Diverse T cell receptor gene usage in HLA-DQ8-associated celiac disease converges into a consensus binding solution

Jan Petersen[1,2], Yvonne Kooy-Winkelaar[3], Khai Lee Loh[1,2], Mai Tran[1,2], Jeroen van Bergen[3], Frits Koning[3*] Jamie Rossjohn[1,2,4,5*] & Hugh H. Reid[1,2*]

[1]Infection and Immunity Program & The Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria 3800, Australia.

[2]Australian Research Council Centre of Excellence in Advanced Molecular Imaging, Monash University, Clayton, Victoria 3800, Australia.

[3]Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden 2333 ZA, The Netherlands.

[4]Institute of Infection and Immunity, Cardiff University School of Medicine, Heath Park, Cardiff CF14 4XN, UK.

[*] Joint senior and corresponding authors. E-mail: hugh.reid@monash.edu, f.koning@lumc.nl, jamie.rossjohn@monash.edu

[5] Lead Contact

Running title: DQ8-mediated Celiac Disease

Abstract

In HLA-DQ8-associated celiac disease, TRAV26-2$^+$-TRBV9$^+$ and TRAV8-3$^+$-TRBV6$^+$ T cells recognize the immunodominant DQ8-glia-α1 epitope, whereupon a non-germline encoded arginine residue played a key role in binding HLA-DQ8-glia-α1. Whether distinct TCR recognition modes exist for gliadin epitopes remains unclear. TCR repertoire analysis revealed populations of HLA-DQ8-glia-α1 and HLA-DQ8.5-glia-γ1 restricted TRAV20$^+$-TRBV9$^+$ T cells that did not possess a non-germline encoded arginine residue. The crystal structures of a TRAV20$^+$-TRBV9$^+$ TCR-HLA-DQ8-glia-α1 complex and two TRAV20$^+$-TRBV9$^+$ TCR-HLA-DQ8.5-glia-γ1 complexes were determined. This revealed the differential specificity towards DQ8-glia-α1 and DQ8.5-glia-γ1 was governed by CDR3β loop mediated interactions. Surprisingly, a germline-encoded arginine residue within the CDR1α loop of the TRAV20$^+$ TCR substituted for the role of the non-germline encoded arginine in the TRAV26-2$^+$-TRBV9$^+$ and TRAV8-3$^+$-TRBV6$^+$ TCRs. Thus in celiac disease, the responding TCR repertoire is driven by a common mechanism that selects for structural elements within the TCR that have convergent binding solutions in HLA-DQ8-gliadin recognition.

Introduction

Celiac disease (CD) is a chronic inflammatory disease of the small intestine that affects ~1.5% of Caucasians (Abadie et al., 2011; Di Sabatino and Corazza, 2009; Koning et al., 2015; Sollid and Jabri, 2013; Stamnaes and Sollid, 2015). A small number of gliadin and glutenin peptides derived from gluten elicit this inflammatory CD4$^+$ T cell response (Abadie et al., 2011; Di Sabatino and Corazza, 2009; Tye-Din et al., 2010). CD is predominantly limited to genetically predisposed individuals, namely those who express HLA-DQ2 (DQA1*0501-DQB1*0201) and/or HLA-DQ8 (DQA1*0301-DQB1*0302) (Abadie et al., 2011; Karell et al., 2003). A third haplotype associated with CD is found in patients who are *HLA-DQ2*$^+$ and *HLA-DQ8*$^+$(Kooy-Winkelaar et al., 2011). These individuals can produce the trans dimer, HLA-DQ8.5 (DQA1*0501-DQB1*0302), which is formed by the HLA-DQ2.5 α–chain and the HLA-DQ8 β–chain, and has a unique peptide-binding repertoire (van Lummel et al., 2012).

The use of HLA-DQ tetramers has enabled an understanding of the responding T cell repertoire towards these HLA-DQ-restricted gliadin determinants. Namely, the HLA-DQ2-gliadin and HLA-DQ8-gliadin response in CD is characterized by biased T cell receptor (TCR) gene usage in the expanded CD4$^+$ population, but not in healthy subjects (Broughton et al., 2012; Dahal-Koirala et al., 2016; Petersen et al., 2014; Petersen et al., 2015; Qiao et al., 2014; Qiao et al., 2011)

(Christophersen et al., 2016). In HLA-DQ8^{+} CD patients the T cell population recognizing the immunodominant epitope, HLA-DQ8-glia-α1, is characteristically enriched in TRAV26-2^{+}-TRBV9^{+} T cells and, to a lesser extent, TRBV6^{+} T cells (Broughton et al., 2012; Petersen et al., 2015). Consistent with the biased TCR repertoire, structural and biophysical analysis of the HLA-DQ8-glia-α1 interaction with three representative TRAV26-2^{+}-TRBV9^{+} TCRs and one TRAV8-3^{+}-TRBV6^{+} TCR revealed a high degree of structural and functional conservation (Broughton et al., 2012; Petersen et al., 2015). Here, hypervariability within the CDR3 regions correlated with the TCRs avidity towards the gliadin determinants and their reactivity to its deamidation states. Of note was the presence of a non-germline encoded arginine residue in the CDR3α and CDR3β loops of these TRBV9^{+} and TRBV9^{-} TCRs, respectively. These CDR3-derived arginine residues were crucial for recognition of HLA-DQ8-glia-α1 (Broughton et al., 2012; Petersen et al., 2015). Similarly, the T cell response towards the immunodominant DQ2.5-glia-α2 epitope was characterized by biased TRAV26-1^{+}-TRBV7-2^{+} TCR usage, and was accompanied by a conserved non-germline encoded Arg residue within the CDR3β loop (Dahal-Koirala et al., 2016; Qiao et al., 2014). Structural studies on the TRBV7-2^{+} TCR-HLA-DQ2.5-glia-α2 complexes revealed that this arginine played a central role in contacting the glia-α2 determinant and the HLA-DQ2.5 molecule (Petersen et al., 2014). Different TCR biases were also observed towards distinct HLA-DQ2.5 gliadin epitopes. Namely, the T cell response to DQ2.5-glia-α1a and DQ2.5-glia-ω2 determinants exhibited frequent usage of the TRAV4 gene, but no conserved usage in the Arg residue within the CDR3β loop was apparent (Petersen et al., 2014; Qiao et al., 2014; Qiao et al., 2011). The crystal structure of a TRAV4^{+} TCR-HLA-DQ2.5-glia-α1a complex provided insight into the DQ2.5-glia-α1a specificity and indicated that different gliadin epitopes are associated with distinct patterns of TCR usage (Petersen et al., 2014).

While the immunodominant epitopes in CD are derived from the α-gliadin fraction of wheat, a sub-population of gliadin reactive T cells respond to epitopes from the abundant γ-gliadin fraction and contributes to the pathogenic T cell response (Kooy-Winkelaar et al., 2011) (Molberg et al., 1998). HLA-DQ8 and HLA-DQ8.5 can present both the immunodominant DQ8-glia-α-1 (SG**EGSFQPSQE**NP) and DQ8.5-glia-γ-1 (QQ**PQQSFPEQE**RP) epitopes from their respective gliadin fractions (Kooy-Winkelaar et al., 2011). Furthermore T cell clones display cross-reactivity to HLA-DQ8 and HLA-DQ8.5 as they respond to their respective epitopes when presented by either HLA molecule (Kooy-Winkelaar et al., 2011). This study included one TRAV20^{+}-TRBV9^{+} HLA-DQ8.5-glia-γ-1 restricted T cell clone that contained an arginine in the CDR3β loop (Kooy-Winkelaar et al., 2011). How TCRs can interact with the DQ8.5-glia-γ-1 determinant remains unclear however.

To ascertain the extent of CDR3 arginine residue selection in HLA-DQ8-gliadin recognition, we expanded our TCR repertoire and structural analysis of TCRs from T cell clones isolated from small intestinal biopsies of CD patients. We provide a structural basis for the recognition of DQ8-glia-α-1 and DQ8.5-glia-γ-1 in the context of HLA-DQ8 and HLA-DQ8.5 by TRAV20^{+}-TRBV9^{+} TCRs. We show that the functional element of the canonical CDR3 arginine is substituted by a germline encoded arginine within the TCR α-chain that acts as an adaptor for the two distinct gliadin peptides. Thus, diverse TCR gene usage in CD converges to a consensus mode of recognition, which has implications for the potential druggability of this interaction as a novel means of immunotherapy to treat CD.

Results

TRAV20^{+}-TRBV9^{+} TCR recognition of HLA-DQ8-glia-α1 and HLA-DQ8/8.5-glia-γ1

TCR sequence analysis of HLA-DQ8 restricted TRBV9^{+} T cell clones revealed two TRAV20^{+}-TRBV9^{+} TCRs from one individual (patient Bel; HLA-DQ8; HLA-DQA1*03/*03-DQB1*0301/*0302), which each recognized a different gliadin epitope. Namely, T cell clone Bel502 recognised HLA-DQ8-glia-α1 (SG**EGSFQPSQE**NP), while T cell clone Bel602 bound HLA-DQ8-glia-γ1 (QQ**PQQSFPEQE**RP) (**Table 1**). The Bel502 and Bel602 TCRs differed only in their CDR3 sequences, and thus CDR3 variability accounted for recognition of these two sequence distinct gliadin determinants. Inspection of the CDR3 sequences revealed that both Bel502 and Bel602 TCRs lacked the non-germline arginine present in the CDR3α loop of the majority of HLA-DQ8-glia-α1 restricted TRAV26^{+}-TRBV9^{+} TCRs (Broughton et al., 2012; Petersen et al., 2015). Although, from a previous study (Kooy-Winkelaar et al., 2011) we had identified an HLA-DQ8.5-glia-γ1 restricted TRAV20^{+}-TRBV9^{+} T cell clone (T15, Patient T; HLA-DQ8.5, DQA1*0505/*0401 DQB1*0302/*0402), which carried an arginine in the CDR3β □□□□ (**Table 1**). Accordingly, the TRAV20-TRBV9 pairing provided the framework for two distinct modes of peptide specificity that were solely determined by the CDR3 regions.

Epitope specificity of the TRAV20^{+}-TRBV9^{+} TCRs

To probe the differences in peptide recognition by the three TRAV20^{+}-TRBV9^{+} TCRs (Bel502, Bel602 and T15) we measured the impact of single alanine substitutions in each position of the peptide on recognition by Bel502 and Bel602 and T15 in a T cell proliferation assay (**Figure 1A-C**). Similar to HLA-DQ8-glia-α1 restricted TRBV9^{+} T cell clones tested previously (Broughton et al., 2012; Petersen et al., 2015), Bel502 was very sensitive to alanine substitution and effectively failed to recognize peptides with a substitution in positions p3 to p6 and p8 to p10 (**Figure 1A**).

Conversely, the HLA-DQ8-glia-γ1 restricted Bel602 appeared to be more tolerant to changes in the peptide, as it only failed to recognize peptides with substitutions in p3, p5 and p7 (**Figure 1B**). As reported previously, the T15 TCR was sensitive to substitution of p3 and p5 to p9, (Kooy-Winkelaar et al., 2011) (**Figure 1C**).

Next, we used surface plasmon resonance (SPR) measurements to determine the affinities of the purified TCRs from Bel502, Bel602 and T15 clones, towards HLA-DQ8-glia-α1, HLA-DQ8.5-glia-α1, HLA-DQ8-glia-γ1 and HLA-DQ8.5-glia-γ1 (**Figure 1D-F**). The Bel502 bound to HLA-DQ8-glia-α1 and HLA-DQ8.5-glia-α1 with a K_D of 2.8 ± 0.1μM and 19.5 ± 0.7 μM, respectively (**Figure 1D**), but had no measurable affinity for DQ8/8.5-glia-γ1 (data not shown). Conversely, the Bel602 TCR and T15 TCR showed no measurable binding to HLA-DQ8-glia-α1 or HLA-DQ8.5-glia-α1 (data not shown). The affinities for HLA-DQ8-glia-γ1 and HLA-DQ8.5-glia-γ1 were K_D = 13.9 ± 0.4 μM and 4.7 ± 0.2 μM, respectively, for Bel602 TCR (**Figure 1E**), and K_D = 3.6 ± 0.2 μM and 2.0 ± 0.1 μM, respectively, for the T15 TCR (**Figure 1F**). Therefore, while the three TCRs could recognize the cognate epitopes bound to HLA-DQ8 and HLA-DQ8.5, there was no cross-reactivity between the gliadin epitopes.

Structure of TRAV20$^+$-TRBV9$^+$ Bel502 TCR–HLA-DQ8-glia-α1 complex

Next, to establish how the TRAV20$^+$-TRBV9$^+$ TCR (Bel502) engaged HLA-DQ8-glia-α1, we determined it's ternary complex to 2.6Å resolution (**Figures 2A, 2B, 3, Table 2**). This provided us an opportunity to compare it to a previously determined TRAV26-2$^+$-TRBV9$^+$ TCR-HLA-DQ8-glia-α1 complex (**Figures 2C & 2D**; (Broughton et al., 2012)). Of interest, while an Arg within the CDR3α loop of the SP3.4 TRAV26-2$^+$-TRBV9$^+$ TCR was critical in binding HLA-DQ8-glia-α1, the Bel502 TCR did not possess an Arg within its CDR3α loop. Thus, despite its common TRBV9 usage with the TRAV26-2$^+$-TRBV9$^+$ TCRs, it was unclear how Bel502 TCR would interact with HLA-DQ8-glia-α1.

The Bel502 TCR engaged HLA-DQ8-glia-α1 approximately 70° with respect to the long axis of the antigen binding cleft (**Figure 2B**). In comparison to the TRAV26-2$^+$-TRBV9$^+$ TCR, the Bel502 TCR docked 2-3Å towards the glia-α1 peptide N-terminus (Petersen et al., 2015). Despite this shifted docking, the footprint of the TCR β-chain, and associated key contacts with HLA-DQ8 and the DQ8-glia-α1 peptide, was very similar to that observed in the TRAV26-2$^+$-TRBV9$^+$ ternary complexes (**Figures 2A-D**; (Broughton et al., 2012)). For example, the positioning of the TRBV9

germline encoded residues Leu37β and Tyr57β, clustered within 1-1.5 Å in the TRAV26-2$^+$-TRBV9$^+$ and TRAV20$^+$-TRBV9$^+$ complexes (**Figures 3B, 4B**) (Petersen et al., 2015).

The buried surface area (BSA) upon complexation by the Bel502 TCR was approximately 1120Å^2, which was higher than that typically observed for the TRAV26$^+$-2-TRBV9$^+$ TCR (circa 900Å^2) (Broughton et al., 2012). This was principally related to the footprint made by the Bel502 TCR α-chain being larger (56% BSA) than that mediated by the TCR α-chain (average BSA of 45%) within the TRAV26-2$^+$ TCR-HLA-DQ8-glia-α1 complexes (Broughton et al., 2012). Indeed, the conformation of the TRAV20$^+$ CDRα loops were distinct from those in the TRAV26$^+$ TCR, with the functional involvement of each CDRα loop appearing to be reassigned (**Figures 3A, 3C**).

The CDR3α loop in the Bel502 TCR mostly avoided the central region of the Ag-binding cleft, and thus was quite distinct from the TRAV26$^+$-TRBV9$^+$ TCRs, whereby a conserved CDR3α arginine played a critical role in binding to P5-Gln and Phe58α from HLA-DQ8 (**Figure 3E**). Instead, the N terminal region of the CDR3α loop of the Bel502 TCR was oriented towards the N-terminus of the peptide while its C-terminal portion made extensive van der Waals contacts with the N-terminal end of the HLA-DQ8 α-helix, including contacts with Phe58α (**Figure 3C, 3F**). Although the CDR3α loop of the Bel502 TCR did not possess an Arg residue, this was compensated by an Arg residue encoded within the CDR1α loop. Here, Arg37α reached into the peptide-binding cleft, where its guanidinium group H-bonded to p3-Ser and stacked against the aromatic sidechain HLA-DQ8 Phe58α (**Figure 3F**). Notably, the interaction of Arg37α was analogous to the CDR3α arginine residue in the TRAV26-2$^+$-TRBV9$^+$ TCRs (**Figure 3E**) (Broughton et al., 2012; Petersen et al., 2015). Thus, TCR interaction with HLA-DQ8-glia-α1 appears to be driven by convergent binding solutions.

TCR recognition of HLA-DQ8.5-glia-γ1

To understand how the responding T-cell repertoire bound HLA-DQ8.5-glia-γ1, we determined the crystal structures of the Bel602 TCR-HLA-DQ8.5-glia-γ1 and T15 TCR-HLA-DQ8.5-glia-γ1 ternary complexes with resolutions of 2.0 Å and 2.9Å, respectively (**Figures 2E, 2F, 2G, 2H, 4, 5 and Table 2**). Here, the DQ8.5-glia-γ1 (PQQSFPEQE) determinant differed from the HLA-DQ8/8.5-glia-α1 sequence (EGSFQPSQE). The majority of differences between the α-chain sequences of HLA-DQ8 (DQA1*0301/DQB1*0302) and HLA-DQ8.5 (DQA1*0501/DQB1*0302) were located distal to the peptide binding cleft (**Figure 6**). However, three differences were located within the peptide-binding cleft, creating an altered environment for the peptide N- and C-termini.

The corresponding residues Glu31α, Arg52α and Ile72α in HLA-DQ8, were Gln31α, a deletion in position 52 and Ser72α in HLA-DQ8.5 In HLA-DQ8-glia-α1 Arg52α and Glu31α formed a part of the p1-pocket, which imparts a preference for an acidic p1 anchor residue (Henderson et al., 2007b). In HLA-DQ8.5-glia-γ1 this position is occupied by the adjacent Phe51α (**Figure 6**), which contributes to a more hydrophobic p1 pocket (Kim et al., 2004; van Lummel et al., 2012). The DQ8.5-glia-γ1 peptide was bound in an extended conformation with residues p1-Pro, p4-Ser, p6-Pro, p9-Glu occupying the anchor positions in HLA-DQ8.5, with p3-Gln and p7-Glu acting as ancillary anchors (**Figures 4E, 5E**).

Compared to the Bel502 TCR-HLA-DQ8-glia-α1 structure, the Bel602 TCR docking atop HLA-DQ8.5-glia-γ1 was shifted by 3-4Å towards the peptide C-terminus (**Figures 2B, 2F**). Nevertheless, Leu37β and Tyr57β from the CDR1β and CDR2β loops participated in similar interactions in the TRBV9^{+} TCR ternary structures (**Figures 3B, 4B, 5B**). Likewise, the interactions of CDR2α Tyr57 and of the proximal α-framework residue Lys66 were largely maintained across the TRAV20^{+} TCR ternary structures (**Figures 3A, 4A, 5A**). Hence, the interactions of both TRAV20 and TRBV9 encoded residues were maintained despite moderate shifts in the TCR docking position. Within the Bel602 TCR ternary complex, the CDR1α-mediated interactions were dominated by Arg37α, which lay in an extended conformation against the HLA-DQ8.5 β-chain residues Thr77β and Val78β, formed a salt bridge with Glu74β (**Figure 4A**), and interacted with peptide residues, p5-Phe, and p3-Gln (**Figure 4E**). Thus, Arg37α occupied a region with a striking resemblance to that engaged by the arginine in the TCR-HLA-DQ8- glia-α1 structures (**Figure 3E**; (Broughton et al., 2012; Petersen et al., 2015)).

The CDR3α of Bel602 interacted with both the HLA-DQ8.5 α-chain and the peptide, and contacted the HLA-DQ8.5 β-chain to a lesser extent (**Figure 4C**). The CDR3α loop adopted an S-shaped conformation, placing Asn112α and Tyr113α on either side of Phe58α. Moreover, Tyr113α H-bonded to Asp55α at the N-terminus of the α-chain helix (**Figure 4C**), whilst Asn112α was situated centrally above the peptide and H-bonded to the peptide residues p2-Gln and p3-Gln (**Figure 4E**), as well as occupying the space between the aromatic sidechains of Phe58α (**Figure 4C**) and p5-Phe (**Figure 4E**). Since the DQ8-glia-α1 and DQ8-glia-γ1 peptides differ in p2, p3 and p5, Asn112α occupied a key position that underpins the peptide specificity of Bel602. The CDR3β sat atop the peptide and interacted with p5-Phe, p7-Glu, and p8-Gln (**Figure 4E**), and moreover, Val108β and Tyr114β, contacted the HLA-DQ8.5 β-chain (**Figure 4D**). Notably, Tyr114β formed H-bonds to both the ancillary peptide anchor p7-Glu (**Figures 4E**) and Arg70β (**Figures 4D**).

