The protein fold of the von Willebrand factor type A domain is predicted to be similar to the open twisted β-sheet flanked by α-helices found in human ras-p21

Yvonne J.K. Edwards, Stephen J. Perkins*

Department of Biochemistry and Molecular Biology, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2PF, UK

Received 5 December 1994

Abstract The von Willebrand Factor type A domain is the prototype for a protein superfamily. It possesses no significant sequence similarity to any known protein structure. Secondary structure predictions indicate a largely alternating pattern of six α-helices and six β-strands. A protein fold for this domain is proposed to correspond to a doubly-wound open twisted β-sheet structure flanked by α-helices. Close agreement was found with the GTP-binding domain of human ras-p21, provided that an extra α-helix was inserted. The structure of the predicted fold showed high compatibility with the proximate location of two Mg^{2+}-binding Asp residues, two disulphide-bridged Cys residues, and other known functional attributes of this domain.

Key words: von Willebrand factor; Doubly-wound α/β fold; ras-p21; Secondary structure prediction; Protein fold recognition

1. Introduction

The von Willebrand factor is a large glycoprotein found in blood plasma. Mutant forms are involved in the aetiology of bleeding disorders [1,2]. In von Willebrand factor, the type A domain (vWF) is the prototype for a protein superfamily containing at least 75 proteins that are similar in sequence [3,4]. This domain is found within other plasma proteins, at least four collagen types, several α-subunits of integrins, and other extracellular proteins. Fourier transform infrared spectroscopy on a recombinant vWF domain from human complement factor B revealed 31% α-helix and 36% β-sheet [4]. The calculation of an averaged secondary structure prediction from a sequence alignment offers improved statistical accuracies [5–7]. Such work lead to the assignment of the vWF fold to the α/β class of protein structures [4].

It is hypothesized that a protein crystal structure that matches this predicted secondary structure already exists in the Brookhaven Protein Data Bank. The prediction was compared against published topologies of known protein folds [8–10]. The vWF sequence compatibility with the structural features present in known three-dimensional structures was also assessed computationally using knowledge-based approaches [11–12]. We show that the GTP-binding fold of human ras-p21 [13,14] which contain a doubly wound open twisted β-sheet flanked by α-helices is readily identified as a good model for the vWF fold.

*Corresponding author. Fax: (44) (171) 794 9645.
E-mail: steve@rfhsm.ac.uk

Abbreviations: vWF domain, von Willebrand factor type A domain; EF-Tu, elongation factor Tu.

2. Materials and methods

The multiple sequence alignment for 75 vWF sequences was taken from [4]. Further alignments were generated by use of the multiple alignment program MULTAL with a range of fixed and variable gap penalties [15], and subsequently refined by eye. Alignments maximized the occurrence of conserved or chemically similar residues [4,5,16] and minimized the number of gaps. Protein coordinates were analysed for the location of their secondary structure elements using DSSP (Define Secondary Structure in Proteins) [17]. Protein structures were visualised using INSIGHT II (Biosym Technologies plc) or SETOR software [18].

3. Results and discussion

3.1. Averaged secondary structure predictions for 75 vWF sequences

The previous prediction of the averaged secondary structure from 75 aligned vWF sequences utilised the unbiased GOR I and Chou-Fasman methods [19,20]. This revealed a largely alternating sequence of α-helices (A) and β-strands (B), which was identified to be B1, A2, B3, AB4, A5, AB6, B7, A8, B9, A10, B11 and A12 in that order [4] (AB means that either α-helix or β-strand was present). AB4 and AB6 were oppositely assigned to α-helix and β-strand in the earlier predictions [4]. This resulted from closely similar probabilities for α-helix and β-strand conformations. More recent algorithms PHD and SAPIENS [6,7] were now employed in order to predict the consensus secondary structure directly from the sequence alignment and clarify the AB4 and AB6 predictions. Both assigned AB4 and AB6 with greater confidence to be B4 and A6, respectively. As B4 was one of the shortest β-strands, this may have influenced the difficulty in assignment. Six α-helices and six β-strands were predicted for vWF in the order BABBAABABABA. The first three and final seven elements showed an alternation of α-helices and β-strands.

