Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 8532–8536, September 1992
Biochemistry

Convergent evolution among immunoglobulin G-binding bacterial proteins

INGA-MARIA FRICK*, MATS WIKSTRÖM†, STURE FORSÉN†, TORBJÖRN DRAKENBERG†‡, HIDEYUKI GOMI§, ULF SJÖBRING¶, AND LARS BJÖRCK*‖

Departments of *Medical and Physiological Chemistry, †Physical, Chemistry 2, and ¶Medical Microbiology, University of Lund, Lund, Sweden; ‡Chemical Laboratory, VTT, POB 202, SF-02151, Espoo, Finland; and §Biotechnology Laboratory, Takarazuka Research Center, Sumitomo Chemical Co., Ltd., Hyogo, Japan

Communicated by Jan Waldenström, June 5, 1992 (received for review April 15, 1992)

ABSTRACT **Protein G, a bacterial cell-wall protein with high affinity for the constant region of IgG (IgGFc) antibodies, contains homologous repeats responsible for the interaction with IgGFc. A synthetic peptide corresponding to an 11-amino acid-long sequence in the COOH-terminal region of the repeats was found to bind to IgGFc and block the interaction with protein G. Moreover, two other IgGFc-binding bacterial proteins (proteins A and H), which do not contain any sequences homologous to the peptide, were also inhibited in their interactions with IgGFc by the peptide. Finally, a decapeptide based on a sequence in IgGFc blocked the binding of all three proteins to IgGFc. This unusually clear example of convergent evolution emphasizes the complexity of protein–protein interactions and suggests that bacterial surface-protein interaction with host protein adds selective advantages to the microorganism.**

Several immunoglobulin-binding bacterial-surface proteins have been identified and characterized (for review, see ref. 1). Most of these proteins show affinity for the Fc region of IgG (IgGFc), and in this group of molecules protein A of *Staphylococcus aureus* (2) and protein G of group C and G streptococci (3, 4) are the most well characterized. Interestingly, proteins A and G appear to interact with the same part of IgGFc (3, 5, 6), although they show no sequence homology within their Fc-binding domains (7–9). More recently, another IgGFc-binding protein was isolated from a group A streptococcal strain (10). The gene encoding this molecule, named protein H, was sequenced (11), and again no sequence homology was detected when protein H was compared with the Fc-binding domains of proteins A and G.

The crystal structure of the complex between IgGFc and one of the five homologous Fc-binding domains of protein A (fragment B) has been solved (12). Synthetic peptides were constructed from the primary structure of areas within Fc proposed to contact fragment B. Such peptides were also synthesized from the amino acid sequence of the Fc-binding domains of protein G. The various peptides were used in competitive binding experiments and NMR studies in further analyses of the IgGFc-binding properties of proteins A, G, and H. The results were then related to the structure and evolution of these bacterial cell-wall proteins.

MATERIALS AND METHODS

Proteins and Peptides. The protein G used comprised two IgGFc-binding domains (CDC; two C domains and one D domain) and was purchased from Pharmacia. The protein A used also included two Fc-binding domains (ED) and was provided by J. Sjöquist (Uppsala University, Uppsala, Sweden). Intact, *Escherichia coli*-produced protein H was isolated as described (10). Albumin and polyclonal IgG were isolated from human plasma by ion-exchange chromatography and gel filtration. IgGFc fragments were prepared by papain digestion of IgG followed by gel filtration and affinity chromatography on protein G-Sepharose (Pharmacia).

Peptides, based on sequences in protein G and IgGFc, were synthesized and analyzed for purity and correct sequence, as described (13, 14). NH_2-terminal amino acid sequencing was done on a pulsed-liquid sequencer (model 477A; Applied Biosystems) with the standard NORMAL-1 program supplied by Applied Biosystems.

Proteins were radiolabeled with ^{125}I by using the Bolton and Hunter (15) reagent (Amersham).