Therefore both the CDR3α and CDR3β loops contributed to the peptide specificity of the Bel602 TCR by directly binding to peptide residues and HLA residues with conformations that distinguish HLA-DQ8-glia-γ1 from HLA-DQ8-glia-α1. Conversely, the lack of affinity of the Bel602 TCR for HLA-DQ8-glia-α1 can be attributed to CDR3 interactions that would sterically hinder the positioning the key residues, such as Arg37α, relative to the DQ8-glia-α1 peptide.

T15 TCR-HLA-DQ8.5-glia-γ1 complex

Unlike the Bel602 TCR, the T15 TCR possessed an Arg within its CDR3β loop, and it was of interest to compare how the T15 TCR and Bel602 TCR bound the DQ8.5-glia-γ1 determinant. The T15 TCR (**Figure 2H**) engaged HLA-DQ8.5-glia-γ1 in a similar docking position to the Bel602 TCR (**Figure 2F**), however its docking angle differed by approximately 8°. The salient features at the T15 TCR-HLA-DQ8.5-glia-γ1 and Bel602 TCR-HLA-DQ8.5-glia-γ1 interfaces were similar, including the role of the germline-encoded Arg37α from the TCR α-chain (**Figures 4A, 4F, 5A, 5E**) as well as Leu37β and Tyr57β from the TCR β-chain (**Figures 4B, 4F, 5B, 5E**). However, differences within the respective CDR3 loops manifested in some differing features at the TCR-HLA-DQ8.5-glia-γ1 interfaces.

The SGGS sequence in the T15 CDR3α loop (**Table 1**) made multiple van der Waals contacts with Phe58α and H-bonded to Asp55α of HLA-DQ8.5 (**Figure 5C**). Interactions with the peptide were mediated chiefly by Tyr113α, which H-bonded to p2-Gln and occupied the space between Phe58α of HLA-DQ8.5 and p5-Phe of the peptide (**Figure 5C, 5E**). Thus, Asn112α of the Bel602 TCR and Tyr113α of the T15 TCR played analogous structural roles. The T15 CDR3β loop mirrored the interactions of the Bel602 CDR3β, despite being structurally distinct. The T15 CDR3β formed an intricate interface with the peptide and the HLA-DQ8.5 β-chain, whilst peripherally contacting the HLA-DQ8.5 α-chain (**Figures 5D, 5E**). Its peptide contacts involved p5-Phe, p6-Pro, p7-Glu and p8-Gln (**Figure 5E**), and therefore it contributed to peptide specificity in much the same way as CDR3β in the Bel602 TCR (**Figure 4E**). The CDR3β arginine, Arg109, occupied a central position underneath the loop, where it H-bonded to the peptide backbone of p6-Pro, formed a salt bridge with p7-Glu and interacted with Arg70β of the HLA-DQ8.5 β-chain (**Figures 5D, 5E**). The electrostatic repulsion between Arg109β and Arg70β was compensated by Asp114β, forming a salt bridge with Arg70β (**Figure 5D, 5E**). Therefore, T15 CDR3β residues Arg109β and Asp114β formed part of a charged network that fulfilled the same function as Tyr114β from the CDR3β loop of the Bel602 TCR, namely to lock on to the ancillary anchor residue p7-Glu.

Energetics in TCR-DQ8-glia-α1 and TCR-DQ8.5-glia-γ1 interactions

To determine the energetic contributions of individual TCR residues to pHLA recognition we undertook a broad based mutagenesis/SPR approach, as conducted previously in other TCR-peptide-MHC (pMHC) systems (Borg et al., 2005; Gras et al., 2012). Namely, we generated nine point mutants of the Bel502 TCR (Arg37αAla, Tyr57αAla, Lys66αAla, Asn109αAla, Asn110αAla, Asn114αAla, Leu37βAla, Tyr57βAla, Arg66βAla) and 11 of the Bel602 TCR (Arg37αAla, Tyr57αAla, Lys66αAla, Met109αAla, Asn112αAla, Tyr113αAla, Leu37βAla, Tyr57βAla, Arg66βAla, Tyr114βPhe, Tyr114βAla) and subsequently determined their affinities for HLA-DQ8-glia-α1, HLA-DQ8.5-glia-α1 and HLA-DQ8.5-glia-γ1 (**Figure 7, Figure S1**). Two of the three Bel502 TCR β-chain substitutions, Leu37β and Tyr57β, were essential for recognition of HLA-DQ8-glia-α1, consistent with the role these positions played in mediating TRAV26-2^{+}-TRBV9^{+} recognition of HLA-DQ8-glia-α1. Alanine substitution of the Bel502 TCR α-chain residues Arg37α, Tyr57α, Lys66α, Asn109α, and Asn114α had a significant impact on recognition of HLA-DQ8-glia-α1 (> 5x increased K_D) (**Figures 7A, 7B**), thereby forming one "energetic hot spot" of this interaction. A second energetic hotspot of the Bel502 TCR α-chain was formed by Tyr57α from the CDR2α loop, and the α-framework residue Lys66α, which both interacted with HLA-DQ8 β-chain helix. The affinities of the Bel502 mutants for HLA-DQ8.5-glia-α1 were altogether lower than for HLA-DQ8-glia-α1, however the relative energetic contributions of individual Bel502 residues on recognition of HLA-DQ8-glia-α1 HLA-DQ8.5-glia-α1 followed the same pattern, suggesting that the trans encoded α-chain of HLA-DQ8.5 did not interfere with individual TCR-pMHC interactions, but rather had a global effect on affinity (**Figures 7A**).

The residues most important for recognition of HLA-DQ8.5-glia-γ1 by Bel602 were the TCR α-chain residues Arg37α, Tyr57α, Asn112α and Tyr113α, and the TCR β-chain residues Tyr57β and Tyr114β. Moreover, substitution of either α-framework residue Lys66α and CDR1β Leu37β had a significant impact (**Figures 7C, 7D**). Therefore, Bel602 appeared to depend on the same pattern of germline encoded residues for recognition of HLA-DQ8.5-glia-γ1 as Bel502 used for recognition of HLA-DQ8-glia-α1.

The energetically important non-germline encoded residues in Bel602 were Asn112α, Tyr113α and, to a lesser degree, Met109α from the CDR3α loop as well as Tyr114β from the CDR3β loop, which broadly reflected that found for the Bel502 TCR. To exclude the possibility that the loss of affinity of the Tyr114βAla by was caused by indirect effects, we generated the Tyr114βPhe mutant and

observed the same loss of recognition. Therefore, recognition of the ancillary anchor point p7-Glu by CDR3β Tyr114 constituted an energetic hotspot absent in Bel502.

Interestingly, the key energetic contacts made by non-germline encoded residues of Bel502 TCR and of Bel602 TCR generally involved peptide residues that differed between DQ8-glia-α1 and DQ8.5-glia-γ1, thereby providing a basis for understanding the epitope specificity of these TCRs. Therefore, the selection of TRAV20^{+}-TRBV9^{+} TCRs by two different gliadin epitopes can be rationalized through the tolerance of the respective TCR germline- HLA-DQ8/8.5 interactions to moderate changes in the TCR docking geometry. This positional flexibility facilitates selection of fitting non-germline residues that interact with both the peptide and HLA-DQ8 and HLA-DQ8.5.

Discussion

Although humans have been consuming wheat for thousands of years and about 40% of the Western population are HLA DQ2^{+}, HLA-D8^{+}, or both, the incidence of CD is only ~1.5%. Hence, as is proposed for autoimmune diseases, whilst *HLA alleles* are necessary they are not sufficient to precipitate disease. In ‘‘multiple hit models” of autoimmune disease pathogenesis, *HLA alleles* along with other genes and environmental factors/events combine to precipitate disease onset (Koning, 2014). Although gluten is a non-self antigen, the genetic pathways and susceptibility loci of CD have many aspects in common with autoimmune disease (Abadie et al., 2011; Jabri and Sollid, 2009; Koning et al., 2015) and anti-tissue transglutaminase (TG2) positive antisera is a diagnostic criteria for CD. Nevertheless, central to CD is the interaction between the TCR and the dietary antigen presented by HLA DQ2/8, and this study has shed further fundamental insight into this process.

Here we undertook a structural and functional analysis of the TCRs Bel502, Bel602 and T15 with the aims to elucidate the structural basis for 1) the selection of the TRAV20^{+}-TRBV9^{+} pairing by different epitopes; 2) the non-germline driven epitope specificity; 3) the cross-reactivity between HLA-DQ8 and HLA-DQ8.5 and 4) the apparent departure from the archetypical blueprint of CDR3-derived arginine containing TRBV9^{+} TCRs (Broughton et al., 2012; Petersen et al., 2015). Such studies provide important insight into the pathogenesis of CD, and more generally illuminates the fundamental properties of TCR-pMHC interactions (Rossjohn et al., 2015).

We have previously shown that gliadin specific T cells can cross-react with their respective antigenic peptide presented both by HLA-DQ8 (DQA1*0301/DQB1*0302) and by HLA-DQ8.5 (DQA1*0501/DQB1*0302) (Kooy-Winkelaar et al., 2011). The present data allowed us to examine

the structural basis for the HLA-DQ8/HLA-DQ8.5 cross-reactivity of the DQ8-glia-α1 specific TCRs Bel502 and T316, and the DQ8.5-γ1 specific TCRs Bel602 and T15 isolated from two *HLA-DQ8*$^{+}$ and *HLA-DQ8.5*$^{+}$ individuals. Consistent with the observed HLA cross-reactivity, the TCR footprints did not overlap with the differences in the HLA-DQ α-chains in any of the structures, showing that this cross-reactivity is attributable to "missing the differences", and represents a form of molecular mimicry (Macdonald et al., 2009).

We previously reported that in HLA-DQ8 associated CD, TRBV9 germline encoded elements drive selection of HLA-DQ8-glia-α1 restricted TRAV26-2^{+}-TRBV9^{+} TCRs with a non-germline encoded arginine in CDR3α (Broughton et al., 2012; Petersen et al., 2015). The present data adds more weight to this concept by extending it to the selection of TRAV20^{+}-TRBV9^{+} TCRs by the same TRBV9 encoded elements. Intriguingly, TRAV20 has the capacity to use a germline-encoded Arg in an analogous functional role to that of the non-germline encoded Arg in the TRAV26-2^{+}-TRBV9^{+} TCRs. Moreover, Arg37α can adapt to the spatial requirements of two different peptides through subtle repositioning of the TCR docking, and allows for positional variability in the CDR3 loops that determine the peptide specificity of the interaction.

Notably, we and others observed a similar mechanism in HLA-DQ2.5 associated CD, where the immunodominant epitope HLA-DQ2.5-glia-α2 is recognized by a T cell population enriched in TRBV7-2^{+} TCRs with an important arginine residue in the CDR3β region (Petersen et al., 2014; Qiao et al., 2014; Qiao et al., 2011). Based on the combined structural and functional data we hypothesized that the selection of TCRs carrying a non-germline encoded arginine in either the CDR3α or in CDR3β region by immunodominant gliadin epitopes is a defining feature of the T cell immune response in CD and may represent an important clue to understanding disease onset and/or progression (Petersen et al., 2014). Collectively, our present data amounts to a striking convergence of structural elements involved in TCR of different immunodominant gliadin epitopes. Specifically, each epitope favours the selection of TCRs with a functionally conserved arginine in a particular structural position.

The relatively high affinity of HLA-DQ2/8-gliadin restricted TCRs is more in keeping with that observed for TCRs recognising microbial derived antigens (Rossjohn et al., 2015). It has recently been reported that TCRs encoding hydrophobic residues at positions 6 and 7 (P6 and P7) of their CDR3β loops are predisposed to being self-reactive (Stadinski et al., 2016). Using the "Self-reactivity Index" described by Stadinski et al. (2016), only 20% (8 of 40) of the gliadin reactive TCRs we have described (Broughton et al 2012, Petersen et al 2014,2015, this study) have a P6, P7

doublet indicative of a TCR that promotes self-reactivity. Thus, differing TCR-centric mechanisms exist that promote self-reactivity. Moreover, our study represents a salient example of how the germline-encoded regions of a TCR can determine peptide-specificity. Given that a sizeable number of TCRs exhibit germline-encoded interactions with the HLA-bound peptide (Rossjohn et al., 2015), it indicates that the prevailing view in the field that non-germline-encoded interactions (CDR3 loops) determine peptide specificity needs revision (Garcia, 2012).

Considering the broader view of CD, the consistent use of these structural elements in the context of different gliadin epitopes, suggests the presence of a common overarching mechanism in the selection process of biased TCR repertoires observed in this disease. The convergence of the TCR-HLA-DQ8 docking footprints suggests that this site of interaction is potentially druggable, thereby providing an opportunity for the development of novel therapeutics to treat CD.

Methods

Gluten-specific T cell clones

The HLA-DQ8/DQ8.5-glia-α1 specific T cell clone T316 and the HLA-DQ8/DQ8.5-glia-γ1 specific T cell clone T15 have been previously described and were isolated from Patient T, HLA-DQA1*0501/*0401- HLA-DQB1*0302/*0402 (HLA-DQ8.5/DQ4 heterozygous (Kooy-Winkelaar et al., 2011; Petersen et al., 2015). The HLA-DQ8/DQ8.5-glia-α1 specific T cell clone Bel502 has been described previously (Petersen et al., 2015); and was just like the HLA-DQ8/DQ8.5-glia-γ1 specific T cell clone Bel602 T cell clones isolated from patient Bel, HLA-DQA1*03/*03-DQB1*0301/*0302. The study was approved by the Medical Ethics Committees of the Free University Medical Center and the Leiden University Medical Center. Written informed consent was obtained from each subject before enrolment.

Peptides

Peptides were synthesized as described (Kooy-Winkelaar et al., 2011)

T cell proliferation assays

T cell proliferation assays were performed as described (Kooy-Winkelaar et al., 2011). Briefly, assays were performed in triplicate in 150 ml IMDM supplemented with glutamine (Life Technologies/Invitrogen, Grand Island, NY) and 10% human serum in 96-well flat-bottom plates. 30,000 mitomycin C-treated Peptide loaded APC were mixed with 15,000 T cells. Synthetic peptides were used at a final concentration of 6 mg/ml. After 48 h at 37°C, cultures were pulsed with 0.5 mCi [3H]thymidine and harvested 18 h later.

Isolation and sequencing of TCRs

Isolation and sequencing of TCRs was performed as described (Broughton et al., 2012)

Mutagenesis, protein expression and purification.

HLA-DQ8/8.5-glia-α1 (DQ8/8.5-glia-α1: PSG**EGSFQPSQE**NPQ), HLA-DQ8/8.5-glia-γ1 (DQ8/8.5-glia-γ1: Q**PQQSFPEQE**), and HLA-DQ2.5-glia-α1 (DQ2.5-glia-α1: Q**PFPQPELPY**P) with epitopes (HLA binding register indicated in bold) covalently linked to the β-chain N-terminus were produced via baculovirus mediated insect cell expression as described (Broughton et al., 2012; Henderson et al., 2007a; Petersen et al., 2014; Petersen et al., 2015). TCR DNA sequences were codon optimized for expression in *E. coli* and synthesized (gBlocks; Integrated DNA Technologies), and cloned into the pET30 expression vector. TCRs containing an introduced interchain disulfide bond between their constant domains were produced via refolding of TCR α- and β-chain E. coli inclusion body preparations followed by purification as described previously (Boulter et al., 2003;

Broughton et al., 2012; Garboczi et al., 1996; Petersen et al., 2014; Petersen et al., 2015). pET30 DNA constructs encoding TCR point mutants were produced using gBlock DNA fragments synthesised by Integrated DNA Technologies (Coralville, IA).

The α- and β-chains of the HLA-DQ2.5/DQ8/8.5-gliadin constructs had C-terminal enterokinase cleavage sites separating a *fos-jun* leucine zipper domain, respectively. The C-terminus of the HLA-DQ8-gliadin β-chain also contained BirA recognition sequence for biotinylation followed by a polyhistidine tag for purification. The *fos-jun* leucine zipper and tags were removed via enterokinase digestion (Genscript) prior to co-purification of the HLA-DQ8/8.5-gliadin protein with TCRs on a Superdex 200 size-exclusion chromatography column (GE healthcare), concentration to 10mg/ml in TBS (10 mM Tris, pH 8.0, 150 mM NaCl) and subsequent crystallization (Broughton et al., 2012; Petersen et al., 2014; Petersen et al., 2015).

Crystallization, data collection and processing.

Crystallisation was carried out by the hanging drop vapor diffusion method using a protein to mother liquor ratio of 1:1. The mother liquors used for crystallization were 0.2M Na-Acetate, 22% PEG8000, 0.1M Tris pH 8.0 for the Bel502 TCR-HLA-DQ8-glia-α1 complex, 0.1 M Ca-Acetate, 22-18 % PEG3350, 0.1M Tris pH 8.0 for the Bel602 TCR-HLA-DQ8.5-glia-γ1 complex, and 8% Tacsimate pH 7.5, 20% PEG3350 for the T15 TCR-HLA-DQ8.5-glia-γ1 complex. Crystals were transferred into mother liquor supplemented with 18% PEG400 or 18% ethylene glycol in the case of Bel602 TCR-HLA-DQ8.5-glia-γ1, and frozen in liquid N2. Data was collected at the microfocus beamline (MX2) of the Australian Synchrotron (Clayton, Victoria) (McPhillips et al., 2002), and processed using MOSFLM and SCALA from the CCP4 program suite (Winn et al., 2011) (refer to **table S1** for data collection and refinement statistics).

Structure determination, refinement and validation.

The structures were solved by molecular replacement in PHASER using the S13 TCR and HLA-DQ8-glia-α1 from the corresponding ternary complex (PDB entry 4Z7U) as separate search models (Storoni et al., 2004). Iterative rounds of model building in Coot (Emsley and Cowtan, 2004) and restrained refinement using the programs, PHENIX (Adams et al., 2010) and BUSTER (http://www.globalphasing.com/buster/) were carried out as described earlier (Broughton et al., 2012; Petersen et al., 2015).

Surface plasmon resonance.

Surface plasmon resonance (SPR) measurements were carried out as described earlier (Broughton et al., 2012; Petersen et al., 2014; Petersen et al., 2015) using a BIACORE 3000 instrument (GE

healthcare). Briefly, biotinylated pMHCs were surface immobilized on a CAP chip (GE healthcare) to between 900 and1200 response units of immobilized protein. In each experiment three of the four pMHCs HLA-DQ8.5-glia-α1, HLA-DQ8.5-glia-α1, HLA-DQ8-glia-γ1 or HLA-DQ8.5-glia-γ1 were measured in parallel, and HLA-DQ2.5-glia-α1 was used as negative control for background correction. Dilution series of each TCR were passed over the surface in 20mM HEPES, 150mM NaCl, 1mM EDTA and 0.01 % TWEEN20 at a flow rate of 10 µl/ml. Background correction and data analysis was subsequently carried out using Sigmaplot 12.5 (Systat Software).

ACCESSION CODES

The X-ray crystal structures of Bel502 TCR-DQ8-glia-α1, Bel602 TCR-DQ8.5-glia-γ1, and T15 TCR-DQ8.5-glia-γ1 have been deposited in the PDB with the following accession codes: 5KS9, 5KSA, and 5KSB, respectively.

AUTHOR CONTRIBUTIONS

J.P. Y.K-W. K.L.L. M.T. J.vB. and H.H.R. conducted the experiments, J.P., J.vB., F.K., J.R., and H.H.R designed the experiments and wrote the paper.

ACKNOWLEDGEMENTS

We thank Allan Thompson for assistance in sequencing. We thank the staff at the National Australian Synchrotron for assistance with data collection and staff at the Monash Macromolecular Crystallization Facility. This work was supported by Australian National Health and Medical Research Council (NHMRC) and Australian Research Council (ARC) funding. J.R. is an NHMRC Australia Fellow (AF50).

References

Abadie, V., Sollid, L.M., Barreiro, L.B., and Jabri, B. (2011). Integration of genetic and immunological insights into a model of celiac disease pathogenesis. Annu Rev Immunol *29*, 493-525.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Borg, N.A., Ely, L.K., Beddoe, T., Macdonald, W.A., Reid, H.H., Clements, C.S., Purcell, A.W., Kjer-Nielsen, L., Miles, J.J., Burrows, S.R., *et al.* (2005). The CDR3 regions of an immunodominant T cell receptor dictate the 'energetic landscape' of peptide-MHC recognition. Nat Immunol *6*, 171-180.