3.2. Protein fold recognition from the vWF prediction

Sequence database searches failed to reveal significant sequence similarity at the >30% level with sequences other than those within the vWF superfamily. Protein fold recognition procedures were thus used to assess the compatibility of the vWF sequence and prediction with known protein structures. An alternating pattern of α-helix and β-strand as observed for B1–B3 and A6–A12 is characteristic of a repeating β–α–β motif [8]. Two β–α–β motifs can be connected by an α-helix to form a parallel β-sheet with four β-strands in two ways. If the β-strands are numbered sequentially, the 1–2–3–4 arrangement will produce a singly-wound closed β-sheet barrel as in triosephosphate isomerase. As no closed barrel with fewer than eight β-strands is known [10], this structure can be ruled out for

SSDI 0014-5793(94)01447-7

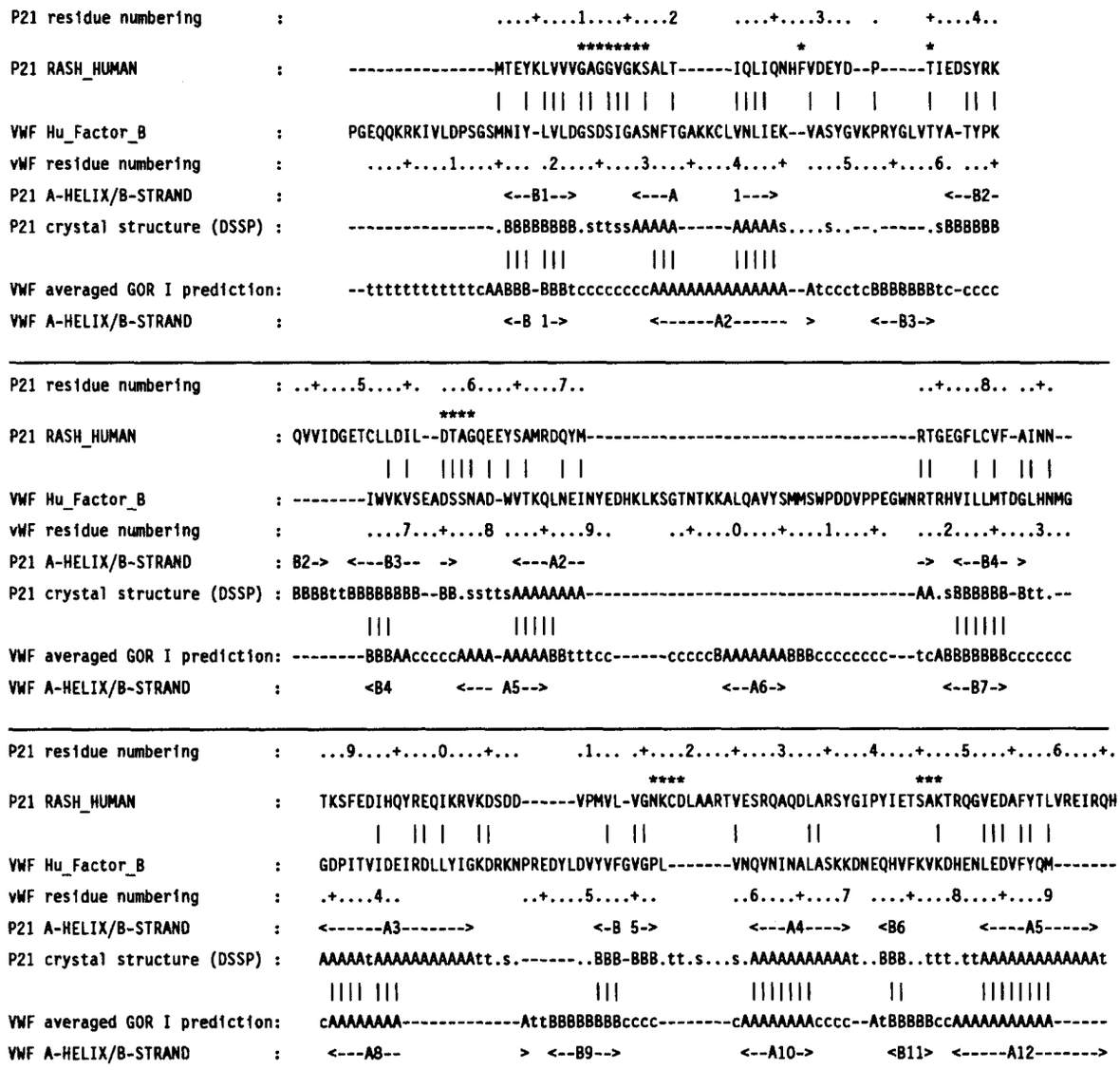

Fig. 1. Alignment of the human ras-p21 sequence with that for the vWF domain in human complement factor B to show the degree of correlation between the observed ras-p21 structure with that from the average of 75 GOR I predictions. Residue numberings from both ras-p21 and the consensus vWF sequence are shown. Similiar sequence and structure identities are denoted by vertical strokes. This gave 60 highly similar residue and 55 structural identities. Dashes indicate sequence deletions in ras-p21 or <50% occupancy in the 75 vWF sequences. Dots, s and t in the DSSP output identifies loop regions in ras-p21. The GTP-binding residues in ras-p21 are asterisked.

