

# Localization of Bound Water in the Solution Structure of the Immunoglobulin Binding Domain of Streptococcal Protein G

## Evidence for Solvent-induced Helical Distortion in Solution

G. Marius Clore and Angela M. Gronenborn

Laboratory of Chemical Physics, Building 2  
National Institute of Diabetes and Digestive and Kidney Diseases  
National Institutes of Health, Bethesda, MD 20892, U.S.A.

(Received 17 September 1991; accepted 7 November 1991)

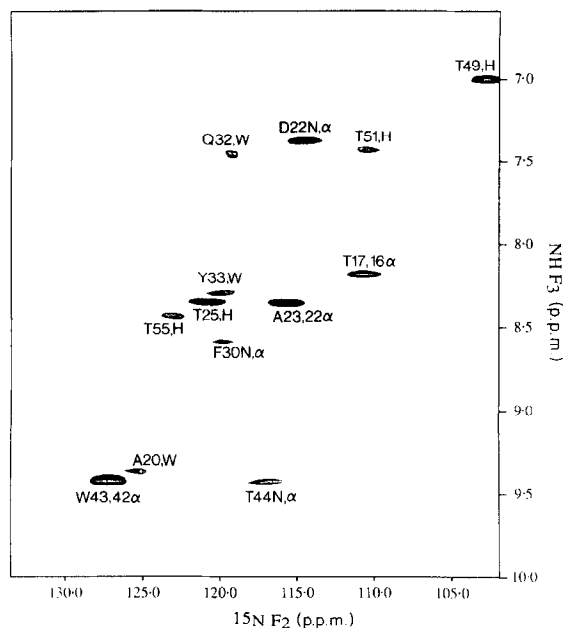
The presence of bound water in the solution structure of the IgG binding domain of streptococcal protein G has been investigated by nuclear magnetic resonance using three-dimensional  $^1\text{H}$  rotating frame Overhauser  $^1\text{H}$ - $^{15}\text{N}$  multiple quantum coherence spectroscopy. The backbone amide protons of three residues, Ala20, Gln32 and Tyr33, are found to be in close proximity to bound water. Examination of the three-dimensional structure of the IgG binding domain indicates that in the vicinity of these three residues there are no backbone groups that do not already participate in hydrogen bonding and there are no suitably placed side-chain groups available for hydrogen bonding with water. As the lifetime of the bound water detected in this nuclear magnetic resonance experiment is greater than about one nanosecond, it is likely that the two bound water molecules participate in a bifurcating hydrogen bonding network comprising a CO–NH hydrogen bonded pair, such that the water molecule accepts a hydrogen bond from the NH proton and donates one to the carbonyl oxygen with the result that the amide proton is involved in a three center hydrogen bond. On the basis of the structure, one water molecule participates in such an interaction with the Ala20(NH)–Met1(CO) hydrogen bonded pair at the beginning of an anti-parallel  $\beta$ -sheet, and the other with the Tyr33(NH)–Val29(CO) hydrogen bonded pair in the single  $\alpha$ -helix. The latter, which is external and solvent accessible, is associated with a distortion in the  $\alpha$ -helix centered around Tyr33 which consists of a significant increase in the CO(i-4)–N(i) and CO(i-4)–NH(i) distances relative to those in the rest of the helix, as well as a significant departure in the  $\phi$ ,  $\psi$  angles of Tyr33 relative to regular helical geometry. Such solvent induced distortions in  $\alpha$ -helices have been previously noticed in crystal structures and were postulated as possible folding intermediates for helical structures. The present observation of this phenomenon in solution indicates, however, that these water molecules are tightly bound and represent an integral part of the protein framework.

**Keywords:** IgG binding domain; bound water; solution structure; 3D heteronuclear n.m.r.; ROE

The immunoglobulin (IgG $\dagger$ ) binding domain of streptococcal protein G is thought to play an important role in the mechanism whereby group G *Streptococcus* evades the host defenses (for a recent

review, see Fahnstock *et al.*, 1990). This small domain of only 55/56 residues is unusual as it is extremely thermostable with a reversible melting transition of 87°C (Gronenborn *et al.*, 1991). In a recent paper, we presented the high resolution solution structure of the IgG binding domain and showed that it comprises a four-stranded mixed parallel/anti-parallel  $\beta$ -sheet, with a +1, –3x, +1 topology, on top of which lies a long  $\alpha$ -helix (Gronenborn *et al.*, 1991). In this communication, we extend our previous study to the localization of bound water using 3D  $^1\text{H}$ - $^{15}\text{N}$  ROESY-HMQC spectroscopy.