Boulter, J.M., Glick, M., Todorov, P.T., Baston, E., Sami, M., Rizkallah, P., and Jakobsen, B.K. (2003). Stable, soluble T-cell receptor molecules for crystallization and therapeutics. Protein Eng *16*, 707-711.

Brochet, X., Lefranc, M.P., and Giudicelli, V. (2008). IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res *36*, W503-508.

Broughton, S.E., Petersen, J., Theodossis, A., Scally, S.W., Loh, K.L., Thompson, A., van Bergen, J., Kooy-Winkelaar, Y., Henderson, K.N., Beddoe, T., *et al.* (2012). Biased T cell receptor usage directed against human leukocyte antigen DQ8-restricted gliadin peptides is associated with celiac disease. Immunity *37*, 611-621.

Christophersen, A., Risnes, L.F., Bergseng, E., Lundin, K.E., Sollid, L.M., and Qiao, S.W. (2016). Healthy HLA-DQ2.5+ Subjects Lack Regulatory and Memory T Cells Specific for Immunodominant Gluten Epitopes of Celiac Disease. J Immunol *196*, 2819-2826.

Dahal-Koirala, S., Risnes, L.F., Christophersen, A., Sarna, V.K., Lundin, K.E., Sollid, L.M., and Qiao, S.W. (2016). TCR sequencing of single cells reactive to DQ2.5-glia-[alpha]2 and DQ2.5-glia-[omega]2 reveals clonal expansion and epitope-specific V-gene usage. Mucosal Immunol *9*, 587-596.

Di Sabatino, A., and Corazza, G.R. (2009). Coeliac disease. Lancet *373*, 1480-1493.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Garboczi, D.N., Utz, U., Ghosh, P., Seth, A., Kim, J., VanTienhoven, E.A., Biddison, W.E., and Wiley, D.C. (1996). Assembly, specific binding, and crystallization of a human TCR-alphabeta with an antigenic Tax peptide from human T lymphotropic virus type 1 and the class I MHC molecule HLA-A2. J Immunol *157*, 5403-5410.

Garcia, K.C. (2012). Reconciling views on T cell receptor germline bias for MHC. Trends in Immunology *33*, 429-436.

Gras, S., Wilmann, P.G., Chen, Z., Halim, H., Liu, Y.C., Kjer-Nielsen, L., Purcell, A.W., Burrows, S.R., McCluskey, J., and Rossjohn, J. (2012). A structural basis for varied alphabeta TCR usage against an immunodominant EBV antigen restricted to a HLA-B8 molecule. J Immunol *188*, 311-321.

Henderson, K.N., Reid, H.H., Borg, N.A., Broughton, S.E., Huyton, T., Anderson, R.P., McCluskey, J., and Rossjohn, J. (2007a). The production and crystallization of the human leukocyte antigen class II molecules HLA-DQ2 and HLA-DQ8 complexed with deamidated gliadin peptides implicated in coeliac disease. Acta Crystallogr Sect F Struct Biol Cryst Commun *63*, 1021-1025.

Henderson, K.N., Tye-Din, J.A., Reid, H.H., Chen, Z., Borg, N.A., Beissbarth, T., Tatham, A., Mannering, S.I., Purcell, A.W., Dudek, N.L., *et al.* (2007b). A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. Immunity *27*, 23-34.

Jabri, B., and Sollid, L.M. (2009). Tissue-mediated control of immunopathology in coeliac disease. Nat Rev Immunol *9*, 858-870.

Karell, K., Louka, A.S., Moodie, S.J., Ascher, H., Clot, F., Greco, L., Ciclitira, P.J., Sollid, L.M., Partanen, J., and European Genetics Cluster on Celiac, D. (2003). HLA types in celiac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the European Genetics Cluster on Celiac Disease. Hum Immunol *64*, 469-477.

Kim, C.Y., Quarsten, H., Bergseng, E., Khosla, C., and Sollid, L.M. (2004). Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. Proc Natl Acad Sci U S A *101*, 4175-4179.

Koning, F. (2014). Pathophysiology of celiac disease. J Pediatr Gastroenterol Nutr *59 Suppl 1*, S1-4.

Koning, F., Thomas, R., Rossjohn, J., and Toes, R.E. (2015). Coeliac disease and rheumatoid arthritis: similar mechanisms, different antigens. Nat Rev Rheumatol *11*, 450-461.

Kooy-Winkelaar, Y., van Lummel, M., Moustakas, A.K., Schweizer, J., Mearin, M.L., Mulder, C.J., Roep, B.O., Drijfhout, J.W., Papadopoulos, G.K., van Bergen, J., *et al.* (2011). Gluten-specific T cells cross-react between HLA-DQ8 and the HLA-DQ2alpha/DQ8beta transdimer. J Immunol *187*, 5123-5129.

Macdonald, W.A., Chen, Z., Gras, S., Archbold, J.K., Tynan, F.E., Clements, C.S., Bharadwaj, M., Kjer-Nielsen, L., Saunders, P.M., Wilce, M.C.J., *et al.* (2009). T Cell Allorecognition via Molecular Mimicry. Immunity *31*, 897-908.

McPhillips, T.M., McPhillips, S.E., Chiu, H.J., Cohen, A.E., Deacon, A.M., Ellis, P.J., Garman, E., Gonzalez, A., Sauter, N.K., Phizackerley, R.P., *et al.* (2002). Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. J Synchrotron Radiat *9*, 401-406.

Molberg, O., McAdam, S.N., Korner, R., Quarsten, H., Kristiansen, C., Madsen, L., Fugger, L., Scott, H., Noren, O., Roepstorff, P., *et al.* (1998). Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. Nat Med *4*, 713-717.

Petersen, J., Montserrat, V., Mujico, J.R., Loh, K.L., Beringer, D.X., van Lummel, M., Thompson, A., Mearin, M.L., Schweizer, J., Kooy-Winkelaar, Y., *et al.* (2014). T-cell receptor recognition of HLA-DQ2-gliadin complexes associated with celiac disease. Nat Struct Mol Biol *21*, 480-488.

Petersen, J., van Bergen, J., Loh, K.L., Kooy-Winkelaar, Y., Beringer, D.X., Thompson, A., Bakker, S.F., Mulder, C.J., Ladell, K., McLaren, J.E., *et al.* (2015). Determinants of gliadin-specific T cell selection in celiac disease. J Immunol *194*, 6112-6122.

Qiao, S.W., Christophersen, A., Lundin, K.E., and Sollid, L.M. (2014). Biased usage and preferred pairing of alpha- and beta-chains of TCRs specific for an immunodominant gluten epitope in coeliac disease. Int Immunol *26*, 13-19.

Qiao, S.W., Raki, M., Gunnarsen, K.S., Loset, G.A., Lundin, K.E., Sandlie, I., and Sollid, L.M. (2011). Posttranslational modification of gluten shapes TCR usage in celiac disease. J Immunol *187*, 3064-3071.

Rossjohn, J., Gras, S., Miles, J.J., Turner, S.J., Godfrey, D.I., and McCluskey, J. (2015). T cell antigen receptor recognition of antigen-presenting molecules. Annu Rev Immunol *33*, 169-200.

Sollid, L.M., and Jabri, B. (2013). Triggers and drivers of autoimmunity: lessons from coeliac disease. Nat Rev Immunol *13*, 294-302.

Sollid, L.M., Qiao, S.W., Anderson, R.P., Gianfrani, C., and Koning, F. (2012). Nomenclature and listing of celiac disease relevant gluten T-cell epitopes restricted by HLA-DQ molecules. Immunogenetics *64*, 455-460.

Stadinski, B.D., Shekhar, K., Gomez-Tourino, I., Jung, J., Sasaki, K., Sewell, A.K., Peakman, M., Chakraborty, A.K., and Huseby, E.S. (2016). Hydrophobic CDR3 residues promote the development of self-reactive T cells. Nature immunology.

Stamnaes, J., and Sollid, L.M. (2015). Celiac disease: Autoimmunity in response to food antigen. Seminars in Immunology *27*, 343-352.

Storoni, L.C., McCoy, A.J., and Read, R.J. (2004). Likelihood-enhanced fast rotation functions. Acta Crystallogr D Biol Crystallogr *60*, 432-438.

Tye-Din, J.A., Stewart, J.A., Dromey, J.A., Beissbarth, T., van Heel, D.A., Tatham, A., Henderson, K., Mannering, S.I., Gianfrani, C., Jewell, D.P., *et al.* (2010). Comprehensive, quantitative mapping of T cell epitopes in gluten in celiac disease. Sci Transl Med *2*, 41ra51.

van Lummel, M., van Veelen, P.A., Zaldumbide, A., de Ru, A., Janssen, G.M., Moustakas, A.K., Papadopoulos, G.K., Drijfhout, J.W., Roep, B.O., and Koning, F. (2012). Type 1 diabetes-associated HLA-DQ8 transdimer accommodates a unique peptide repertoire. J Biol Chem *287*, 9514-9524.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., *et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr *67*, 235-242.

Figure Legends

Figure 1. Determination of TRAV20^{+}-TRBV9^{+} TCR Specificity and Affinity

(A-C) Effect of alanine substitutions in the antigenic peptides on T cell recognition. Variants of DQ8–glia-α1 (SGEGSFQPSQENP) and DQ2-glia-γ1 (QQPQQSFPEQERP), carrying an alanine instead of the original amino acid at the indicated positions, were tested for recognition by clones Bel502 (A), Bel602 (B), and T15 (C) in the presence of DQ8-homozygous EBV-LCL. The bar on the left hand side represents the response to the unmodified peptide. The data are representative of at least two experiments. (D-F) SPR affinity measurements with the TCRs Bel502 (D), Bel602 (E) and T15 (F). TCR affinities for immobilised HLA-DQ8/8.5–glia-α1 and HLA-DQ8/8.5–glia- γ1 were determined using dilution series of TCRs with a maximal concentration of 48 μM (and 96 μM depending on protein availability). Two independent experiments were performed for Bel502 and Bel602 (D, E) and one experiment for T15 (F). Data for TCRs with no measureable affinity for HLA-DQ-gliadins not shown.

Figure 2. Structural overview of TCR-pMHC complexes.

(A, C, E, G) Cartoon representation of TCR-pMHC complexes. (B, D, F, H) Surface representation of TCR footprints and TCR docking. (A, B) Bel502 TCR-DQ8-glia-α1. (C, D) SP3.4 TCR-DQ8-glia-α1. (E, F) Bel602 TCR-DQ8.5-glia-γ1. (G, H) T15 TCR-DQ8.5-glia-γ1. Vα and Vβ centre of mass positions are shown as connected black circles. The HLA-DQ8 α- and β-chain are coloured light green and light yellow, respectively, and the peptide is shown as grey sticks. The CDR loops 1α, 2α, 3α, 1β, 2β and 3β are colored red, pink, cyan, orange, magenta and blue, respectively, and framework residues are coloured green. TCR footprint colours are in accordance with nearest TCR contact residue.

Figure 3. Structure of the Bel502-HLA-DQ8-glia-α1 complex and comparison with the SP3.4-HLA-DQ8-glia-α1 complex.

(A-D) Interactions between HLA-DQ8 and Bel502 TCR and SP3.4 TCR: (A) CDR1α and CDR2α of Bel502; (B) CDR1β and CDR2β of Bel502 and SP3.4 (transparent) overlay; (C) CDR3α of Bel502; (D) CDR3β of Bel502. (E) Interactions between SP3.4 and the DQ8-glia-α1 peptide. (F) Interactions between Bel502 and the DQ8-glia-α1 peptide. The HLA-DQ8 α-chain, β-chain and peptide are colored light green, light yellow and grey, respectively, and the CDR loops 1α, 2α, 3α, 1β, 2β and 3β are colored red, pink, cyan, orange, magenta and blue, respectively. (A) CDR1α Arg37 reaches across the peptide and interacts with Phe58α. CDR2α Tyr57 and framework residue Lys66 form an adaptable interface with the HLA-DQ8 β-chain helix. (B) CDR1β Leu37 and

CDR2β Tyr57 form interactions with the HLA-DQ8 α-chain helix that are conserved between the Bel502 and SP3.4 TCRs as well as across other TRBV9^{+} TCRs. (C) CDR3α forms an extensive interface with HLA-DQ8 Phe58α. (D) The CDR3β HLA-DQ8 spans the peptide binding cleft and contacts both HLA-DQ8 α- and β-chain. (E) The SP3.4 TCR forms H-bonds with the DQ8-glia-α1 peptide residues p1-E, p3-S, p5-Q and p8-Q. CDR3□ Arg110 H-bonds to p3-S and p5-Q. (F) The Bel502 TCR forms H-bonds with the DQ8-glia-α1 peptide residues p-1-Gly, p3-Ser, p5-Gln and p8-Gln. CDR1α Arg37 H-bonds to p3-Ser and p5-Gln. The peptide interactions involving TRBV9 germline encoded residues CDR1β Leu37 and CDR2β Tyr57 are conserved across TRBV9^{+} TCRs.

Figure 4. Structure of the Bel602-DQ8.5-glia-γ1 complex.

(A-D) Interactions between HLA-DQ8.5 and (A) CDR1α and CDR2α, (B) CDR1β and CDR2β, (C) CDR α, and (D) CDR3β of Bel602. (E) Interactions between Bel602 and the DQ8.5-glia-α1 peptide (grey sticks). The HLA-DQ8.5 α- and β-chain are colored light green and light yellow, respectively, and the CDR loops 1α, 2α, 3α, 1β, 2β and 3β are colored red, pink, cyan, orange, magenta and blue, respectively. A CDR1α□Arg37 is positioned close to the HLA-DQ8.5 β-chain and forms a salt bridge with HLA-DQ8.5 Asp76β.□□□□2a Tyr57 forms an adaptable interface with the HLA-DQ8.5 β-chain helix. (B) CDR1β Leu37 and CDR2β Tyr57 form conserved interactions with the HLA-DQ8 α-chain helix. (C) CDR3α adopts an S-shaped conformation and forms extensive contacts with HLA-DQ8.5 Phe58α□□□□□□□□□□□□□□□□□□□Asp□□α. (D) CDR3β Tyr114 H-bonds to HLA-DQ8.5 Arg70β. (E) The Bel602 TCR forms H-bonds to the DQ8.5-glia-γ1 peptide residues p2-Gln, p3-Gln, p7-Glu and p8-Gln. CDR3α and CDR3β form an extensive van der Waals interface with p5-Phe. CDR1α Arg37 H-bonds to and p2-Gln, p3-Gln and p5-Gln and contacts p5-Phe.

Figure 5. Structure of the T15-DQ8.5-glia-γ1 complex.

(A-D) Interactions between HLA-DQ8.5 and (A) CDR1α and CDR2α, (B) CDR1β and CDR2β, (C) CDR3α, (D) CDR3β. (E) Interactions between T15 and the DQ8.5-glia-γ1 peptide (grey sticks). The HLA-DQ8.5 α- and β-chain are colored light green and light yellow, respectively, and the CDR loops 1α, 2α, 3α, 1β, 2β and 3β are colored red, pink, cyan, orange, magenta and blue, respectively.

Figure 6. Impact of polymorphisms in HLA-DQ8 and HLA-DQ8.5 on peptide specificity and TCR recognition.

(A, B) The differential peptide specificity of HLA-DQ8 and HLA-DQ8.5 is determined by

polymorphisms within the peptide binding cleft (shown as orange sticks). Positions 31 and 52 of the HLA-DQ8/8.5 α□□□□□□□ form part of the p1 pocket and govern peptide specificity for the anchor residue at this position. Differences between HLA-DQ8 and HLA-DQ8.5 (red) do not overlap with TCR contact residues (blue). The lack of overlap explains why each TCR can recognize its cognate peptide in the context of both HLA-DQ8 and HLA-DQ8.5. (A) Overlay of HLA-DQ8-glia-α1 in ternary complex structures with the Bel502 and T316 TCRs. In HLA-DQ8 the charged residues Glu31α and Arg52α □□□□□□□□□□□□□□□□□e□□□□□□□□ a □□□□□□□□□□□□□□□□□□□d□peptide □□□□□□□□□□□□□□□ p1 pocket. (B) Overlay of HLA-DQ8.5-glia-α1 in ternary complex structures with the Bel602 and T15 TCRs. Position 31α in HLA-DQ8.5 is a glutamine residue whereas□□□□□□□□□□□□α□is absent, and its position occupied in the p1 pocket by Phe51α. This creates a more hydrophobic p1 pocket.

Figure 7. Energetic hot spots

Effect of TCR point mutations at the pMHC interface. (A, C) Measured affinities determined by SPR. Error bars represent the standard error of the fit to all data (derived from two independent experiments for each TCR mutant). The graphs were truncated at a K_D value of 150 μM. TCR mutants with undetectable binding therefore do not have error bars. (B, D) Position of mutations tested. Residues were coloured according to their impact on K_D values: >10x: red, >5x K_D orange, >3x K_D yellow, <3x K_D green. (A) Affinities of the Bel502 TCR and Bel502 interface mutants for surface bound HLA-DQ8-glia-α1 (filled bars) and HLA-DQ8.5-glia-α1 (striped bars). (B) Mutants at the Bel502-pMHC interface. (C) Affinities of the Bel602 TCR and Bel602 interface mutants for surface bound HLA-DQ8.5-glia-γ1. (D) Mutants at the Bel602-pMHC interface. See also Figure S1.

Figure 8. Clustering of arginine residues at the interface of TCRs and HLA-gliadin epitopes.

Each epitope, HLA-DQ8-glia-α1 and HLA-DQ8.5-glia-γ1 is recognized by TCRs with an arginine in a specific position relative to the peptide. In the case of HLA-DQ8-glia-α1 the arginine can be provided by either CDR1α□□□□□□, CDR3α (cyan) or CDR3β□□blue□.

Figure S1, related to Figure 7. SPR affinity measurements

SPR affinity measurements with TCR point mutants. K_D values were calculated by fitting normalized data from two independent measurements with maximal TCR concentrations of 48 or 96 μM. RU – response units.

Table 1 TRAV and TRBV gene usage and CDR3 sequences of T cell clones

Clone	Restriction[a]	TRAV[b]	TRAJ	CDR1α	CDR2α	CDR3α	TRBV	TRBD	TRBJ	CDR1β	CDR2β	CDR3β
Bel502[c]	DQ8-glia-α1	20*01	39*01	VSGLRG	LYSAGEE	C AVALNNNAGNMLT FG	9*01	1*01	2-3*01	SGDLS	SNEGSKA	CAS SVAPGSDTQ YFG
Bel602	DQ8-glia-γ1	20*01	33*01	VSGLRG	LYSAGEE	C AVQFMDSNYQLI WG	9*01	1*01	2-7*01	SGDLS	SNEGSKA	CAS SVAGTPSYEQ YFG
T316[d]	DQ8.5-glia-α1	8-3*01	36*01	YGATPY	YGATPY	C AVGETGANNLF FG	6-1*01	2*02	2-1*01	MNHNS	SASEGT	CAS SEARRYNEQ FGP
T15[d]	DQ8.5-glia-γ1	20*02	6*01	VSGLRG	LYSAGEE	C AVQASGGSYIPT FG	9*01	1*01	2-3*01	SGDLS	SNEGSKA	CAS SNRGLGTDTQ YFG

[a]DQ8/8.5 epitope nomenclature according to Sollid et al (Sollid et al., 2012);

[b]TCR variable gene usage as described in IMGT-V-QUEST database (Brochet et al., 2008);

[c]TRAV and TRBV usage and CDR sequences previously published (Petersen et al., 2015);

[d] TRAV and TRBV usage and CDR sequences previously published (Broughton et al., 2012)

Table 2. Data collection and refinement statistics.

