatially close to the active site, His^{10}. This information is based on the drastically reduced sugar phosphorylation activities (< 6%) of the two mutants, W12F and S72C, as compared to that of wild-type $IIAB^{man}$ [11]. Further, the activity of the double mutant, W12,69F, was similar to that of the wild-type protein, demonstrating that the low activity of a single W12F mutant can be 'restored' by the second mutation at residue 69 (Z. Marković-Housley, unpublished results). This strongly suggested an interaction between residues 12 and 69. Secondly, the secondary structure determined by preliminary NMR studies indicated an alternating β/α secondary structure consisting of 4 α-helices and 5 β-strands (Fig. 1). The results from a secondary structure prediction, using the method of Chou and Fasman [21] (GCG package program), are also shown in Fig. 1 for comparison. The NMR secondary structure assignments place the active site, His^{10}, at the carboxyl-terminal end of β-strand 1 (loop 1), and residues 69 and 72 in the loop at the carboxyl-terminal end of β-strand 3 (loop 3). This result is consistent with the observation that the active site residues in nearly all commonly observed α/β structures, both the closed eight-stranded parallel β-barrels and the open twisted parallel β-sheets, are located in the loop regions at the carboxyl ends of the β-strands [12]. In principle, an α/β protein can assume either an open β-sheet structure, which is not limited in size, or an eight-stranded β-barrel of limited size (a minimum of about 200 residues). Thus, the IIA dimer could assume either of these two structures while the IIA monomer ($\approx$130 residues) is only compatible with an open β-sheet structure. For the individual IIA monomer a particularly attractive topology was immediately very suggestive, that is, an open β-sheet in which strands β-1 and β-3 are adjacent. This would permit the spatial proximity of loops 1 and 3, as suggested by the mutation studies. Such an arrangement with the reversed strand order would be possible only in an open parallel β-sheet but not in a parallel β-barrel due to the right-handed nature of the $\beta\alpha\beta$ motif [13]. A survey of 20 different open-sheet α/β structures made by Brändén [12] showed that the reversed strand order is always present and is indicative of active site crevices formed by the oppositely directed loops that connect the C-terminal ends of parallel and adjacent β-strands to the N-terminal ends of α-helices on both sides of the β-sheet (a so-called

topological switch point) [13]. In almost all reported structures of this kind the active site, or part of it, is found in such crevices. This pattern recurs so regularly that the position of the active site can be predicted with high probability by identifying a topological switch point in the topology diagram. Here the converse reasoning was applied in an attempt to deduce the local topology of IIA from the presumed positions of the active site residues (loop 1 and 3), as suggested by mutant studies. Considering the above arguments, together with the observation that the strand order 4 3 1 2 is frequently found in open-sheet α/β proteins [13], we suggest that the probable topology of IIA is a parallel β-sheet with this strand order. However, without further information, alternative β-strand arrangements cannot be excluded.

3.2. Tertiary structure prediction

To test the above hypothesis and to gain an independent prediction of the topology, the IIA sequence was subjected to tertiary structure prediction based on 3D structure profiles derived from known atomic structures [4,7]. These methods are particularly successful if there is a significant sequence similarity between a test sequence and a protein of known structure. Unfortunately this is not the case with the IIA domain. However, it has been shown that these methods can, in some cases, identify sequences that adopt a similar fold, even in the absence of significant sequence similarity. In view of these observations and knowing that IIA is an α/β protein we have sequentially matched the IIA sequence, as well as 200 randomly selected protein sequences, to 3D structure profiles derived from 20 α/β protein structures, surveyed by Brändén [12]. Using the program, PROFILE, matching of the IIA sequence to the 3D profiles showed that the best alignment was consistently found with a template constructed from clostridial flavodoxin [22], although the Z-score was relatively low. The highest scores (calculated with penalties for a gap opening of 4.0 and a gap elongation of 0.04) were obtained with other flavodoxins (Z = 12.2 and 5.9), but the next highest scoring protein was IIA (Z-score = 3.1). To account for the strong bias of these methods in selecting proteins of similar lengths (138 vs. 136 residues for flavodoxin and IIA, respectively), 14 α/β sequences of known structure were truncated in such a way as to keep a minimal fold of similar size close to that of flavodoxin. Their lengths ranged from 126 to 164 with an average value of 145. Matching the IIA to truncated profiles showed that the second highest score, after flavodoxin, was again observed with IIA. Thus, a similar polypeptide length is not a predominant selection criterion. Interestingly, among the tested truncated proteins of similar size the only non-flavodoxin protein which had a Z-score similar to that of IIA was carboxypeptidase. The truncated carboxypeptidase has the same strand order (4 3 1 2) as flavodoxin. This result indicates the potential of the 3D profile method for selecting for proteins which have similar strand order.

```
        bbbbbbbb     hhhhhhhhhhhhhhh  bbbbbb
AII   MTIAIVIGTHGWAAEQLLKTAEMLLGEQENVGWIDFVPGE
FXN   ..MKIVYWSGTGNTEKMAELIAKGIIESGKDVNTINVSDV   40
        bbbbbb    hhhhhhhhhhhhhhhhh  bbbbbb

        hhhhhhhhhh          bbbbbb          hh
AII   NAETLIEKYNAQLAKLDTTKGVLFLVDTWGGSP.....FN
FXN   ...NIDELLNED........ILILGCSAMGDEVLEESEFE   80
         hhhhhhhh           bbbbbb          hhh

      hhhhhhh     bbbbbb          bbbbbbbbbbbbb
AII   AASRIVVDKEHYEVIAGVNIPMLVETLMARDDDPSFDELV
FXN   PFIEEISTKISGKKVALFGSYGWGD.....GKWMRDFEER  120
      hhhhhhh     bbbbbbbbb         hhhhhhh

      hhhhh      bbbb
AII   ALAVETGREGVKALKAKPV
FXN   MNGYGCVVVETPLIVQN
           bbbbbbbbbbbb
```

