ELSEVIER

FEBS Letters 344 (1994) 191–195

FEBS 13998

FEBS LETTERS

A parallel three stranded α-helical bundle at the nucleation site of collagen triple-helix formation

Hans-Jürgen Hoppe[a,*], Paul N. Barlow[b], Kenneth B.M. Reid[a]

[a]*MRC Immunochemistry Unit and* [b]*Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK*

Received 30 March 1994

Abstract

A short stretch of 35 amino acids is identified as the structural motif responsible for the tight parallel association and trimerization of the three identical polypeptide chains of lung surfactant protein D, which contains both collagen regions and C-type lectin domains. This 'neck-region' is located at the nucleation site at which the collagenous sequences fold into a staggered triple-helix and is shown, by CD, NMR, and cross-linking of recombinant peptides, to consist of a triple-stranded parallel α-helical bundle in a non-staggered, and extremely strong, non-covalent association. This type of association between three polypeptide chains may represent a common structural feature immediately following the C-terminal end of the triple-helical region of collagenous proteins.

Key words: Collectin; Collagen; Lung surfactant protein; Coiled-coil; Triple-helix

1. Introduction

The biosynthesis of collagen molecules requires the correct alignment of three polypeptides consisting of Gly-Xaa-Yaa triplets to form the triple-helix [1]. Each chain assumes a left-handed helical structure in the right-handed triple-helix, which is stabilized by inter-chain hydrogen bonds. In the folding of collagens, the globular C-terminal regions, often involving inter-chain disulphide bridges [2], are considered to play a fundamental role in recognition and association of three polypeptide chains, resulting in registered nucleation of triple-helix formation, which then proceeds in a zipper-like fashion [3]. However, clear principles governing all of the requirements for collagen triple-helix formation have yet to be identified.

Refolding experiments on collagen type III indicated that specific inter-chain disulphide bridges formed between C-terminal globular protein structures, can be sufficient to function as a nucleus for the refolding of a triple-helix in vitro, whereas reduction abrogates this process completely [4]. However, the molecular mechanism guiding association and registered alignment of collagens has remained elusive since the family of proteins containing collagenous sequences is large and sequence comparison of the different types of C-terminal, non-collagen-like, regions did not reveal a common motif shared by FACITs (fibril associated collagens with interrupted triple-helix, types IX, XII, XIV, AND XVI), the collagens of striated fibrils (types I, II, III, V, and XI), or the collagens with C1q-like C-terminal domains (types VIII and X) [5]. The frequent formation of inter-chain disulphide bonds has further complicated the search for protein modules involved in the inter-chain association and subsequent nucleation of triple-helix formation.

The collectin family of collagenous proteins is composed of the serum proteins mannan-binding protein (MBP), collectin-43 and bovine conglutinin as well as the lung surfactant proteins SP-D and SP-A [6]. Each polypeptide chain in the collectins contains an N-terminal collagen-like region (of between 20 and 59 Gly-Xaa-Yaa triplets) linked, by a short stretch of 34–39 amino acids (which from the 'neck' region) to a C-terminal, C-type lectin domain (of 113–118 amino acids) (Fig. 1a). As no direct association between the lectin domains can be observed [7] we propose that, in the collectin family of proteins, the functions of inter-chain recognition, trimerization, and registered alignment of the three collagenous polypeptide chains is mediated by the 'neck-region', i.e. the short stretch of amino acids, which links the collagen region to the globular domains.

2. Materials and Methods

2.1. PCR, cloning and sequencing

A human lung λgt11 cDNA library (Clontech Labs Inc., Palo Alto, CA, USA), previously shown to contain the SP-D cDNA [8], wasused as a template for a polymerase chain reaction (1 μl/100 μl PCR). Two oligonucleotides were synthezised on the basis of the published cDNA sequence, engineering an *Eco*RI restriction site at each end of the coding region (5′-TAACAAAAAGAATTCTGCCATGCTG-3′ and 5′-GCCCCACCGAATTCAGTT GGCTCAG-3′). PCR reactions con-

*Corresponding author. Fax: (44) (865) 2757291.