Competitive Binding Assay. Polyclonal IgGFc fragments were coupled to polyacrylamide beads (Immunobeads; Bio-Rad). ^{125}I-labeled protein A, G, or H in 0.1 ml of Veronal buffer, pH 7.35/0.15 M NaCl/0.1% gelatin (VBS), 0.1 ml of Immunobead-coupled IgGFc fragments in VBS, and 0.2-ml samples of unlabeled protein A, G, H, or synthetic peptides (in VBS) were mixed in dilution series with known concentrations and incubated at 20°C for 2 hr. Beads were washed and centrifuged, and the radioactivity of the pellets was measured. To determine the amounts of IgGFc beads yielding sensitive assays, constant amounts of ^{125}I-labeled protein A, G, or H were incubated with various amounts of beads. The bound radiolabeled protein was measured, and a dilution of the IgGFc beads was chosen that bound the tested radiolabeled protein just below maximum. Relative affinities of various protein and peptide preparations were measured from displacement curves, where the amount of unlabeled sample was plotted against the observed inhibition of radioactivity bound to the beads. Finally, affinity constants were calculated from competitive binding experiments, as described (16).

NMR Experiments. Intact IgG or IgGFc was dissolved in 0.067 M phosphate-buffered saline, pH 7.4 (PBS), containing 10% ^2H_2O for the lock signal and then concentrated by centrifugation on an Amicon Centriprep-10 concentrator. Solutions (0.45 ml) of IgG or IgGFc were added to freeze-dried protein G peptides (PG5 or PG8), yielding a final concentration of 0.25 mM IgG or IgGFc and 2.5 mM peptide. Samples with free peptide PG5 or PG8 (2.5 mM) were prepared by adding PBS buffer to freeze-dried peptide. The pH was adjusted to 5.0 without correcting for isotope effects. Concentrations of the NMR samples were confirmed by quantitative amino acid analysis. NMR spectra were acquired on a General Electric Omega 500 spectrometer operating at 500.13 MHz for ^1H. Resonances were assigned by using phase-sensitive correlated spectroscopy (COSY), total

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; IgGFc, constant region of IgG; ROESY, rotating frame Overhauser enhancement spectroscopy.

‖To whom reprint requests should be addressed at: Department of Medical and Physiological Chemistry, University of Lund, P.O. Box 94, S-221 00 Lund, Sweden.

correlated spectroscopy (TOCSY), nuclear Overhauser enhancement spectroscopy (NOESY), and rotating frame Overhauser (ROESY) experiments. The data were transformed with the Omega software and then transferred into a format used by our in-house software MAGNE (to be described elsewhere) for further analysis. For the NOESY and ROESY spectra, a total of 512 free induction decay units (FIDs) was acquired with a spectral width of 5509.6 Hz, 2048 points, and 64 scans per free induction decay unit. The dispersive signal due to the water signal was removed by using a method suggested by Adler and Wagner (17) and included in the MAGNE program. The mixing time of the total correlated spectroscopy experiment was 120 ms and for the NOESY and ROESY experiments was 200 ms; all experiments were done at 278 K.

RESULTS

Analysis of IgGFc-Binding Properties of a C Domain in Protein G with Synthetic Peptides. The 55-amino acid-long C domains of protein G are highly homologous (>90% sequence identity) and are responsible for the interaction with IgGFc (8, 9). A collection of peptides was synthesized from the sequence of the C1 domain (Fig. 1) and tested for inhibitory activity in a competitive binding assay. In this assay, human IgGFc fragments were coupled to polyacrylamide beads. Radiolabeled protein G fragments, comprising two C domains flanking a central D domain, were added, followed by addition of the various synthetic peptides (unlabeled) depicted in Fig. 1. Fig. 1 shows that peptides containing a sequence in the COOH-terminal third of the C1 domain blocked the binding of the radiolabeled CDC protein G fragment. The smallest peptide showing inhibitory activity (PG8) was 11 amino acids long and corresponded to amino acid residues 34–44 in the domain. As compared with the entire C1 domain (PG1) and to peptides PG2 and PG3, larger amounts of peptide PG8 were required to inhibit binding. Thus, 50% inhibition of the binding of CDC to IgGFc was obtained with 267 nmol of PG8, as compared with 20, 84, and 93 nmol of PG1, -2, and -3, respectively. Inhibition with PG4, however, was similar to that of PG8 (312 nmol). The affinity constants for the binding of CDC, PG1, and PG8 to IgGFc were $5.0 \times 10^9\ M^{-1}$, $2.8 \times 10^7\ M^{-1}$, and $5.8 \times 10^5\ M^{-1}$, respectively. The results indicate that the sequence 34–44 participates in the interaction between protein G and IgGFc. However, it should be pointed out that the experimental system used represents an indirect assessment of this protein–protein interaction. Thus, NMR methods were used to more directly analyze a possible physical association between the region of the C1 domain corresponding to PG8 and IgGFc (see below).

