Volume 221, number 1, 37-42 FEB 05031 August 1987

Secondary structure of the variant surface glycoproteins of trypanosomes

Fritz Jähnig, Roland Bülow, Theo Baltz* and Peter Overath

Max-Planck-Institut für Biologie, Corrensstrasse 38, 7400 Tübingen, FRG and **Laboratoire d'Immunologie et Biologie Parasitaire, Université Bordeaux II, 146 rue Léo Saignat, 33076 Bordeaux, France*

Received 28 May 1987

The secondary structure of seven variant surface glycoproteins (VSGs) of trypanosomes has been determined by Raman spectroscopy. They are all predominantly α-helical, the α-helix content varying between 50 and 60%. The β-strand content varies between 20 and 25%, and the content of β-turn and nonregular structures is about 25%. For three VSGs the N-terminal domain obtained by proteolytic cleavage was found to have essentially the same secondary structure as the complete VSGs. For three VSGs a secondary structure prediction has been performed applying the rules of Chou and Fasman. In all cases, two long α-helices extending over about 50 residues or 80 Å are predicted in agreement with the X-ray diffraction data of Freymann et al. [(1984) Nature 311, 167–169] and Metcalf et al. [(1987) Nature 325, 84–86]. The region between the two α-helical segments exhibits a high potential of β-turns, suggesting that this segment may be exposed on the cell surface and carry major antigenic determinants.

Variant surface glycoprotein; Raman spectroscopy; Structure prediction

1. INTRODUCTION

The pathogenic African trypanosomes, exemplified by *Trypanosoma brucei*, are covered by a surface coat that is 12–15 nm thick. For an individual cell, the coat consists of a single species of glycoprotein, the variant surface glycoprotein (VSG), coded by a specific gene from a large repertoire. Activation of different genes from this repertoire leads to the expression of serologically distinct VSGs at the cell surface, thereby allowing the parasite to escape the immune response (reviews [1–4]).

A common structural feature of all VSGs appears to be a glycan-phosphatidylinositol residue connected via ethanolamine [5] to the C-terminus which anchors the protein in the cytoplasmic membrane [6]. In a C-terminal domain comprising about one-third of the molecules, different VSGs show considerable sequence homology while there is little homology in the remaining N-terminal two-thirds [7]. Information on the secondary and tertiary structure of VSGs is limited. Circular dichroism spectra suggested that the α-helix content for several VSGs is relatively low, 28–38%, with the exception of one VSG with an α-helix content of 49% [8]. In contrast, Lalor et al. [9] concluded on the basis of structural predictions that all VSGs have a high potential to form α-helices. Because the overall distribution of segments of α-helical potential was highly variable, the three-dimensional folding pattern of different VSGs appeared unlikely to be the same. On the other hand, electron-microscopic [10], hydrodynamic [11] and X-ray studies [12,13] suggested that the few VSGs

Correspondence address: F. Jahnig, Max-Planck-Institut für Biologie, Corrensstrasse 38, 7400 Tübingen, FRG

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

so far examined have a similar, rod-like shape. In particular, the dimeric N-terminal domains of VSG MITat 1.2 and ILTat 1.24 are 100 Å long, rod-like molecules composed of two 80–90 Å long α-helices per subunit [12,13].

The Raman-spectroscopic experiments reported in this paper show that different VSGs share a similar secondary structure characterized by a high α-helix content. These results in conjunction with structure predictions and X-ray data [12,13] lead to the proposal of a general model for the folding of variant surface proteins.

2. MATERIALS AND METHODS

2.1. *Purification of VSGs and derived N-terminal fragments*

VSGs of the MITat serodeme were purified by DEAE-cellulose chromatography and isoelectric focusing [14]. For isolation of the tryptic N-terminal fragments [15], VSGs purified by DEAE chromatography were used. 10 mg MITat 1.2 was incubated for 10 min at 37°C with 0.5 mg trypsin, MITat 1.4 for 60 min at room temperature. The reaction was stopped by the addition of phenylmethylsulfonyl fluoride (final concentration 2 mM) and the sample was then diluted with the gel suspension. The N-terminal fragment of MITat 1.4 was electrofocused in a gradient from pH 5 to 9, the fragment of MITat 1.2 in a gradient from pH 3.5 to 10. After focusing at 5 W constant power at 10°C overnight VSG-containing regions were cut out of the gel and separated from ampholines by ultrafiltration using an Amicon PM10 membrane. The purified proteins were lyophilized, resuspended in water and transferred to glass capillaries.