SSDI 0014-5793(94)00383-7

taining 20 pmol of each primer, 0.25 mM dNTPs (Pharmacia), and 1 unit *Taq* polymerase (Promega) in the supplied reaction buffer, were performed with 28 cycles in a Perkin-Elmer/Cetus DNA Thermal Cycler (Buckinghamshire, UK), using a cycle of denaturation at 94°C for 1 min, annealing at 55°C for 30 s, and extension at 72°C for 2 min. The PCR product of the expected size was purified by electroelution after electrophoresis on a 1% Sea-Plaque (FMC Bioproducts, Vallensbaek, Denmark) agarose gel. The DNA fragment was digested with *Eco*RI (Promega) and ligated into the *Eco*RI site of pBluescript SK using T4 ligase (Amersham, Buckinghamshire, UK). The ligation mixture was transformed into competent cells of the *E. coli* strain NM554 using the $CaCl_2$ method, and plasmids isolated from single colonies were subjected to double-stranded sequencing using the T7 sequencing kit (Amersham). A 168 bp DNA fragment encoding residues 182–237 of human SP-D was generated using the unique *Sma*I and *Msc*I restriction sites and inserted into the *Sma*I restriction site of the bacterial expression vector pGex-2T [9]. The plasmid was transformed into the *E. coli* strain BL21.

Expression of the glutathione *S*-transferase neck-peptide fusion protein was achieved by adding 0.3 mM isopropyl-β-D-thiogalactopyranoside (NOVA Biochem) to the bacterial culture medium at OD of 0.8 at 600 nm. Cells were grown in 2 × TY medium at 37°C for 6 h, harvested by centrifugation, resuspended in buffer A (200 mM Tris-Cl, pH 7.8, 300 mM NaCl, 50 mM Na-EDTA), and lysed by mild sonication in the presence of 1% (v/v) Triton X-100. After centrifugation at 16,000 rpm for 30 min. the supernatant was passed over a glutathione-agarose (Sigma) affinity column, and after washing of the resin in buffer A the bound fusion protein was eluted with 10 mM glutathione (Sigma) in buffer A, followed by digestion with thrombin (Sigma) for 60 min at 37°C. The pH was then adjusted to 3.0 with citric acid. Precipitated glutathione *S*-transferase was removed by centrifugation at 15,000 rpm and the supernatant was subjected to anion exchange chromatography on a Pharmacia HighLoad-S FPLC column equilibrated at pH 3.0 in 100 mM Na-citratre buffer. The neck-peptide was further purified by reverse-phase chromatography on a C4-HPLC column. Fractions containing the neck-peptide were freeze-dried twice and easily dissolved in water. ^{15}N labelled peptides were generated by substituting the 2 × TY medium described above with M9 minimal medium containing $(^{15}NH_4)_2SO_4$ as the only nitrogen source.

2.3. Cross-linking and SDS-PAGE analysis

The peptide concentration was adjusted to 100 μg/ml in 100 mM sodium phosphate buffer, 100 mM NaCl, pH 7.4. Cross-linking wasperformed by adding bis-(sulfosuccinimidyl)suberate (Pierce Chemical Co., UK) to the samples (on ice). The samples were then incubated at 37°C for 20 min and the reactions were stopped by boiling in Tris-containing SDS-PAGE loading buffer. Aliquots of 25 μl were electrophoresed on a 15% Tris-tricine-glycerol SDS-polyacrylamide gel and stained with Coomassie blue R-250.

2.4. Circular dichroism and NMR

CD measurements were carried out on a JASCO 720 spectrometer equipped with a thermoelectric temperature controller. Curves were recorded digitally and fed through the data processor for signal-averaging and baseline subtraction. The peptide solution was adjusted to an OD of 1.0 at 210 nm (1 cm light path) and spectra were recorded at 25° C in 30 mM phosphate buffer pH 7.4 over 190–300 nm with a 0.1 mm path-length quartz cuvette. Cuvettes were allowed to equilibrate to the temperatures selected for 15 min. The heteronuclear single quantum coherence HSQC (^1H,^{15}N) NMR spectrum was recorded according to [10] on a sample of the neck-region peptide (~ 2 mM) that was isotopically labelled with ^{15}N, at 30°C, pH 4.5, in 90% H_2O/10% D_2O.