NMR Analysis of a Protein G Peptide (PG8) that Blocks the Interaction Between Protein G and IgGFc. NMR methods can

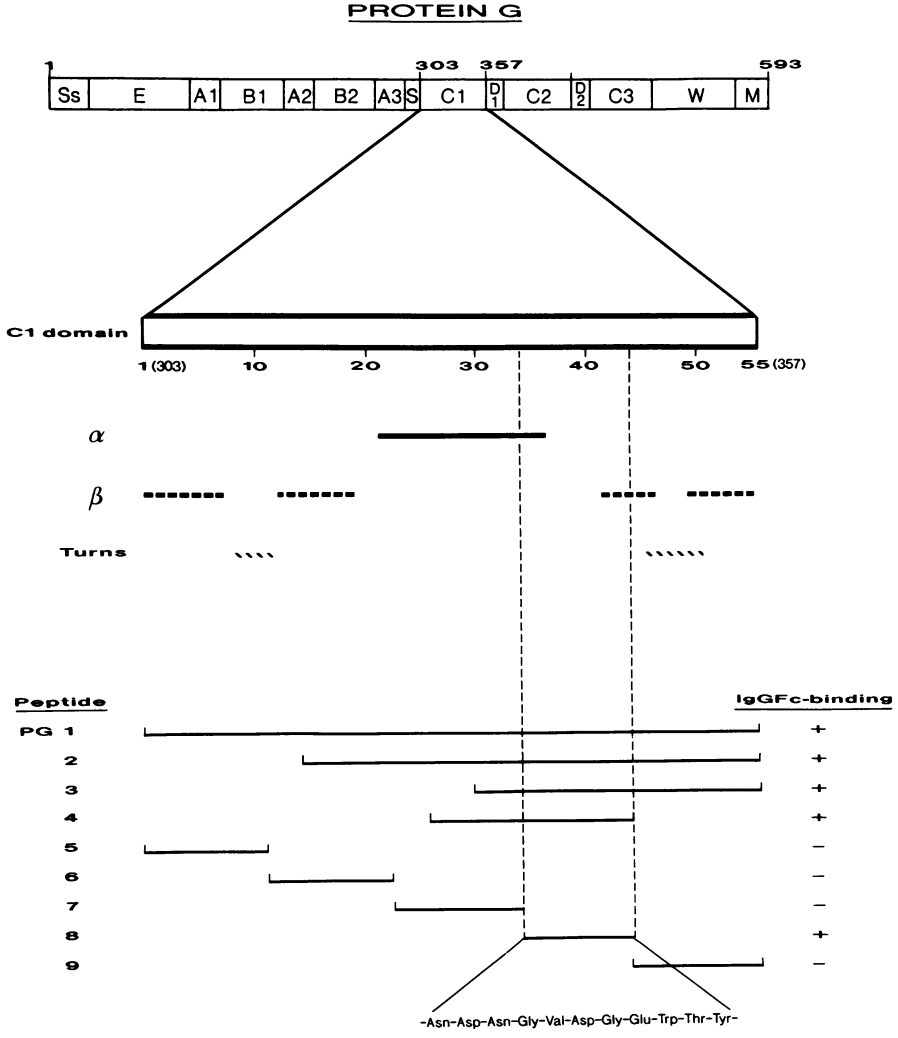

FIG. 1. Schematic representation of protein G using the nomenclature of Olsson *et al.* (18). α-Helical structure, β-sheets, and turns within the C1 domain are indicated according to Gronenborn *et al.* (19). Bars represent synthetic peptides (PG1–9) corresponding to various sequences in the C1 domain. Their IgGFc-binding activity was measured in a competitive binding assay.