VSGs of the BOTat serodeme were purified by affinity chromatography on concanavalin A-Sepharose [16] and the tryptic N-terminal fragment of BOTat 1 was obtained as described in [17].

2.2. *Raman measurements*

The Raman measurements and data analysis were performed as described [18]. Typically, 100 scans across the region of the amide I band were recorded. The fluorescence background was about 10-times higher than the amide I band, but relatively smooth so that it could be subtracted as a flat baseline. For the data analysis in terms of secondary structure we used the second method of Williams [19] and $k_{rank} = 3$ (in his notation). This implies that the error in the percentages of secondary structure classes is $\pm 5\%$ for a given protein. However, the error in the difference between the secondary structure of two proteins is only about $\pm 2\%$.

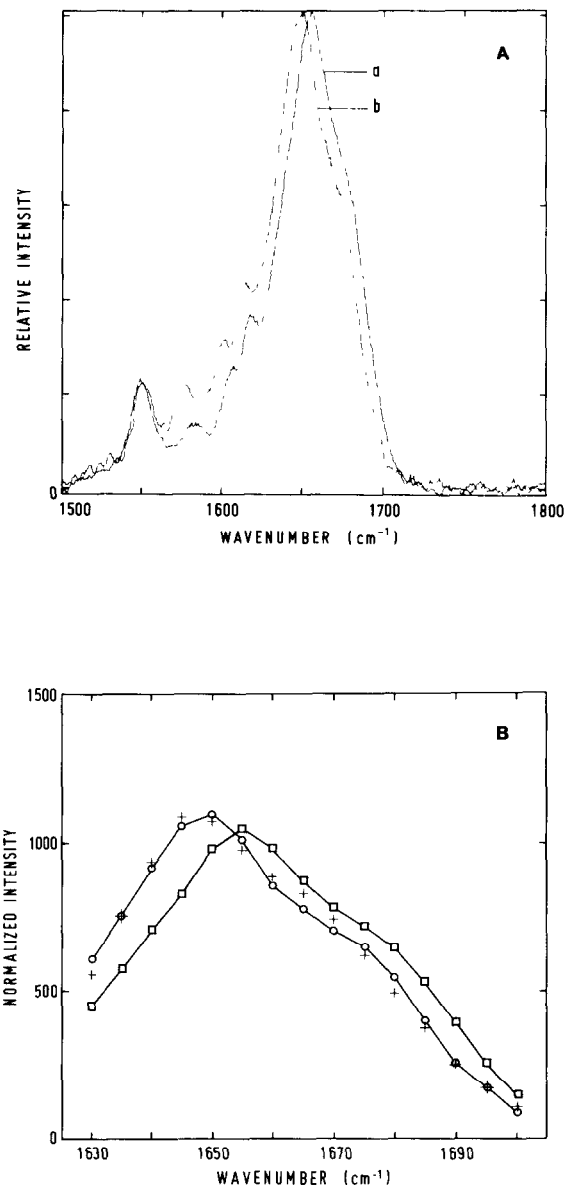

Fig.1. (A) Raman spectrum in the range 1500–1800 cm^{-1} of MITat 1.2 (a) and MITat 1.4 (b). (B) 15-point representation of the amide I band of MITat 1.2 (□) and MITat 1.4 (○) together with a fit of the latter (+).

2.3. *Structure predictions*

For the prediction of secondary structure the rules of Chou and Fasman [20] were employed. The hydrophobicity profile $H(i)$ was calculated according to Kyte and Doolittle [21] as an average of seven neighbors.

The amino acid sequence of MITat 1.4 was taken from Boothroyd et al. [22]. The sequences of MITat 1.2 and BOTat 1 were kindly provided by M.J. Turner (Merck Sharp and Dohme Research Laboratories, Rahway, NJ) and J.E. Donelson (Department of Biochemistry, University of Iowa, Iowa City), respectively.

Comparison of amino acid sequences was performed by using the program ALIGN [23].