vWF. The 2–1–3–4 arrangement results in an open, twisted β-sheet, which is doubly wound. The minimum number of β-strands known for an open twisted β-sheet with α-helices on both sides of the β-sheet is four, and this is compatible with the vWF prediction.

The April 1992 Brookhaven databank contains 208 protein folds, of which 38 are doubly-wound α/β folds [10]. Four groups of single domain α/β folds were identified [10], namely ras-p21 and EF-Tu (5 α-helices and 6 β-strands in the order BAB-BABABABA), dihydrofolate reductase (4 α and 8 β; BABAB-BABABBB), adenylate kinase (10 α and 5 β; ABABAAABA-BAAABA) and phophoglycerate mutase (8 α and 5 β; BAABA-BAAAABAB). The ras-p21 group came the closest to the 6 α/6 β and BABBAABABABA predicted for vWF. Two protein fold recognition algorithms THREAD [11] and QSLAVE [12] were used to score sequence compatibility with a library of known protein folds. The vWF sequence of factor B was used as a search template. The α/β fold with the lowest pairwise and solvation pseudoenergy terms was identified by THREAD (Brookhaven codes and energy scores in brackets) to be ras-p21 (5p21; −399 kcal/mol), which was followed by thioredoxin reductase (1trb; −328 kcal/mol), guanylate kinase (1gky; −316 kcal/mol), D-ribose binding protein (1dri; −300 kcal/mol), and others. QSLAVE gave similar scores for dihydrofolate reductase, EF-Tu, adenylate kinase, ras-p21 and others. THREAD suggested a high similarity of the vWF sequence with the GTP-binding fold of human ras-p21.

3.3. *Comparison of the fold topology of ras-p21 with that predicted for vWF*

The alignment of the predicted vWF secondary structure against that observed for ras-p21 is shown in Fig. 1. The sec-