	Bel502-HLA-DQ8-glia-α1	Bel502-HLA-DQ8.5-glia-γ1	T15-HLA-DQ8.5-glia-γ1
Wavelength (Å)	0.9686	0.9537	0.9537
Resolution (Å)	46.9 - 2.55 (2.641 - 2.55)	45.75 - 2.0 (2.071 - 2.0)	85 - 2.9 (3.004 - 2.9)
Space group	P 1 21 1	P 1 21 1	P 21 21 21
Unit cell a b c α β γ	74.56 56.87 232.05 90 92.77 90	62.548 98.83 80.175 90 95.69 90	115.89 125.05 165.63 90 90 90
Total reflections	123381 (12135)	130970 (13040)	105795 (10287)
Unique reflections	63405 (6276)	65534 (6527)	53672 (5253)
Multiplicity	1.9 (1.9)	2.0 (2.0)	2.0 (2.0)
Completeness (%)	98.84 (99.04)	99.98 (99.97)	99.38 (98.96)
Mean I/sigma(I)	5.52 (1.78)	12.52 (2.71)	8.70 (2.06)
Wilson B-factor	25.13	31.38	42.96
R-merge	0.097 (0.409)	0.0299 (0.297)	0.079 (0.454)
R-meas	0.1377	0.04224	0.1121
CC1/2	0.958 (0.781)	0.998 (0.856)	0.994 (0.802)
CC*	0.989 (0.937)	1 (0.96)	0.999 (0.944)
R-work (%)	21.20 (27.92)	0.16.75 (24.74)	0.23.05 (33.07)
R-free (%)	23.82 (24.78)	0.20.62 (28.20)	27.00 (45.22)
Number of non-hydrogen atoms	13552	7248	13111
macromolecules	12756	6579	12836
ligands	60	15	112

water	736	634	163
Protein residues	1615	820	1623
RMS(bonds)	0.014	0.013	0.012
RMS(angles)	1.67	1.72	1.69
Ramachandran favoured (%)	97	97	97
Ramachandran outliers (%)	0	0.38	0.19
Clashscore	3.69	1.86	4.64
Average B-factor	36.20	45.60	54.60
macromolecules	36.30	44.90	54.50
ligands	79.20	69.40	103.40
solvent	30.30	52.30	29.20

Figure1

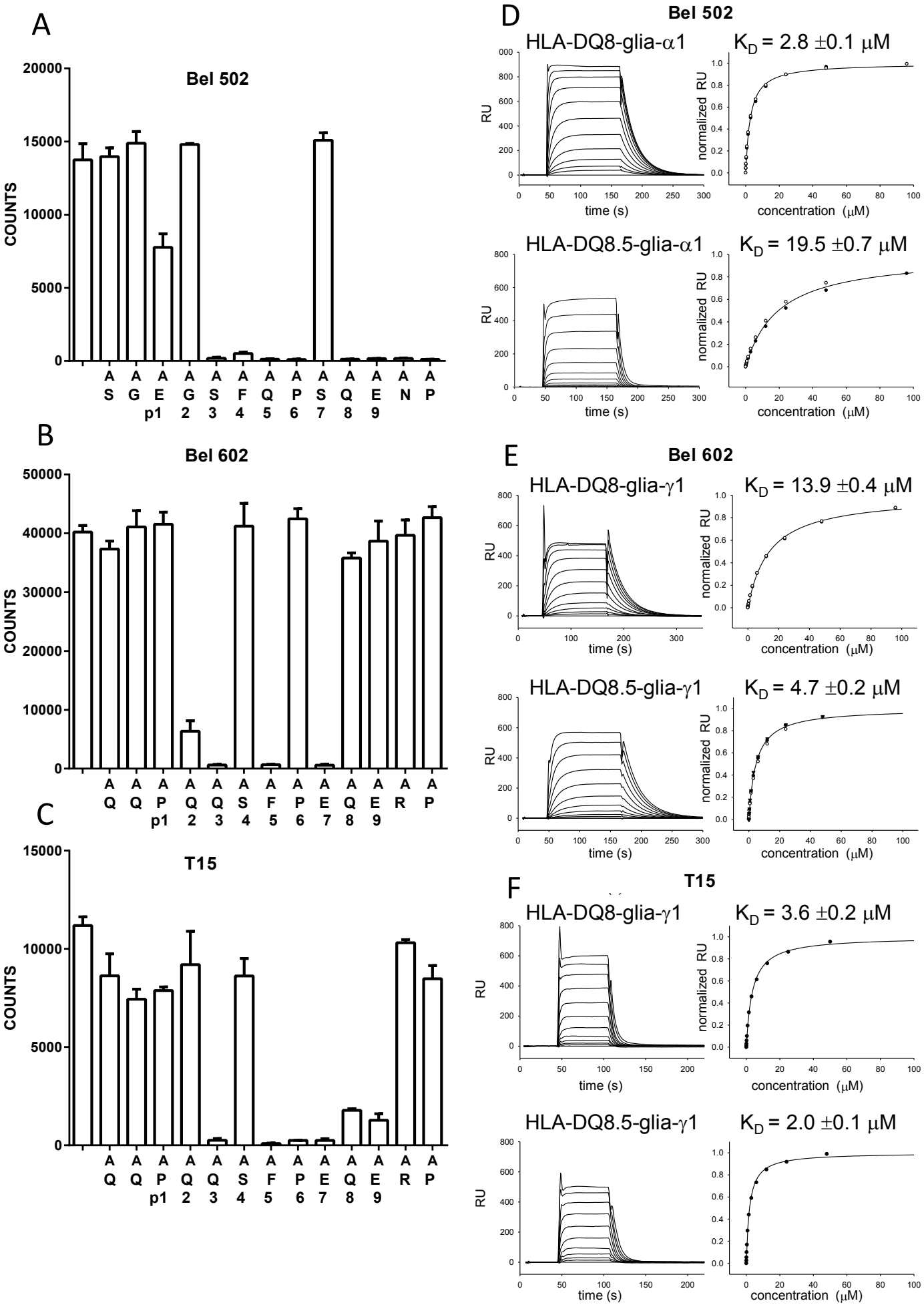
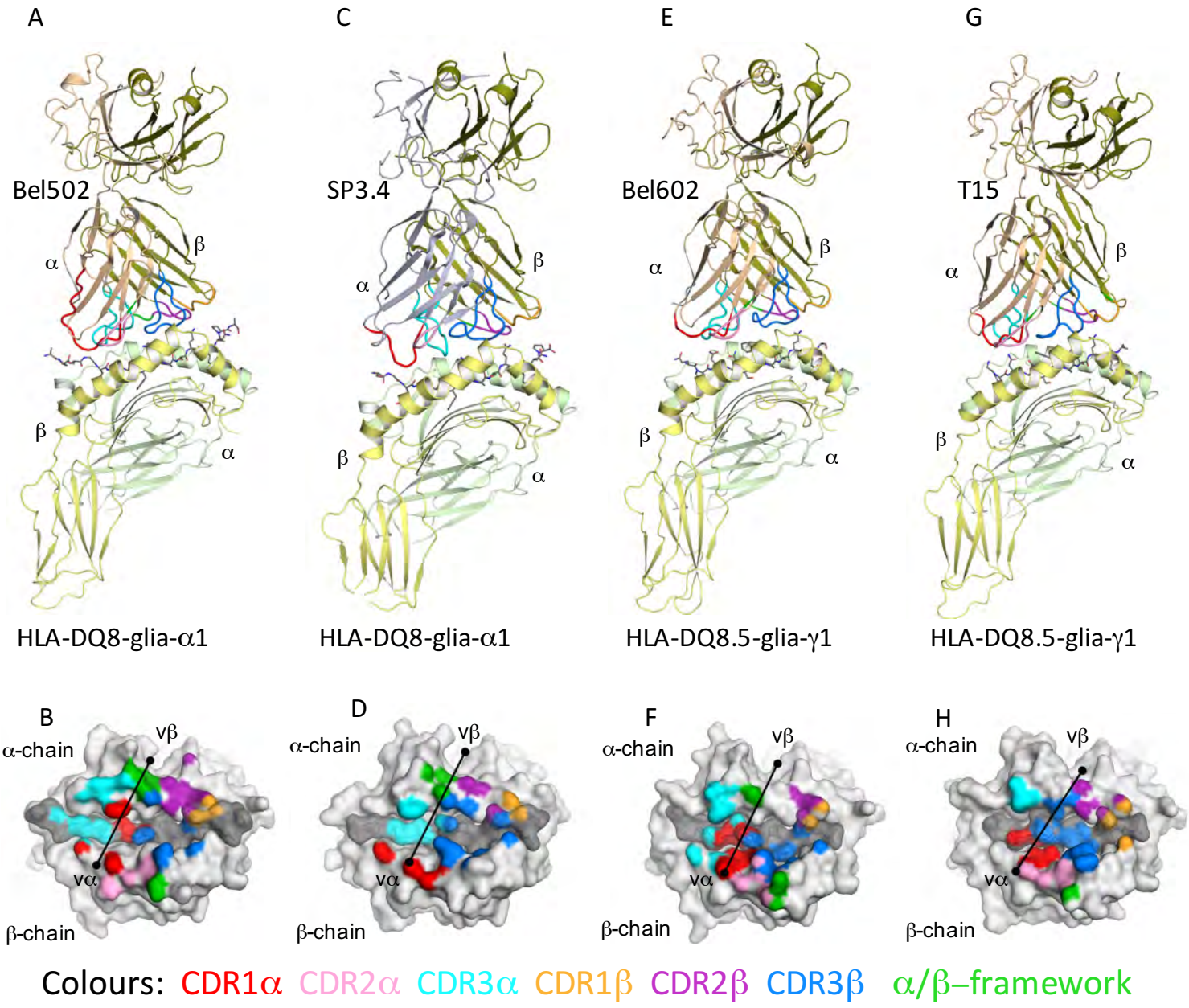