3. RESULTS

3.1. *Raman spectroscopy*

The Raman spectra in the range 1500–1800 cm^{-1} of two VSGs, MITat 1.2 and MITat 1.4, are shown in fig.1A. It is evident that the amide I bands around 1650 cm^{-1} of the two proteins differ, whereas the tryptophan bands around 1550 cm^{-1} essentially coincide. The difference in the amide I bands indicates a different secondary structure of the two proteins. For the quantitative analysis, the amide I bands of the two proteins were fitted by a superposition of the amide I bands of 15 reference proteins of known three-dimensional structure (fig.1B). The result for the secondary structure is presented in table 1. MITat 1.2 has a total α-helix content of about 50% and a total β-strand content of 26%, whereas MITat 1.4 has a higher α-helix content of 60% and a lower β-strand content of 18%. The difference in the percentages of secondary structure classes between two independent preparations of MITat 1.4 was maximally 3% which leads to an error of $\pm 2\%$ due to sample preparation. This value is of the same order of magnitude as the error for the difference of the percentages of secondary structure of two proteins due to the data analysis. Hence, the difference in secondary structure between MITat 1.2 and MITat 1.4 is slightly larger than the total error from sample preparation and data analysis and, therefore, is probably significant.

In total, seven VSGs and the N-terminal segments of three of them were investigated by this technique (table 1). All VSGs are predominantly α-helical, their α-helix content varying from 49% for MITat 1.2 to 62% for BOTat 1. The two proteins MITat 1.2 and MITat 1.4 discussed above represent extreme cases. The N-terminal segments of

Table 1

Secondary structure of VSGs and derived amino-terminal segments determined from their Raman amide I spectra

	Molecular mass (kDa)	H_{tot} (%)	S_{tot} (%)	$T + U$ (%)
VSGs				
MITat 1.2 (221)	58	49	26	25
MITat 1.4 (117)	62	60	18	22
MITat 1.5 (118)		55	19	26
MITat 1.6 (121)		53	22	25
BOTat 1	58	62 (49)[a]	14 (25)[a]	24 (26)[a]
BOTat 28		55 (43)[a]	20 (29)[a]	25 (43)[a]
BOTat 201		55	20	25
N-terminal segments				
MITat 1.2	45	50 (50)[b]	25	25
MITat 1.4	48	61	15	24
BOTat 1	40	60	15	25

[a] Value in parentheses was obtained by CD measurements [8]

[b] Value in parentheses was estimated from X-ray diffraction data [12]

Molecular mass, as determined by SDS-polyacrylamide gel electrophoresis; H_{tot}, total α-helix content; S_{tot}, total β-strand content; T, β-turn content; U, undefined

VSGs have either a slightly lower or the same α-helix content as the complete VSGs.

3.2. *Structure predictions*

Fig.2 shows the structure prediction plots for the surface coat protein of variant MITat 1.4. The profiles of the α-helix potential P_α, β-strand potential P_β, and turn potential P_t (fig.2A) lead to a structure prediction as expressed by the function S (fig.2B): subscripts α, β and t designate prediction of α-helix, β-strand, and turn structure, respectively. The method predicts an α-helix content of 40% and a β-strand content of 18%. The predicted β-strand content agrees with the result from Raman spectroscopy, but the predicted α-helix content is much lower than the experimental value. This implies that predicted regions of turns and undefined structure must, in fact, be α-helical.

Two regions of α-helical structure extending over about 50 residues can be predicted: helix I from residue 58 to 103 and helix II from residue 245 to 288. In the middle of helix II the Chou-Fasman rules predict a short region of β-strand structure corresponding to six consecutive hydrophobic residues (^{266}Leu Leu Ala Val Leu Val^{271}). Accordingly, this region exhibits a very high hydrophobicity in the Kyte-Doolittle plot (fig.2D). Since in the Chou-Fasman analysis hydrophobic residues such as Val, Ile and Phe have a high potential for β-structure, any hydrophobic sequence is predicted to have β-structure. However, such residues may alternatively form a hydro-

