

The serum albumin-binding domain of streptococcal protein G is a three-helical bundle: a heteronuclear NMR study

Per J. Kraulis^{a,*}, Per Jonasson^b, Per-Åke Nygren^b, Mathias Uhlén², Lena Jendeberg^{a,b}, Björn Nilsson^a, Johan Kördel^a

^aDepartment of Structural Biochemistry, Preclinical R&D, Pharmacia Biopharmaceuticals, S-112 87 Stockholm, Sweden

^bDepartment of Biochemistry and Biotechnology, Royal Institute of Technology, S-100 44 Stockholm, Sweden

Received 26 September 1995; revised version received 22 November 1995

Abstract Streptococcal protein G (SPG) is a cell surface receptor protein with a multiple domain structure containing tandem repeats of serum albumin-binding domains (ABD) and immunoglobulin-binding domains (IgBD). In this paper, we have analysed the fold of ABD. Far-UV circular dichroism analysis of ABD indicates high helical content (56%). Based on an analysis of nuclear magnetic resonance ¹³C secondary chemical shifts, sequential and short-range NOEs, and a few key nuclear Overhauser effects, we conclude that the ABD is a three-helix bundle. The structure of the ABD is, thus, quite different from the IgBD of protein G [Gronenborn, A.M. et al. (1991) *Science* 253, 657–661]. This strongly suggests that the ABD and the IgBD of SPG have evolved independently from each other. However, the fold of ABD is similar to that of the IgBD of staphylococcal protein A, possibly indicating a common evolutionary ancestor, despite the lack of sequence homology.

Key words: Protein G (*Streptococcus*); Serum albumin; Secondary structure; Global fold; Three-helix bundle; Nuclear magnetic resonance

1. Introduction

Protein G (SPG) is a multidomain protein present on the cell surface of *Streptococcus* of groups C and G. It displays binding to immunoglobulins from several species by interaction with the constant Fc region of IgG or the CH1 domain of certain Fab fragments [1], which is thought to help the organism evade the host defenses by creating a coat of host proteins [2]. A second activity of SPG is the binding to serum albumin from various species, including human serum albumin (HSA) [3,4]. The immunoglobulin and serum albumin-binding activities have been shown to reside within small and functionally separate domains of the SPG molecule [5,6] (Fig. 1a). Other bacterial surface proteins that bind immunoglobulins include protein A (SPA) from *Staphylococcus aureus* [7,8] and protein L from *Peptostreptococcus magnus* which binds to immunoglobulin light chains [9], while the PAB protein from *P. magnus* is another example of a serum albumin-binding protein [10].

Depending on the *Streptococcus* strain, SPG contains two or three immunoglobulin-binding domains of approximate 55 residues, denoted B1 through B3 or C1 through C3 by different authors [11–13] (Fig. 1a). 3D structures have been determined for single immunoglobulin-binding domains by NMR [14,15] and by X-ray crystallography [16,17]. In addition, also the

structures of SPG immunoglobulin-binding domains in complex with Fab [18] or Fc [19] have been described. The structure of an individual domain consists of a mixed parallel/antiparallel β -sheet of four strands, with a single α -helix packed onto the sheet. In contrast, NMR studies have shown that a single SPA-derived immunoglobulin-binding domain consists of three helices in a roughly antiparallel arrangement [20,21]. A medium-resolution X-ray determination of domain B of SPA in complex with the constant region of immunoglobulin G showed the presence of only the two N-terminal helices [22].

Since the bacterial receptors consist of multiple repetitive domains it is, from an evolutionary, biological and biochemical point of view, interesting to compare the structure of a serum albumin-binding domain with the known structure of the immunoglobulin-binding domains present in the same SPG receptor molecule. Here, we report the assignment of the NMR ¹H, ¹⁵N and ¹³C backbone resonances and secondary structure, as well as a tentative model of the global fold, of one of the three serum albumin-binding motifs present in SPG, denoted ABD for albumin-binding domain. Surprisingly, the fold of ABD is quite different from that of the SPG immunoglobulin-binding domain.

2. Materials and methods

2.1. DNA constructions and bacterial strains

All DNA manipulations were done essentially as described by Sambrook and coworkers [23]. PCR amplifications were performed on a Techne PHC-II thermocycler (Techne, UK) using standard protocols and buffers [24]. The cloning work was done using *E. coli* strain RR1ΔM15 [25] as host while the production of the ¹⁵N/¹³C-labelled protein was done in strain RV308 [26].