$\dagger$  Abbreviations used: IgG, immunoglobulin G; ROE, rotating frame Overhauser effect; ROESY, rotating frame Overhauser spectroscopy; HMQC, heteronuclear multiple quantum coherence;  $^1\text{H}$ - $^{15}\text{N}$  ROESY-HMQC,  $^1\text{H}$  rotating frame Overhauser  $^{15}\text{N}$ - $^1\text{H}$  multiple quantum coherence spectroscopy; 2D, two-dimensional; 3D, three-dimensional; h.p.l.c., high pressure liquid chromatography; n.m.r., nuclear magnetic resonance; p.p.m., parts per million.

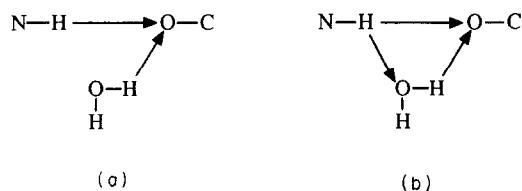


**Figure 1.** The  $\text{NH}(\text{F}_3)\text{-}^{15}\text{N}(\text{F}_2)$  plane at the  $^1\text{H}(\text{F}_1)$  frequency of the water resonance (4.76 p.p.m.) of the 600 MHz  $^1\text{H}\text{-}^{15}\text{N}$  ROESY-HMQC spectrum of 1.5 mM-IgG binding domain at 25°C and pH 4.3. The spectrum was recorded as described by Clore *et al.* (1990) with a 77 ms ROE mixing time, approximately equal to the  $T_{1\rho}$  of the amide protons for optimal sensitivity. Only negative cross-peaks are plotted; positive cross-peaks that arise from chemical exchange are not shown. Peaks in the spectrum are labeled *i, j*, where *i* refers to the residue name and number of the amide  $^{15}\text{N}$  and NH atoms, and *j* refers to either a water proton (W), a protein hydroxyl or water proton (H) or an aliphatic proton. The IgG binding domain was expressed in *Escherichia coli* grown on minimal medium with  $^{15}\text{NH}_4\text{Cl}$  as the sole nitrogen source to obtain uniform (>95%)  $^{15}\text{N}$  labeling, and partially purified as described previously (Gronenborn *et al.*, 1991). This purification procedure yields two species of protein, with and without the N-terminal methionine, in a ~7 : 3 mixture. The major species with the N-terminal methionine was then purified to homogeneity by reverse phase h.p.l.c. and was greater than 99% pure as judged by N-terminal sequencing and n.m.r. spectroscopy. The  $^{15}\text{N}$  spectrum was assigned in a straightforward manner by reference to the  $^1\text{H}$  assignments from 2D  $^1\text{H}\text{-}^{15}\text{N}$  Overbodenhausen correlation (Bodenhausen & Ruben, 1980; Bax *et al.*, 1990; Norwood *et al.*, 1990) and  $^1\text{H}\text{-}^{15}\text{N}$  HMQC-NOESY (Gronenborn *et al.*, 1989) spectra. The complete list of  $^{15}\text{N}$  assignments, together with the  $^1\text{H}$  chemical shifts, have been deposited in the Brookhaven Protein Data Bank (accession code number R1GB1MR). The spectral width in the  $^1\text{H}(\text{F}_1)$ ,  $^{15}\text{N}(\text{F}_2)$  and  $^1\text{H}(\text{F}_3)$  dimensions were 12.82 p.p.m., 32.12 p.p.m. and 13.44 p.p.m., respectively, with the  $^1\text{H}$  carrier placed at the water frequency (4.76 p.p.m.) and the  $^{15}\text{N}$  carrier at 118.5 p.p.m. The spectrum was recorded with 128 complex  $\times$  16 complex  $\times$  1024 real points in  $t_1$ ,  $t_2$  and  $t_3$ , respectively. The number of points in the  $^{15}\text{N}(t_2)$  dimension was extended to 32 complex points by mirror image linear prediction (Zhu & Bax, 1990). After zero-filling in all dimensions, the absorptive part of the final processed data matrix comprised 256 ( $\text{F}_1$ )  $\times$  64 ( $\text{F}_2$ )  $\times$  1024 ( $\text{F}_3$ ) points. The spectrum was processed on a Sun Sparc workstation using in-house routines for the Fourier transform (Kay *et al.*, 1989) and linear prediction (Zhu & Bax,