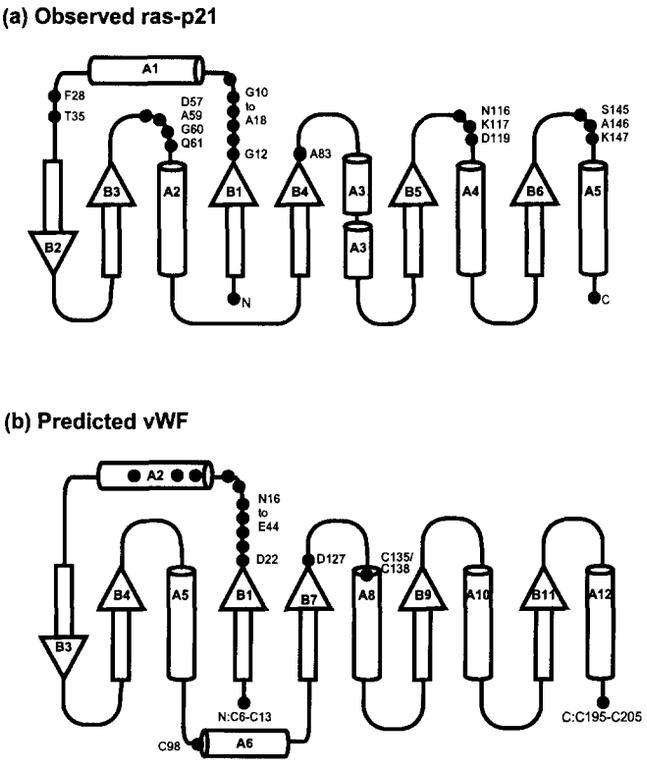

Fig. 2. Supersecondary structure topologies for the open twisted β-sheet and α-helices of ras-p21 and vWF. α-Helices are represented as cylinders and β-strands are shown as arrows. Loops connect consecutive elements of secondary structure. Both topologies contain the same number of β-strands, but the vWF domain contains an extra α-helix A6 which precedes β-strand B7 that forms part of the binding crevice with B1. (a) The secondary structure of ras-p21 shows the location (●) of residues on five loops at one side of the β-sheet that constitutes the GTP-binding site of ras-p21, the N- and C-termini, and G12 and A83 that are replaced by Asp in vWF. (b) The predicted topology for the vWF domain shows the location (●) of the GP Ib receptor recognition peptide in human vWF between residues N16 to E44, as well as those of Asp^{22} and Asp^{127}, the N- and C-terminal Cys residues, and the residues involved in the putative disulphide bridge between Cys^{98}–Cys^{135}/Cys^{8}.

ondary structure numbering follows [4,13]. MULTAL sequence alignment for 12 vWF and 12 ras-p21 sequences taken from [4,14] showed that B1, A1, B2, B3 and A2 of ras-p21 could be correlated with B1, A2, B3, B4 and A5 of vWF. THREAD showed that B4, A3, B5, A4, B6 and A5 of ras-p21 could be aligned with B7, A8, B9, A10, B11 and A12 of vWF. The combination of both analyses suggested that A6 is the extra α-helix in vWF that is not present in ras-p21. A5 is not favoured as the extra α-helix as it exhibited greater sequence similarities with A2 in ras-p21 in Fig. 1.

The topology of ras-p21 is compared with that proposed for vWF in Fig. 2. The vWF β-sheet is proposed to contain the same number of β-strands as ras-p21. A6 could be readily incorporated. The observed lengths of the α-helices (10–17 residues) and β-strands (3–10 residues) in ras-p21 agree to within 3 residues of the predicted lengths in vWF, with the exceptions of A2, B4 and A8 in vWF. The lengths of the connecting loops agreed to within 1–6 residues, except at the insertion point for A6 which is flanked by loops of 13 and 14 residues. These lengths are consistent with the proposed topology of Fig. 2.

3.4. Validation of the predicted vWF fold by molecular graphics

Almost all substrates or cofactor binding sites in a doubly wound α/β domain are located at a crevice lined by the loops that connect the central β-strands with the α-helices [8,9]. In the proposed vWF fold, Fig. 2b shows that such a crevice is formed between B1 and B7, where the loops from the carboxy edge of two adjacent β-strands lead into opposite faces of the β-sheet (Fig. 3). Interestingly Asp^{22} and Asp^{127} which are highly conserved in many vWF sequences [4] have been implicated in the physiologically relevant role of Mg^{2+} binding in the vWF of human CR3 [21]. These are located at the carboxy-ends of B1 and B7 in the crevice. Despite their separation in the sequence, Fig. 3 predicts that they are adjacent in three-dimensional space. The three other groups of doubly-wound α/β folds could not reproduce this proximity for reason of the connectivity of the α-helices and β-strands within their protein topologies.

In vWF, 77% of Cys residues occur before residue 16 or after residue 194 [4]. Cys^{509}–Cys^{695} and $Cys^{923}Cys^{1109}$ are disulphide bridges which enclose the A1 and A3 domains in human von Willebrand factor [22,23]. In the predicted vWF fold, the N- and C-termini are at the same side of the protein (Figs. 2 and 3). Sequence comparisons shows that Cys^{6}/Cys^{13} (vWF) precedes ras-p21 by approximately 6 amino acids, and likewise Cys^{195}/Cys^{204} (vWF) follows ras-p21. The distance between the α-carbons of the N- and C-termini in ras-p21 is 1.9 nm (Fig. 3). The predicted vWF topology is thus consistent with disulphide bridge formation between Cys^{6}/Cys^{13} and Cys^{195}/Cys^{204} when these are present.

