Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallographic analysis of the NNA7 Fab and proposal for the mode of human blood-group recognition

Kefang Xie,[a]‡ Shuh-Chyung Song,[b]‡ Steven L. Spitalnik[b] and Joseph E. Wedekind[a]*

[a]Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642, USA, and [b]Department of Pathology, College of Physicians and Surgeons of Columbia University, Room P&S 15-408, 630 West 168th Street, NY, NY 10032, USA

‡ These authors contributed equally to this work.

Correspondence e-mail: joseph_wedekind@urmc.rochester.edu

Received 25 June 2005
Accepted 26 July 2005

PDB References: NNA7, 1t2q, r1t2qsf; NNA7-G91S, 2d03, r2d03sf.

The NNA7 Fab antibody fragment recognizes the human N-type blood-group antigen comprised of the N-terminal glycopeptide of glycophorin A (GPA). A mutant form of this Fab fragment, NNA7-G91S, exhibits markedly reduced antigen binding. To provide insight into how these Fab fragments recognize this glycopeptide antigen, the crystal structures of NNA7 and NNA7-G91S were solved and refined to 1.83 and 1.97 Å resolution, respectively. Both molecules are composed of the same heavy (H) chain Fd fragment, but each contains a slightly different light (L) chain owing to the G91S substitution. Specifically, the G91S mutation pushes the backbone of the neighboring H chain away from complementarity-determining region 3 (CDR3) of the L-chain variable region, allowing an additional glycerol cryoprotectant molecule to enter the antigen-combining site near the putative location of O-linked glycosylation. Each Fab fragment also possesses a well defined 2-(*N*-morpholino)ethanesulfonic acid (MES) molecule trapped in its antigen-combining site, as well as a crystallographic symmetry-related molecule comprising an amino-acid sequence that is virtually identical to the N-terminus of GPA. The MES molecule interacts with the H-chain CDR in a manner reminiscent of antibody–carbohydrate complexes. These results suggest a model for recognition of the glycopeptide antigen that accounts for the deleterious effect of the G91S substitution.

1. Introduction

Glycophorin A (GPA) is the most abundant glycoprotein on the human red blood cell membrane and is of clinical importance in blood transfusion because it harbors the allelic blood-group M and N antigens (Telischi *et al.*, 1976; Sancho *et al.*, 1998). These blood-group antigens are localized to the N-terminus of the GPA polypeptide, which exhibits amino-acid polymorphisms at positions 1 and 5 (Blumenfeld & Adamany, 1978), thereby differentiating the M and N blood types (Fig. 1). In addition, there are O-linked disialylated tetrasaccharides at the Ser and Thr residues at positions 2 and 3, as well as the Thr residue at position 4 (Lisowska *et al.*, 1980). Many mouse monoclonal antibodies (MoAb) and human polyclonal antisera recognizing the M or N antigens have been described (Anstee & Edwards, 1982; Bigbee *et al.*, 1983, 1984; Rubocki & Milgrom, 1986; Lisowska *et al.*, 1987; Wasniowska *et al.*, 1988; Anstee & Lisowska, 1990; Jaskiewicz *et al.*, 1990, 1994; Duk *et al.*, 1994), most of which will only bind the relevant target if both the correct amino-acid sequence and the attached oligosaccharides are present (Judd, 1979; Sadler *et al.*, 1979; Bigbee *et al.*, 1983; Rubocki & Milgrom, 1986; Wasniowska *et al.*, 1988; Jaskiewicz *et al.*, 1990; Duk *et al.*,

1994). Thus, these antibodies recognize glycopeptide epitopes and therefore provide a model system for studying the molecular basis of the humoral immune response in which the antigenic specificity is dictated by both the underlying peptide sequence and the attached carbohydrates (Hanisch *et al.*, 1995; Karsten *et al.*, 1998). At present, no crystal structures of such glycopeptide-specific antibodies have been determined. Nonetheless, detailed characterization of individual antigenic determinants is of clinical importance for the identification of blood that is compatible for transfusion, as well as for diagnosing hemolytic disease of the newborn.

Efforts to elucidate the respective contributions of the peptide and carbohydrate moieties of the MN antigens in eliciting a humoral immune response led to the sequencing of the IgG2a anti-N N92 mouse MoAb derived from a hybridoma cell line (Wasniowska *et al.*, 1988; Czerwinski *et al.*, 1994). Molecular biological and Fab-phage display techniques were then employed to construct and purify the N92 Fab fragment (Czerwinski *et al.*, 1999). N92 recognizes a carbohydrate-dependent but sialic acid-independent epitope of GPA in which the Leu residue at position 1 (Fig. 1*a*) is the immunodominant amino acid (Wasniowska *et al.*, 1976; Blackall *et al.*, 1994). Studies using light-chain shuffling to produce high-affinity anti-M or anti-N Fab fragments led to the identification of a particular L chain obtained from a Fab-phage display library that, in combination with the parental N92 H-chain Fd fragment (Czerwinski *et al.*, 1998), resulted in the high-affinity NNA7 anti-N Fab fragment. Interestingly, the N92 Fab fragment and the library-derived NNA7 Fab fragment had identical L-chain CDR1 and CDR2 sequences and only four amino-acid differences in their L-chain CDR3 sequences. The sequences of the NNA7 and N92 Fab fragments also differed at L-chain positions 1–7. Despite these subtle differences, the NNA7 Fab had significantly higher affinity for the N-antigen compared with the parental N92 Fab (Czerwinski *et al.*, 1998). To identify the basis for enhanced antigen-binding affinity by the NNA7 Fab fragment, its L-chain CDR3 sequence was mutated from Gly to Ser at position 91 (*i.e.* the G91S mutation) to match that of the parental, lower affinity N92 Fab fragment. This single change dramatically decreased the affinity for N antigen of the resulting NNA7-G91S Fab fragment as determined by ELISA and by its failure to agglutinate human red blood cells (Song, Czerwinski *et al.*, 2004).

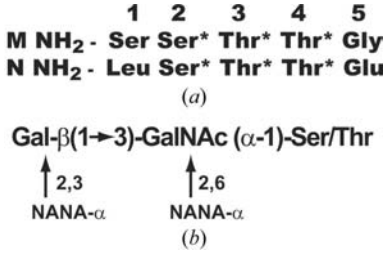


Figure 1
Schematic depiction of the primary sequence of the antigenic GPA glycopeptide. (*a*) The polypeptide sequence of the N-terminus depicting the amino-acid polymorphisms that differentiate the M and N antigens. O-Glycosylated residues are marked with asterisks. (*b*) Structures of the O-glycans at positions 2, 3 and 4 that contribute to MN antigenicity. The sugars are abbreviated as follows: Gal, galactose; GalNAc, *N*-acetylgalactosamine; NANA, *N*-acetylneuraminic (sialic) acid.