FIGURE 1

Figure



Colours: CDR1α CDR2α CDR3α CDR1β CDR2β CDR3β α/β–framework

FIGURE 2

Figure

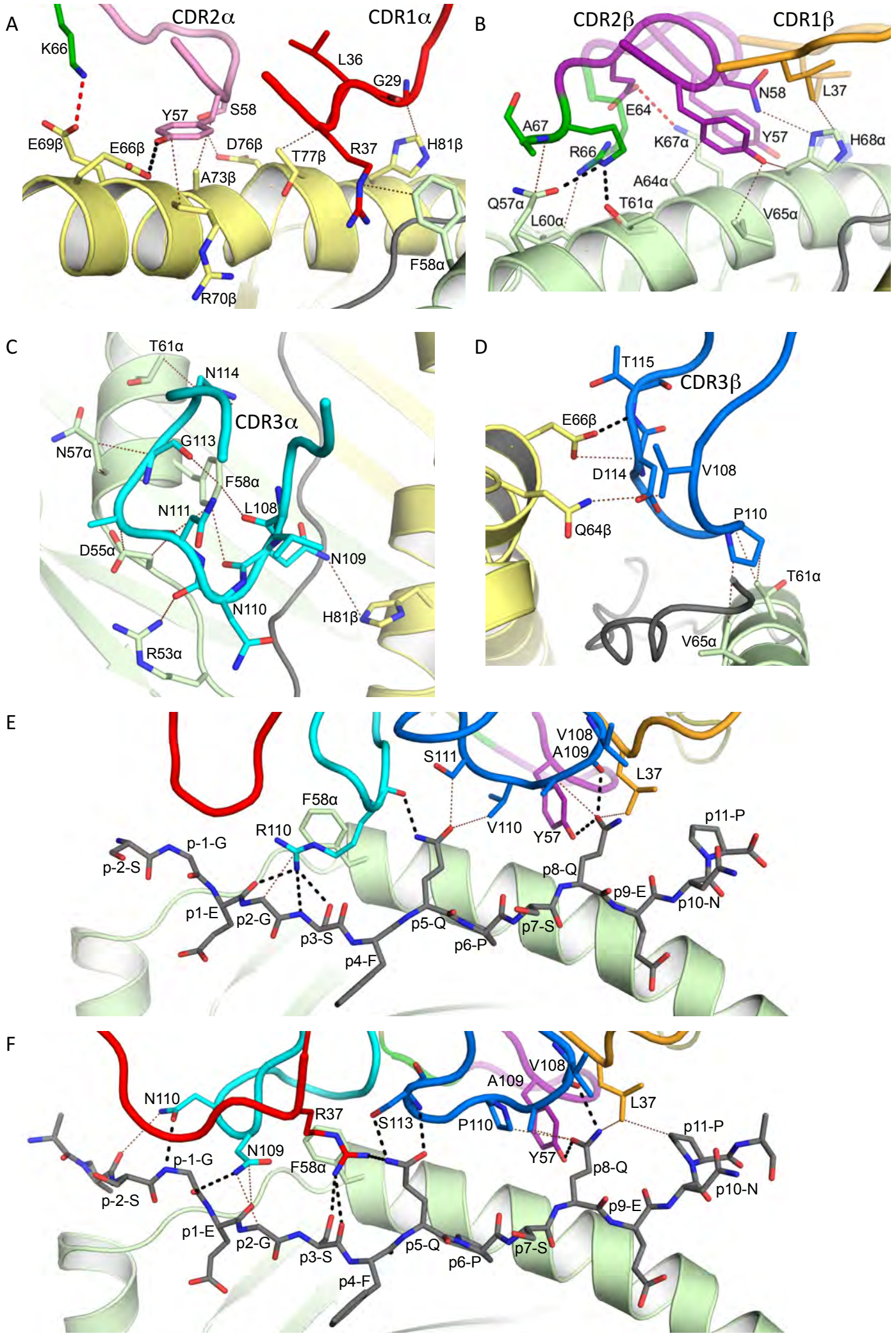


FIGURE 3

Figure

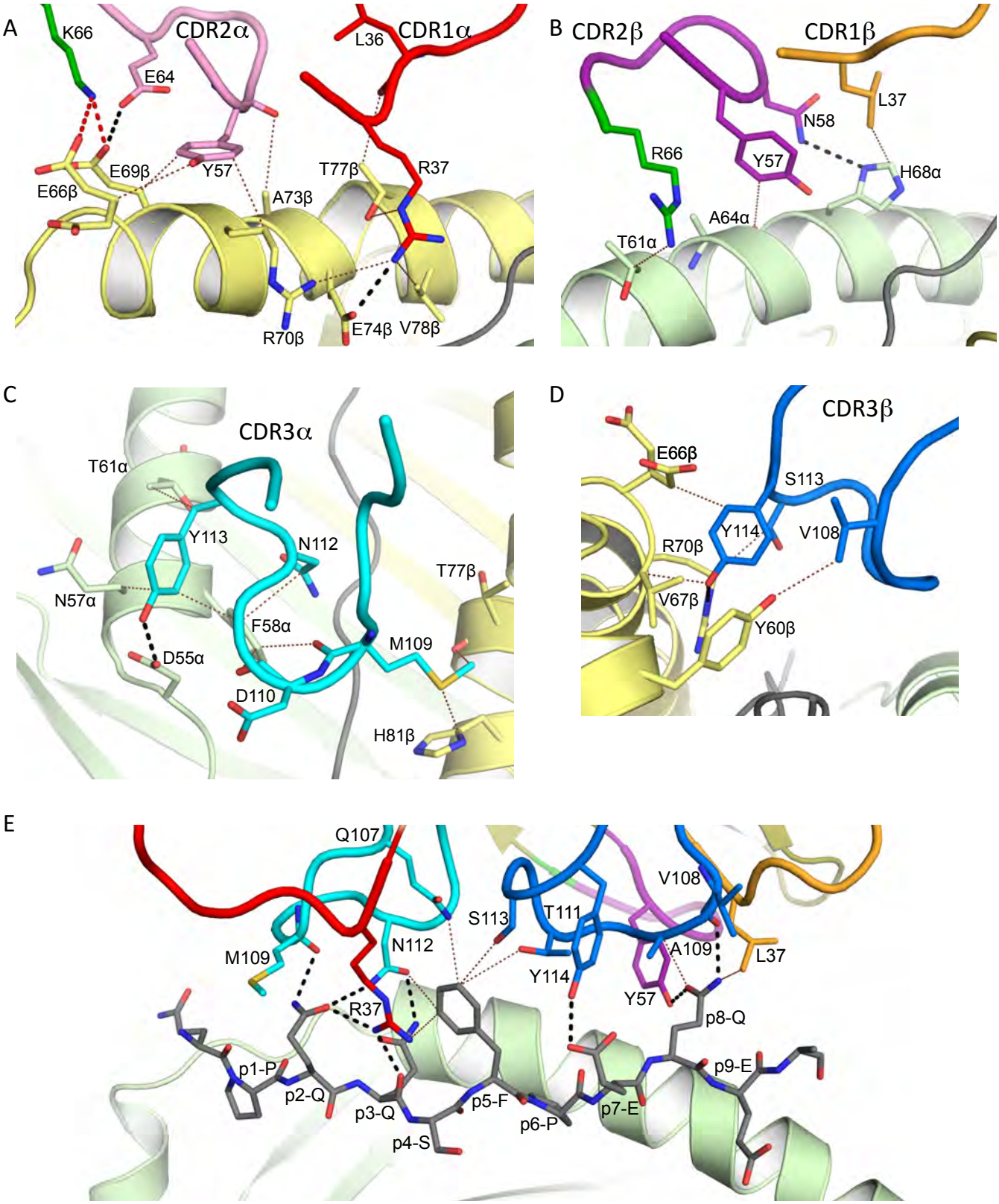


FIGURE 4

Figure

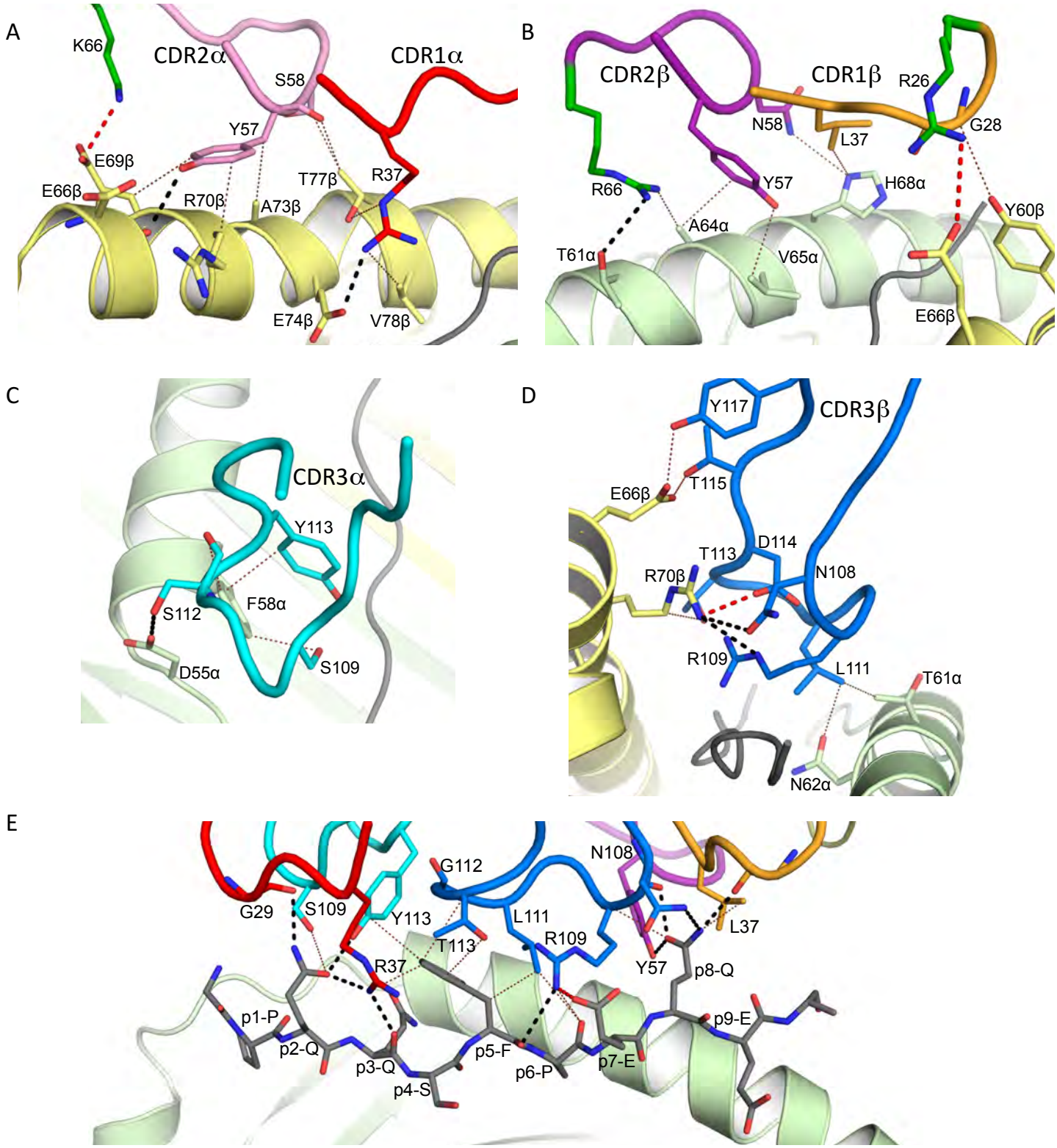


FIGURE 5

Figure

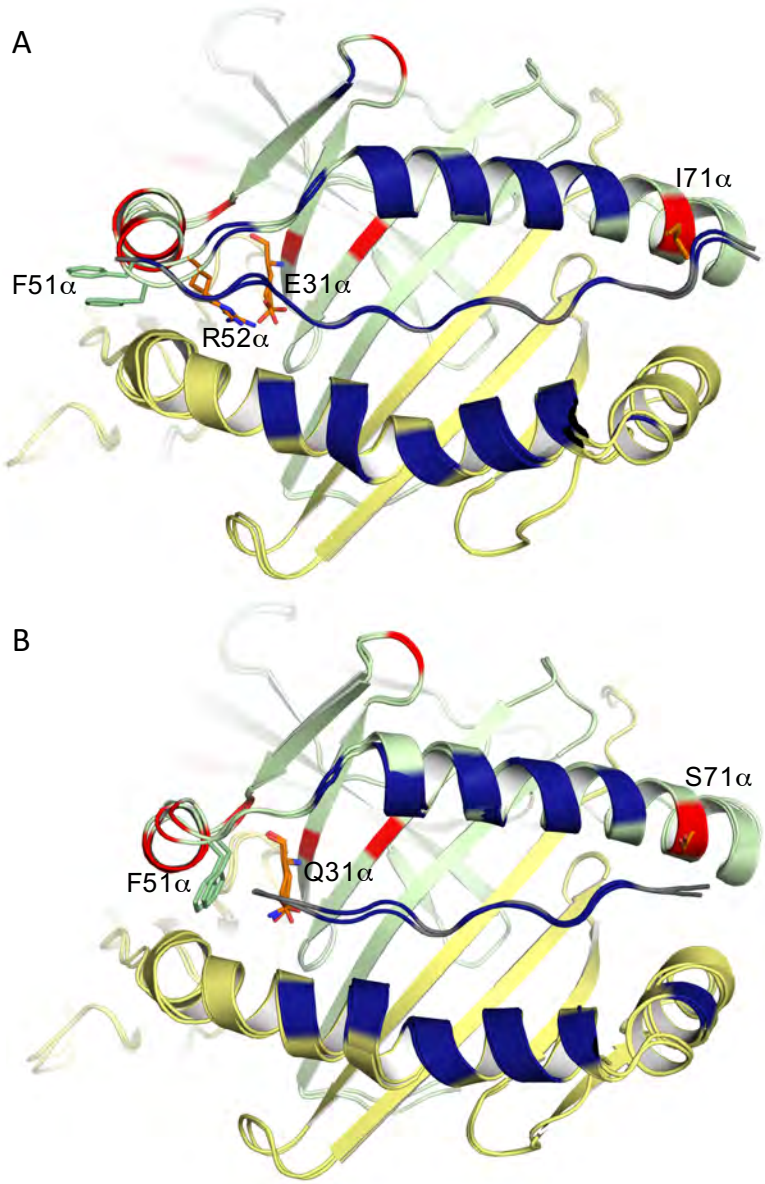


FIGURE 6

Figure7

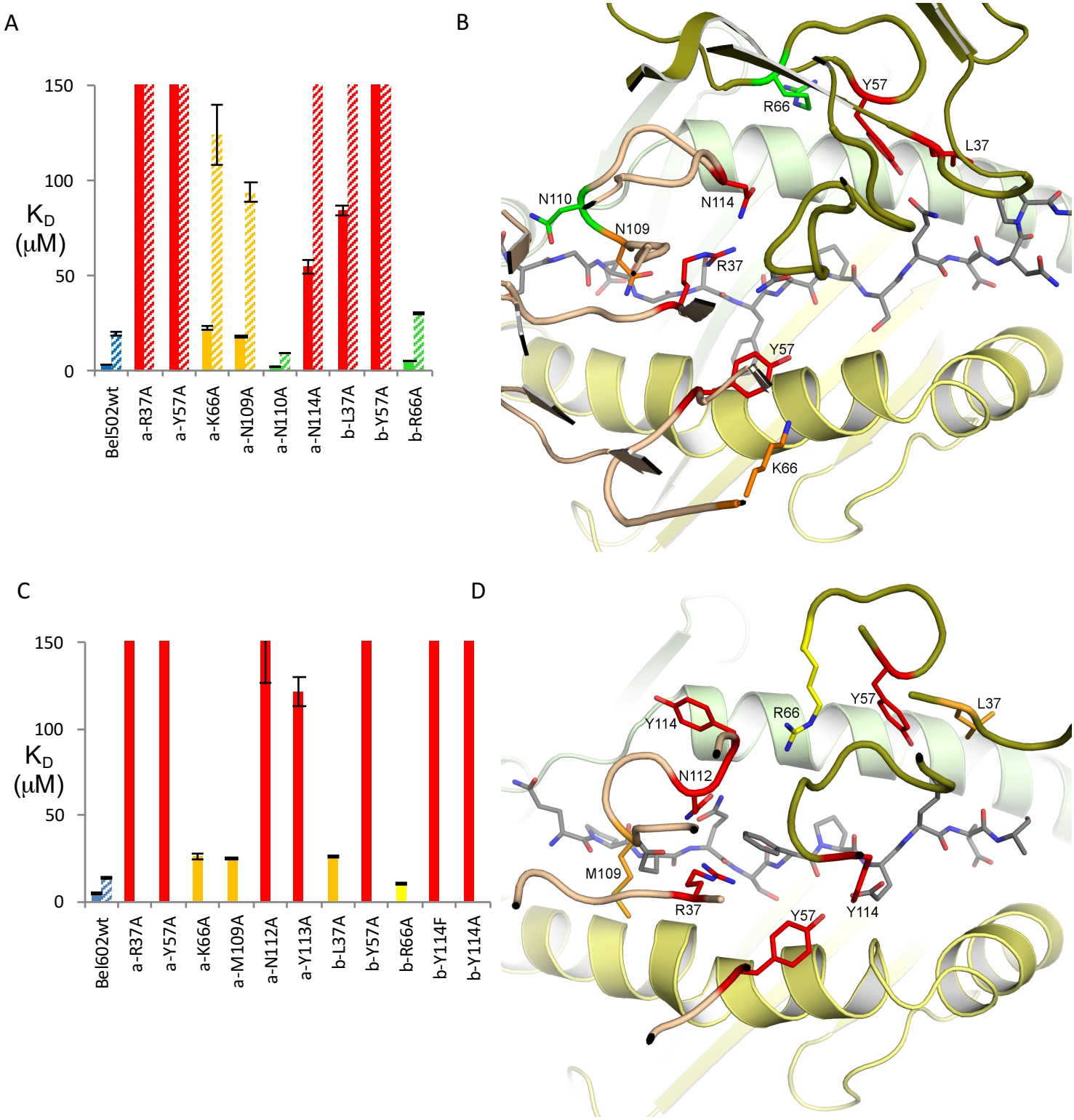


FIGURE 7

Figure

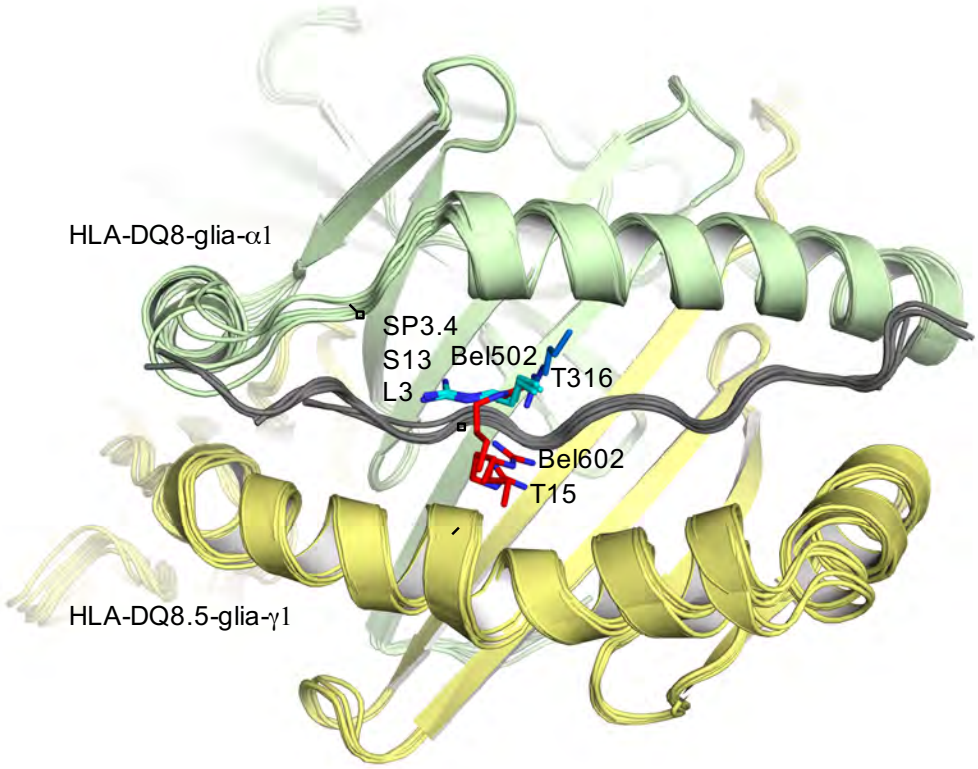


FIGURE 8

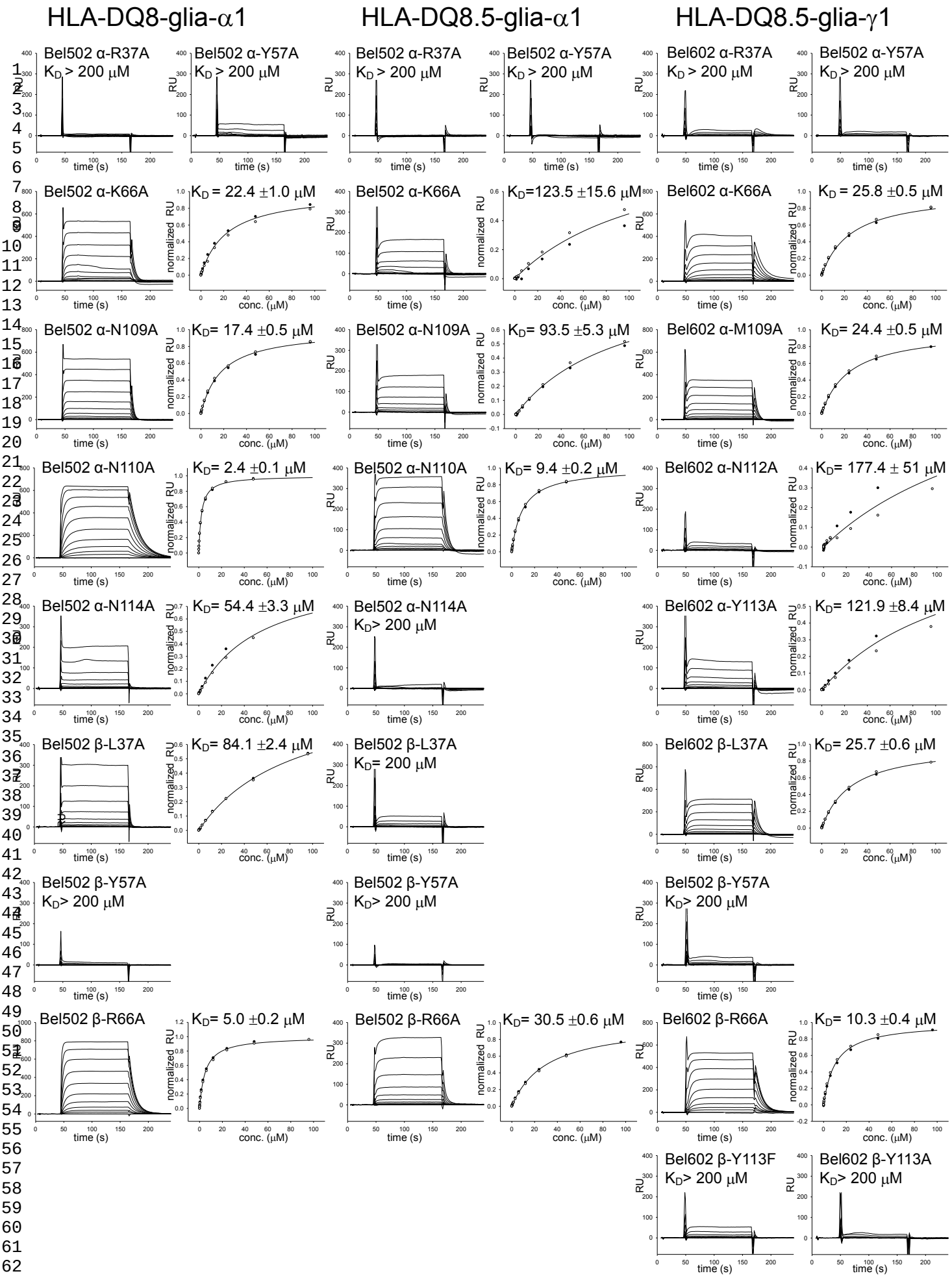

HLA-DQ8-glia-α1
HLA-DQ8.5-glia-α1
HLA-DQ8.5-glia-γ1
Bel502 α-R37A
KD > 200 μM
Bel502 α-Y57A
KD > 200 μM
Bel502 α-R37A
KD > 200 μM
Bel502 α-Y57A
KD > 200 μM
Bel602 α-R37A
KD > 200 μM
Bel502 α-Y57A
KD > 200 μM
Bel502 α-K66A
KD = 22.4 ±1.0 μM
Bel502 α-K66A
KD=123.5 ±15.6 μM
Bel602 α-K66A
KD = 25.8 ±0.5 μM
Bel502 α-N109A
KD= 17.4 ±0.5 μM
Bel502 α-N109A
KD= 93.5 ±5.3 μM
Bel602 α-M109A
KD= 24.4 ±0.5 μM
Bel502 α-N110A
KD= 2.4 ±0.1 μM
Bel502 α-N110A
KD= 9.4 ±0.2 μM
Bel602 α-N112A
KD= 177.4 ± 51 μM
Bel502 α-N114A
KD= 54.4 ±3.3 μM
Bel502 α-N114A
KD> 200 μM
Bel602 α-Y113A
KD= 121.9 ±8.4 μM
Bel502 β-L37A
KD= 84.1 ±2.4 μM
Bel502 β-L37A
KD= 200 μM
Bel602 β-L37A
KD= 25.7 ±0.6 μM
Bel502 β-Y57A
KD> 200 μM
Bel502 β-Y57A
KD> 200 μM
Bel502 β-Y57A
KD> 200 μM
Bel502 β-R66A
KD= 5.0 ±0.2 μM
Bel502 β-R66A
KD= 30.5 ±0.6 μM
Bel602 β-R66A
KD= 10.3 ±0.4 μM
Bel602 β-Y113F
KD> 200 μM
Bel602 β-Y113A
KD> 200 μM
RU
normalized RU
time (s)
conc. (μM)

FIGURE S1

suggested cover image

Click here to download Supplemental Text & Figures
toast06_v06.jpeg

Preliminary Full wwPDB X-ray Structure Validation Report (i)

Jun 2, 2016 – 08:43 PM EDT

This is a Preliminary Full wwPDB X-ray Structure Validation Report.
This report is produced by the wwPDB validation pipeline
before deposition or annotation of the structure.
This is not an official wwPDB validation report and is not a proof of deposition.
This report should not be submitted to journals.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

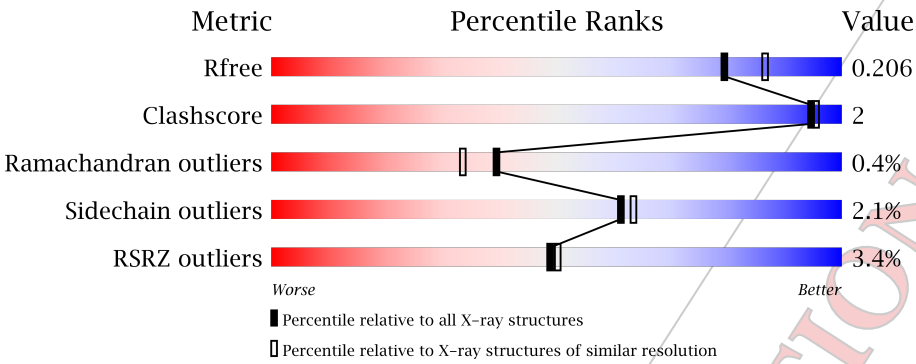
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027674
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027674

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	6249 (2.00-2.00)
Clashscore	102246	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)
RSRZ outliers	91569	6262 (2.00-2.00)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	182	94% 6%
2	B	190	6% 96% • •
3	C	195	3% 91% 7% •
4	D	242	5% 95% 5%
5	J	11	91% 9%

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 7228 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	182	1455	936	237	280	2	0	0	0

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	190	1569	992	277	293	7	0	3	0

- Molecule 3 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	195	1525	959	251	307	8	0	0	0

- Molecule 4 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	242	1940	1225	337	373	5	0	2	0

- Molecule 5 is a protein called GLN-PRO-GLN-GLN-SER-PHE-PRO-GLU-GLN-GLU-ALA.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
5	J	11	90	55	15	20	0	0	0

- Molecule 6 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: unknown).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
6	A	1	14	8	1	5	0	0

- Molecule 7 is CALCIUM ION (three-letter code: CA) (formula: unknown).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
7	K	1	1	1	0	0

- Molecule 8 is water.

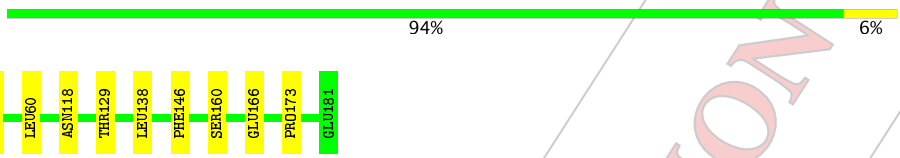
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
8	S	410	410	410	0	0
			Total	O		
8	W	224	224	224	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

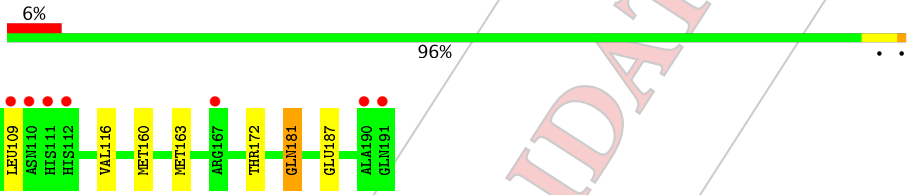
- Molecule 1:

Chain A:



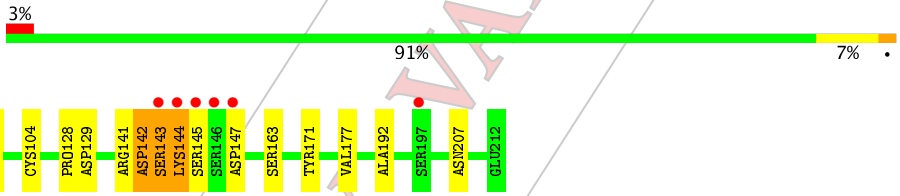
- Molecule 2:

Chain B:



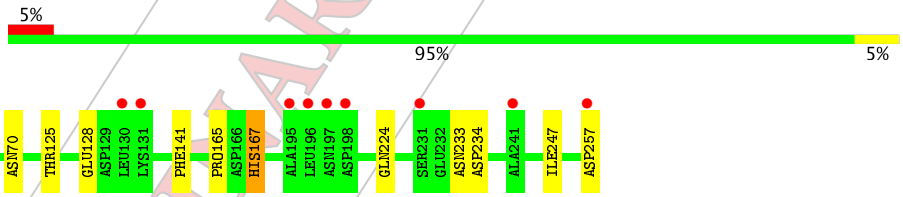
- Molecule 3:

Chain C:



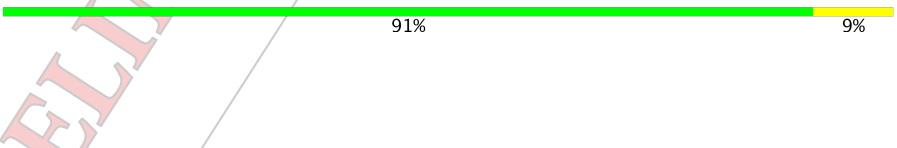
- Molecule 4:

Chain D:



- Molecule 5: GLN-PRO-GLN-GLN-SER-PHE-PRO-GLU-GLN-GLU-ALA

Chain J:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	62.55Å 98.83Å 80.17Å 90.00° 95.69° 90.00°	Depositor
Resolution (Å)	45.75 – 2.00 42.35 – 2.00	Depositor EDS
% Data completeness (in resolution range)	100.0 (45.75-2.00) 100.0 (42.35-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$	-	Xtriage
Refinement program	BUSTER 2.10.1	Depositor
R, R_{free}	0.169 , 0.206 0.170 , 0.206	Depositor DCC
R_{free} test set	3290 reflections (5.29%)	DCC
Wilson B-factor (Å^2)	(Not available)	Xtriage
Anisotropy	(Not available)	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 52.9	EDS
L-test for twinning[1]	$< \|L\| >$ = (Not available), $< L^2 >$ = (Not available)	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	7228	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *(Not available)*

[1]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: CA, NAG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| > 5$	RMSZ	$\#\|Z\| > 5$
1	A	0.52	0/1497	0.63	0/2043
2	B	0.52	0/1618	0.66	0/2208
3	C	0.50	0/1557	0.74	1/2107 (0.0%)
4	D	0.47	0/2000	0.62	0/2724
5	J	0.69	0/92	0.75	0/124
All	All	0.50	0/6764	0.66	1/9206 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	144	LYS	C-N-CA	5.11	134.48	121.70

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1455	0	1403	5	0
2	B	1569	0	1501	3	0
3	C	1525	0	1460	11	0
4	D	1940	0	1839	4	0
5	J	90	0	79	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	A	14	0	13	0	0
7	K	1	0	0	0	0
8	S	410	0	0	0	0
8	W	224	0	0	0	0
All	All	7228	0	6295	22	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 2.