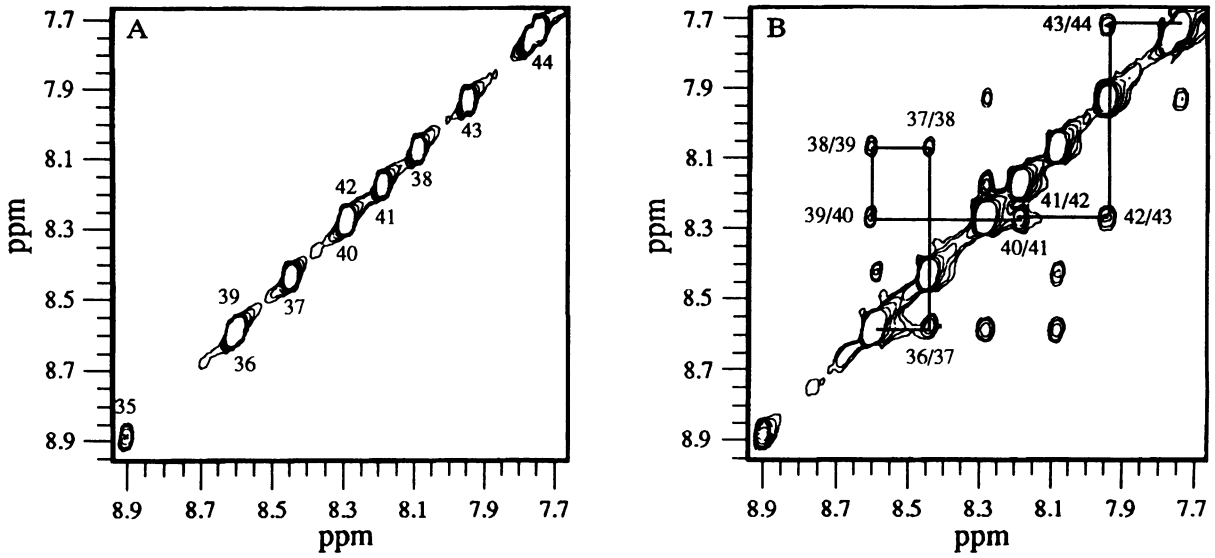

FIG. 2. NOESY spectra of protein G peptide PG8 corresponding to amino acids 34–44 of the C1 domain. The two spectra were run, processed, and plotted under identical conditions. (*A*) PG8 (2.5 mM) alone in PBS, pH 5.0. (*B*) PG8 (2.5 mM) plus IgGFc (0.25 mM) in PBS, pH 5.0.

be used to obtain information on molecular structure by performing nuclear Overhauser enhancement (NOE) experiments (20, 21). NOEs arise from cross-relaxation between spins (protons) close in space (<5 Å). In a two-dimensional NOESY such cross-relaxation is manifested as off-diagonal cross-peaks with positions given by the resonance frequencies of the interacting protons. The magnitude of such cross-peaks depends not only on the distance between the protons but also on the correlation time for the vector connecting them. Molecules, with the size of PG8 (1.4 kDa), normally have correlation times such that the cross-peak intensities are close to zero, independent of distance. In larger molecules like IgGFc, intense cross-peaks may be expected. This result is used in the so-called transferred NOE experiments (22, 23), where the efficient cross-relaxation of a small molecule bound to a larger molecule is used. The cross-relaxation occurring in the bound molecule is then transferred to the free molecule via chemical exchange, and such cross-peaks can be seen even when a large excess of small molecules is present. Fig. 2*A* shows an expanded portion of a NOESY spectrum from a sample containing PG8 alone, whereas Fig. 2*B* shows the same view of a NOESY spectrum in the presence of IgGFc. The molar ratio PG8/IgGFc was 10:1. The same cross-peaks were seen in PG8 in the presence of intact IgG (the same molar ratios), whereas another protein G peptide, PG5, did not give rise to such cross-peaks. ROESY, which is a modified version of the NOESY experiment (24, 25), also brings out cross-relaxation between neighboring protons. However, the cross-peak intensities depend less on the correlation time. In a ROESY experiment on free PG8, cross-peaks similar to those in Fig. 2*B*, although weaker, were seen (data not shown). This result indicated that no major conformational changes occurred in PG8 upon binding to IgGFc. In a control experiment human serum albumin (0.25 mM) was used instead of IgGFc together with PG8. The NOESY spectrum showed the same result as for the free peptide, excluding the possibility that the cross-peaks seen in the presence of IgGFc were from increased viscosity. In summary, the results show an association between PG8 and IgGFc in solution.