A DNA fragment encoding a 46-residue serum albumin-binding domain (ABD) derived from streptococcal protein G was PCR amplified from the plasmid template pTrpBB [27] using oligonucleotides MI-13: 5'-CCGAATTCATTAGC TGAAGCTAAAGTC-3' and TCS-2: 5'-CCGGATCCACTAGTCTCGAGTTAAGGTA ATGCAGCTAAA-ATTTC-3'. The PCR product was first inserted into pRIT28 adapted for solid phase DNA sequencing [24]. By the use of flanking *Eco*R1 and *Xho*I restriction sites, the ABD-encoding fragment was subsequently inserted into the expression vector pE318 resulting in pEABD, used for secreted expression and functional analysis. The pE318 vector is a derivative of pEZZ18 [28] in that the two IgG-binding Z domains are deleted through excision using flanking *Acl*I restriction sites, followed by religation of the vector. Subsequently, the ABD-encoding gene was excised from pEABD with *Fsp*I and *Xho*I followed by insertion into the *Fsp*I and *Sal*I restricted expression vector pTrpBBT1T2 constructed from pTrpBB by insertion into the *Pst*I and *Hind*III sites of the strong *rrnB* ribosomal RNA transcription terminator PCR amplified from pKK223-3 (Pharmacia Biotech, Sweden) using oligonucleotides HEAN19: 5'-CCCCCTGCAGCTCGAGCGCTTAAGTGT TTTGG-CGGATG-3' and HEAN20: 5'-CCCCAAGCTTAGAGTTTGTAG-AAACGC-3'. The resulting expression vector pTrpABDT1T2, thus, encodes a fusion protein containing a *trp* operon-derived leader

*Corresponding author. Fax: (46) (8) 6954084.

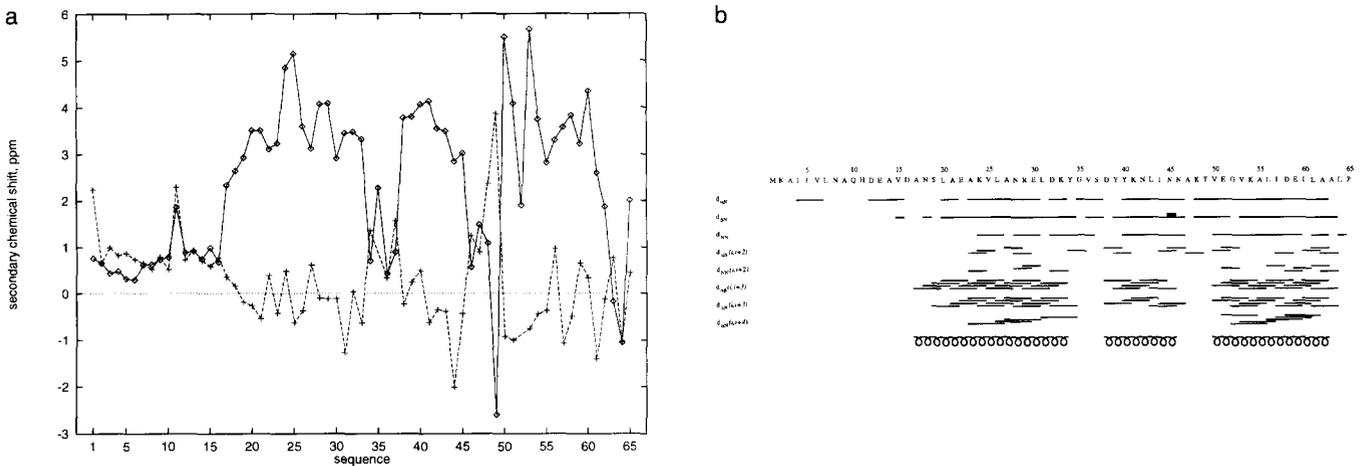


Fig. 3. (a) Secondary chemical shifts in ABD for $^{13}\text{C}^\alpha$ (solid line) and $^{13}\text{C}^\beta$ (dashed line) as a function of the sequence. The secondary chemical shift plots indicate the presence of three α -helical segments separated by coil regions. The first 16 residues have values indicating random coil. (b) Summary of the sequential and short-range NOE connectivities observed in the simultaneous ^{13}C - and ^{15}N -edited 3D NOESY spectrum. The schematic coils at the bottom of the figure depict the locations of the three α -helices in the sequence.