All (22) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:141:ARG:HG3	3:C:142:ASP:H	1.39	0.86
2:B:181:GLN:CD	2:B:181:GLN:H	1.94	0.70
3:C:128:PRO:HG3	3:C:177:VAL:HG11	1.77	0.65
2:B:172:THR:HG22	2:B:187:GLU:HG2	1.81	0.60
3:C:163:SER:H	3:C:207:ASN:CB	2.15	0.59
4:D:125:THR:HG21	4:D:165:PRO:HB3	1.83	0.59
3:C:141:ARG:HG3	3:C:142:ASP:N	2.16	0.57
1:A:39:LYS:HG2	1:A:60:LEU:HD11	1.93	0.51
3:C:141:ARG:HH12	4:D:141:PHE:HE2	1.59	0.50
2:B:116:VAL:HG22	2:B:160:MET:HG3	1.94	0.49
4:D:224:GLN:HG3	4:D:247:ILE:HG23	1.94	0.49
3:C:163:SER:OG	3:C:207:ASN:HB3	2.13	0.49
1:A:118:ASN:HB2	1:A:166:GLU:HB2	1.97	0.47
1:A:138:LEU:HD12	1:A:146:PHE:CE2	2.50	0.46
3:C:145:SER:O	3:C:147:ASP:HA	2.16	0.45
4:D:125:THR:HG1	4:D:167[A]:HIS:CE1	2.36	0.43
1:A:53:ARG:O	5:J:-1:GLN:HB3	2.19	0.42
3:C:142:ASP:O	3:C:143:SER:HB2	2.19	0.42
3:C:40:PHE:O	3:C:104:CYS:HA	2.21	0.41
3:C:16:GLY:HA2	3:C:93:ALA:HA	2.02	0.41
3:C:171:TYR:O	3:C:192:ALA:HA	2.19	0.41
1:A:166:GLU:HG2	1:A:173:PRO:HB3	2.03	0.40

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	180/182 (99%)	177 (98%)	3 (2%)	0	100	100
2	B	191/190 (100%)	182 (95%)	8 (4%)	1 (0%)	34	26
3	C	193/195 (99%)	182 (94%)	9 (5%)	2 (1%)	19	11
4	D	242/242 (100%)	238 (98%)	4 (2%)	0	100	100
5	J	9/11 (82%)	9 (100%)	0	0	100	100
All	All	815/820 (99%)	788 (97%)	24 (3%)	3 (0%)	39	33

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	142	ASP
3	C	143	SER
2	B	109	LEU

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	166/166 (100%)	163 (98%)	3 (2%)	66	69
2	B	172/175 (98%)	169 (98%)	3 (2%)	68	71
3	C	170/174 (98%)	168 (99%)	2 (1%)	78	81
4	D	212/210 (101%)	204 (96%)	8 (4%)	40	36
5	J	10/10 (100%)	10 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	730/735 (99%)	714 (98%)	16 (2%)	61 62

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	11	ASN
1	A	129	THR
1	A	160	SER
2	B	104	SER
2	B	163	MET
2	B	181	GLN
3	C	129	ASP
3	C	144	LYS
4	D	13	THR
4	D	70	ASN
4	D	128	GLU
4	D	167[A]	HIS
4	D	167[B]	HIS
4	D	233	ASN
4	D	234	ASP
4	D	257	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 1 is modelled with single atom - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
6	NAG	A	1000	1	14,?,?	0.27	0	15,?,?	0.89	1 (6%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	NAG	A	1000	1	-	0/6/?/?	0/1/?/?

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	A	1000	NAG	C1-O5-C5	3.25	116.92	112.14

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	182/182 (100%)	-0.30	0 100 100	22, 38, 67, 90	0
2	B	190/190 (100%)	0.23	11 (5%) 26 28	23, 36, 77, 119	0
3	C	195/195 (100%)	-0.11	6 (3%) 52 53	26, 44, 78, 98	0
4	D	242/242 (100%)	-0.01	11 (4%) 37 38	28, 46, 84, 103	0
5	J	11/11 (100%)	-0.09	0 100 100	24, 27, 55, 73	0
All	All	820/820 (100%)	-0.05	28 (3%) 49 50	22, 41, 79, 119	0

All (28) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	106	THR	10.1
2	B	109	LEU	8.7
2	B	111	HIS	7.6
2	B	108	ALA	7.6
2	B	110	ASN	6.4
2	B	112	HIS	5.5
3	C	145	SER	5.1
2	B	191	GLN	4.7
2	B	107	GLU	4.1
3	C	147	ASP	3.5
3	C	143	SER	3.5
4	D	257	ASP	3.3
4	D	197	ASN	3.2
4	D	198	ASP	3.1
3	C	146	SER	3.1
2	B	104	SER	3.1
2	B	190	ALA	2.9
3	C	197	SER	2.9
4	D	47	ASP	2.9
4	D	195	ALA	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
4	D	231	SER	2.6
4	D	48	GLN	2.6
3	C	144	LYS	2.6
4	D	131	LYS	2.4
4	D	241	ALA	2.3
4	D	196	LEU	2.3
4	D	130	LEU	2.1
2	B	167	ARG	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
6	NAG	A	1000	14/?	0.86	0.21	-	56,69,82,82	0
7	CA	K	1	1/?	0.96	0.04	-	59,59,59,59	0

6.5 Other polymers (i)

There are no such residues in this entry.

Preliminary Full wwPDB X-ray Structure Validation Report (i)

Jun 2, 2016 – 08:22 PM EDT

This is a Preliminary Full wwPDB X-ray Structure Validation Report.
This report is produced by the wwPDB validation pipeline
before deposition or annotation of the structure.
This is not an official wwPDB validation report and is not a proof of deposition.
This report should not be submitted to journals.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

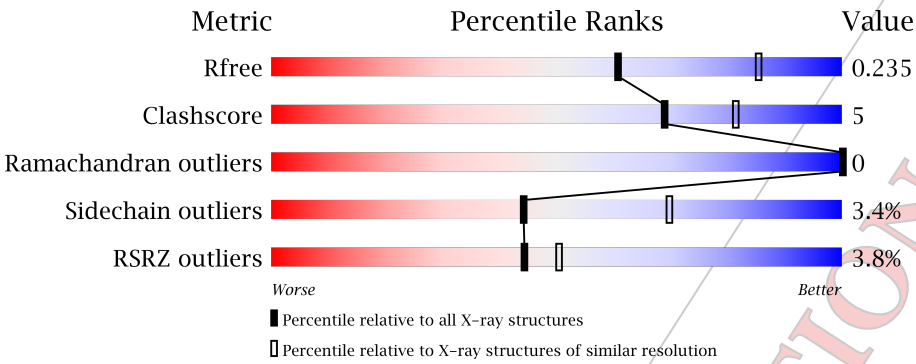
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027674
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027674

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.55 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	4549 (2.58-2.50)
Clashscore	102246	5292 (2.58-2.50)
Ramachandran outliers	100387	5194 (2.58-2.50)
Sidechain outliers	100360	5196 (2.58-2.50)
RSRZ outliers	91569	4561 (2.58-2.50)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	182	3% 89% 10% •
1	C	182	3% 88% 12%
2	B	181	3% 85% 13% ••
3	D	173	6% 85% 14% •
4	E	193	5% 84% 14% ••

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	F	240	.% 92% 8%
5	H	240	3% 90% 9%
6	G	192	7% 84% 15%
7	I	16	6% 88% 13%
7	J	16	6% 100%

2 Entry composition (i)

There are 10 unique types of molecules in this entry. The entry contains 13517 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	182	1464	944	240	278	2	0	0	0
1	C	182	1464	944	240	278	2	0	0	0

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	181	1493	946	262	278	7	0	1	0

- Molecule 3 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	D	173	1395	890	243	255	7	0	0	0

- Molecule 4 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	E	193	1484	929	252	294	9	0	0	0

- Molecule 5 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	240	1887	1193	327	362	5	0	0	0
5	H	240	1886	1193	330	358	5	0	0	0

- Molecule 6 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	G	192	1466	918	247	292	9	0	0	0

- Molecule 7 is a protein called ALA-PRO-SER-GLY-GLU-GLY-SER-PHE-GLN-PRO-SER-GLN-GLU-ASN-PRO-GLN.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
7	I	16	112	67	19	26	0	0	0
7	J	16	112	67	19	26	0	0	0

- Molecule 8 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: unknown).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
8	A	1	14	8	1	5	0	0
8	B	1	14	8	1	5	0	0
8	C	1	14	8	1	5	0	0
8	D	1	14	8	1	5	0	0

- Molecule 9 is CALCIUM ION (three-letter code: CA) (formula: unknown).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
9	K	1	1	1	0	0
9	K	1	1	1	0	0
9	K	1	1	1	0	0
9	K	1	1	1	0	0

- Molecule 10 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
10	S	445	445	445	0	0

Continued on next page...

Continued from previous page...

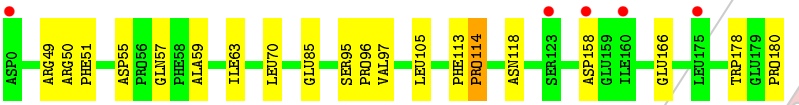
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
10	W	249	249	249	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

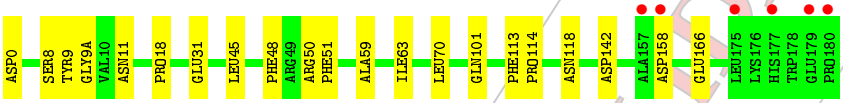
- Molecule 1:

Chain A:



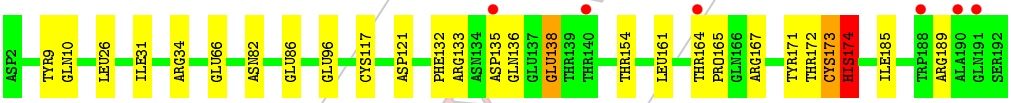
- Molecule 1:

Chain C:



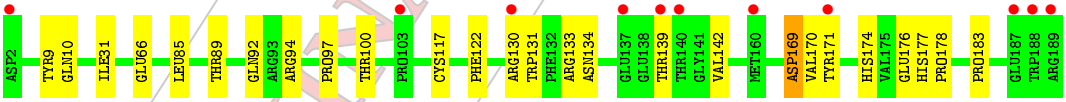
- Molecule 2:

Chain B:



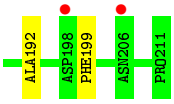
- Molecule 3:

Chain D:



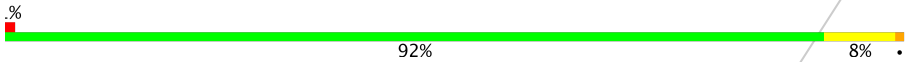
- Molecule 4:

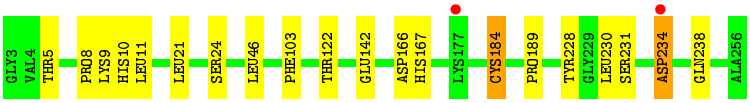
Chain E:



- Molecule 5:

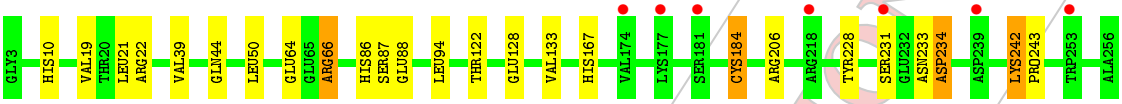
Chain F:





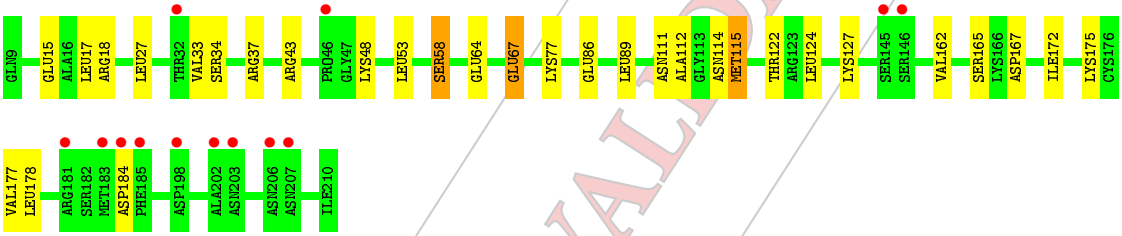
- Molecule 5:

Chain H:



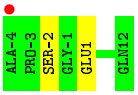
- Molecule 6:

Chain G:



- Molecule 7: ALA-PRO-SER-GLY-GLU-GLY-SER-PHE-GLN-PRO-SER-GLN-GLU-ASN-PRO-GLN

Chain I:



- Molecule 7: ALA-PRO-SER-GLY-GLU-GLY-SER-PHE-GLN-PRO-SER-GLN-GLU-ASN-PRO-GLN

Chain J:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	74.56Å 56.87Å 232.05Å 90.00° 92.77° 90.00°	Depositor
Resolution (Å)	46.90 – 2.55 46.84 – 2.55	Depositor EDS
% Data completeness (in resolution range)	98.8 (46.90-2.55) 98.8 (46.84-2.55)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$	-	Xtriage
Refinement program	BUSTER 2.10.1	Depositor
R, R_{free}	0.199 , 0.224 0.214 , 0.235	Depositor DCC
R_{free} test set	1263 reflections (2.03%)	DCC
Wilson B-factor (Å^2)	(Not available)	Xtriage
Anisotropy	(Not available)	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 57.2	EDS
L-test for twinning[1]	$< \|L\| >$ = (Not available), $< L^2 >$ = (Not available)	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	13517	wwPDB-VP
Average B, all atoms (Å^2)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *(Not available)*

[1]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: CA, NAG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| >5$	RMSZ	$\#\|Z\| >5$
1	A	0.66	1/1508 (0.1%)	0.66	0/2061
1	C	0.83	3/1508 (0.2%)	0.73	1/2061 (0.0%)
2	B	0.82	3/1534 (0.2%)	0.79	3/2092 (0.1%)
3	D	0.78	0/1430	0.71	0/1951
4	E	0.76	2/1514 (0.1%)	0.69	0/2050
5	F	0.60	0/1938	0.66	0/2644
5	H	0.56	0/1937	0.64	1/2642 (0.0%)
6	G	0.60	0/1495	0.64	0/2026
7	I	0.43	0/115	0.63	0/156
7	J	0.40	0/115	0.57	0/156
All	All	0.70	9/13094 (0.1%)	0.69	5/17839 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	B	0	1

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	174[A]	HIS	CA-C	5.81	1.68	1.52
2	B	174[B]	HIS	CA-C	5.81	1.68	1.52
1	C	31	GLU	CD-OE1	-5.79	1.19	1.25
1	A	114	PRO	N-CD	5.49	1.55	1.47
4	E	14	PRO	N-CD	5.48	1.55	1.47
4	E	115	MET	SD-CE	-5.32	1.48	1.77
1	C	31	GLU	CD-OE2	-5.20	1.20	1.25
2	B	165	PRO	N-CD	5.12	1.55	1.47

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	18	PRO	N-CD	5.01	1.54	1.47

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	173	CYS	CA-CB-SG	7.79	128.02	114.00
1	C	142	ASP	CB-CG-OD1	6.09	123.78	118.30
5	H	242	LYS	C-N-CD	5.51	139.98	128.40
2	B	96	GLU	C-N-CD	5.43	139.80	128.40
2	B	164	THR	C-N-CD	5.41	139.77	128.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	174[B]	HIS	Mainchain

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1464	0	1395	15	0
1	C	1464	0	1395	12	0
2	B	1493	0	1424	19	1
3	D	1395	0	1307	18	0
4	E	1484	0	1417	18	0
5	F	1887	0	1784	11	1
5	H	1886	0	1789	15	0
6	G	1466	0	1385	18	0
7	I	112	0	94	2	0
7	J	112	0	94	0	0
8	A	14	0	13	0	0
8	B	14	0	13	0	0
8	C	14	0	13	0	0
8	D	14	0	13	0	0
9	K	4	0	0	0	0
10	S	445	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
10	W	249	0	0	0	0
All	All	13517	0	12136	114	1

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 5.