3. Results and discussion

The cDNA sequencing of human lung surfactant protein D has allowed the prediction of the 356 amino acids of the mature protein, and a 20 residue-long leader sequence [8]. Polymerase chain reaction, on a λgt11 cDNA library from human lung, generated a 1.2 kbp product corresponding to the complete coding region of the human SP-D cDNA. Sequencing confirmed the correct amplification of the coding region, without PCR errors (data not shown). Digestion with the restriction enzymes *Sma*I and *Msc*I generated a 168 bp DNA fragment, which encoded 7 Gly-Xaa-Yaa triplets and the 35 non-collagen like residues of the neck-region leading up to the C-type lectin domain. It was cloned into the pGEX-2T bacterial expression vector and the correct orientation of the insert was checked by restriction digestion. High levels of expression of the glutathione-*S*-transferase/neck-region peptide fusion-protein were obtained after 6 h of induction with IPTG. Thrombin digestion of the affinity purified fusion-protein resulted in two polypeptides, the glutathion-*S*-transferase and the neck-region peptide, carrying an additional Gly-Ser-Pro triplet at the N-terminus and the residues Gly-Ile-Pro-His-Arg-Asp at the C-terminal end, representing the polylinker present in pGex-2T (Fig. 1). 14 mg of the recombinant peptide were purified per litre culture in the three-step purification procedure. The peptide elutes in a single peak from the HighLoad-S column, at 800 mM NaCl, and minor contaminants are removed in the reverse-phase step. The purity of the peptide was confirmed by SDS-PAGE analysis, N-terminal sequencing of residues 1–46, and laser desorption mass spectroscopy (data not shown).

Using size exclusion chromatography, under non-dissociating conditions, the 65-residue-long peptide run as a single peak having an apparent molecular weight of 21–24 kDa. SDS-PAGE analysis showed single chain size of 6 kDa, however upon reaction with a cross-linking reagent, a single protein species of 21 kDa was detected when the reaction went to completion, while protein bands corresponding to 6, 13, and 21 kDa were seen in partially cross-linked reactions (Fig. 2). Higher oligomers were never seen. Thus, the region expressed is sufficient to form a trimer.

In order to determine if the 7 Gly-Xaa-Yaa triplets at the N-terminal third of the peptide made any contribution to the formation of the trimer, collagenase digestion was carried out, and the molecular weight of the resulting peptide was reduced to 4 kDa. It was shown, by N-terminal sequencing, that all the collagen triplets had been removed. This did not, however, reduce the ability of the remaining peptide to form stable trimers in solution.

Both peptides were also found to re-assemble into trimeric complexes even after heat-denaturation (98° C for 20 min in phosphate-buffered saline) and mixing varying portions of collagenase digested and intact neckpeptide followed by heat-denaturation and cooling resulted in hetero-trimerization to complexes in the expected stochiometric amounts (data not shown). Therefore, the C-terminal 35 residues were sufficient to mediate the stable non-covalent reversible association into trimeric complexes.