$^{15}\text{N}/^{13}\text{C}$ -double-labelled protein sample was used for this investigation. The measured atomic mass of the protein was 7497.0 Da, which indicates that the incorporation of ^{15}N and ^{13}C was in excess of 98.0%. This allows the use of multidimensional heteronuclear NMR experiments to increase the spectral resolution. The NMR resonances were assigned using a minimal set of triple-resonance spectra for a problem of this size.

The resonance assignments of the backbone and $^{13}\text{C}^\beta$ nuclei were obtained by performing a sequential walk in the HNCACB and CACB(CO)NH spectra. The assignment walk was started at several different spin systems exhibiting characteristic combinations of $^{13}\text{C}^\alpha/^{13}\text{C}^\beta$ chemical shifts. The placement of the sequential fragments in the sequence of the protein could usually be made unambiguously when 3 or 4 spin systems had been connected [43]. Occasionally, reference was made to the 2D HSQC spectra (^1H - ^{15}N and ^1H - ^{13}C) and to the HNCB and CBCA(CO)HA spectra to interpret regions with severe overlap of cross-peaks. Aliasing was used in some 3D spectra to optimize the digital resolution, and was resolved by reference to the 2D HSQC spectra and by comparison to typical chemical shifts in amino-acid residues [44]. The backbone assignments were subsequently confirmed by analysis of the HCC(CO)NH-TOCSY and CC(CO)NH-TOCSY spectra, which gave nearly complete assignments of all ^1H , ^{15}N and ^{13}C nuclei in the side-chains. Almost all cross-peaks in all spectra were subsequently assigned to ensure consistency. The HN resonances of residues Met-1 and Lys-2 could not be assigned, presumably because the signals are attenuated by solvent exchange effects. The resonance assignments are available upon request.

It has been shown that the ^{13}C chemical shifts of C^α and C^β nuclei correlate with the conformation of the backbone in proteins [45]. When the random-coil values of the C^α and C^β nuclei for each amino-acid residue type [46] are subtracted from the chemical shifts of these nuclei in the residues of the sequence, the remainder (the secondary chemical shift) shows a characteristic pattern which allows an accurate delineation of the secondary structure in terms of α -helix, β -strand or random coil [47]. To obtain further data on the secondary structure, the simultaneous ^{15}N - and ^{13}C -edited 3D NOESY was investigated for the

presence of sequential and short-range NOEs indicative of secondary structure.

The secondary chemical shift as a function of the sequence is shown for ABD in Fig. 3a. The sequential and short-range NOEs are shown in a summary in Fig 3b. This data shows unambiguously the presence of three helices separated by turns or random-coil regions. The first 16 or 17 residues are in a random-coil structure, or more likely, are flexible in solution. This part of the sequence is not present in the protein G sequence, but is rather derived from the cloning construct. The first 16 residues in the protein studied here are the same as in the F30A mutant [41] of the Z domain of staphylococcal protein A [42]. The chemical shifts for these residues are essentially identical in the two proteins (data not shown), and the lack of sequential and short-range NOEs, supports the conclusion that this N-terminal sequence is flexible in both proteins. The first helix starts at Ala-17 or Asn-18 and ends with Lys-33. The protein G sequence starts at Leu-20. The second helix begins with Asp-38 and ends with Asn-45. The third helix spans Val-50 to Ala-62. It is possible that helix 3 continues beyond residue Ala-62 in the intact protein, since fraying effects at the C-terminal of the construct studied here may cause artifacts. The sharp reversal of the trend of the secondary chemical shift at residue Thr-49 indicates that this residue probably forms a so-called capping box for helix 3 [48], i.e. the main-chain amide group of residue Gly-52 is hydrogen-bonded to the hydroxyl group of Thr-49. Although the typical N-cap contains a Thr or Ser at position i and a Glu or Gln at residue $i+3$, in ABD the residue at $i+3$ is Gly, which is conserved among the different sequences, except in the sequence from ALB8 which does have a Glu at this position (Fig. 1b). The region 47 Ala to 51 Glu (the AKTVE motif) is strictly conserved in the homologous sequences (Fig. 1b) and this region forms most of the loop connecting the second and third helices.