To elucidate the molecular basis for NNA7 binding to the glycopeptide N antigen, as well as a reason for the markedly decreased affinity caused by the G91S substitution, we solved the crystal structures of the NNA7 and NNA7-G91S Fab fragments by molecular replacement. The structures revealed that the antigen-combining site forms a large crescent-shaped cleft consistent with the need to accommodate glycopeptide antigens. Owing to the absence of antigenic ligand during crystallization, an MES buffer molecule was trapped within the antigen-combining site and interacted with residues derived from the H-chain CDRs. The six-membered heterocyclic ring of the MES molecule adopted a chair conformation which resembles a carbohydrate hexose subunit. Comparison of the structure of the NNA7–MES complex with the binding modes of natural protein–carbohydrate interactions suggested that this buffer molecule provides insights into the recognition determinants of the NNA7 Fab fragment for the glycan moiety of its target antigen. The L-chain G91S mutation falls within the crescent-shaped antigen-combining cleft and correlates well with the location of antigen binding. In addition, the G91S substitution in the L chain has structural consequences resulting in a deflection of the adjacent peptide backbone of the H-chain CDR3. The steric clash changes the variable-region structure at the confluence of the H-chain and L-chain CDRs, which is manifested by changes in crystal packing. These results are discussed in light of the crystal contacts with the Ser′-Ser′-Thr′-Lys′-Val′ sequence, which may mimic the natural GPA-encoded peptide (Fig. 1*a*), and nearby GOL molecules occupying the antigen-combining pocket, which are possible mimics of the O-glycans of the glycopeptide N antigen.

2. Methods

2.1. Protein purification and crystallization

Protein expression and purification methods have been described previously (Song, Xie *et al.*, 2004). Using hanging-drop vapor-diffusion methods, crystals of NNA7 grew after 6 d at 293 K from 22–24%(*w*/*v*) polyethylene glycol 5000 monomethylether, 0.10 *M* MES buffer pH 6.5, 0.30 *M* $(NH_4)_2SO_4$ and 2–8 m*M* YCl_3. Crystals of the NNA7-G91S mutant were obtained under comparable conditions, but were significantly smaller (50 × 50 × 100 µm) and required longer than three weeks to reach maximum size.

2.2. X-ray data collection, processing and structure determination

X-ray diffraction data were collected on a single crystal of NNA7 to 1.83 Å resolution using an in-house X-ray source as described elsewhere (Song, Xie *et al.*, 2004). Diffraction data for the NNA7-G91S mutant were collected at station A1 of the Cornell High Energy Synchrotron Source (CHESS, Ithaca, NY, USA). Intensities were recorded on a Quantum 210 CCD detector (ADSC) using a crystal-to-detector distance of

220 mm and an exposure time of 60 s per degree. Data were processed and scaled with the *Crystal Clear* software package (Pflugrath, 1999). Crystals of NNA7 and NNA7-G91S belonged to space group $P2_12_12_1$, with unit-cell parameters a = 57.9, b = 77.1, c = 118.1 Å and a = 57.60, b = 71.12, c = 122.9 Å, respectively, with one Fab molecule per asymmetric unit. Data-reduction statistics are summarized in Table 1.

Each Fab fragment structure was solved by molecular replacement using *AMoRe* (Navaza, 1994). The NNA7 structure was solved using the 48G7 Fab fragment (PDB code 1aj7) as a search model (Wedemayer *et al.*, 1997), whereas the structure of NNA7-G91S was solved using the NNA7 structure as a search model. Both search models were divided into their respective constant (C) and variable (V) regions, with the intervening hinge sequences removed. Rotation- and translation-function searches were conducted using structure-factor amplitudes between 3.0 and 15.0 Å resolution and resulted in solutions with a correlation coefficient of 45.5 and an R factor of 47.8% for the NNA7 structure and a correlation coefficient of 52.6 and an R factor of 46.2 for the NNA7-G91S structure.

2.3. Structure refinement

The molecular-replacement solution of NNA7 was subjected to 7500 K torsional annealing as implemented in *CNS* (Brünger *et al.*, 1998) (final R factor of 37.1% and $R_{\rm free}$ of 44% for data between 2.65 and 25 Å resolution). The resulting model was utilized in automated model building using diffraction data from 1.90 to 25 Å resolution as implemented in *ARP*/*wARP* v.6 (Perrakis *et al.*, 1999). The final R factor and $R_{\rm free}$ of the model were 20.6 and 30.4%, respectively. The resulting electron-density maps were of sufficient quality that a majority of the side chains and their rotamers could be assigned by inspection of $(2F_{\rm o} - F_{\rm c})$ and $(F_{\rm o} - F_{\rm c})$ maps in *O* (Jones *et al.*, 1991). During the refinement process it was discovered that nearly 20% of the protein sequence in C_H1 differed from the expected sequence (Song, Xie *et al.*, 2004). DNA sequencing and conceptual translation into an amino-acid sequence corroborated the electron-density observations and the final sequence was defined accordingly. The structure was refined by cycles of manual rebuilding in *O* with cycles of simulated annealing, positional and individual B-factor refinement in *CNS* (Brünger *et al.*, 1998) to a final resolution of 1.83 Å. A comparable strategy was used to refine the NNA7-G91S structure. The final refinement statistics are summarized in Table 1.

Throughout this paper, the amino-acid numbering system of the L chain of the NNA7 Fab fragment follows the standard immunological approach (Wu & Kabat, 1970; Kabat *et al.*, 1977), which was used in the initial description of NNA7 (Czerwinski *et al.*, 1998). Similarly, the sequence and numbering of the NNA7 H chain were provided in the initial description of the N92 MoAb (Czerwinski *et al.*, 1994); the N92 H-chain Fd fragment was used to construct NNA7 (Song, Czerwinski *et al.*, 2004). Additional details for the numbering scheme were derived from http://www.bioinf.org.uk courtesy of Dr A. C. R. Martin. The complete amino-acid sequences of the NNA7 L chain and of the NNA7 H-chain Fd fragment, along with the Kabat numbering, are provided as supplementary material[1].

Table 1
Intensity and refinement statistics for NNA7 and NNA7-G91S structures.

Values in parentheses are for the highest resolution shell.

	NNA7	NNA7-G91S
Intensity statistics		
Resolution range (Å)	1.83–36.7 (1.83–1.90)	1.97–24.4 (1.97–2.04)
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a = 57.9, b = 77.1, c = 118.1	a = 57.6, b = 71.1, c = 122.9
Molecules per AU	1	1
Solvent content (%)	56.4	50.7
No. of measurements	161696	147694
No. of unique reflections	38051 (953)	35107 (2741)
Average redundancy	3.7 (1.4)	4.21 (3.34)
Completeness (%)	92.4 (53.0)	96.4 (88.1)
$R_{\rm sym}$† (%)	6.5 (29.3)	7.0 (42.4)
$I/\sigma(I)$	12.9 (2.4)	9.6 (2.6)
Refinement statistics		
Resolution range (Å)	1.83–36.7	1.97–24.4
Total No. of reflections	45378	35104
No. of protein atoms	3889 [430 amino acids]	3296 [429 amino acids]
No. of water atoms	612	246
No. of glycerols	2	5
No. of MES molecules	1	1
R factor/$R_{\rm free}$/$R_{\rm cryst}$‡ (%)	20.4/24.7/20.1	21.8/26.2/22.0
R.m.s. deviations from ideality		
Bond lengths (Å)	0.005	0.009
Bond angles (°)	1.4	1.6
Ramachandran plot§		
Most favored (%)	91.9	90.5
Additionally allowed (%)	7.6	9
Temperature factors		
⟨B factor⟩, protein (Å^2)	29.0	41.0
⟨B factor⟩, waters (Å^2)	61.4	63.3