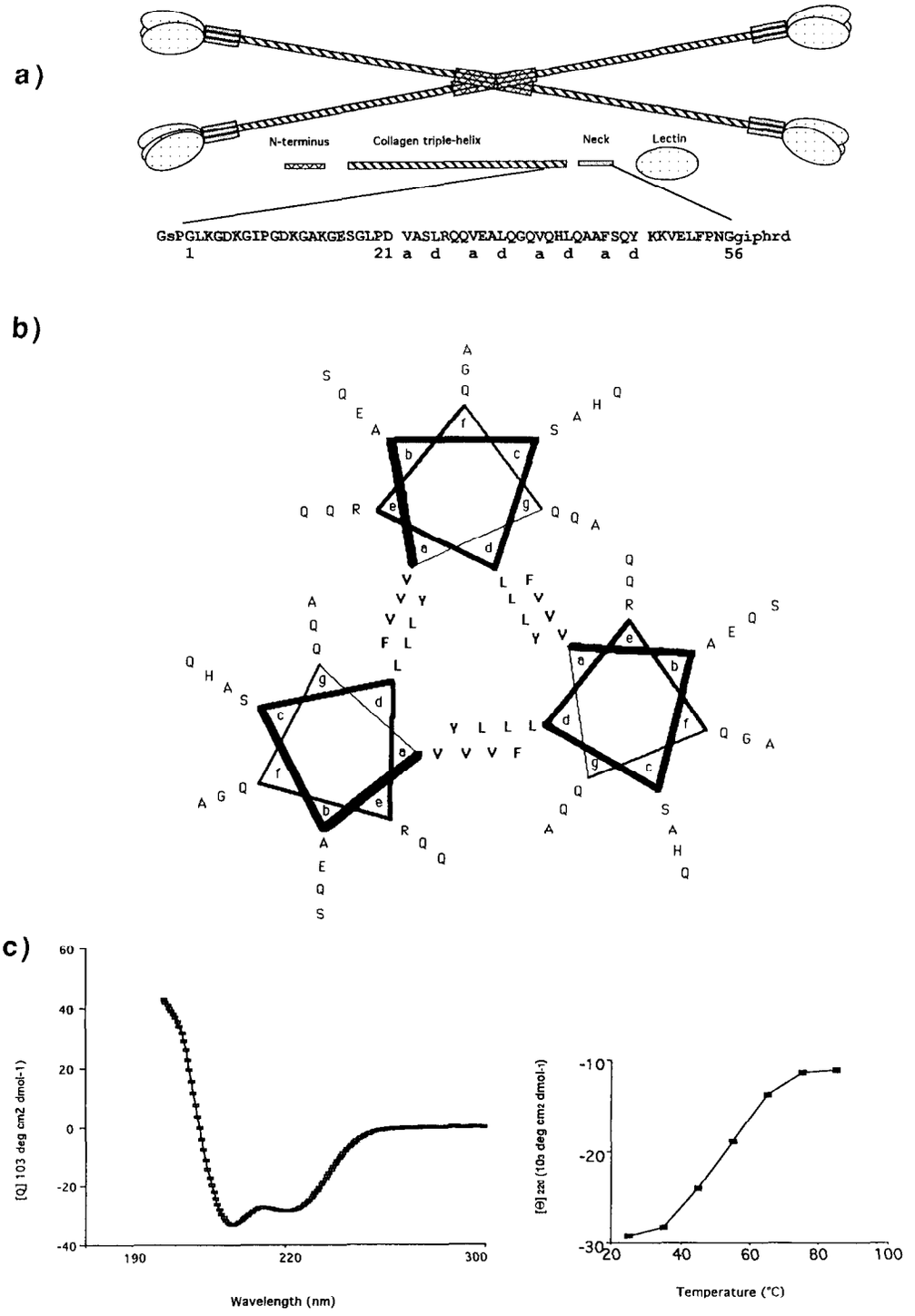

Fig. 1. (a) Schematic drawing of the location of the neck-region peptide. Human SP-D consists of 12 identical polypeptide chains (each of 356 amino acids) which assemble into 4 rod-like structures, each composed of three chains, which form triple-helical collagenous structures over residues 26–202. The C-terminal ends of the molecule contain C-type lectin domains linked to the collagenous domains via the neck-region, whereas the N-terminus is involved in oligomerization of the trimers into a tetramer. (b) Projection of residues 22–48 of the neck-region peptide onto a helical wheel. The view is down the helical axis starting at the C-terminal end of the helices from each of three neck-region peptide sequences. The predominance of valines in position d and leucines in position a reveals the hydrophobic interactions holding the helices together. (c) CD measurements on the collagenase digested neck-region peptide. The CD spectrum at 25°C indicates the presence of α-helical structure. The CD signal at 220 nm shows a thermal unfolding transition at 55°C.

To determine the secondary structure of the peptide, far-ultraviolet CD measurements were carried out on the collagenase-treated neck-peptide (Fig. 1c). The spectra show a strong positive value at 193 nm with two negative