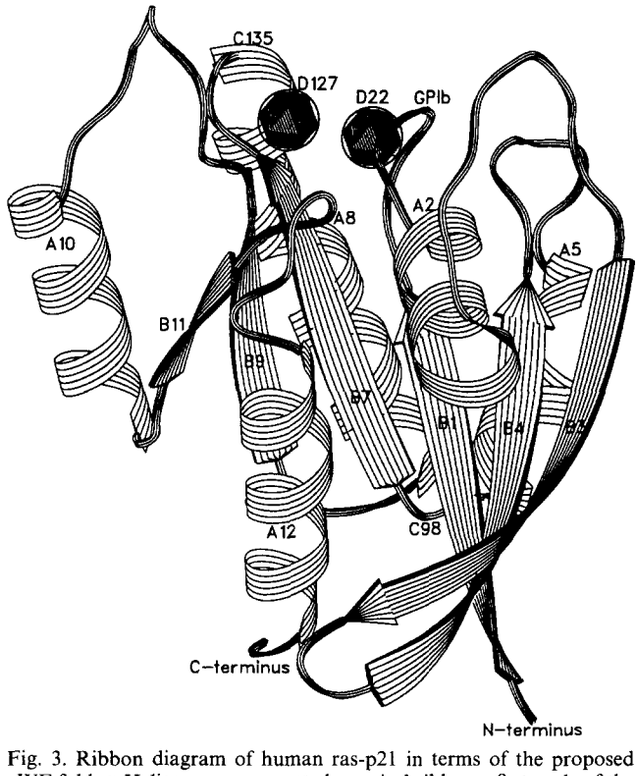

Fig. 3. Ribbon diagram of human ras-p21 in terms of the proposed vWF fold. α-Helices are represented as spiral ribbons, β-strands of the twisted β-sheet are shown as arrows, and loops are shown as ropes. The protein fold shows the open twisted β-sheet with six β-strands (B11, B9, B7, B1, B4 and B3 from left to right) which is flanked by five α-helices (A12 and A2 in front of the β-sheet, and A8 and A5 behind it). The positions of the GP Ib-binding peptide, Asp^{22}/Asp^{127}, the location of A6 at Cys^{98} (see text), and the N-and C-termini are labelled.

The remaining Cys residues are poorly conserved in vWF [4]. The only pair which appear and disappear simultaneously at equivalent positions in the vWF alignment occurs in 4 sequences at Cys^{98} and Cys^{135}/Cys^{138} in human and chicken CoVI α1-N1 and CoVI α2-N1. Interestingly, this involves the extra A6 helix, and is suggestive that A6 and A8 are disulphide linked (Fig. 2b). There are 34 residues between A5 and B7 in vWF (Fig. 1). This length is estimated to be sufficient for the polypeptide to extend towards A8 for disulphide bridge formation, then to pass back (Fig. 2b). Molecular graphics show that A6 can be inserted between A5 and A8 of vWF as required (Fig. 3).

The carboxyl-edge of the parallel β-sheet in ras-p21 (Fig. 2a) contains the GTP-binding site. Highly conserved residues in GTP-binding proteins include Phe^{28}, the NKxD motif at Lys^{117}, and the SAK motif at Lys^{147}, all of which interact with guanine [13,14]. Two of these correspond to deletions in the vWF sequence, while SAK becomes KVK (Fig. 1). Thr^{35} and the GxxxxGK(S/T) and DxxG motifs, which are variously involved in Mg^{2+}/phosphate interactions and switching between GDP and GTP binding, are also poorly conserved in the vWF sequences [4]. These observations are consistent with the non-binding of GTP to vWF. Interestingly, the GxxxxGK(S/T) site for GTP in Fig. 2a is replaced by a vWF sequence in Fig. 2b that has been implicated in vWF recognition by the human platelet GP Ib–IX complex [1,2,24].