Fig. 2. Alignment of the amino acid sequences of the IIA domain and flavodoxin according to their secondary structures. Symbols b and a stand for β-strands and α-helices, respectively. This alignment was used to model the IIA sequence on the 3D structure of flavodoxin.

To verify the results obtained with the program, PROFILE, and to account for the small number of tertiary templates used, the program, QSLAVE, based on an algorithm of tertiary structure prediction developed by Blundell and colleagues, was also applied [7]. Out of 248 protein templates derived from families of homologous protein structures, 101 templates (α, β and α/β) of relatively similar size (134 ± 50 residues) were used for matching the IIA sequence. The two highest scoring templates were both flavodoxins (the scores for 3fxn was 1775 and score for 1fx1 was 1878, although 1fx1 gets a lower score when multiplied by the length-dependent weighting factor). The selection of the flavodoxin as the highest scoring template out of 101 templates is significant, despite the relatively low score, and indicates a structural relationship between the IIA domain and flavodoxin.

3.3. Construction of a 3D model

In order to model the IIA sequence on the 3D flavodoxin structure, the amino acid sequences of IIA and flavodoxin (16% identity) were aligned manually with respect to their secondary structures. The NMR-determined secondary structure of IIA was used for this alignment. The optimal alignment was obtained by introducing gaps in the loop regions of each protein (Fig. 2). The alignment of α-helices and β-strands is surprisingly good; 82% of the α-helix and β-sheet residues are aligned (67/82). It is also remarkable that the active site residues located in loop 1 of each protein are involved in phosphate binding. Residues in loop 3 of IIA are also involved in phosphate binding while residues in loops 2 and 3 of flavodoxin interact primarily with the isoalloxazine ring of the flavin.

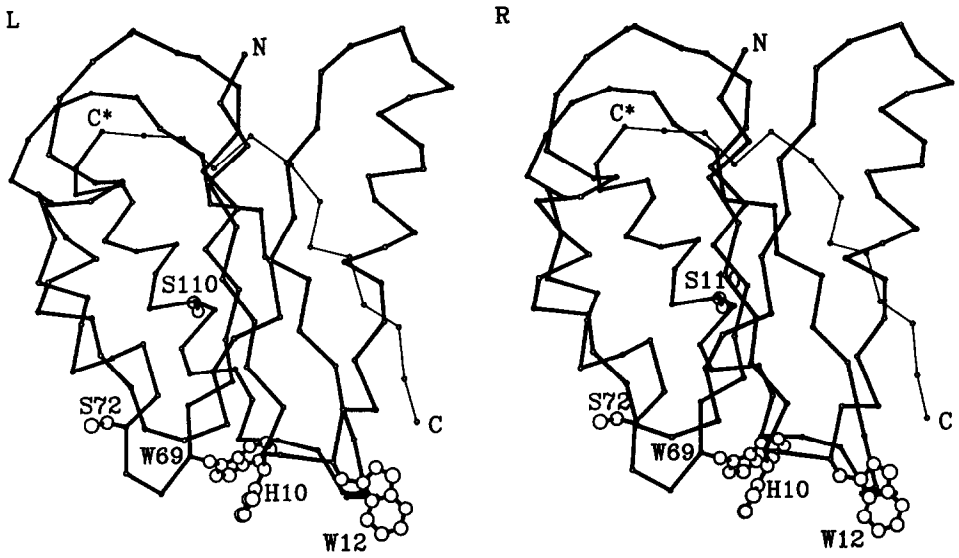

Fig. 3. Stereoview of the C_α-backbone of the energy-minimized IIA model. The side chains of the functionally and structurally relevant residues are indicated. Symbol C stands for the C-terminal part of the molecule, and C* indicates the last residue of the correctly predicted part of the IIA domain (solid line). The thin line represents the part of the model which could not be predicted.