† $R_{\rm sym} = [\sum |I(h)_j - \langle I(h)\rangle| / \sum I(h)_j] \times 100$, where $I(h)_j$ is the observed intensity of the jth measurement of reflection h and $\langle I(h)\rangle$ is the mean intensity of reflection h. ‡ $R = (\sum ||F_{\rm o}| - |F_{\rm c}|| / \sum |F_{\rm o}|) \times 100$, where $F_{\rm o}$ and $F_{\rm c}$ are the observed and calculated structure-factor amplitudes. $R_{\rm free}$ is calculated using a test set of reflections that were randomly selected and excluded from the refinement, $R_{\rm work}$ is calculated using the remaining reflections and $R_{\rm cryst}$ is calculated using all reflections in the resolution range of refinement. § As described by Laskowski *et al.* (1993)

3. Results and discussion

3.1. Structure determination and fold of the NNA7 Fab fragment

The structures of the NNA7 and NNA7-G91S Fab fragments were solved by molecular replacement. The quality of each structure is indicated by its high-resolution diffraction data, refinement statistics (Table 1), $R/R_{\rm free}$ values of 20.4/24.7% and 22.2/26.3%, respectively, and representative electron-density maps (Figs. 2*a* and 4*a*). There is one Fab fragment (one L-chain and one H-chain Fd fragment) per asymmetric unit. Of the total of 220 amino acids in the L chain, 217 were observed for NNA7 and the G91S mutant. Of the 230

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference BE5042). Services for accessing these data are described at the back of the journal.

amino acids in the heavy chain, 213 were observed for NNA7 and 212 were observed for the G91S mutant. The main-chain electron density for both structures is well defined and contiguous with exception of a disordered loop (Fig. 2*c*) within the C-terminal portion of the H chain (residues H130–H133 of NNA7 and H128–H133 of the G91S mutant). A C^{α} superposition of the refined NNA7 Fab fragment and the molecular-replacement search model indicated that the principal axis of rotation relating the L and H chains deviated by 16° (Fig. 2*b*). Although the individual domains superimposed with an r.m.s.d. of 0.8 Å, the relative constant (C) and variable (V) region orientations exhibited a significant rigid-body movement relative to the search model, which implied that the linkers (Fig. 2*c*) possess considerable flexibility; this is relevant in documenting the effects of the site-directed G91S mutation, which impacts the L-chain/H-chain interface. Nonetheless, the buried surface area at the L-chain/H-chain interface remains fairly constant despite changes in conformation and sequence, with 11 296 $Å^2$ buried for the search model compared with 11 598 $Å^2$ for NNA7 (Fraczkiewicz & Braun, 1998). Notably, crystal contacts at the V and C regions of NNA7 promote ordered linkers (Fig. 2*c*) and this region is devoid of conformational heterogeneity in both the wild-type and mutant structures.

The global fold of NNA7 comprises L and H chains, each with V and C regions that are joined through a well structured linker (Fig. 2*c*). Each region represents a conserved immunoglobulin Greek-key β-barrel structure whose walls are formed by three-stranded and five-stranded antiparallel β-sheets. Additional short α-helices were found in the loop regions of the V and C regions of the L chain (V_L and C_L), as well as in the V region of the H-chain Fd fragment (V_H); these represent atypical secondary-structural elements, since antibodies traditionally comprise primarily β-sheet structures (Schiffer *et al.*, 1973). The respective helices of the L and H