The preliminary analysis of the simultaneous ^{13}C - and ^{15}N -edited 3D NOESY spectrum showed the presence of two sets of well-resolved, easily identified NOEs which give some information on the relative arrangement of the three α -helices. The first set involves NOEs between protons in the side-chain of

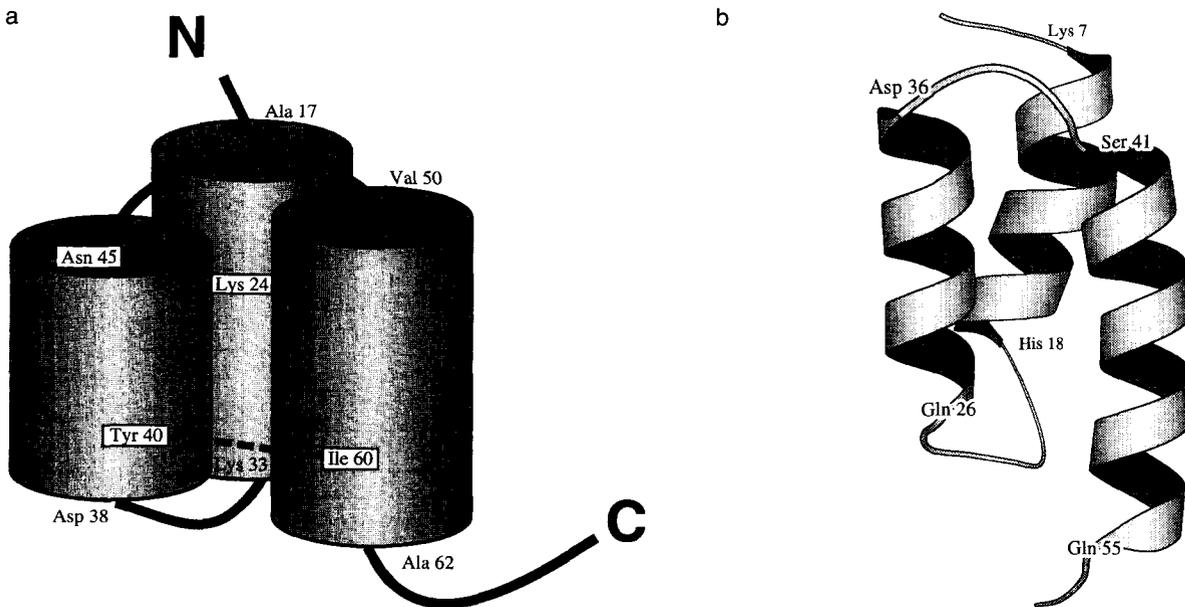


Fig. 4. Schematic comparison of the structural folds of (a) the albumin-binding domain of streptococcal protein G and (b) domain Z of staphylococcal protein A [42]. The relative arrangement of the α -helices in the ABD model is based on a few key NOEs and is, therefore, approximate. The overall chirality of the fold is not determined by the available data. The figure was generated using MolScript [52].

Lys-24 and the backbone of Asn-45, while the second set involves NOEs between protons in the aromatic ring of Tyr-40 and the side-chain of Ile-60. The restraints given by these NOEs were used to construct a schematic model (Fig. 4a). The data is consistent with a roughly antiparallel bundle of α -helices. This model only depicts a likely arrangement of the α -helices, since the available data does not allow any statements about the exact relative angles and positions of the α -helices, nor about the overall chirality of the fold.

The fold is clearly different from the IgBD of protein G, which has a β -sheet of four strands with a single α -helix packed onto it. However, a three-helix bundle similar to that in ABD is present in the IgBD of staphylococcal protein A [20,21]. A comparison of the amino-acid sequences of ABD with the immunoglobulin-binding domain of staphylococcal protein A indicated no detectable significant sequence homology, not even when attempting to aid the sequence comparison with the positions of the helices. As is evident from Fig. 4, there is, however, a clear structural similarity, although the lengths of the helices are somewhat different.