The ROE enables one to distinguish through-space interactions from chemical exchange as these two effects have opposite sign in the spectrum (Bothner-By *et al.*, 1984). Hence, it provides a means of resolving short interproton contacts between NH protons and bound water from exchange effects involving water (Otting & Wüthrich, 1989; Clore *et al.*, 1990). By spreading the ROE interactions in a third dimension corresponding to the  $^{15}\text{N}$  chemical shift of the directly bonded backbone nitrogen atoms, the 3D  $^1\text{H}\text{-}^{15}\text{N}$  ROESY-HMQC spectrum permits one to readily identify interactions at the water frequency by inspection of the  $\text{NH}(\text{F}_3)\text{-}^{15}\text{N}(\text{F}_2)$  plane at the  $^1\text{H}(\text{F}_1)$  frequency of the water resonance (Clore *et al.*, 1990). This plane through the 3D  $^1\text{H}\text{-}^{15}\text{N}$  ROESY-HMQC spectrum of the IgG binding domain, uniformly labeled with  $^{15}\text{N}$ , is shown in Figure 1. Of the 13 peaks present, six are assigned to  $\text{C}^{\alpha}\text{H}(i)\text{-NH}(i)$  and  $\text{C}^{\alpha}\text{H}(i)\text{-NH}(i+1)$  ROEs and seven can be attributed to direct or indirect ROEs with water. Of the latter, four involve threonine residues and can be potentially explained by an indirect pathway comprising an ROE from the backbone NH to the hydroxyl side-chain proton and exchange between the hydroxyl proton and water. In the case of the other three ROEs, involving Ala20, Gln32 and Tyr33, there are no neighboring exchangeable protons so that they can be unambiguously attributed to direct interactions with bound water.

In the two previous cases that we examined, namely interleukin- $1\beta$  (Clore *et al.*, 1990, 1991) and human thioredoxin (Forman-Kay *et al.*, 1991), the water molecules were internal and involved in stabilizing interactions with backbone atoms which would otherwise have unsatisfied hydrogen bonds. In the case of the IgG binding domain, however, there are no backbone carbonyl or NH groups in the vicinity of Ala20, Gln32 or Tyr33 that do not already participate in hydrogen bonds and there are no suitable neighboring side-chain groups in this region of the structure available for hydrogen bonding with water. This leaves one with two possible hydrogen bonding schemes for the bound water molecules. The first is one in which only the carbonyl oxygen of the CO-NH hydrogen bonding pair is involved in a bifurcating hydrogen bond with a water molecule (Fig. 2(a)); the second is one in which the bound water molecule participates in bifurcating interaction with both partners of the hydrogen bonded CO-NH pair where the amide proton is involved in a three-center hydrogen bond, as shown in Figure 2(b) (Sundaralingam & Sekharudu, 1989). Of these two possibilities, the latter is the more likely, as the observation of a ROE between a protein proton and water implies that the lifetime of the bound water is longer than

1990) in  $\text{F}_2$ , together with the commercially available software package NMR2 (New Methods Research, Inc, Syracuse, NY) for processing the  $\text{F}_1\text{-F}_3$  planes.