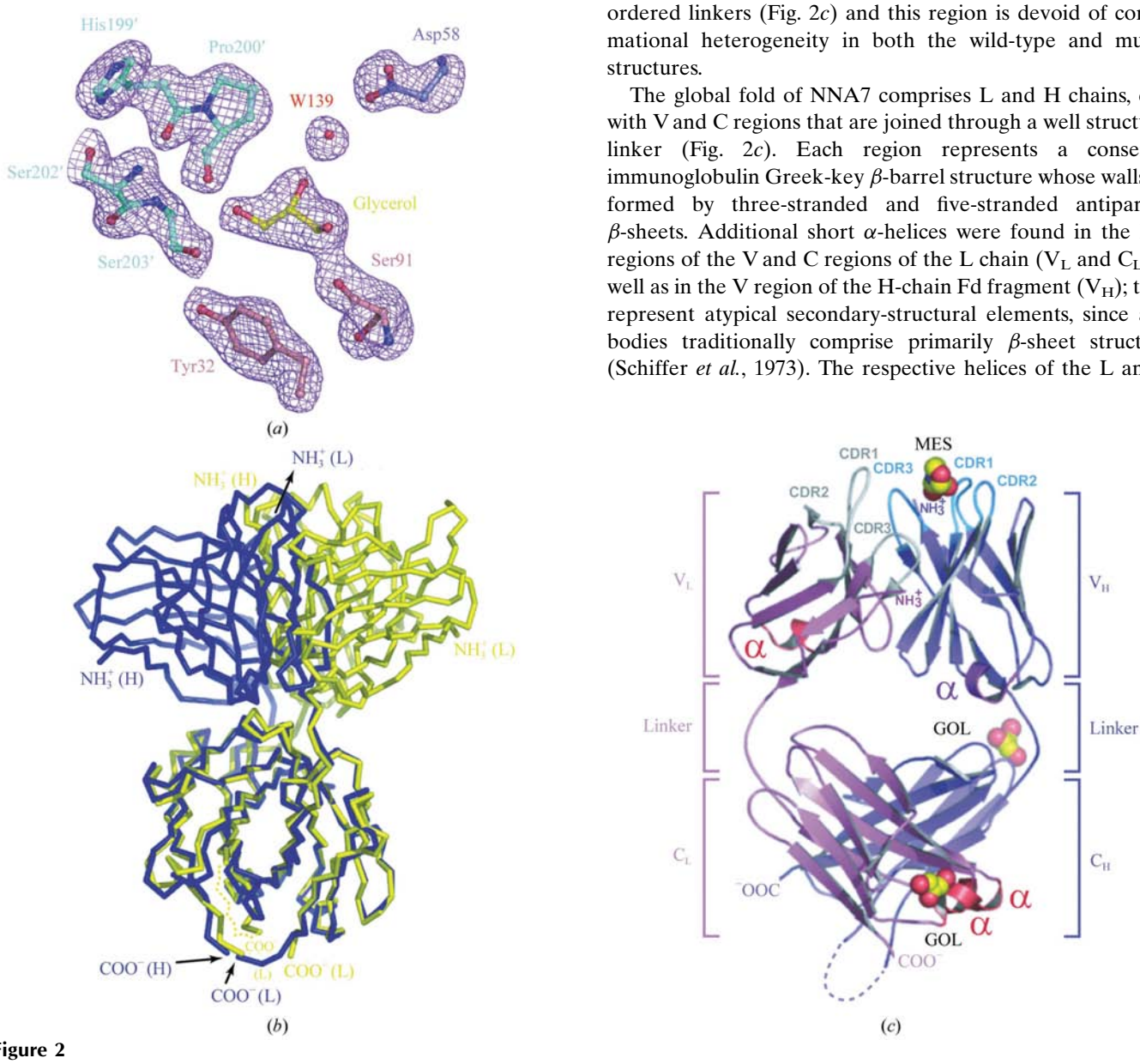


Figure 2
Representative electron density and schematic diagrams of the Fab fragment structure. (*a*) Omit electron-density map for the NNA7-G91S mutant Fab fragment with coefficients ($F_o - F_c$). The map is contoured at the 3.5σ level. The H-chain (Asp58) and L-chain (Tyr32 and Ser91) residues are colored blue and pink, respectively, with the symmetry-related H-chain residues (His199′, Pro200′, Ser202′ and Ser203′) in cyan. Glycerol (yellow) and water molecules (*i.e.* W139) are shown in the antigen-combining pocket. (*b*) C^{α} superposition of the NNA7 Fab fragment and the 48G7 search model. The alignment was performed by superimposing the constant regions. The NNA7 Fab fragment is shown in yellow and 48G7 is depicted in blue. (*c*) Ribbon representation depicting the global fold of the NNA7 Fab fragment. The L and H chains are shown in pink and blue, respectively. The CDRs from L and H chains are shown in light grey and light blue, with respective text labels in grey and blue. The α-helices of the L chain are in red (denoted α) and those of the H chain are purple. The MES and GOL molecules are represented by yellow and red space-filling models. The dashed blue line indicates a break in the chain arising from disordered electron density.

chains are: α1 (L79–L83, H83–H87), α2 (L121–L127, H97–H100B) and α3 (L183–L187, H200–H203) (Fig. 2*c*).

3.2. Structural changes resulting from the G91S mutation

It became clear during the initial crystallographic analysis that the NNA7-G91S mutant exhibited significantly altered unit-cell parameters relative to the wild-type NNA7 Fab fragment (Table 1). This suggested that the mutant Fab structure underwent conformational changes as a result of the single amino-acid change in the L-chain CDR3. A superposition analysis of the global wild-type and mutant structures revealed that the overall structures differed in a significant way, with an overall r.m.s.d. of 0.82 Å distributed over all C^{α}-atom pairs (Table 2). A comparison of the structures resulting from a C-region superposition revealed that the V regions underwent a modest rigid-body change that propagated through the linker residues (Figs. 3*a* and 3*b*). To discover additional differences in the individual domains or chains, further paired superpositions were performed to reveal the largest backbone differences. The results demonstrated that the N-termini of the chains were most different and that changes were greater in the H chain compared with the L chain (Table 2). However, differences were greater in V_L compared with V_H. Furthermore, the greatest r.m.s. differences were observed when comparing the local CDR3 sequences of either the H or L chains (Table 2).

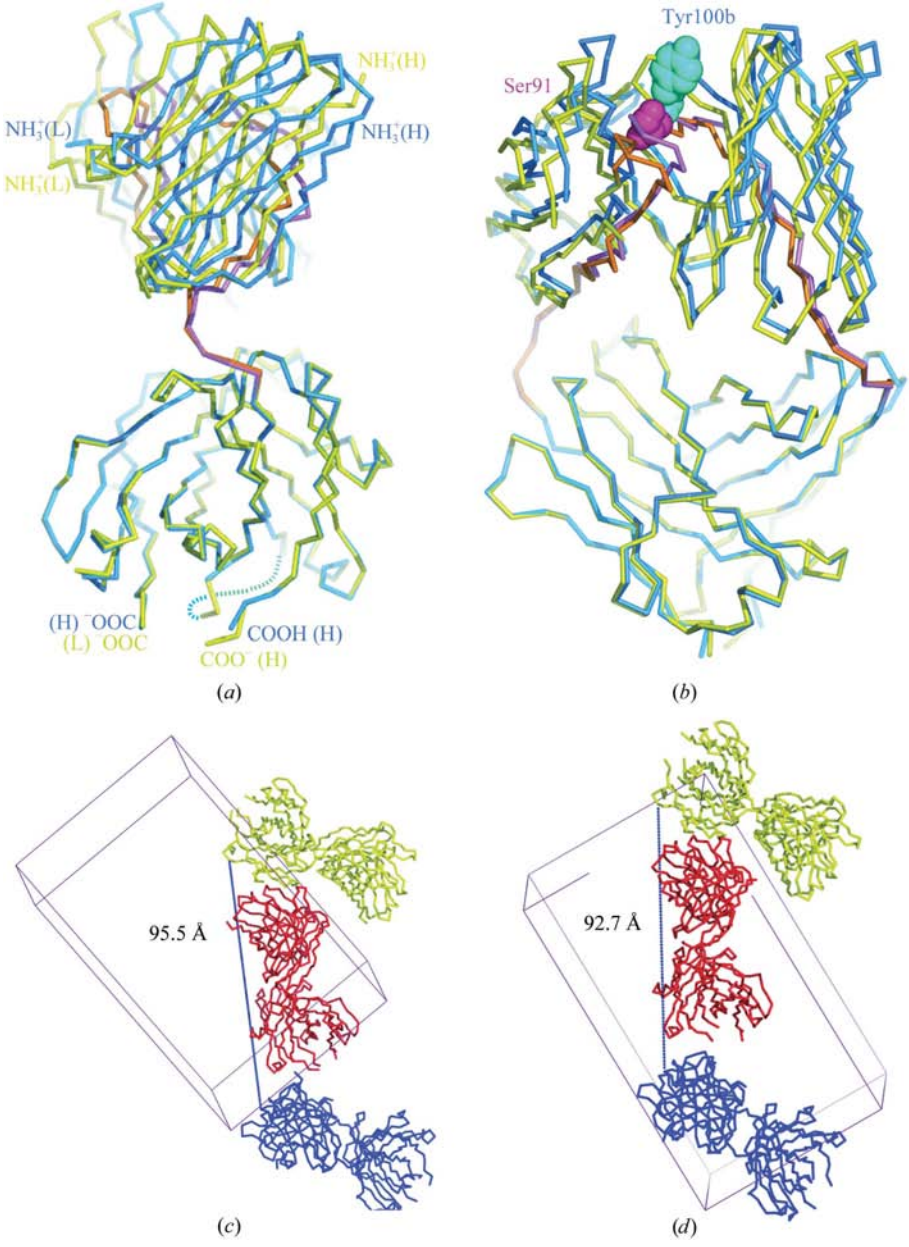


Figure 3
'Side' view (*a*) and 90° rotated 'front' view (*b*) of the superposition of the NNA7 and NNA7-G91S structures. NNA7 is yellow and NNA7-G91S is blue. The L chain containing Ser91 is pink and the H chain containing Tyr100b is cyan; both residues are shown as space-filling models. In the NNA7 and NNA7-G91S structures, L-chain CDR3 sequences and their associated linkers are shown in orange and magenta, respectively. (*c*) Crystal packing diagram of NNA7 and (*d*) of NNA7-G91S. Structures are in similar orientations. Yellow and blue molecules represent symmetry-related objects. Only the packing partners flanking the top and bottom of the red molecule are shown.