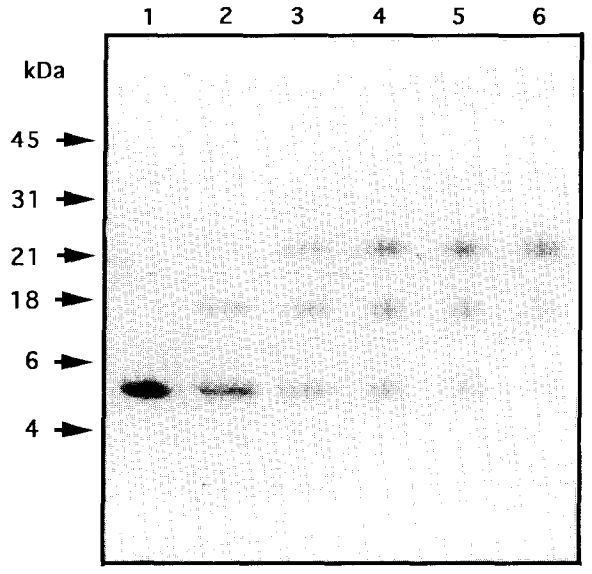

Fig. 2. Chemical cross-linking of the neck-region peptide. Cross-linking was performed by adding bis-(sulfosuccinimidyl)suberate to the samples (on ice) to yield final concentrations of 0 mM, lane 1; 2 mM, lane 2; 2 mM, lane 3; 3 mM, lane 4; 4 mM, lane 5; 5 mM, lane 6; 7 mM, lane 8. The samples were then incubated at 37°C for 20 min and the reactions were stopped by boiling in Tris-containing SDS-PAGE loading buffer. The samples (25 μl) were run on a 15% Tris-tricine-glycerol SDS-polyacrylamide gel and stained with Coomassie blue R-250. The position of the standards is indicated with arrows.

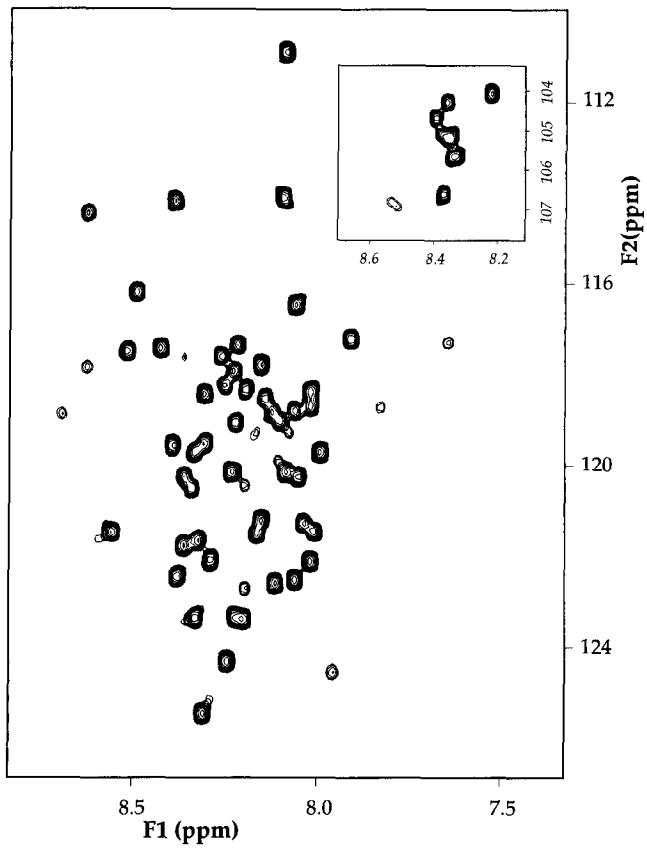

Fig. 3. A portion of the HSQC (^1H,^{15}N) NMR spectrum recorded on a sample of the neckregion peptide (~2 mM) that was isotopically labelled with ^{15}N, at 30°C, pH 4.5, in 90% H_2O/10% D_2O. The region contains cross-peaks between backbone nitrogens and attached protons (Crosspeaks in the inset correspond to glycine residues). The number of cross-peaks corresponds to the number of residues present in the peptide, consistent with a single magnetic environment for each residue.

values at 208 and 223 nm, consistent with the expected profile of α-helical structure [11]. The structure disappears reversibly with increasing temperature and a thermal unfolding transition at 55°C was observed.

As the location of the neck-region within the SP-D protein suggests a parallel orientation of the α-helices and the amino acid sequence of the peptide contains hydrophobic residues, in a repeating heptad pattern, the three α-helices could associate in a coiled-coil with the hydrophobic residues forming the interface between the helices (Fig. 1b) [12]. Also, the presence of valine residues at the a position and leucines at the d position of the heptad repeats in this homotrimer is consistent with observations by others [13] that demonstrated the involvement of β-branched side-chains at position a in the assembly of synthetic α-helical peptides into trimers rather than dimers or tetramers.