Studies on the Relation Between the IgGFc-Binding Properties of Proteins A, G, and H. Proteins A and G both contain repeated IgGFc-binding domains, whereas protein H has a single Fc-binding region (7–9, 11). As mentioned above, the Fc-binding regions of these proteins show no sequence homology, although at least proteins A and G appear to bind to the same part of IgGFc (3, 5, 6). It seemed that this somewhat unexpected observation could be studied in more detail with the synthetic peptides described here.

In a first set of experiments the binding of radiolabeled CDC protein G fragments to Fc beads was inhibited with unlabeled CDC, protein A, or protein H. The protein A used also contained two Fc-binding domains (ED), whereas the protein H represented the entire molecule. Fig. 3 shows that the inhibition curves obtained had similar shapes and that the amount of protein H needed to obtain 50% inhibition was ≈100-fold more as compared with the protein A and protein G preparations. The results suggest closely located binding sites on Fc for the three proteins and a lower affinity for protein H. Additional experiments demonstrated that the binding of protein A and protein H to Fc beads could also be blocked with protein G peptides PG1–4 and PG8—i.e., the same peptides that inhibited binding of protein G to Fc (Fig. 1). Fig. 4 shows the inhibition curves obtained when the binding of radiolabeled protein H to Fc beads was inhibited with PG8, the sequence of which showed no homology with either protein A or protein H.

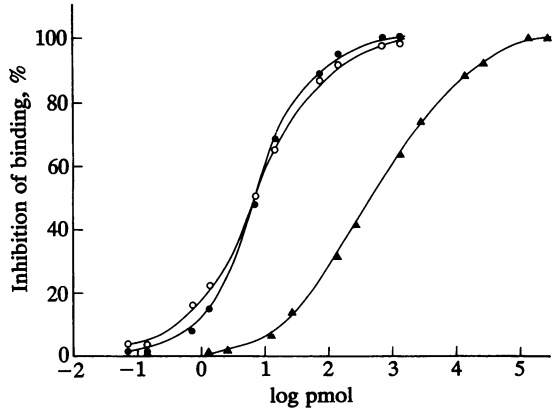

FIG. 3. Binding of ^{125}I-labeled protein G fragments, containing two C and one D domain (CDC), to IgGFc coupled to polyacrylamide beads was inhibited with different amounts of unlabeled protein preparations as follows: ○, CDC protein G fragments; ●, ED protein A fragments; ▲, intact protein H.

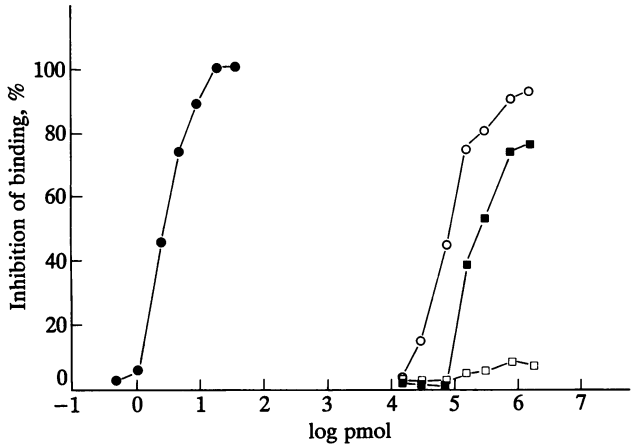

FIG. 4. Binding of ^{125}I-labeled protein H to IgGFc beads was inhibited with different amounts of unlabeled protein H and peptides as follows: ●, protein H; ○, protein G peptide PG8; □, IgGFc peptide 1; ■, IgGFc peptide 2.