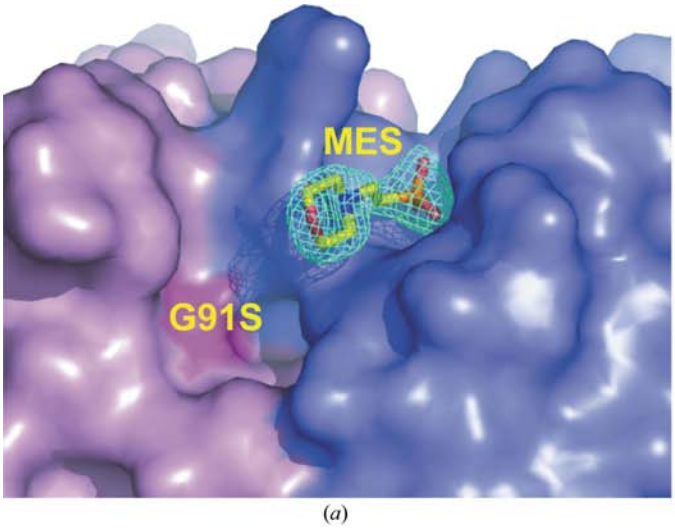
The antigen-combining site of a Fab fragment comprises six CDRs: three in V_L and three in V_H (Fig. 2*c*). The CDRs comprise genetically diversified loops and strands located on the periphery of the β-barrel of each paired V_L and V_H. The remaining portions of the V regions consist primarily of β-sheets that provide a framework to support the CDRs. Hence, the CDRs determine both the antigen-specificity and affinity of the antibody, which are fine-tuned by clonal selection in germinal B-cell centers (French *et al.*, 1989). Prior studies of antigen–antibody interactions showed that CDRs contribute differently to antigen binding; among L-chain CDRs, CDR3 residues dominated these contacts, especially the amino acids at positions 91, 94 and 96 (Wedemayer *et al.*, 1997). Our observation that the CDRs of NNA7 and NNA7-G91S have relatively different backbone agreement (maximum r.m.s.d. values of 0.45 and 0.40 Å, respectively; Table 2) supports the prior evidence documenting that a single amino-acid change at L-chain position 91 markedly modifies the affinity of the NNA7 Fab fragment (Song, Czerwinski *et al.*, 2004).

The largest changes in the antigen-binding pockets occur at position 91 of the L-chain CDR3 (the site of the single amino-acid change) and at residue Tyr100b of the H-chain CDR3. In

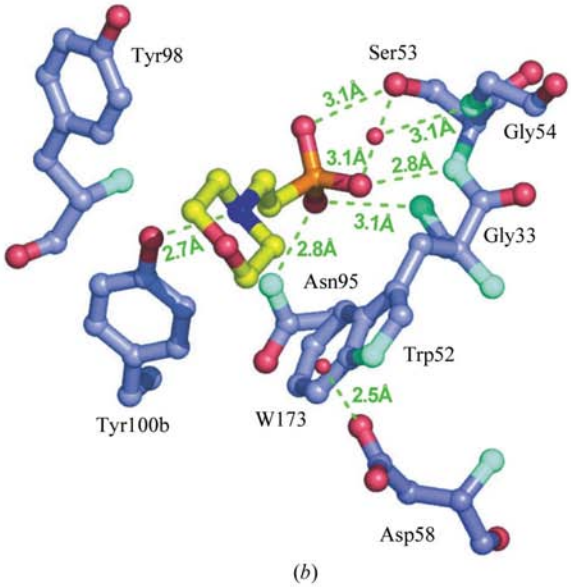
addition, changes at the confluence of these two critical CDR3 sequences are manifested as changes in the backbone that propagate to and terminate at the respective linker regions (Figs. 3*a* and 3*b*), which absorb the movement owing to their inherent flexibility (Fig. 2*b*). The net result is a rigid-body motion of the V region relative to the C region (Figs. 3*a* and 3*b*) that impacts not only the V-region alignment between wild-type and mutant structures, but also the crystal packing (Figs. 3*c* and 3*d*). Although the crystal packing of both structures is very similar, the G91S mutant displayed a more compact structure in which two symmetry-related objects are closer by approximately 3 Å compared with the wild-type structure (Figs. 3*c* and 3*d*). Such differences provide a plausible explanation of how the G91S mutation causes subtle changes in the molecular shape of the Fab fragment and its interaction with the relevant ligand (see below). Additional structural analysis will be required to determine the exact molecular basis by which these two Fab fragments differentially bind to the authentic N-type glycopeptide antigen.

3.3. A portion of the antigen-combining site is occupied by a molecule of MES

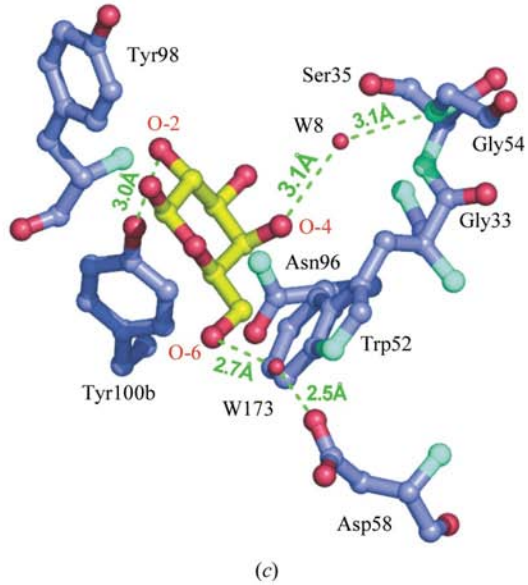
The antigen-combining site of the NNA7 Fab fragment is a semicircular cleft at the apex of the V regions that is approximately 9 Å wide by 12 Å deep (Figs. 2*c* and 4*a*). The dimensions of the pocket are consistent with the ascribed biological function of the antibody, which entails recognition of up to five amino acids at the N-terminus of GPA and at least some of the attached O-glycans (Fig. 1). Although no antigen was included in the crystallization medium, the antigen-combining site is occupied by a MES molecule derived from the crystallization buffer (Fig. 4*a*). The residues interacting with MES are exclusively in V_H and participate in (i) electrostatic interactions, (ii) hydrogen bonds and (iii) hydrophobic contacts (Fig. 4*b*). For example, three H-chain residues donate hydrogen bonds to the sulfonic acid moiety of the MES molecule, including the backbone amide group of Gly33, the amide side chain of Asn95 and the γ-OH group of Ser53. The hydroxyl group of Ser53 engages in a bifurcated hydrogen-bond interaction with the sulfonic acid group and also donates a hydrogen bond from its backbone amide (Fig. 4*b*). Such complementary electrostatic interactions are unlikely to be related to GPA binding since specificity has been reported to be sialic acid-independent. Other hydrogen bonds include involvement of the hydroxyl group of Tyr100b, which interacts with the tertiary amine of the morpholino ring. Finally,



(*a*)



(*b*)



(*c*)

Figure 4
Close-up views of the antigen-combining site of the NNA7 and NNA7-G91S structures. (*a*) Surface representation of the antigen-combining site of the NNA7 Fab fragment complexed with MES (ball-and-stick model). The L chain is pink and the H chain is blue. The position of the L-chain G91S mutation is shown as a red patch. Simulated-annealing omit ($F_o - F_c$) electron-density map for the NNA7 MES molecule is shown at the 3σ contour level. (*b*) Schematic diagram of the mode of MES binding to NNA7. (*c*) A theoretical model of galactose binding to the same site as in (*b*). All amino-acid residues are positioned exactly as in the experimental NNA7 structure. Dashed lines indicate possible hydrogen bonds with associated distances; the red spheres indicate water molecules.