The ABD structure is unusual in that it is a small folded domain of only about 45 residues, which has no additional stabilising features such as bound ligands, metal ions, disulfide bridges, or other cross-links. There are few examples of such small folded domains, the IgBD of protein A being one. The three-helix bundles are, therefore, attractive model systems for experimental as well as theoretical folding studies [49].

It is intriguing that the HSA-binding domain of SPG has a similar fold as the IgBD of protein A from *S. aureus*, while the IgBD of protein G has a different fold. The two IgBDs represent different structural solutions to the molecular recognition problem for one class of target molecules. It is tempting to speculate that the three-helix fold of the albumin-binding domain has an evolutionary relationship with the immunoglob-

ulin-binding domain of protein A, although the sequences of the domains do not show any obvious traces of such a relationship. Doolittle [50] has, however, postulated that structure normally is better conserved than sequence, why an evolutionary relationship might still exist.

Acknowledgements: We thank Lewis Kay for continued NMR support, P.-O. Edlund for performing the mass spectrometry analysis, and the staff at the Swedish NMR Centre for technical assistance. This work was supported by the National Swedish Board for Technical Development (NUTEK) and the Swedish National Research Council (NFR).

References

- [1] Björck, L. and Åkerström, B. (1990) in: *Bacterial Immunoglobulin-Binding Proteins* (Boyle, M.P.D., Ed.), pp. 113–126, Academic Press, CA.
- [2] Langone, J.J. (1982) *Adv. Immunol.* 32, 157–252.
- [3] Nygren, P.-Å., Ljungqvist, C., Troeborg, H., Nustad, K. and Uhlén, M. (1990) *Eur. J. Biochem.* 193, 143–148.
- [4] Falkenberg, C., Björck, L. and Åkerström, B. (1992) *Biochemistry* 31, 1451–1457.
- [5] Åkerström, B., Nielsen, E. and Björck, L. (1987) *J. Biol. Chem.* 262, 13388–13391.
- [6] Nygren, P.-Å., Eliasson, M., Abrahmsén, L., Uhlén, M. and Palmcrantz, E. (1988) *J. Mol. Recogn.* 1, 69–74.
- [7] Forsgren, A. and Sjökvist, J. (1966) *J. Immunol.* 97, 822–827.
- [8] Uhlén, M., Guss, B., Nilsson, B., Gatenbeck, S., Philipson, L. and Lindberg, M. (1984) *J. Biol. Chem.* 259, 1695–1702.
- [9] Myhre, E.B. and Erntell, M. (1985) *Mol. Immunol.* 22, 879–885.
- [10] de Chateau, M. and Björck, L. (1994) *J. Biol. Chem.* 269, 12147–12151.
- [11] Fahnestock, S.R., Alexander, P., Nagle, J. and Filpula, D. (1986) *J. Bacteriol.* 167, 870–880.
- [12] Guss, B., Eliasson, M., Olsson, A., Uhlén, M., Frej, A.-K., Jörnvall, H., Flock, J.-I. and Lindberg, M. (1986) *EMBO J.* 5, 1567–1575.
- [13] Olsson, A., Eliasson, M., Guss, B., Nilsson, B., Hellman, U., Lindberg, M. and Uhlén, M. (1987) *Eur. J. Biochem.* 168, 319–324.