Included in Fig. 4 are also the inhibition curves obtained with two decapeptides synthesized to include amino acids proposed to contact the Fc-binding domains of protein A as judged from x-ray crystallography (12). The amino acids of Fc peptide 1 (KDTLMISRTP) are in the NH_2-terminal, and those of Fc peptide 2 (VLHQDWLNGK) are in the COOH-terminal part of the CH2 domain of IgGFc. However, in the three-dimensional structure of the protein A–IgGFc complex, the protein A-interacting residues of these two sequences are exposed close to each other in CH2 domain and are located in the CH2–CH3 interface region (12). As seen in Fig. 4, Fc peptide 2 inhibited the interaction between protein H and Fc, whereas Fc peptide 1 did not inhibit this interaction. This result was also the case when the two Fc peptides were tested for inhibition of binding between protein A or protein G to IgGFc (data not shown).

The combined data of the experiments described in this section suggest that the binding sites for proteins A, G, and H in IgGFc are closely located or even overlapping.

DISCUSSION

The structure of the IgGFc-binding domains of protein G has recently been determined in solution by NMR spectroscopy (19, 26) and was found to comprise a four-stranded β-sheet flanking a central α-helix. The smallest synthetic peptide inhibiting the interaction between protein G and IgGFc in the present work is located in the COOH-terminal part of the Fc-binding protein G domains. The sequence of this peptide, PG8, is found in and around an extended structure connecting the α-helix with the COOH-terminal outer β3 strand of the domain (Fig. 5). Our results suggest that this region is of importance to the IgGFc-binding activity of streptococcal protein G, although a definite mapping of the binding site(s) must await determination of the three-dimensional structure of the protein G–IgGFc complex.

The major finding of this work is that three unrelated genes of different bacterial species have evolved to encode surface proteins capable of interacting with the same region in human IgGFc. This observation represents an unusually clear case of convergent evolution, where the results could be explained were the three-dimensional structure of the Fc-binding regions of proteins A, G, and H similar. However, from the data so far available, identification of such related regions is not possible. On the contrary, the x-ray crystallographic and NMR data on protein A (12, 27) and the NMR data on protein G (19, 26) have indicated major structural differences between these proteins, which underline the complexity of protein–protein interactions.

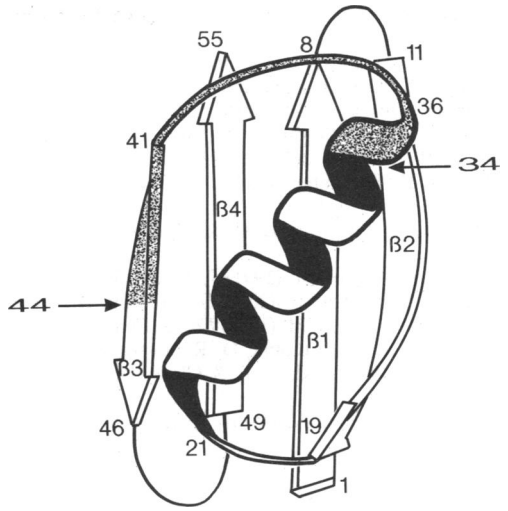

FIG. 5. Schematic drawing of the polypeptide fold of the C1 protein G domain; data are taken from Gronenborn *et al.* (19). The shaded area between amino acids 34 and 44 shows the location of protein G peptide PG8.

Although the structural basis for the convergent evolution of IgGFc-binding bacterial proteins remains unclear, it definitely suggests that these proteins are connected with essential functions that add selective advantages to the microorganisms. *In vivo*, Fc-binding proteins will enable the microorganism to cover itself with host proteins, which could help it evade the defense mechanisms of the infected host. Moreover, in group A streptococci a proteolytic enzyme (a cysteine proteinase), which is active both intra- and extracellularly (28), has been shown to be of vital importance to the bacteria by promoting their growth (29). In this context it is noteworthy that protein H acts as a substrate for this proteinase at the bacterial-cell surface (L.B., unpublished observation) and that Fc-binding bacterial proteins also show affinity for human proteinase inhibitors (13). Thus, Fc-binding proteins could influence proteolytic mechanisms associated with bacterial growth. Were these proteins conclusively found to participate in molecular events critical to basic microbial functions, such as growth or colony formation or to the host–parasite relationship, more light would be shed on the case of convergent evolution described here.

We are indebted to Drs. Henry Franzén and Jerzy Trojnar for synthesis of peptides. This work was supported by grants from the Swedish Medical Research Council (Projects 7480 and 10434), the Swedish National Board for Technical Development (Project 90-00262), King Gustav V:s 80-Years Foundation, the Medical Faculty of the University of Lund, and the Foundations of Kock, Österlund, and Crafoord.