Table 2
Superposition of C^{α} atoms between NNA7 and NNA7-G91S.

Superposition type	R.m.s. deviation (Å)	Maximum displacement (Å)	Location/residue
Overall	0.822	2.58	H chain, 212 (C-terminus)
L chains	0.685	1.94	187
H chains	0.817	2.70	212 (C-terminus)
Constant regions	0.369	1.08	L chain, 187
Variable regions	0.212	1.01	L chain, 2 (N-terminus)
L-chain variable	0.245	—	—
H-chain variable	0.129	—	—
CDR1 L chain	0.183†	0.307	27B
CDR2 L chain	0.165†	0.209	54
CDR3 L chain	0.264†	0.451	91
CDR1 H chain	0.157†	0.218	27
CDR2 H chain	0.146†	0.175	54
CDR3 H chain	0.237†	0.397	100b

† Based on the mean r.m.s.d. values obtained only from CDR atoms, whose positions were obtained from a superposition between all C^{α} atoms of the wild type and mutant in the variable region.

hydrophobic contacts and complementary fit contributed to MES binding. In particular, the side chains of Trp52 and Tyr98 engage in edge-to-face contacts with each side of the morpholino ring (Figs. 4*a* and 4*b*). Additional evidence supporting the relevance of the MES-binding site for blood-group N-antigen recognition is suggested by the observation that the NNA7 Fab fragment differed from the parental N92 MoAb at positions 1–7 at the N-terminus of its L chain (Song, Czerwinski *et al.*, 2004). Although the actual distance between the MES molecule and the N-terminus is 22 Å, a simple conformational change in CDR3 could result from changes to the N-terminal sequence, which makes direct contacts to CDR3 in the NNA7 structure (Fig. 2*c*). This spatial proximity reveals a rationale for how modest changes in the N-terminus of the L chain could influence the antigen-combining site as suggested previously (Song, Czerwinski *et al.*, 2004).

3.4. MES as a structural mimic for galactose

Structurally, the morpholino group of an isolated MES molecule is a molecular mimic for hexose sugar subunits, such as those of the O-glycans attached to GPA. In contrast, the sulfonic acid group is spatially equivalent to a β-glycosidic linkage, although stereochemically distinct. To explore whether the interaction of this buffer molecule with the Fab fragment serves as a mimic for the natural mode of glycopeptide antigen recognition, a galactose molecule from the human galectin-3 carbohydrate-recognition domain (PDB code 1a3k; Seetharaman *et al.*, 1998) was manually superimposed upon the MES molecule in the antigen-combining site of the NNA7 Fab. Interestingly, the morpholino ring possessed a similar pucker to galactose and several hydrogen-bond interactions were possible between the Fab fragment and the sugar (Figs. 4*b* and 4*c*). The protein residues involved were exclusively located in the H-chain CDR sequences. For example, the O-2 hydroxyl group of the galactose could form a hydrogen bond with the Tyr100b hydroxyl group (Fig. 4*c*). In addition, carbohydrates are known to participate in hydrogen bonds with water molecules that are pre-bound to the protein (Cygler *et al.*, 1991). As such, it is possible that the O-4 and O-6 hydroxyl groups of the superimposed galactose could form hydrogen bonds with water molecules W8 and W173 that are observed in the NNA7 structure (Fig. 4*c*). W8 is located within hydrogen-bonding distance of the Gly54 backbone amide group, whereas W173 binds to the Asp58 carboxyl group. Such indirect contacts are present in the galectin-3 carbohydrate-recognition domain structure, although in this case the O-4 and O-6 hydroxyl groups of galactose not only contacted certain amino-acid residues through interactions with water molecules, but also formed direct hydrogen bonds with specific side chains (Seetharaman *et al.*, 1998). The hypothetical MES-derived polar interaction with the O-4 hydroxyl group of the galactose model is especially important

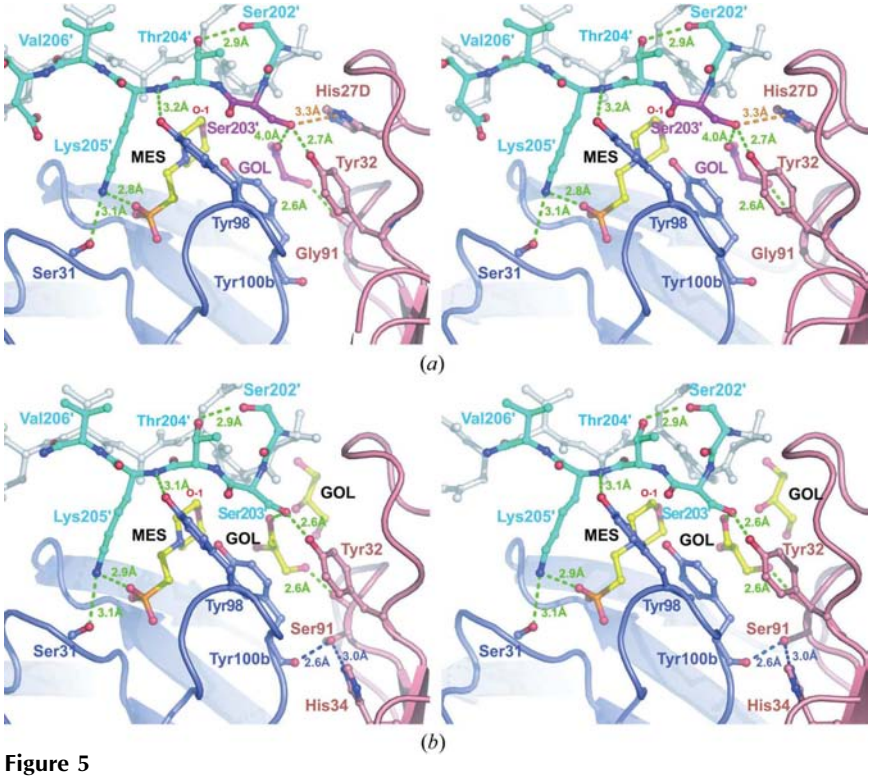


Figure 5
Stereoviews of crystal contacts at the antigen-combining site. (*a*) NNA7 Fab fragment, (*b*) NNA7-G91S Fab fragment. The H-chain and L-chain residues are blue and pink, respectively. Residues derived from the symmetry-related H-chain crystal contact are shown in cyan; the remaining residues on the symmetry-related H chain are shown in white. MES and GOL molecules are depicted as ball-and-stick models. Conserved hydrogen bonds are shown as dashed green lines. The residues that exhibit noticeable structural changes as a result of the G91S mutation in the L chain are depicted in purple. The particular hydrogen bonds that are unique to the wild-type and mutant structures are shown in orange (*a*) and blue (*b*).

because the orientation of the O-4 hydroxyl group is often a key feature for the selectivity of recognition of specific carbohydrates by proteins (Bush *et al.*, 1999).