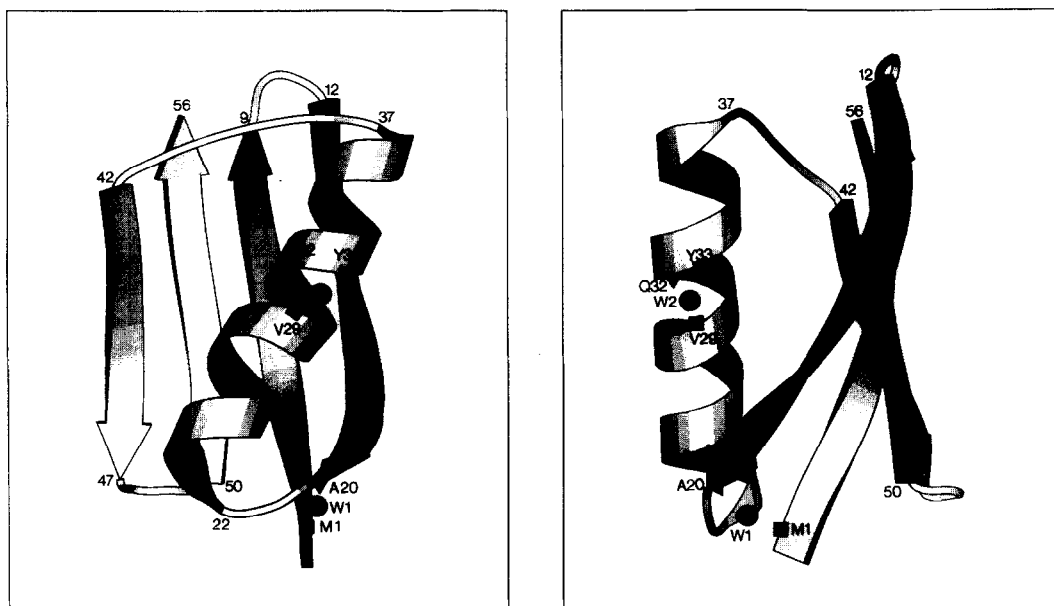


**Figure 2.** Hydrogen bonding schemes involving bound water and a CO-NH hydrogen bonded pair. In (a) the water donates a hydrogen bond to the carbonyl oxygen, whereas in (b) it donates a hydrogen bond to the carbonyl oxygen and accepts one from the amide proton atoms so that the amide proton is involved in a 3-center hydrogen bond. The hydrogen bonds and their directionality are indicated by the arrows.

about one nanosecond (Clore *et al.*, 1990). This suggests that the bound waters detected in this manner by n.m.r. are involved in at least two hydrogen bonding interactions (Clore *et al.*, 1990; Forman-Kay *et al.*, 1991).

In the structure of the IgG binding domain, the amide proton of Ala20 is hydrogen bonded to the carbonyl of Met1 at the beginning of an anti-parallel  $\beta$ -sheet, and the amide protons of Gln32 and Tyr33 participate in CO(i-4)-NH(i) hydrogen bonding in the helix with the carbonyls of Lys28 and Val29, respectively (Fig. 3). Thus, water molecule W1 would be hydrogen bonded to both the carbonyl of Met1 and the amide of Ala20. From inspection of the structure, it seems most likely that only a single

water molecule is close to Gln32 and Tyr33 and that this water (W2) is hydrogen bonded to the carbonyl of Val29 and the amide of Tyr33, rather than to the carbonyl of Lys28 and the amide of Gln32. The reason for this choice is based on an examination of the CO(i-4)-N(i) and CO(i-4)-NH(i) distances within the helix. Specifically the CO(Val29)-N(Tyr33) and CO(Val29)-NH(Tyr33) distances have values of  $3.34(\pm 0.03)$  Å and  $2.53(\pm 0.016)$  Å ( $1 \text{ Å} = 0.1 \text{ nm}$ ), respectively (averaged over 60 calculated structures), which are significantly longer than that of the other CO(i-4)-N(i) and CO(i-4)-NH(i) distances within the helix which have average values of  $2.99(\pm 0.12)$  Å and  $2.17(\pm 0.19)$  Å, respectively. In addition, the  $\phi$ ,  $\psi$  angles of Tyr33 have values of  $-74.9(\pm 3.7)^\circ$ ,  $-59.2(\pm 2.9)^\circ$ , which deviate significantly from the average  $\phi$ ,  $\psi$  values of  $-61.8(\pm 7.1)^\circ$ ,  $-39.2(\pm 10.0)^\circ$  for the other residues in the helix and fall outside the bounds seen in regular  $\alpha$ -helices ( $\phi$ ,  $\psi$  values of  $-66^\circ$ ,  $-41^\circ$  and  $-59^\circ$ ,  $-44^\circ$  in hydrophobic and hydrophilic environments, respectively; Blundell *et al.*, 1983). Thus, there is a small degree of unwinding at this point within the helix. The overall helix (residues 23 to 37), however, remains straight. In high resolution crystal structures, these distortions from classical helical geometry are associated with the presence of a water molecule hydrogen bonded to either the backbone carbonyl of the i-4 residue alone or to both the backbone carbonyl of the i-4 residue and amide proton of residue i (Blundell *et al.*, 1983; Sundaralingam & Sekharudu, 1989).



**Figure 3.** A ribbon drawing of 2 views of the solution structure of the IgG binding domain illustrating the approximate location of the bound water molecules and the groups with which they interact. The residues whose NH protons give rise to ROEs with bound water are indicated by the triangles (Ala20, Gln32 and Tyr33), and the residues whose backbone carbonyls are potentially involved in hydrogen bonding interactions with the bound water are indicated by squares (Met1 and Val29). The co-ordinates for the solution structure of the IgG binding domain have been deposited in the Brookhaven protein data bank with accession numbers 1GB1 and 2GB1 for the restrained minimized mean structure and the 60 individual simulated annealing structures, respectively (Gronenborn *et al.*, 1991). The schematic ribbon diagram was produced with the program MOLSCRIPT (Kraulis, 1991).