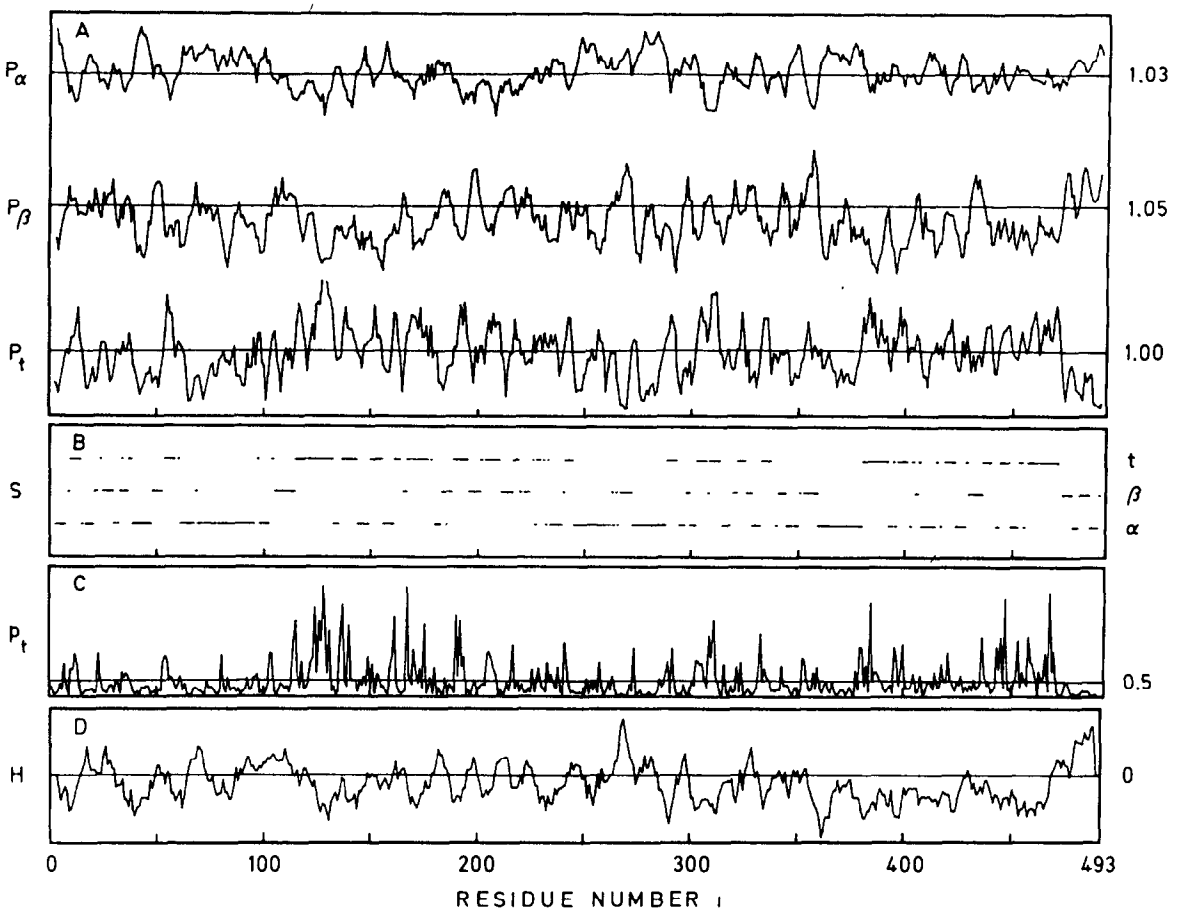

Fig.2. Structure-prediction plots for MITat 1.4. (A) The α-helix potential P_α, β-strand potential P_β and turn potential P_t, according to Chou and Fasman [20]. (B) Structural prediction of α-helix ($S = \alpha$), β-strand ($S = \beta$), or turn ($S = t$). (C) Position-dependent turn potential p_t, according to Chou and Fasman [20]. (D) Hydrophobicity H, according to Kyte and Doolittle [21].

phobic or amphipathic α-helix. For example, the hemagglutinin of influenza virus has a long α-helix of about 50 residues as determined by X-ray diffraction [24], which includes a segment of five hydrophobic residues. The Chou-Fasman analysis wrongly predicts β-structure for this region. Analogously, the VSG segment around residue 270 may actually be part of a long α-helix.

The plot of the position-dependent turn potential p_t (fig.2C) predicts turns at the ends of each long α-helix. This plot, furthermore, shows that turns occur with high frequency in the region between residues 100 and 200.