The molecular basis for mutant forms of human von Willebrand factor in Type II von Willebrand Disease has been defined [1–3]. Type IIB disease is characterised by an increased affinity for GP Ib, and has been related to eight mutations involving residues M540, R543, R545, W550, V551, V553, P574 and R578 in the A1 domain of human von Willebrand factor. The first six are located in the loop between A2 and B3 on the carboxyl-edge of the β-sheet, while the last two relate to the end of the loop between B4 and A5 at the same edge. Fig. 2b shows that all eight residues occur in regions similar to the GTP site in ras-p21. Type IIA disease involves a decreased affinity for GP Ib, and involves mutations at S743, R834, V844, S850, I865 and E875. Those at S743, V844 and S850 are also located at the carboxy-edge of the β-sheet in Fig. 2b, while the others are distributed between A6 and B7 (R834) and the centres of B9 (I865) and A10 (E875). That involving R834 is particularly susceptible to proteolytic degradation [2], and is predicted to be on a surface-exposed loop in the vWF model which explains this susceptibility. While the situation with some of the Type IIA mutations is less clear-cut than that with Type IIB, the proposed vWF fold can be correlated with the increased affinity observed in Type IIB disease.

4. Conclusions

Of interest is that the structural similarity with ras-p21 has been identified for the vWF domain which is extracellular, even though ras-p21 is an intracellular protein. This predicted model is able to account for the location of known functional groups, disulphide bridge connectivities and functional binding sites in vWF. The vWF model will be of value for functional and structural studies, although it should be noted that an atomic resolution structure determination will be required to verify it.

Acknowledgements: We thank the Wellcome Trust and Royal Society for support. Drs. M.S. Johnson, D.T. Jones, W.R. Taylor and Professor T.L. Blundell are thanked for the use of their software. This study was presented at the 'Genes, Proteins and Computers III' Conference, September 1994 [25].

References

[1] Ruggeri, Z.M. and Ware, J. (1993) FASEB J. 7, 308–316.
[2] Sadler, J.E. (1991) J. Biol. Chem. 266, 22777–22780.
[3] Colombatti, A. and Bonaldo, P. (1991) Blood 77, 2305–2315.
[4] Perkins, S.J., Smith, K.F., Williams, S.C., Haris, P.I., Chapman, D. and Sim, R.B. (1994) J. Mol. Biol. 238, 104–119.
[5] Perkins, S.J., Haris, P.I., Sim, R.B. and Chapman, D. (1988) Biochemistry 27, 4004–4012.
[6] Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599.
[7] Wako, H. and Blundell, T.L. (1994) J. Mol. Biol. 238, 693–708.
[8] Brändén, C. and Tooze, J. (1991) Introduction to Protein Structure, Garland Press, New York.
[9] Richardson, J.S. (1981) Adv. Prot. Chem. 34, 167–339.
[10] Orengo, C.A., Flores, T.P., Taylor, W.R. and Thornton, J.M. (1993) Prot. Eng. 6, 485–500.
[11] Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) Nature 358, 86–89.
[12] Johnson, M.S., Overington, J.P. and Blundell, T.L. (1993) J. Mol. Biol. 231, 735–752.
[13] Pai, E.F., Kabsch, W., Krengel, U., Holmes, K.C., John, J. and Wittinghofer, A. (1989) Nature 341, 209–214.
[14] Valencia, A., Chardin, P., Whittinghofer, A. and Sander, C. (1991) Biochemistry 30, 4648–4654.
[15] Taylor, W.R. (1988) J. Mol. Evol. 28, 161–169.
[16] Taylor, W.R. (1986) J. Mol. Biol. 188, 233–258.
[17] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.
[18] Evans, S.V. (1993) J. Mol. Graphics 11, 134–138.
[19] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[20] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. Relat. Areas Mol. Biol. 47, 45–148.
[21] Michishita, M., Videm, V. and Arnaout, M.A. (1993) Cell 72, 857–867.
[22] Marti, T., Rösselet, S.J., Titani, K. and Walsh, K. (1987) Biochemistry 26, 8099–8109.
[23] Mohri, H., Fujimura, Y., Shima, M., Yoshioka, A., Houghten, R.A., Ruggeri, Z.M. and Zimmerman, T.S. (1988) J. Biol. Chem. 263, 17901–17904.
[24] Berndt, M.C., Ward, C.M., Booth, W.J., Castaldi, P.A., Mazurov, A.V. and Andrews, R.K. (1992) Biochemistry 31, 11144–11151.
[25] Edwards, Y.J.K. and Perkins, S.J. (1994) Genes, Proteins and Computers III: An International Conference on Bioinformatics,