3.5. A model for glycopeptide recognition

As described above, the high-resolution diffraction results obtained in this study identified the presence of a bound MES molecule in the antigen-combining site. However, in the crystal structures, the antigen-combining site is also occupied by a β-strand from the H-chain constant region (C_H1) of a symmetry-related Fab molecule (Fig. 5*a*). Significantly, the sequence of this β-strand, Ser202′-Ser*′-Thr*′-Lys*′-Val206′, is very similar to that of the natural GPA N-terminal peptide sequence (Fig. 1*a*), with two of the three residues that correspond to O-glycosylation sites of GPA (marked with asterisks) occupied by either GOL or MES (Fig. 5*a*). In particular, Ser203′ and Lys205′ interact with and are adjacent to GOL and MES, respectively. GOL was used as a cryoprotectant for the crystal and was observed during structure refinement (Table 1 and Fig. 2*a*); its hydroxyl-rich structure has been noted previously to be analogous to that of a sugar (Unligil & Rini, 2000).

Collectively, the 'surrogate ligands' described above provide the basis for a model of glycopeptide binding by NNA7 that also accounts for the loss in affinity resulting from the G91S mutation (Fig. 5*b*). In this model, the β-strand and GOL molecule serve as putative mimics of the respective peptide and sugar moieties of the GPA antigen. Specifically, these ligands form site-directed and unique contacts only to CDR residues derived from the L (*e.g.* His27D, Tyr32 and Gly91) and H (*e.g.* Ser31 and Tyr98) chains (Fig. 5*a*). As such, the details and location of these interactions were highly suggestive of a glycopeptide 'readout' model. Notably, no comparable interactions were observed at any other location in the structure. With regard to the interaction details, both wild-type and mutant NNA7 Fab structures indicated that Ser203′ O^{γ} hydrogen bonds with a GOL molecule that mediates a hydrogen-bond contact to the carbonyl O atom of the L chain at Gly/Ser91 (Figs. 5*a* and 5*b*). In GPA the corresponding Ser2* (Fig. 1*b*) is glycosylated, suggesting that the observed hydrogen-bonding pattern may reflect how this O-linked oligosaccharide interacts at Gly91. The interpretation of GOL as a glycan mimic is supported by solution evidence suggesting that disruption of the Gly91 local structure results in a loss in antigen affinity for the mutated NNA7 (Song, Czerwinski *et al.*, 2004). Similarly, the backbone of Lys205′ may be a mimic for Thr4* of GPA, with MES serving as a mimic of the attached glycan (Figs. 1*a* and 5*a*). This model is based on the observation that the Lys205′ amide backbone accepts a hydrogen bond from the H-chain hydroxyl group of Tyr98, *i.e.* a side-chain-independent readout mechanism. Furthermore, if the Lys205′ side chain were substituted with Thr (as in GPA), the peptide would be approximately positioned for an O-glycosidic linkage to a monosaccharide at position O-1 (Fig. 4*c*). These observations further strengthen the hypothesis that MES is a conformational mimic of a hexose. Finally, this proposed model suggests that if Thr204′ of the β-strand crystal contact were equivalent to Thr3 on GPA (Figs. 1*a* and 5*a*), then the O-glycan attached to this residue of GPA points 'away' from the antigen-binding cleft of NNA7 and does not contribute to antigen–antibody binding. Although the specific details of carbohydrate recognition have not been examined for each GPA residue, the model provides a testable hypothesis that can be addressed in future experimental studies.

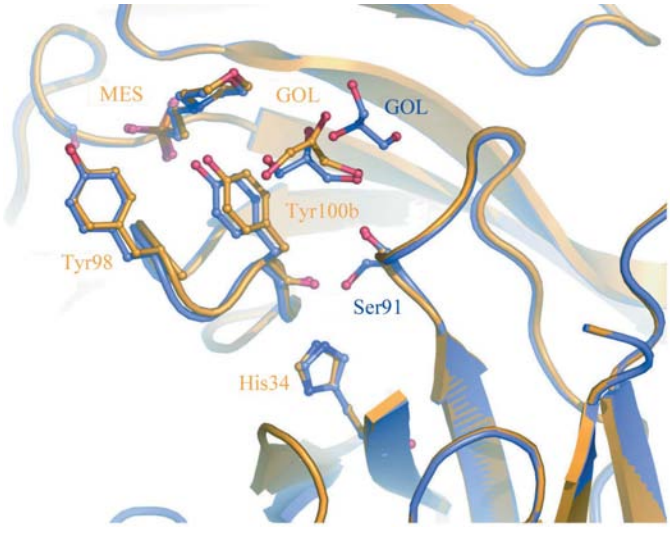


Figure 6
Superposition of the structures of the antigen-combining sites of NNA7 (yellow) and NNA7-G91S (blue). Differences in the structures are accentuated by changes in the label color.

3.6. The effect of the G91S mutation on antigen binding

The Gly-to-Ser substitution at L-chain position 91 significantly lowers the binding affinity of the mutated NNA7 Fab fragment for the N antigen (Song, Czerwinski *et al.*, 2004). A comparison of the wild-type and mutant structures (Figs. 5*a* and 5*b*) revealed two immediate effects of this mutation upon the antigen-combining site. Firstly, a single hydrogen-bond contact is lost between His27D of the L chain and Ser203′ in C_H1, the putative mimic of the N antigen (Figs. 5*a* and 5*b*). Secondly, the longer side chain of L-chain Ser91 in NNA7-G91S establishes two new hydrogen-bond contacts: with the His34 imidazole ring of the L chain and the carbonyl O atom of Tyr98 of the H chain (blue lines in Fig. 5*b*). The latter two contacts are absent in the wild-type structure and appear to make the L-chain CDR3 more rigid. Consequently, the larger conformationally restricted side chain of Ser91 pushes the H-chain CDR3 outwards, causing a significant shift at the L-chain/H-chain interface. This effect accounts for the significant r.m.s.d.s of 0.45 and 0.40 Å in the pairwise superposition of CDR3s at amino acids Ser91 and Tyr100b (Table 2). The overall effect of the Gly-to-Ser substitution is the generation of a slightly wider antigen-binding cleft in the mutant, resulting in the trapping of an additional GOL molecule in the antigen-combining site (Fig. 6). In addition, the movement of the H chain is greater than that of the L chain, as shown by the r.m.s.d.s for these regions of 0.82 and 0.69 Å, respectively. The

net effect is that the movement of the H chain is propagated in both directions throughout V_H (Figs. 3*a* and 3*b*), which is likely to have significant consequences for antigen binding.