All (114) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:132:PHE:CD1	2:B:174[B]:HIS:CD2	2.17	1.33
4:E:15:GLU:OE1	4:E:123:ARG:NH2	1.86	1.07
5:H:234:ASP:O	5:H:242:LYS:NZ	1.88	1.07
2:B:132:PHE:CE1	2:B:174[B]:HIS:CD2	2.48	1.02
2:B:132:PHE:CD1	2:B:174[B]:HIS:HD2	1.68	0.97
2:B:132:PHE:CD1	2:B:174[B]:HIS:NE2	2.35	0.95
2:B:132:PHE:HD1	2:B:174[B]:HIS:CD2	1.70	0.95
2:B:117:CYS:CB	2:B:173:CYS:SG	2.54	0.94
5:H:231:SER:HB2	5:H:233:ASN:OD1	1.76	0.85
4:E:111:ASN:HB3	4:E:115:MET:CE	2.11	0.81
2:B:132:PHE:HD1	2:B:174[B]:HIS:HD2	1.11	0.79
6:G:111:ASN:HB3	6:G:115:MET:HE3	1.64	0.78
2:B:172:THR:HG21	2:B:174[B]:HIS:CE1	2.20	0.76
6:G:178:LEU:HB3	5:H:184:CYS:HB3	1.66	0.76
3:D:139:THR:O	3:D:142:VAL:HG12	1.86	0.74
4:E:67:GLU:HG3	4:E:77:LYS:HG3	1.71	0.73
6:G:33:VAL:HG12	6:G:34:SER:N	2.03	0.73
6:G:33:VAL:HG12	6:G:34:SER:H	1.54	0.72
1:A:95:SER:HB2	1:A:96:PRO:HD2	1.70	0.71
2:B:172:THR:CG2	2:B:174[B]:HIS:CE1	2.73	0.71
6:G:111:ASN:HB3	6:G:115:MET:CE	2.20	0.70
6:G:58:SER:O	6:G:64:GLU:HB2	1.93	0.69
5:H:21:LEU:HD22	5:H:122:THR:HG21	1.74	0.68
6:G:67:GLU:HG3	6:G:77:LYS:HG3	1.79	0.64
1:A:50:ARG:HG2	1:A:51:PHE:CE1	2.33	0.64
4:E:111:ASN:HB3	4:E:115:MET:HE2	1.81	0.63
6:G:67:GLU:HG3	6:G:77:LYS:CG	2.28	0.62
4:E:178:LEU:HB3	5:F:184:CYS:HB3	1.80	0.61
2:B:132:PHE:CE1	2:B:174[B]:HIS:HD2	2.05	0.60
1:C:50:ARG:HD2	1:C:51:PHE:CZ	2.36	0.60
3:D:169:ASP:N	3:D:169:ASP:OD1	2.34	0.60
1:A:50:ARG:HG2	1:A:51:PHE:CD1	2.36	0.60

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:48:PHE:CE1	3:D:89:THR:HB	2.37	0.60
4:E:58:SER:O	4:E:64:GLU:HB2	2.02	0.59
3:D:130:ARG:HB2	3:D:174:HIS:HB3	1.85	0.59
5:H:22:ARG:HG2	5:H:88:GLU:HG2	1.85	0.58
1:A:57:GLN:OE1	5:H:66:ARG:NH2	2.36	0.58
4:E:67:GLU:HG3	4:E:77:LYS:CG	2.34	0.58
3:D:85:LEU:HD21	7:I:-2:SER:HB3	1.86	0.57
2:B:10:GLN:HB2	2:B:31:ILE:HB	1.85	0.57
4:E:33:VAL:HG12	4:E:34:SER:N	2.20	0.57
6:G:18:ARG:HH21	6:G:127:LYS:HE2	1.71	0.56
3:D:142:VAL:O	3:D:142:VAL:HG13	2.04	0.56
2:B:133:ARG:HD3	2:B:171:TYR:CE2	2.41	0.55
6:G:165:SER:HB2	6:G:172:ILE:HG12	1.88	0.55
6:G:27:LEU:HD22	6:G:122:THR:HG21	1.88	0.54
4:E:27:LEU:HD22	4:E:122:THR:HG21	1.90	0.54
1:A:118:ASN:HB2	1:A:166:GLU:HB2	1.90	0.54
1:C:118:ASN:HB2	1:C:166:GLU:HB2	1.88	0.53
3:D:176:GLU:HG3	3:D:183:PRO:HB3	1.89	0.53
3:D:117:CYS:HB2	3:D:131:TRP:CZ2	2.43	0.53
5:H:44:GLN:HB2	5:H:50:LEU:HD13	1.91	0.53
1:C:50:ARG:HD2	1:C:51:PHE:CE2	2.44	0.53
2:B:172:THR:HG22	2:B:174[B]:HIS:CE1	2.46	0.51
1:A:50:ARG:NE	1:A:51:PHE:CE1	2.79	0.51
6:G:37:ARG:HH21	6:G:114:ASN:HD21	1.58	0.50
1:A:97:VAL:HG11	1:A:180:PRO:HB3	1.93	0.50
5:H:234:ASP:OD1	5:H:234:ASP:N	2.27	0.50
1:C:45:LEU:HB2	1:C:48:PHE:CD2	2.47	0.50
4:E:111:ASN:HB3	4:E:115:MET:HE1	1.93	0.50
1:C:48:PHE:CD1	3:D:89:THR:HB	2.47	0.49
3:D:85:LEU:HD13	7:I:1:GLU:OE2	2.13	0.49
5:F:231:SER:N	5:F:234:ASP:OD1	2.46	0.49
1:C:70:LEU:HD13	3:D:9:TYR:HB2	1.96	0.48
1:C:59:ALA:O	1:C:63:ILE:HG12	2.15	0.47
4:E:141:ARG:HB2	5:F:142:GLU:HB2	1.95	0.47
5:H:167:HIS:HB3	5:H:228:TYR:HB2	1.97	0.47
1:A:85:GLU:HG3	2:B:34:ARG:NH2	2.29	0.47
3:D:97:PRO:HB3	3:D:122:PHE:HB3	1.96	0.47
1:C:48:PHE:HE1	3:D:89:THR:HB	1.80	0.47
5:F:5:THR:OG1	5:F:24:SER:HB2	2.15	0.47
1:A:59:ALA:O	1:A:63:ILE:HG12	2.16	0.46
2:B:121:ASP:HA	2:B:154:THR:HB	1.97	0.46

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:138:GLU:HG3	2:B:161:LEU:HD11	1.96	0.46
5:F:167:HIS:HB3	5:F:228:TYR:HB2	1.97	0.46
3:D:133:ARG:O	3:D:134:ASN:HB2	2.16	0.46
5:H:39:VAL:HG21	5:H:87:SER:HB2	1.98	0.46
1:A:105:LEU:HD21	1:A:178:TRP:CD2	2.50	0.46
3:D:10:GLN:HB2	3:D:31:ILE:HB	1.98	0.45
1:C:8:SER:C	1:C:9(A):GLY:HA2	2.36	0.45
5:F:166:ASP:HB2	5:F:189:PRO:HG2	1.98	0.45
5:F:8:PRO:HG3	5:F:11:LEU:HD13	1.99	0.45
6:G:33:VAL:CG1	6:G:34:SER:N	2.73	0.45
1:A:70:LEU:HD13	2:B:9:TYR:HB2	1.98	0.45
3:D:133:ARG:HG3	3:D:171:TYR:HE1	1.80	0.45
4:E:33:VAL:CG1	4:E:34:SER:N	2.79	0.45
4:E:149:SER:HB2	4:E:199:PHE:HD2	1.81	0.44
1:A:50:ARG:HD3	1:A:51:PHE:CZ	2.53	0.44
6:G:43:ARG:HB3	6:G:53:LEU:HD11	1.99	0.44
6:G:17:LEU:HB3	6:G:124:LEU:HD12	2.00	0.44
4:E:171:TYR:O	4:E:192:ALA:HA	2.17	0.44
1:A:95:SER:HB2	1:A:96:PRO:CD	2.46	0.44
1:C:113:PHE:CG	1:C:114:PRO:HA	2.52	0.44
4:E:49:GLY:HA2	5:F:103:PHE:CE1	2.53	0.44
5:H:10:HIS:HB3	5:H:167:HIS:HD1	1.83	0.43
4:E:96:PRO:HG2	4:E:186:LYS:HE2	2.00	0.43
5:H:133:VAL:O	5:H:243:PRO:HG3	2.18	0.43
6:G:33:VAL:CG1	6:G:34:SER:H	2.26	0.42
5:H:19:VAL:HG22	5:H:94:LEU:HD11	2.01	0.42
6:G:27:LEU:HD12	6:G:89:LEU:HD23	2.00	0.42
5:H:233:ASN:CG	5:H:234:ASP:N	2.73	0.42
3:D:177:HIS:CD2	3:D:178:PRO:HD2	2.54	0.42
5:F:21:LEU:HD22	5:F:122:THR:HG21	2.02	0.42
5:F:234:ASP:OD1	5:F:234:ASP:N	2.53	0.42
2:B:82:ASN:O	2:B:86:GLU:HG2	2.20	0.41
4:E:102:TYR:CE1	4:E:124:LEU:HD23	2.55	0.41
1:A:55:ASP:CG	6:G:112:ALA:HA	2.40	0.41
1:C:9:TYR:N	1:C:9(A):GLY:HA2	2.34	0.41
5:H:233:ASN:N	5:H:233:ASN:OD1	2.52	0.41
2:B:172:THR:HG23	2:B:185:ILE:HG23	2.02	0.41
3:D:133:ARG:HG3	3:D:171:TYR:CE1	2.55	0.41
5:F:9:LYS:HB3	5:F:10:HIS:CD2	2.55	0.41
1:A:113:PHE:CG	1:A:114:PRO:HA	2.56	0.40
4:E:43:ARG:HG2	4:E:53:LEU:HD21	2.03	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:167:ARG:NH2	5:F:238:GLN:O[1_546]	1.05	1.15

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	180/182 (99%)	178 (99%)	2 (1%)	0	100	100
1	C	180/182 (99%)	175 (97%)	5 (3%)	0	100	100
2	B	178/181 (98%)	169 (95%)	9 (5%)	0	100	100
3	D	167/173 (96%)	160 (96%)	7 (4%)	0	100	100
4	E	191/193 (99%)	185 (97%)	6 (3%)	0	100	100
5	F	238/240 (99%)	230 (97%)	8 (3%)	0	100	100
5	H	238/240 (99%)	229 (96%)	9 (4%)	0	100	100
6	G	190/192 (99%)	185 (97%)	5 (3%)	0	100	100
7	I	14/16 (88%)	14 (100%)	0	0	100	100
7	J	14/16 (88%)	14 (100%)	0	0	100	100
All	All	1590/1615 (98%)	1539 (97%)	51 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	165/167 (99%)	163 (99%)	2 (1%)	78	92
1	C	165/167 (99%)	161 (98%)	4 (2%)	57	81
2	B	163/167 (98%)	157 (96%)	6 (4%)	41	66
3	D	147/161 (91%)	141 (96%)	6 (4%)	37	61
4	E	163/170 (96%)	156 (96%)	7 (4%)	35	59
5	F	204/208 (98%)	200 (98%)	4 (2%)	63	85
5	H	203/208 (98%)	196 (97%)	7 (3%)	44	70
6	G	159/169 (94%)	148 (93%)	11 (7%)	19	34
7	I	12/13 (92%)	12 (100%)	0	100	100
7	J	12/13 (92%)	12 (100%)	0	100	100
All	All	1393/1443 (96%)	1346 (97%)	47 (3%)	44	70

All (47) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	49	ARG
1	A	158	ASP
2	B	26	LEU
2	B	66	GLU
2	B	135	ASP
2	B	136	GLN
2	B	138	GLU
2	B	189	ARG
1	C	0	ASP
1	C	11	ASN
1	C	101	GLN
1	C	158	ASP
3	D	66	GLU
3	D	92	GLN
3	D	94	ARG
3	D	100	THR
3	D	169	ASP
3	D	170	VAL
4	E	10	VAL
4	E	15	GLU
4	E	48	LYS
4	E	58	SER
4	E	66	LYS
4	E	115	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	E	144	LYS
5	F	46	LEU
5	F	184	CYS
5	F	230	LEU
5	F	234	ASP
6	G	15	GLU
6	G	48	LYS
6	G	58	SER
6	G	67	GLU
6	G	86	GLU
6	G	115	MET
6	G	162	VAL
6	G	167	ASP
6	G	175	LYS
6	G	177	VAL
6	G	184	ASP
5	H	64	GLU
5	H	66	ARG
5	H	86	HIS
5	H	128	GLU
5	H	184	CYS
5	H	206	ARG
5	H	234	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	134	ASN
2	B	156	GLN
3	D	64	GLN
3	D	156	GLN
5	F	132	ASN
6	G	114	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 4 are modelled with single atom - leaving 4 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
8	NAG	A	1000	1	14,?,?	0.29	0	15,?,?	0.51	0
8	NAG	B	1000	2	14,?,?	0.27	0	15,?,?	0.50	0
8	NAG	C	1000	1	14,?,?	0.27	0	15,?,?	0.47	0
8	NAG	D	1000	3	14,?,?	0.29	0	15,?,?	0.50	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
8	NAG	A	1000	1	-	0/6/?/?	0/1/?/?
8	NAG	B	1000	2	-	0/6/?/?	0/1/?/?
8	NAG	C	1000	1	-	0/6/?/?	0/1/?/?
8	NAG	D	1000	3	-	0/6/?/?	0/1/?/?

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	182/182 (100%)	-0.06	5 (2%)	58	63	14, 30, 60, 71	0
1	C	182/182 (100%)	0.05	6 (3%)	50	56	8, 29, 69, 95	0
2	B	181/181 (100%)	-0.03	6 (3%)	50	56	13, 30, 74, 113	0
3	D	173/173 (100%)	0.11	11 (6%)	23	26	7, 25, 99, 130	0
4	E	193/193 (100%)	0.29	9 (4%)	35	41	10, 37, 71, 87	0
5	F	240/240 (100%)	0.02	2 (0%)	87	89	8, 26, 53, 81	0
5	H	240/240 (100%)	0.13	7 (2%)	55	61	12, 34, 77, 104	0
6	G	192/192 (100%)	0.24	13 (6%)	20	23	15, 43, 73, 96	0
7	I	16/16 (100%)	0.00	1 (6%)	23	27	11, 19, 57, 64	0
7	J	16/16 (100%)	0.19	1 (6%)	23	27	16, 24, 64, 72	0
All	All	1615/1615 (100%)	0.09	61 (3%)	44	50	7, 32, 74, 130	0

All (61) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
7	J	-4	ALA	6.4
4	E	198	ASP	5.8
2	B	164	THR	4.4
6	G	146	SER	4.4
5	H	218	ARG	4.1
1	C	158	ASP	3.9
1	C	157	ALA	3.8
4	E	206	ASN	3.7
6	G	181	ARG	3.6
4	E	147	ASP	3.5
5	F	234	ASP	3.5
4	E	181	ARG	3.4
6	G	183	MET	3.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
4	E	146	SER	3.3
3	D	140	THR	3.3
6	G	32	THR	3.3
2	B	135	ASP	3.2
6	G	203	ASN	3.1
3	D	137	GLU	3.0
3	D	2	ASP	3.0
5	H	231	SER	3.0
1	A	175	LEU	2.9
3	D	103	PRO	2.8
6	G	145	SER	2.8
1	C	177	HIS	2.8
1	A	158	ASP	2.7
1	A	0	ASP	2.6
6	G	184	ASP	2.6
4	E	142	ASP	2.6
6	G	198	ASP	2.5
5	H	253	TRP	2.5
3	D	139	THR	2.5
5	H	177	LYS	2.5
7	I	-4	ALA	2.5
3	D	187	GLU	2.5
1	C	175	LEU	2.4
3	D	171	TYR	2.4
5	F	177	LYS	2.4
2	B	190	ALA	2.4
3	D	130	ARG	2.4
1	C	180	PRO	2.3
4	E	168	SER	2.3
6	G	206	ASN	2.3
1	A	160	ILE	2.3
2	B	188	TRP	2.3
2	B	140	THR	2.2
6	G	46	PRO	2.2
5	H	239	ASP	2.2
5	H	181	SER	2.2
6	G	207	ASN	2.1
1	C	179	GLU	2.1
2	B	191	GLN	2.1
6	G	202	ALA	2.1
3	D	189	ARG	2.1
4	E	136	ALA	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
4	E	144	LYS	2.1
5	H	174	VAL	2.1
6	G	185	PHE	2.1
3	D	160	MET	2.0
3	D	188	TRP	2.0
1	A	123	SER	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
9	CA	K	4	1/?	0.64	0.27	-	121,121,121,121	0
9	CA	K	3	1/?	0.95	0.07	-	74,74,74,74	0
8	NAG	A	1000	14/?	0.70	0.32	-	82,88,92,94	0
8	NAG	B	1000	14/?	0.74	0.17	-	80,83,86,86	0
9	CA	K	1	1/?	0.87	0.07	-	66,66,66,66	0
9	CA	K	2	1/?	0.85	0.18	-	66,66,66,66	0
8	NAG	C	1000	14/?	0.78	0.23	-	66,75,81,82	0
8	NAG	D	1000	14/?	0.80	0.19	-	66,69,71,74	0

6.5 Other polymers (i)

There are no such residues in this entry.

Preliminary Full wwPDB X-ray Structure Validation Report (i)

Jun 2, 2016 – 08:30 PM EDT

This is a Preliminary Full wwPDB X-ray Structure Validation Report.
This report is produced by the wwPDB validation pipeline
before deposition or annotation of the structure.
This is not an official wwPDB validation report and is not a proof of deposition.
This report should not be submitted to journals.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

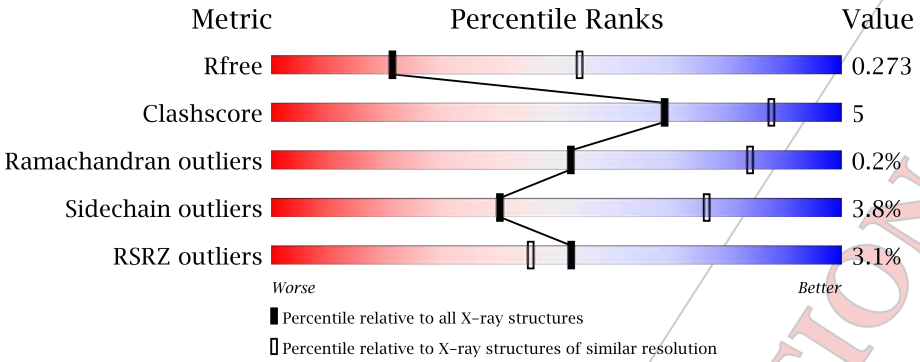
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027674
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027674

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1451 (2.90-2.90)
Clashscore	102246	1668 (2.90-2.90)
Ramachandran outliers	100387	1630 (2.90-2.90)
Sidechain outliers	100360	1632 (2.90-2.90)
RSRZ outliers	91569	1456 (2.90-2.90)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	181	9% 84% 14% •
1	C	181	7% 87% 10% •
2	B	179	.% 87% 12% •
3	D	178	5% 82% 17% •
4	E	200	.% 91% 9% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	F	241	.% 90% 10%
5	H	241	.% 89% 10%
6	G	200	2% 85% 15%
7	I	11	82% 18%
7	J	11	91% 9%

2 Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 13111 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	181	1432	922	234	274	2	0	0	0
1	C	181	1432	922	234	274	2	0	0	0

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	179	1454	923	253	271	7	0	0	0

- Molecule 3 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	D	178	1453	922	254	270	7	0	0	0

- Molecule 4 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	E	200	1543	971	251	314	7	0	0	0

- Molecule 5 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	241	1911	1205	336	365	5	0	0	0
5	H	241	1908	1204	336	363	5	0	0	0

- Molecule 6 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	G	200	1533	966	251	309	7	0	0	0

- Molecule 7 is a protein called GLY-PRO-GLN-GLN-SER-PHE-PRO-GLU-GLN-GLU-ALA.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
7	I	11	85	52	14	19	0	0	0
7	J	11	85	52	14	19	0	0	0

- Molecule 8 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: unknown).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
8	A	1	14	8	1	5	0	0
8	A	1	14	8	1	5	0	0
8	A	1	14	8	1	5	0	0
8	B	1	14	8	1	5	0	0
8	C	1	14	8	1	5	0	0
8	C	1	14	8	1	5	0	0
8	C	1	14	8	1	5	0	0
8	D	1	14	8	1	5	0	0

- Molecule 9 is water.