1. Boyle, M. D. P., ed. (1990) *Bacterial Immunoglobulin Binding Proteins* (Academic, San Diego)
2. Forsgren, A. & Sjöquist, J. (1966) *J. Immunol.* **97**, 822–827.
3. Reis, K. J., Ayoub, E. M. & Boyle, M. D. P. (1984) *J. Immunol.* **132**, 3091–3097.
4. Björck, L. & Kronvall, G. (1984) *J. Immunol.* **133**, 969–974.
5. Åkerström, B. & Björck, L. (1986) *J. Biol. Chem.* **261**, 10240–10247.
6. Stone, G. C., Sjöbring, U., Björck, L., Sjöquist, J., Barber, C. V. & Nardella, F. A. (1989) *J. Immunol.* **143**, 343–351.
7. Sjödahl, J. (1977) *Eur. J. Biochem.* **73**, 343–351.
8. Fahnestock, S. R., Alexander, P. J., Nagle, J. & Filipula, D. R. (1986) *J. Bacteriol.* **167**, 870–880.
9. Guss, B., Eliasson, M., Olsson, A., Uhlén, M., Frej, A. K., Jörnvall, H., Flock, J.-I. & Lindberg, M. (1986) *EMBO J.* **5**, 1567–1575.
10. Åkesson, P., Cooney, J., Kishimoto, F. & Björck, L. (1990) *Mol. Immunol.* **27**, 523–531.

11. Gomi, H., Hozumi, T., Hattori, S., Tagawa, C., Kishimoto, F. & Björck, L. (1990) *J. Immunol.* **144,** 4046–4052.
12. Deisenhofer, J. (1981) *Biochemistry* **20,** 2361–2370.
13. Sjöbring, U., Trojnar, J., Grubb, A., Åkerström, B. & Björck, L. (1989) *J. Immunol.* **143,** 2948–2954.
14. Nilsson, B. H. K., Solomon, A., Björck, L. & Åkerström, B. (1992) *J. Biol. Chem.* **267,** 2234–2239.
15. Bolton, A. E. & Hunter, W. M. (1973) *Biochem. J.* **133,** 529–539.
16. Åkerström, B. & Björck, L. (1989) *J. Biol. Chem.* **264,** 19740–19746.
17. Adler, M. & Wagner, G. (1991) *J. Magn. Reson.* **91,** 450–454.
18. Olsson, A., Eliasson, M., Guss, G., Nilsson, B., Hellman, U., Lindberg, M. & Uhlén, M. (1987) *Eur. J. Biochem.* **168,** 319–324.
19. Gronenborn, A. M., Filipula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T. & Clore, G. M. (1991) *Science* **253,** 657–661.
20. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (Wiley, New York).
21. Wagner, G. (1990) *Prog. Nucl. Magn. Reson. Spectrosc.* **22,** 101–139.
22. Clore, G. M. & Gronenborn, A. M. (1982) *J. Magn. Reson.* **48,** 402–417.
23. Campbell, A. P. & Sykes, B. D. (1991) *J. Magn. Reson.* **93,** 77–92.
24. Bothner-By, A. A., Jeanloz, R. W., Lee, J., Stephens, R. L. & Warren, C. D. (1984) *J. Am. Chem. Soc.* **106,** 811–813.
25. Bax, A. & Davis, D. (1985) *J. Magn. Reson.* **65,** 355–360.
26. Lian, L.-Y., Yang, J. C., Derrick, J. P., Sutcliffe, M. J., Roberts, G. C. K., Murphy, J. P., Goward, C. R. & Atkinson, T. (1991) *Biochemistry* **30,** 5335–5340.
27. Torigoe, H., Shimada, I., Saito, A., Sato, M. & Arata, Y. (1990) *Biochemistry* **29,** 8787–8793.
28. Lo, S.-S., Liang, S.-M. & Liu, T.-Y. (1984) *Anal. Biochem.* **136,** 89–92.
29. Björck, L., Åkesson, P., Bohus, M., Trojnar, J., Abrahamson, M., Olafsson, I. & Grubb, A. (1989) *Nature (London)* **337,** 385–386.