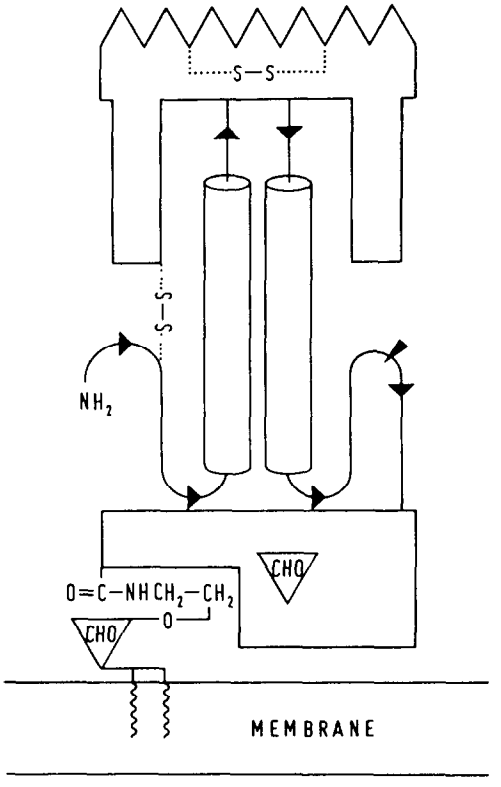

Fig.3. Schematic model for the folding of VSGs as exemplified for MITat 1.4. Two long α-helices are represented as cylinders, the zig-zag line on top indicating the region of predominant turn structure supposed to provide the major antigenic determinants. The long arrow indicates the major tryptic cleavage site specifying the N- and C-terminal fragments; CHO denotes carbohydrate chains linked to the protein or as part of the C-terminal glycolipid which anchors the protein in the cytoplasmic membrane.

Structure predictions were also performed for MITat 1.2 and BOTat 1 (not shown). For both proteins, two α-helices, about 50 residues in length, can be predicted. For MITat 1.2, they extend from residues 26 to 75 and 228 to 267, and for BOTat 1 from residues 23 to 77 and 277 to 330.

3.3. *Homology determination*

The three VSGs MITat 1.2, MITat 1.4, and BOTat 1 show some amino acid sequence homology in a domain of about 100 residues at the C-terminus. In the remaining N-terminal region of about 350 residues essentially no homology is found, except for four conserved cysteine residues at positions 15, 121, 140 and 182 (numbers for MITat 1.4 with slight alteration for the other two proteins) and two conserved prolines at positions 137 and 157. For MITat 1.4, disulfide bridges have been shown to connect residues 15 with 140 and 121 with 182 [25]. Thus, the conservation of these residues suggests conservation of S-S bridges which together with the conservation of the proline residues is indicative of a similar tertiary structure of the VSGs.

4. DISCUSSION

The results of our Raman measurements show that the VSGs of trypanosomes comprise a family of proteins with similar secondary structure in spite of the fact that, apart from a segment of about 100 amino acids at the C-terminus, their primary sequences show littly homology. Their predominant structural motif is the α-helix, in contrast to the results of Duvillier et al. [8]. A high α-helix content is also expected by the application of structure-prediction methods. This has been recognized previously [9].

Another characteristic feature of VSGs is the conservation of the pattern of cysteine residues, not only in the C-terminal segment but also in the N-terminal segment. Although the intramolecular linkage of these cysteine residues to form disulfide bridges has been determined only for MITat 1.4 [25], they may be of decisive importance for a similar three-dimensional folding of VSGs.

In an attempt to propose general features of VSG structure by combining the experimental results and the structure predictions, we have been guided by the X-ray result for MITat 1.2 of

Freymann et al. [12]. The salient point of this structure is a bundle of four long α-helices in the dimeric N-terminal domain. Bundles of α-helices are likewise found in other surface proteins such as hemagglutinin [24]. Fig.3 shows a schematic model for the folding of VSGs as exemplified for MITat 1.4. The two cylinders are α-helical segments which are predicted from the primary sequence. The approx. 100 residues between the two helices may form the globular head of VSGs. A globular domain at the end of a rod-like structure has been demonstrated by electron microscopy [10]. This segment between the two α-helices is rich in β-turn potential and thus well suited to provide the immunological epitopes. Recently, an epitope of lysozyme was found to be formed by β-turns creating a planar surface region [26]. It is proposed that the two α-helices provide a common structural motif for the three-dimensional structure of all VSGs although the peptide segments forming these α-helices will vary both in sequence and in the relative location along the chain. The helices may form a rigid core around, above and below which the rest of the polypeptide chain as well as the carbohydrate is arranged. The α-helices will dictate an elongated shape to the entire molecule which may be advantageous for lateral packing in the surface coat.