- [14] Gronenborn, A.M., Filpula, D.R., Essig, N.Z., Achari, A., Whitlow, M., Wingfield, P.T. and Clore, G.M. (1991) *Science* 253, 657–661.
- [15] Lian, L.-Y., Derrick, J.P., Sutcliffe, M.J., Yang, J.C. and Roberts, G.C.K. (1992) *J. Mol. Biol.* 228, 1219–1234.
- [16] Achari, A., Hale, S.P., Howard, A.J., Clore, G.M., Gronenborn, A.M., Hardman, K.D. and Whitlow, M. (1992) *Biochemistry* 31, 10449–10457.
- [17] Gallagher, T., Alexander, P., Bryan, P. and Gilliland, G.L. (1994) *Structure* 3, 4721–4729.
- [18] Derrick, J.P. and Wigley, D.B. (1994) *J. Mol. Biol.* 243, 906–918.
- [19] Sauer-Eriksson, A.E., Kleywegt, G.J., Uhlén, M. and Jones, T.A. (1995) *Structure* 3, 265–278.
- [20] Gouda, H., Torigoe, H., Saito, A., Sato, M., Arata, Y. and Shimada, I. (1992) *Biochemistry* 31, 9665–9672.
- [21] Lyons, B.A., Tashiro, M., Cedergren, L., Nilsson, B. and Montelione, G.T. (1993) *Biochemistry* 32, 7839–7845.
- [22] Deisenhofer, J. (1981) *Biochemistry* 20, 2361–2370.
- [23] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- [24] Hultman, T., Ståhl, S., Hornes, E. and Uhlén, M. (1989) *Nucleic Acids Res.* 17, 4937–4946.
- [25] Rütger, U. (1982) *Nucl. Acids Res.* 10, 5765–5772.
- [26] Maurer, R., Meyer, B.J. and Ptashne, M. (1980) *J. Mol. Biol.* 139, 147–161.
- [27] Öberg, U., Rundström, G., Grönlund, H., Uhlén, M. and Nygren, P.-Å. (1994) in: *Proceedings of the 6th European Congress on Biotechnology* (Alberghina et al., Eds.), pp. 179–182, Elsevier, Amsterdam, The Netherlands.
- [28] Löwenadler, B., Jansson, B., Plaeus, S., Holmgren, E., Nilsson, B., Moks, T., Palm, G., Josephsson, S., Philipsson, L. and Uhlén, M. (1986) *Gene* 58, 87–97.
- [29] Manavalan, P. and Johnson, W.C. (1987) *Anal. Biochem.* 167, 76–85.
- [30] Muhandiram, D.R., Farrow, N.A., Xu, G.Y., Smallcombe, S.H. and Kay, L.E. (1993) *J. Magn. Reson.* B102, 317–321.
- [31] Santoro, J. and King, G.C. (1992) *J. Magn. Reson.* 97, 202–207.
- [32] Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson.* 89, 496–514.
- [33] Wittekind, M. and Mueller, L. (1993) *J. Magn. Reson.* B101, 201–205.
- [34] Grzesiek, S. and Bax, A. (1992a) *J. Am. Chem. Soc.* 114, 6291–6293.
- [35] Kay, L.E. (1993) *J. Am. Chem. Soc.* 115, 2055–2057.
- [36] Logan, T.M., Olejniczak, E.T., Xu, X.R. and Fesik, S.W. (1992) *FEBS Lett.* 314, 413–418.
- [37] Montelione, G.T., Lyons, B.A., Emerson, S.D. and Tashiro, M. (1992) *J. Am. Chem. Soc.* 114, 10974–10975.
- [38] Pascal, S.M., Muhandiram, D.R., Yamazaki, T., Forman-Kay, J.D. and Kay, L.E. (1994) *J. Magn. Reson.* B103, 197–201.
- [39] Kraulis, P.J. (1989) *J. Magn. Reson.* 84, 627–633.
- [40] Kraulis, P.J., Domaille, P.J., Campbell-Burk, S.L., Van Aken, T. and Laue, E.D. (1994) *Biochemistry* 33, 3515–3531.
- [41] Cedergren, L., Andersson, R., Jansson, B., Uhlén, M. and Nilsson, B. (1993) *Prot. Eng.* 6, 441–448.
- [42] Jendeberg, L., Tashiro, M., Tejero, R., Lyons, B.A., Uhlén, M., Montelione, G.T. and Nilsson, B. (1995) *Biochemistry*, in press.
- [43] Grzesiek, S. and Bax, A. (1992b) *J. Biomol. NMR* 3, 185–204.
- [44] Kraulis, P.J. (1994) *J. Mol. Biol.* 243, 696–718.
- [45] Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.* 113, 5490–5492.
- [46] Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995) *J. Biomol. NMR* 5, 67–81.
- [47] Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR* 4, 171–180.
- [48] Gronenborn, A.M. and Clore, G.M. (1994) *J. Biomol. NMR* 4, 455–458.
- [49] Boczeko, E.M. and Brooks, C.L., III (1995) *Science* 269, 393–396.
- [50] Doolittle, R.F. (1992) *Prot. Sci.* 1, 191–200.
- [51] Sjöbring, U. (1992) *Infect. Immun.* 60, 3601–3608.
- [52] Kraulis, P.J. (1991) *J. Appl. Cryst.* 24, 946–950.