In order to obtain a complete structure determination of the neck-region peptide heteronuclear single quantum coherence (^1H,^{15}N) NMR spectra on ^{15}N-labelled peptide were collected and showed only one magnetic environment for each residue (Fig. 3). As the peptide exists as a trimer and as each residue within any one of the three α-helices shows the same magnetic environment as the corresponding residues in the other chains the structure of the α-helical bundle must have a 3-fold symmetry. Thus, the neck-peptide assumes the same oligomeric structure as the trimeric stalk of influenza hemagglutinin [14], but, unlike the virus stalk region peptide, the SP-D peptide formed a trimeric structure over a wide range of pH (3.0–9.5). The 3-fold symmetry observed proves the non-staggered and parallel association of the three helices and is contrasted by the staggered alignment of anti-parallel helices demonstrated recently for the spectrin molecule [15]. Surprisingly, therefore, the association of three right-handed α-helices in a parallel and non-staggered left-handed superhelix can serve as the nucleation site for the formation of a right-handed collagen superhelix of left-handed helices. As the three polypeptide chains involved are identical and as the collagen helix and the α-helical bundle are positioned in a direct junction this region of the SP-D molecule should contain a sharp bending of the peptide structure.

The tightly associated trimer of α-helices found at the neck-region of SP-D is, to our knowledge, the first example of a self-assembling structural motif, C-terminal to a collagen triple-helical structure, which does not involve the formation of disulphide bridges. Also, our findings demonstrate that, although collagenous sequences of re-

peating Gly-Xaa-Yaa triplets require additional protein sequences for inter-chain recognition, and association at their C-terminal ends, to initiate folding to an intact triple-helix, this association itself does not have to be in a staggered fashion in order to align the three chains in the correct register to form the staggered collagen helix. This may ease the search for structural motifs at the nucleation site of other proteins containing collagenous sequences. This particular motif might be useful in a model system to study the folding pathway of the collagen triple-helix in vitro and should allow the trimerization of any type of modular structure (e.g. Fv regions of antibodies) where stability and multivalency may be of importance.

Acknowledgements: We thank Tony Willis for peptide sequencing and Dr. S.K.A. Law for critical reading of the manuscript. H.-J.H. was a Rhodes Scholar and also wishes to thank the Hermann-und-Lilly-Schillingstiftung im Stifterverband für die deutsche Wissenschaft for generous personal support.

References

[1] Traub, W. and Piez, K.A. (1971) Adv. Protein Chem. 25, 243–352.
[2] Labourdette, L. and Rest, M.v.d. (1993) FEBS Lett. 320, 211–214.
[3] Engel, J. and Prockop, D.J. (1991) Annu. Rev. Biophys. Chem. 20, 137–152.
[4] Bèchinger, H.P., Bruckner, P., Timpl, R., Prockop, D.J. and Engel, J. (1980) Eur. J. Biochem. 106, 619–632.
[5] Bork, P. (1992) FEBS Lett. 307, 49–54.
[6] Holmskov, U., Malhotra, R., Sim, R.B. and Jensenius, J.C. (1994) Immunol. Today 15, 67–73.
[7] Weis, W.I., Kahn, R., Fourme, R., Drickamer, K. and Hendrickson, W.A. (1988) Science 254, 1608–1615.
[8] Lu, J., Willis, A.C. and Reid, K.B.M. (1993) Biochem. J. 284, 795–802.
[9] Smith, D.B. and Johnson, K.S. (1988) Gene 67, 31–40.
[10] Norwood, T.J., Boyd, J., Heritage, J.E., Soffe, N. and Campbell, I.D. (1990) J. Magn. Reson. 87, 9638–9644.
[11] Chen, Y.-H., Yang, J.T. and Chau, K.H. (1974) Biochemistry 13, 3350–3359.
[12] Pauling, L., Corey, R.B. (1953) Nature 171, 59–61.
[13] Harbury, P.B., Zhang, T., Kim, P.S. and Alber, T. (1993) Science 262, 1401–1407.
[14] Carr, C.M. and Kim, P.S. (1993) Cell 73, 823–832.
[15] Yan, Y., Winograd, E., Viel, A., Cronin, T., Harrison, S.C. and Branton, D. (1994) Science 262, 2027–2030.