The alignment shown in Fig. 2 was used as a guide to superimpose residues of IIA onto the defined coordinates of flavodoxin using the computer program, FRODO, on an ESV graphics workstation. Maintaining the peptide backbone of flavodoxin the side chains of non-identical amino acids in α-helices and β-strands were replaced and the backbone torsion angles, ϕ and ψ, were optimized with the program, InsightII. Remaining amino acids from IIA were modelled in loops connecting secondary structure elements by using InsightII that has facilities for searching the Protein Data Bank for possible loop conformations based on loop sequence homologies. The entire structure was energy minimised (program XPLOR) in order to produce the model that is chemically correct with respect to bond lengths, bond angles, and dihedral angles. The resulting IIA model is shown in Fig. 3. Superposition of IIA model structure onto that of flavodoxin showed good agreement for the entire monomeric sequence of the IIA domain (residues 1–136). The model differed from the structure of the clostridial flavodoxin by an average rms deviation of 2.5 Å for main chain atoms of the 82 residues in the β-strands and α-helices. An analysis of the hydrophobic residues showed that the model, although not precise, does possess a reasonably well-defined hydrophobic core. Strands 1, 3 and 4 are made almost exclusively of hydrophobic residues. These 3 strands are totally buried in the interior of the model and build up an hydrophobic core with the hydrophobic residues of the surrounding helices. However, some of the hydrophobic residues are not buried. This may be due either to the inaccuracy of the model or to the fact that the isolated IIA is a tryptic fragment of $IIAB^{man}$ which physiologically exists as a dimer. Thus, it is likely some of the exposed residues in the model are buried in the dimer interface and others in the IIA–IIB inter-domain interface within the intact molecule. To assess the validity of the proposed model the program, PROFILE, was used to match the 3D model with its own amino acid sequence. A good compatibility of the 3D model with its 1D sequence is shown by the calculated score of Z = 9.1, which compares well with the scores obtained by matching the FXN structure with the sequences of flavodoxin from other organisms (Z = 12.2 and 5.9).

3.4. *Confirmation of the model*

Recently, a parallel β-sheet with the strand order of 4 3 1 2 of our model (which contains 90 amino acids) was confirmed by preliminary results from long range nuclear Overhauser effect [9]: ten inter-strand contacts below 5 Å were found between strands 1 and 3, five between 1 and 2, and eight between 3 and 4. However, strand 5 (including residues 127–130) at the C-terminal end was found to be antiparallel to strand 4. Incorporation of the new NMR data (hydrogen bonds between the β-strands) into the initial model permitted a more accurate positioning of the strands with respect to one another but clearly did not produce a significant change in the overall β-sheet topology that was predicted. Also, it proved to be quite difficult to connect the spatially close ($\approx$4 Å) antiparallel strands 4 and 5 by the rigid element, α-helix 4. To achieve this, helix 4 had to be relaxed and a hairpin inserted. The exact conformation of the C-terminal part of IIA awaits the 3D structure determination by X-ray crystallography and NMR.

The IIA model is consistent with the functional properties of a number of site-directed mutants, for example, the spatial proximity of the active site residues located

in loops at the carboxyl end of adjacent β-strands 1 and 3, as discussed above (see section 3.1.). It also helped us to interpret crosslinking experiments with cysteine mutants and to propose the possible subunit arrangement in IIA. For example, it has been shown that the single mutants, S72C, H10C and S110C, readily form intersubunit disulfide bridges under oxidising conditions [11], indicating that these residues must be spatially close and thus near the dimer interface. According to our model (Fig. 3) a spatial proximity of Cys^{10} and Cys^{72} would be possible if the two IIA domains were arranged such that the carboxyl ends of the parallel strands of each domain point towards each other. This is the case in many two-domain enzymes which bind two substrates [12]. However, from the model it is not obvious how Cys^{110}, located near the end of the long loop joining strand 4 and helix 4, can be near the putative dimer interface. Further support for the proposed dimer arrangement came from solution studies of the mutants, H10C, W12F and S72C. These mutant molecules are less resistant to dissociation by sodium dodecylsulphate than the wild-type protein, suggesting that residues 10, 12, and 72 are involved in the interactions at the dimer interface [11]. Additional studies replacing residues in the loops at the carboxyl ends of strands 1, 2 and 3 with Cys are planned to verify the validity of this mode of dimerisation.