4. Summary and conclusions

Multiple antibodies recognize glycopeptide antigens comprising peptide and carbohydrate components. However, none of these interactions has yet been characterized at the molecular level. As such, we solved the structure of the NNA7 Fab fragment, which recognizes the human blood-group N antigen, to elucidate the molecular interactions with the clinically relevant GPA-encoded glycopeptide antigen. Structures of the wild-type NNA7 Fab fragment and of the NNA7-G91S loss-of-function mutant were determined by the molecular-replacement method and refined to high resolution. Both crystal structures identified an MES buffer molecule trapped in the antigen-combining site that appeared to be a structural mimic of galactose. Analysis of the crystal-packing interactions suggested a mode of glycopeptide antigen binding that could account for the deleterious effects of the G91S substitution in the L chain. Diminished antigen binding by NNA7-G91S was attributed to steric clashes between the Ser91 side chain of the L chain and the backbone of the H-chain CDR3. The resulting clash caused a subtle backbone rearrangement that propagated through the V regions. This model provides testable hypotheses for future structure–function studies and establishes the structural groundwork for crystallographic analysis of the NNA7 Fab fragment liganded with its *bona fide* glycopeptide antigen, which is necessary to validate the conclusions presented in this study.

We thank Ms Celeste MacElrevey for helpful suggestions and the staff of the A1 beamline at CHESS for assistance with X-ray data collection. CHESS is supported by the National Science Foundation under award DMR-0225180 and the National Institutes of Health through its National Center for Research Resources under award RR-01646. This work was supported in part by NIH grants P01 HL54516 (SLS) and GM63162 (JEW), as well as institutional support from the University of Rochester (JEW and SLS) and Columbia University (SLS). KX is an Elon Huntington Hooker Graduate Fellow of the University of Rochester.

References

Anstee, D. J. & Edwards, P. A. (1982). *Eur. J. Immunol.* **12**, 228–232.
Anstee, D. J. & Lisowska, E. (1990). *J. Immunogenet.* **17**, 301–308.
Bigbee, W. L., Langlois, R. G., Vanderlaan, M. & Jensen, R. H. (1984). *J. Immunol.* **133**, 3149–3155.
Bigbee, W. L., Vanderlaan, M., Fong, S. S. & Jensen, R. H. (1983). *Mol. Immunol.* **20**, 1353–1362.
Blackall, D. P., Ugorski, M., Pahlsson, P., Shakin-Eshleman, S. H. & Spitalnik, S. L. (1994). *J. Immunol.* **152**, 2241–2247.
Blumenfeld, O. O. & Adamany, A. M. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 2727–2731.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Bush, C. A., Martin-Pastor, M. & Imberty, A. (1999). *Annu. Rev. Biophys. Biomol. Struct.* **28**, 269–293.
Cygler, M., Rose, D. R. & Bundle, D. R. (1991). *Science*, **253**, 442–445.
Czerwinski, M., Blackall, D. P., Abrams, W. R., Rubocki, R. J. & Spitalnik, S. L. (1994). *Mol. Immunol.* **31**, 279–288.
Czerwinski, M., Krop-Watorek, A., Siegel, D. L. & Spitalnik, S. L. (1999). *Transfusion*, **39**, 364–371.
Czerwinski, M., Siemaszko, D., Siegel, D. L. & Spitalnik, S. L. (1998). *J. Immunol.* **160**, 4406–4417.
Duk, M., Sticher, U., Brossmer, R. & Lisowska, E. (1994). *Glycobiology*, **4**, 175–181.
Fraczkiewicz, R. & Braun, W. (1998). *J. Comput. Chem.* **19**, 319–333.
French, D. L., Laskov, R. & Scharff, M. D. (1989). *Science*, **244**, 1152–1157.
Hanisch, F. G., Stadie, T. & Bosslet, K. (1995). *Cancer Res.* **55**, 4036–4040.
Jaskiewicz, E., Czerwinski, M., Syper, D. & Lisowska, E. (1994). *Blood*, **84**, 2340–2345.
Jaskiewicz, E., Moulds, J. J., Kraemer, K., Goldstein, A. S. & Lisowska, E. (1990). *Transfusion*, **30**, 230–235.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Judd, W. J. (1979). *Transfusion*, **19**, 768–772.
Kabat, E. A., Wu, T. T. & Bilofsky, H. (1977). *J. Biol. Chem.* **252**, 6609–6616.
Karsten, U., Diotel, C., Klich, G., Paulsen, H., Goletz, S., Muller, S. & Hanisch, F. G. (1998). *Cancer Res.* **58**, 2541–2549.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J.. Appl. Cryst.* **26**, 283–291.
Lisowska, E., Duk, M. & Dahr, W. (1980). *Carbohydr. Res.* **79**, 103–113.
Lisowska, E., Messeter, L., Duk, M., Czerwinski, M. & Lundblad, A. (1987). *Mol. Immunol.* **24**, 605–613.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Pflugrath, J. W. (1999). *Acta Cryst.* D**55**, 1718–1725.
Rubocki, R. & Milgrom, F. (1986). *Vox Sang.* **51**, 217–225.
Sadler, J. E., Paulson, J. C. & Hill, R. L. (1979). *J. Biol. Chem.* **254**, 2112–2119.
Sancho, J. M., Pujol, M., Fernandez, F., Soler, M., Manzano, P. & Feliu, E. (1998). *Br. J. Haematol.* **103**, 268–269.
Schiffer, M., Girling, R. L., Ely, K. R. & Edmundson, A. B. (1973). *Biochemistry*, **12**, 4620–4631.
Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S. H. & Rini, J. M. (1998). *J. Biol. Chem.* **273**, 13047–13052.
Song, S. C., Czerwinski, M., Wojczyk, B. S. & Spitalnik, S. L. (2004). *Transfusion*, **44**, 173–186.
Song, S. C., Xie, K., Czerwinski, M., Spitalnik, S. L. & Wedekind, J. E. (2004). *Acta Cryst.* D**60**, 788–791.
Telischi, M., Behzad, O., Issitt, P. D. & Pavone, B. G. (1976). *Vox Sang.* **31**, 109–116.
Unligil, U. M. & Rini, J. M. (2000). *Curr. Opin. Struct. Biol.* **10**, 510–517.
Wasniowska, K., Drzeniek, Z. & Lisowska, E. (1976). *Biochem. Biophys. Res. Commun.* **76**, 385–390.
Wasniowska, K., Duk, M., Steuden, I., Czerwinski, M., Wiedlocha, A. & Lisowska, E. (1988). *Arch. Immunol. Ther. Exp.* **36**, 623–632.
Wedemayer, G. J., Patten, P. A., Wang, L. H., Schultz, P. G. & Stevens, R. C. (1997). *Science*, **276**, 1665–1669.
Wu, T. T. & Kabat, E. A. (1970). *J. Exp. Med.* **132**, 211–250.