In summary, we have identified two bound water molecules in the solution structure of the IgG binding domain of Streptococcal protein G. Both bound water molecules may play an important role in contributing to the unusual stability of this structure. Thus, water molecule W1 could stabilize the  $\beta$ -sheet at the N terminus of the polypeptide chain, while W2 may stabilize a distortion in the  $\alpha$ -helix around Tyr33 arising as a consequence of the extremely well packed hydrophobic core. While such solvent induced helical distortions have been observed previously in high resolution crystal structures, this is the first time that it has been directly demonstrated in solution. Thus, one can conclude that the presence of such water molecules in X-ray structures should not be regarded as arising from crystal packing forces, but rather as intrinsic components of the protein structure. In addition, both bound waters are located in external, solvent accessible positions (Fig. 3), indicating that bound waters with long lifetimes ( $> 1$  ns) do not necessarily need to be buried and internal. Rather they simply need to be stabilized by an appropriate number of hydrogen bonding interactions.

We thank Drs Marc Whitlow, Nina Essig and David Filpula for providing us with the IgG binding domain clone of Streptococcal protein G. This work was supported by the AIDS Targeted Anti-Viral Program of the Office of the Director of the National Institutes of Health.

### References

- Bax, A., Ikura, M., Kay, L. E., Torchia, D. A. & Tschudin, R. (1990). Comparison of different modes of two-dimensional reverse correlation NMR for the study of proteins. *J. Magn. Reson.* **86**, 304–318.
- Blundell, T., Barlow, D., Borkakoti, N. & Thornton, J. (1983). Solvent-induced distortions and the curvature of  $\alpha$ -helices. *Nature (London)*, **306**, 281–283.
- Bodenhausen, G. & Ruben, D. J. (1980). Natural abundance 15-nitrogen NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Letters*, **69**, 185–189.
- Bothner-By, A. A., Stephens, R. L., Lee, J. T., Warren, C. D. & Jeanloz, R. W. (1984). Structure determination of a tetrasaccharide: transient nuclear Overhauser effect in the rotating frame. *J. Amer. Chem. Soc.* **106**, 811–813.
- Clore, G. M., Bax, A., Wingfield, P. T. & Gronenborn, A. M. (1990). Identification and localization of bound internal water in the solution structure of interleukin- $1\beta$  by heteronuclear three-dimensional  $^1\text{H}$  rotating frame Overhauser  $^{15}\text{N}$ - $^1\text{H}$  multiple quantum coherence NMR spectroscopy. *Biochemistry*, **29**, 5671–5676.
- Clore, G. M., Wingfield, P. T. & Gronenborn, A. M. (1991). High resolution three dimensional structure of interleukin- $1\beta$  in solution by three and four dimensional nuclear magnetic resonance spectroscopy. *Biochemistry*, **20**, 2315–2323.
- Fahnestock, S. R., Alexander, P., Filpula, D. & Nagle, J. (1990). Structure and evolution of the streptococcal genes encoding protein G. In *Bacterial Immunoglobulin Binding Proteins* (Boyle, N. D. P., ed.), vol. 1, pp. 133–148. Academic Press, New York.
- Forman-Kay, J. D., Gronenborn, A. M., Wingfield, P. T. & Clore, G. M. (1991). Determination of the positions of bound water in the solution structure of reduced human thioredoxin by heteronuclear three-dimensional nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **220**, 209–216.
- Gronenborn, A. M., Bax, A., Wingfield, P. T. & Clore, G. M. (1989). A powerful method of sequential proton resonance assignment in proteins using relayed  $^{15}\text{N}$ - $^1\text{H}$  multiple quantum coherence spectroscopy. *FEBS Letters*, **243**, 93–98.
- Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T. & Clore, G. M. (1991). A novel highly stable fold of the immunoglobulin binding domain of Streptococcal protein G. *Science*, **253**, 657–661.
- Kay, L. E., Marion, D. & Bax, A. (1989). Practical aspects of 3D heteronuclear NMR of proteins. *J. Magn. Reson.* **84**, 72–84.
- Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
- Norwood, T. J., Boyd, J., Heritage, J. E., Soffe, N. & Campbell, I. D. (1990). Comparison of techniques for  $^1\text{H}$ -detected heteronuclear  $^1\text{H}$ - $^{15}\text{N}$  spectroscopy. *J. Magn. Reson.* **87**, 488–501.
- Otting, G. & Wüthrich, K. (1989). Studies of protein hydration in aqueous solution by direct NMR observation of individual protein-bound water molecules. *J. Amer. Chem. Soc.* **111**, 1871–1875.
- Sundaralingam, M. & Sekharudu, Y. C. (1989). Water-inserted  $\alpha$ -helical segments implicate reverse turns as folding intermediates. *Science*, **244**, 1333–1337.
- Zhu, G. & Bax, A. (1990). Improved linear prediction for truncated signals of known phase. *J. Magn. Reson.* **90**, 405–410.