To summarize, the overall tertiary fold of the IIA domain is shown to be similar to that of the sequence-unrelated protein, flavodoxin, which is a parallel open β-sheet with a strand order of 5 4 3 1 2 and helices on both sides of the sheet. This structural similarity was identified by applying the predictive method of 3D structure profiles. The flavodoxin fold is consistent with the properties of a number of site-directed mutants and NMR secondary structure. The approximate 3D model of IIA has been constructed using the known atomic structure of flavodoxin and NMR secondary structure of IIA. Recent NMR results confirmed the parallel strand order, 4 3 1 2, of our model (residues 1–120), whereas β-strand 5 was shown to be antiparallel to β-strand 4. Thus, the fold of 90% of the entire amino acid sequence was correctly predicted while the conformation of the C-terminal part of the molecule (residues 120–134, including loop 4 and β-strand 5) was not. The correct prediction of 90% of the overall fold of IIA represent a significant success in terms of the employed predictive methods. Moreover, it demonstrates the value of these methods in identifying similarly folded protein domains lacking complete structural and functional relationships. Our inability to predict the conformation of the C-terminal part of the IIA domain is not a serious limitation in using this model for the design of new functional and structural mutants since the active site residues are all located near the middle of the β-sheet. Also, separate studies with the IIA domain truncated at residue 110 showed that this part of the molecule is redundant for activity in vivo (B. Erni, unpublished results). Although the proposed model is not accurate, it is consistent with our data and suggestive regarding a possible arrangement of the subunits in the dimer. However, it awaits final confirmation from NMR and X-ray diffraction analysis. The X-ray structure determination, using crystals diffracting to 2.0 Å resolution, is presently in progress.

Acknowledgements: The authors gratefully acknowledge the contributions made by Dr. T. Blundell in providing the use of the software facilities at Birbeck College (University of London), and by Dr. R. Sowdhamini in performing the profile search with the QSLAVE program. We also wish to thank Drs. C. Wright and T. Schirmer for critical reading of the manuscript, and Dr. C. Brändén for fruitful discussions. In addition, we are grateful to Drs. J.N. Jansonius, B. Erni and H. Kessler for their support and interest. This study was supported by Swiss National Science Foundation Grants 31-29795.90 (B. Erni) and 31-36432.92 (J.H. Jansonius) and by the Deutsche Forschungsgemeinschaft (DFG) and the Fonds der Chemischen Industrie.

References

[1] Levit, M. and Chothia, C. (1976) Nature 261, 552–558.
[2] Richardson, J.S. (1981) Adv. Prot. Chem. 34, 167–334.
[3] Chothia, C. (1992) Nature 357, 543–544.
[4] Bowie, J.U., Lüthy, R. and Eisenberg, D. (1991) Science 253, 164–170.
[5] Lüthy, R., McLachlan, A. and Eisenberg, D. (1991) Proteins 10, 229–239.
[6] Overington, J.P., Donnelly, D., Johnson, M.S., Sali, A. and Blundell, T.L. (1992) Protein Sci. 1, 216–226.
[7] Johnson, M.S., Overington, J.P. and Blundell, T.L. (1993) J. Mol. Biol. 231, 735–752.
[8] Erni, B., Zanolari, B., Graff, P. and Kocher, H.P. (1989) J. Biol. Chem. 264, 18733–18741.
[9] Seip, S., Balbach, J., Behrens, S., Kessler, H., Flükiger, K., deMeyer, R. and Erni, B. (1994) submitted for publication.
[10] Génovésio-Taverne, J.C., Sauder, U., Pauptit, R.A., Jansonius, J.N. and Erni, B. (1990) J. Mol. Biol. 216, 515–517.
[11] Stolz, B., Huber, M., Marković-Housley, Z. and Erni, B. (1993) J. Biol. Chem. 268, (in press).
[12] Brändén, C-I. (1980) Quart. Rev. Biophys. 13, 317–338.
[13] Brändén, C-I. and Tooze, J. (1991) in: Introduction to Protein Structure, Alpha/Beta Structures, vol. 4, pp. 43–57, Garland, New York, London.
[14] Erni, B., Zanolari, B. and Kocher, H.P. (1987) J. Biol. Chem. 262, 5238–5247.
[15] Gribksov, M., Luthy, R. and Eisenberg, D. (1990) Methods Enzymol. 183, 146–159.
[16] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[17] Deveraux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[18] Jones, T.A. (1985) Methods Enzymol. 115, 157–189.
[19] Insight II User Guide, version 2.2.0. San Diego: Biosym Technologies, 1993.
[20] Brünger, A.T., Clore, G.M., Gronenborn, A.M. and Karplus, M. (1987) Protein Eng. 1, 399–406.
[21] Chou, P.Y. and Fasman, G.D. (1974) Biochemistry 13, 222–245.
[22] Smith, W.W., Burnett, R.M., Darling, G.D. and Ludwig, M.L. (1977) J. Mol. Biol. 117, 195–225.
[23] Blundell, T.L. and Johnson, M.S. (1993) Protein Sci. 2, 877–883.