Even if the overall shape of different VSGs is similar, their detailed surface topology may nevertheless not always allow a sufficiently dense packing at the cell surface to exclude access of lytic high molecular mass components to the cell membrane. Therefore, it is expected that in the mammal trypanosomes with a porous coat due to improper packing of VSGs will be readily eliminated. However, such cells may well be viable in vitro. This leads to the prediction of a class of variants which can be propagated in vitro but are not infective to the mammalian host.

ACKNOWLEDGEMENTS

We thank Klaus Dornmair for many useful discussions on the Raman data, Isolde Riede for help with the program ALIGN, Drs M. Turner and J.E. Donelson for sequence data, and the Fonds der Chemischen Industrie for support.

REFERENCES

[1] Cross, G.A.M. (1984) Phil. Trans. R. Soc. Lond. B 307, 3-12.
[2] Turner, M.J., Cardoso de Almeida, M.L., Gurnett, A.M., Raper, J. and Ward, J. (1985) Curr. Top. Microbiol. Immunol. 117, 23-55.
[3] Vickerman, K. (1985) Br. Med. Bull. 41, 105-114.
[4] Borst, P. (1986) Annu. Rev. Biochem. 55, 701-732.
[5] Holder, A.A. (1983) Biochem. J. 209, 261-262.
[6] Ferguson, M.A.J., Haldar, K. and Cross, G.A.M. (1985) J. Biol. Chem. 260, 4963-4968.
[7] Rice-Ficht, A.C., Chen, K.K. and Dolenson, J.E. (1981) Nature 294, 53-57.
[8] Duvillier, G., Aubert, J.P., Baltz, T., Richet, C. and Degand, P. (1983) Biochem. Biophys. Res. Commun. 110, 491-498.
[9] Lalor, T.M., Kjeldgaard, M., Shimamoto, G.T., Strickler, J.E., Konigsberg, W.H. and Richards, F.F. (1984) Proc. Natl. Acad. Sci. USA 81, 998-1002.
[10] Cohen, C., Reinhardt, B., Parry, D.A.D., Roelants, G.E., Hirsch, W. and Kanwé, B. (1984) Nature 311, 169-171.
[11] Gurnett, A.M., Raper, J. and Turner, M.J. (1986) Mol. Biochem. Parasitol. 18, 141-153.
[12] Freymann, D.M., Metcalf, P., Turner, M. and Wiley, D.C. (1984) Nature 311, 167-169.
[13] Metcalf, P., Blum, M., Freymann, D., Turner, M. and Wiley, D.C. (1987) Nature 325, 84-86.
[14] Cross, G.A.M. (1984) J. Cell. Biochem. 24, 79-90.
[15] Johnson, J.G. and Cross, G.A.M. (1979) Biochem. J. 178, 689-697.
[16] Baltz, T., Baltz, D. and Pautrizel, R. (1976) Ann. Immunol. (Inst. Pasteur) 127C, 761-774.
[17] Baltz, T., Duvillier, G., Giroud, C., Richet, C., Baltz, D. and Degand, P. (1983) FEBS Lett. 158, 174-178.
[18] Vogel, H., Wright, J.K. and Jahnig, F. (1985) EMBO J. 4, 3625-3631.
[19] Williams, R.W. (1983) J. Mol. Biol. 166, 581-603.
[20] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251-276.
[21] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
[22] Boothroyd, J.C., Paynter, C.A., Coleman, S.L. and Cross, G.A.M. (1982) J. Mol. Biol. 157, 547-556.
[23] Dayhoff, M.O., Barker, W.C. and Hunt, T.L. (1983) Methods Enzymol. 91, 534-545.
[24] Wilson, I.A., Skehel, J.J. and Wiley, D.C. (1981) Nature 289, 366-373.
[25] Allen, G. and Gurnett, L.P. (1983) Biochem. J. 209, 481-487.
[26] Amit, A.G., Mariuzza, R.A., Phillips, S.E.V. and Poljak, R.J. (1985) Nature 313, 156-158.