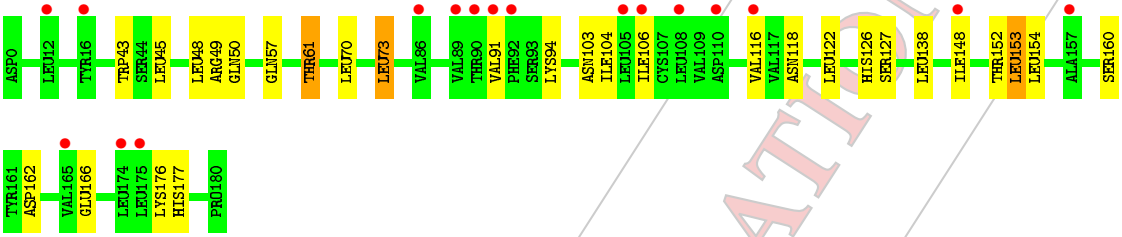
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
9	L	36	36	36	0	0
9	W	127	127	127	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

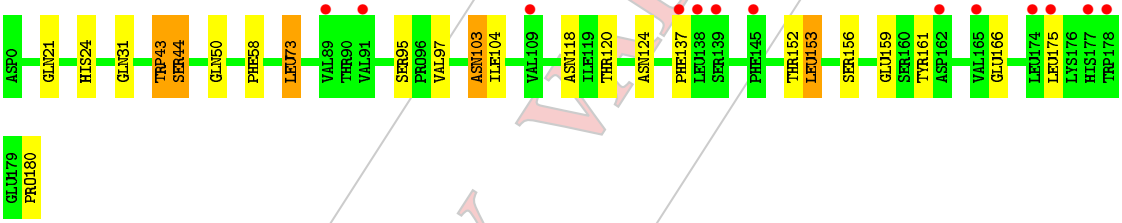
- Molecule 1:

Chain A:



- Molecule 1:

Chain C:

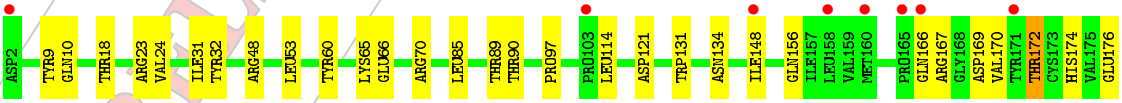


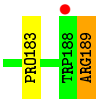
- Molecule 2:

Chain B:

- Molecule 3:

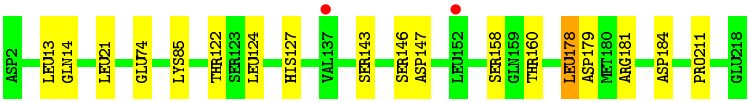
Chain D:





- Molecule 4:

Chain E:

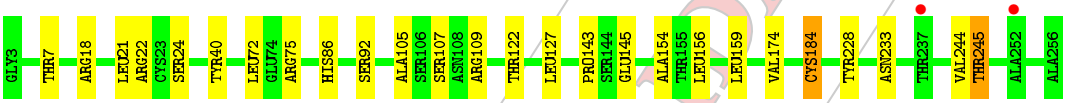


- Molecule 5:

Chain F:

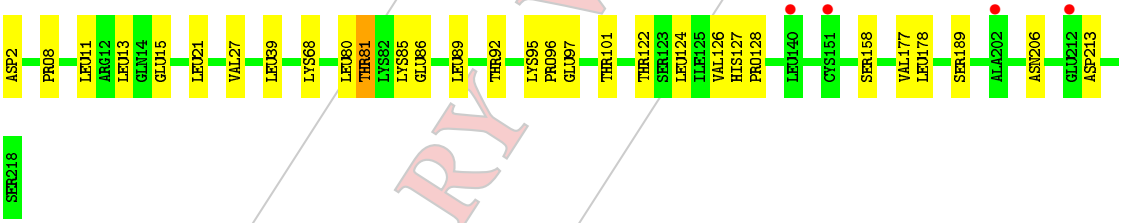
- Molecule 5:

Chain H:



- Molecule 6:

Chain G:



- Molecule 7: GLY-PRO-GLN-GLN-SER-PHE-PRO-GLU-GLN-GLU-ALA

Chain I:

- Molecule 7: GLY-PRO-GLN-GLN-SER-PHE-PRO-GLU-GLN-GLU-ALA

Chain J:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	115.89Å 125.05Å 165.63Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	85.00 – 2.90 59.32 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.4 (85.00-2.90) 99.4 (59.32-2.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$	-	Xtriage
Refinement program	BUSTER 2.10.1	Depositor
R, R_{free}	0.211 , 0.250 0.234 , 0.273	Depositor DCC
R_{free} test set	2710 reflections (5.32%)	DCC
Wilson B-factor (Å^2)	(Not available)	Xtriage
Anisotropy	(Not available)	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 53.2	EDS
L-test for twinning[1]	$< \|L\| >$ = (Not available), $< L^2 >$ = (Not available)	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	13111	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *(Not available)*

[1]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: NAG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| >5$	RMSZ	$\#\|Z\| >5$
1	A	0.39	0/1474	0.61	0/2017
1	C	0.91	2/1474 (0.1%)	0.90	4/2017 (0.2%)
2	B	0.49	0/1491	0.65	1/2035 (0.0%)
3	D	0.43	0/1490	0.63	0/2033
4	E	0.38	0/1579	0.63	0/2145
5	F	0.37	0/1961	0.59	0/2671
5	H	0.37	0/1958	0.58	0/2667
6	G	0.40	0/1569	0.60	0/2132
7	I	0.40	0/87	0.61	0/117
7	J	0.43	0/87	0.62	0/117
All	All	0.49	2/13170 (0.0%)	0.65	5/17951 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	43	TRP	C-N	-30.00	0.65	1.34
1	C	44	SER	C-N	-9.72	1.11	1.34

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	44	SER	O-C-N	-20.99	89.11	122.70
1	C	44	SER	CA-C-N	14.96	150.12	117.20
1	C	44	SER	C-N-CA	12.84	153.80	121.70
1	C	43	TRP	C-N-CA	8.51	142.98	121.70
2	B	164	THR	C-N-CD	6.05	141.11	128.40

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1432	0	1368	22	0
1	C	1432	0	1366	21	0
2	B	1454	0	1375	13	0
3	D	1453	0	1378	25	0
4	E	1543	0	1440	10	0
5	F	1911	0	1820	12	0
5	H	1908	0	1818	11	0
6	G	1533	0	1433	14	0
7	I	85	0	74	2	0
7	J	85	0	74	1	0
8	A	42	0	38	0	0
8	B	14	0	13	0	0
8	C	42	0	38	0	0
8	D	14	0	13	0	0
9	L	36	0	0	0	0
9	W	127	0	0	0	0
All	All	13111	0	12248	117	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 5.

All (117) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:43:TRP:CA	1:C:44:SER:N	1.94	1.29
1:C:43:TRP:C	1:C:44:SER:CA	2.02	1.27
1:C:43:TRP:O	1:C:44:SER:N	1.64	1.26
3:D:166:GLN:NE2	3:D:167:ARG:O	1.88	1.05
1:A:103:ASN:O	1:A:153:LEU:HD12	1.62	1.00
3:D:172:THR:OG1	3:D:174:HIS:CD2	2.19	0.94
3:D:169:ASP:O	3:D:189:ARG:HD2	1.77	0.82
3:D:172:THR:HG1	3:D:174:HIS:CD2	1.97	0.79
1:A:160:SER:OG	1:A:177:HIS:CE1	2.42	0.72
1:C:43:TRP:C	1:C:44:SER:N	0.65	0.70
3:D:89:THR:HG23	3:D:90:THR:H	1.56	0.69

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:H:72:LEU:HD22	5:H:75:ARG:HH11	1.59	0.66
5:F:111:LEU:HD23	7:I:5:PHE:HB3	1.78	0.66
2:B:133:ARG:O	2:B:136:GLN:CB	2.44	0.66
1:C:156:SER:HB3	1:C:159:GLU:HG3	1.78	0.66
2:B:133:ARG:O	2:B:136:GLN:N	2.30	0.64
3:D:172:THR:OG1	3:D:174:HIS:HD2	1.82	0.63
5:H:109:ARG:HG2	7:J:8:GLN:HB2	1.81	0.62
3:D:134:ASN:ND2	3:D:169:ASP:OD1	2.33	0.62
4:E:181:ARG:HG2	6:G:92:THR:HG21	1.80	0.62
2:B:133:ARG:N	2:B:136:GLN:O	2.33	0.61
6:G:13:LEU:HD12	6:G:124:LEU:HD11	1.82	0.60
5:F:21:LEU:HD22	5:F:122:THR:HG21	1.85	0.59
5:F:39:VAL:HG21	5:F:87:SER:HB2	1.86	0.58
2:B:114:LEU:HD22	2:B:163:MET:HG3	1.85	0.58
3:D:131:TRP:CE3	3:D:172:THR:O	2.57	0.58
3:D:18:THR:HB	3:D:23:ARG:HB3	1.87	0.57
5:H:21:LEU:HD22	5:H:122:THR:HG21	1.86	0.57
1:A:104:ILE:HG12	1:A:152:THR:HG22	1.86	0.56
4:E:14:GLN:HE22	4:E:158:SER:HB2	1.70	0.56
2:B:134:ASN:OD1	2:B:169:ASP:HA	2.06	0.56
1:C:95:SER:H	1:C:103:ASN:HD21	1.53	0.56
1:C:21:GLN:HE22	1:C:137:PHE:H	1.55	0.55
1:C:118:ASN:HB2	1:C:166:GLU:HB2	1.88	0.54
3:D:166:GLN:HE22	3:D:167:ARG:HG2	1.72	0.54
4:E:21:LEU:HD22	4:E:122:THR:HG21	1.91	0.52
5:F:22:ARG:HG2	5:F:88:GLU:HG2	1.91	0.52
5:H:228:TYR:HA	5:H:245:THR:HB	1.91	0.52
1:A:122:LEU:HB2	1:A:162:ASP:HB2	1.92	0.52
1:A:103:ASN:O	1:A:153:LEU:CD1	2.46	0.52
5:H:40:TYR:HB2	5:H:105:ALA:HB3	1.91	0.51
5:F:105:ALA:HB1	5:F:116:GLN:HG2	1.93	0.51
3:D:48:ARG:HH22	5:H:244:VAL:HB	1.74	0.51
5:F:170:LEU:HG	5:F:225:VAL:HG22	1.93	0.51
2:B:10:GLN:HB2	2:B:31:ILE:HB	1.94	0.50
5:F:140:VAL:HG23	5:F:250:ALA:HB3	1.94	0.50
1:C:73:LEU:HD22	3:D:32:TYR:HB2	1.94	0.50
1:A:45:LEU:HB2	1:A:48:LEU:HD22	1.94	0.49
3:D:166:GLN:NE2	3:D:167:ARG:N	2.60	0.49
1:C:153:LEU:HD23	1:C:161:TYR:CE2	2.48	0.49
3:D:176:GLU:HG3	3:D:183:PRO:HB3	1.95	0.49
2:B:172:THR:HG22	2:B:187:GLU:HB3	1.93	0.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:D:10:GLN:HB2	3:D:31:ILE:HB	1.94	0.49
1:A:73:LEU:HD22	2:B:32:TYR:HB2	1.94	0.49
3:D:166:GLN:O	3:D:169:ASP:HB2	2.13	0.49
4:E:13:LEU:HD12	4:E:124:LEU:HD11	1.95	0.49
4:E:127:HIS:HB3	4:E:158:SER:HB3	1.95	0.48
6:G:21:LEU:HD22	6:G:122:THR:HG21	1.95	0.48
6:G:127:HIS:HB3	6:G:158:SER:HB3	1.95	0.48
1:C:24:HIS:HB3	1:C:31:GLN:HE21	1.79	0.48
4:E:160:THR:HG21	4:E:211:PRO:HD3	1.94	0.48
1:A:94:LYS:HD2	1:A:104:ILE:HD12	1.95	0.47
3:D:166:GLN:HE22	3:D:167:ARG:CG	2.26	0.47
1:A:160:SER:OG	1:A:177:HIS:NE2	2.47	0.47
5:F:143:PRO:HG2	5:F:154:ALA:HB1	1.96	0.47
1:A:118:ASN:HB2	1:A:166:GLU:HB2	1.97	0.47
1:C:124:ASN:HD21	1:C:159:GLU:HA	1.80	0.47
3:D:66:GLU:O	3:D:70:ARG:HG2	2.15	0.47
1:C:104:ILE:HG12	1:C:152:THR:HG22	1.97	0.47
5:H:40:TYR:HE1	5:H:107:SER:HB3	1.78	0.46
6:G:15:GLU:HG3	6:G:96:PRO:HD3	1.97	0.46
1:C:58:PHE:HB3	7:I:3:GLN:HE22	1.81	0.46
5:F:43:GLN:HB2	5:F:53:LEU:HD11	1.97	0.46
4:E:74:GLU:HG2	6:G:97:GLU:HG2	1.98	0.46
4:E:178:LEU:HB3	5:F:184:CYS:HB3	1.98	0.45
6:G:178:LEU:HB3	5:H:184:CYS:HB3	1.99	0.45
1:C:153:LEU:CD1	1:C:153:LEU:C	2.86	0.45
1:C:73:LEU:HD13	3:D:9:TYR:HE1	1.82	0.45
1:C:43:TRP:O	1:C:44:SER:CA	2.44	0.45
3:D:167:ARG:HG3	3:D:167:ARG:O	2.17	0.44
1:A:153:LEU:CD1	1:A:153:LEU:C	2.86	0.44
5:F:230:LEU:HD12	5:F:243:PRO:HD2	2.00	0.44
1:C:97:VAL:HG11	1:C:180:PRO:HB3	2.00	0.43
1:A:153:LEU:HD13	1:A:153:LEU:O	2.18	0.43
3:D:170:VAL:HA	3:D:189:ARG:HD2	2.00	0.43
3:D:85:LEU:O	3:D:89:THR:HG22	2.18	0.43
4:E:179:ASP:HB3	4:E:181:ARG:HD2	2.00	0.43
3:D:60:TYR:HA	5:H:7:THR:HG21	2.00	0.43
1:A:153:LEU:HD13	1:A:153:LEU:C	2.39	0.43
2:B:165:PRO:HG2	2:B:165:PRO:O	2.19	0.42
1:C:103:ASN:O	1:C:153:LEU:HD12	2.19	0.42
1:A:57:GLN:O	1:A:61:THR:HB	2.19	0.42
2:B:133:ARG:O	2:B:136:GLN:CA	2.67	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:160:SER:HG	1:A:177:HIS:CE1	2.38	0.42
6:G:39:LEU:HD13	6:G:80:LEU:HD13	2.01	0.42
4:E:143:SER:HB3	4:E:146:SER:HB3	2.01	0.42
6:G:21:LEU:HD12	6:G:89:LEU:HD23	2.02	0.42
1:C:156:SER:HB3	1:C:159:GLU:CG	2.49	0.41
2:B:117:CYS:HB2	2:B:131:TRP:CZ2	2.55	0.41
1:C:153:LEU:HD13	1:C:153:LEU:C	2.41	0.41
5:H:18:ARG:HH11	5:H:92:SER:HB3	1.86	0.41
6:G:2:ASP:HA	6:G:27:VAL:HG23	2.03	0.41
1:A:43:TRP:CD1	1:A:49:ARG:HA	2.56	0.41
1:A:73:LEU:HD23	2:B:53:LEU:HD23	2.03	0.41
6:G:128:PRO:HG3	6:G:177:VAL:HG21	2.02	0.41
1:A:122:LEU:HD23	1:A:127:SER:HA	2.02	0.41
1:A:91:VAL:HG23	1:A:176:LYS:HB3	2.02	0.41
1:A:43:TRP:CG	1:A:49:ARG:HA	2.56	0.41
3:D:131:TRP:HE3	3:D:172:THR:O	2.02	0.40
3:D:148:ILE:HB	3:D:156:GLN:HB3	2.02	0.40
5:F:168:VAL:HG12	5:F:227:PHE:HA	2.02	0.40
6:G:96:PRO:HA	6:G:126:VAL:HB	2.02	0.40
6:G:81:THR:HG23	6:G:85:LYS:H	1.86	0.40
5:H:143:PRO:HG2	5:H:154:ALA:HB1	2.02	0.40
1:A:106:ILE:HG23	1:A:148:ILE:HG23	2.04	0.40
1:A:70:LEU:HD13	2:B:9:TYR:HB2	2.02	0.40
6:G:8:PRO:HG3	6:G:11:LEU:HD13	2.03	0.40

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	179/181 (99%)	172 (96%)	7 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	179/181 (99%)	174 (97%)	5 (3%)	0	100	100
2	B	175/179 (98%)	165 (94%)	9 (5%)	1 (1%)	30	67
3	D	174/178 (98%)	169 (97%)	4 (2%)	1 (1%)	30	67
4	E	198/200 (99%)	188 (95%)	10 (5%)	0	100	100
5	F	239/241 (99%)	233 (98%)	5 (2%)	1 (0%)	39	74
5	H	239/241 (99%)	234 (98%)	5 (2%)	0	100	100
6	G	198/200 (99%)	188 (95%)	10 (5%)	0	100	100
7	I	9/11 (82%)	9 (100%)	0	0	100	100
7	J	9/11 (82%)	9 (100%)	0	0	100	100
All	All	1599/1623 (98%)	1541 (96%)	55 (3%)	3 (0%)	52	84

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	121	ASP
3	D	121	ASP
5	F	166	ASP

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	162/165 (98%)	154 (95%)	8 (5%)	31	67
1	C	162/165 (98%)	156 (96%)	6 (4%)	41	77
2	B	156/165 (94%)	149 (96%)	7 (4%)	34	70
3	D	157/165 (95%)	150 (96%)	7 (4%)	34	70
4	E	170/177 (96%)	166 (98%)	4 (2%)	57	86
5	F	207/209 (99%)	204 (99%)	3 (1%)	74	93
5	H	206/209 (99%)	195 (95%)	11 (5%)	28	63
6	G	168/177 (95%)	160 (95%)	8 (5%)	31	67

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
7	I	9/9 (100%)	9 (100%)	0	100	100
7	J	9/9 (100%)	9 (100%)	0	100	100
All	All	1406/1450 (97%)	1352 (96%)	54 (4%)	40	76

All (54) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	50	GLN
1	A	61	THR
1	A	73	LEU
1	A	116	VAL
1	A	126	HIS
1	A	138	LEU
1	A	153	LEU
1	A	154	LEU
2	B	43	ASP
2	B	88	ARG
2	B	114	LEU
2	B	120	THR
2	B	127	ILE
2	B	134	ASN
2	B	137	GLU
1	C	50	GLN
1	C	73	LEU
1	C	103	ASN
1	C	120	THR
1	C	153	LEU
1	C	175	LEU
3	D	24	VAL
3	D	53	LEU
3	D	65	LYS
3	D	97	PRO
3	D	114	LEU
3	D	172	THR
3	D	189	ARG
4	E	85	LYS
4	E	147	ASP
4	E	178	LEU
4	E	184	ASP
5	F	48	GLN
5	F	184	CYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
5	F	237	THR
6	G	68	LYS
6	G	81	THR
6	G	86	GLU
6	G	95	LYS
6	G	101	THR
6	G	189	SER
6	G	206	ASN
6	G	213	ASP
5	H	22	ARG
5	H	24	SER
5	H	86	HIS
5	H	127	LEU
5	H	145	GLU
5	H	156	LEU
5	H	159	LEU
5	H	174	VAL
5	H	184	CYS
5	H	233	ASN
5	H	245	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	21	GLN
1	C	31	GLN
1	C	103	ASN
1	C	124	ASN
3	D	92	GLN
3	D	134	ASN
3	D	166	GLN
4	E	14	GLN
4	E	131	GLN
5	F	219	ASN
5	H	233	ASN
7	I	2	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

8 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
8	NAG	A	1000	1	14,?,?	0.28	0	15,?,?	0.69	1 (6%)
8	NAG	A	1001	1,8	14,?,?	0.25	0	15,?,?	0.67	1 (6%)
8	NAG	A	1002	8	14,?,?	0.27	0	15,?,?	0.37	0
8	NAG	B	1000	2	14,?,?	0.38	0	15,?,?	1.86	2 (13%)
8	NAG	C	1000	1	14,?,?	0.36	0	15,?,?	0.84	1 (6%)
8	NAG	C	1001	1,8	14,?,?	0.27	0	15,?,?	0.63	1 (6%)
8	NAG	C	1002	8	14,?,?	0.30	0	15,?,?	0.71	0
8	NAG	D	1000	3	14,?,?	0.37	0	15,?,?	1.65	1 (6%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
8	NAG	A	1000	1	-	0/6/?/?	0/1/?/?
8	NAG	A	1001	1,8	-	0/6/?/?	0/1/?/?
8	NAG	A	1002	8	-	0/6/?/?	0/1/?/?
8	NAG	B	1000	2	-	0/6/?/?	0/1/?/?
8	NAG	C	1000	1	-	0/6/?/?	0/1/?/?
8	NAG	C	1001	1,8	-	0/6/?/?	0/1/?/?

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
8	NAG	C	1002	8	-	0/6/?/?	0/1/?/?
8	NAG	D	1000	3	-	0/6/?/?	0/1/?/?

There are no bond length outliers.

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
8	C	1001	NAG	C1-O5-C5	2.06	115.16	112.14
8	A	1001	NAG	C1-O5-C5	2.13	115.28	112.14
8	B	1000	NAG	C2-N2-C7	2.35	126.16	123.11
8	A	1000	NAG	C1-O5-C5	2.41	115.69	112.14
8	C	1000	NAG	C1-O5-C5	2.93	116.45	112.14
8	D	1000	NAG	C1-O5-C5	6.06	121.06	112.14
8	B	1000	NAG	C1-O5-C5	6.25	121.32	112.14

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	181/181 (100%)	0.58	17 (9%)	11	6	29, 57, 92, 103	0
1	C	181/181 (100%)	0.51	13 (7%)	18	12	34, 61, 93, 114	0
2	B	179/179 (100%)	0.11	2 (1%)	82	80	30, 54, 115, 143	0
3	D	178/178 (100%)	0.21	9 (5%)	32	25	27, 55, 117, 145	0
4	E	200/200 (100%)	0.15	2 (1%)	84	82	30, 51, 74, 109	0
5	F	241/241 (100%)	-0.05	2 (0%)	87	86	28, 46, 76, 99	0
5	H	241/241 (100%)	0.01	2 (0%)	87	86	29, 44, 74, 95	0
6	G	200/200 (100%)	0.09	4 (2%)	68	64	22, 46, 72, 92	0
7	I	11/11 (100%)	-0.01	0	100	100	28, 34, 47, 60	0
7	J	11/11 (100%)	-0.04	0	100	100	30, 35, 47, 57	0
All	All	1623/1623 (100%)	0.18	51 (3%)	52	45	22, 50, 92, 145	0

All (51) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	2	ASP	6.1
1	A	165	VAL	4.9
3	D	188	TRP	4.4
1	A	174	LEU	3.7
2	B	165	PRO	3.5
1	A	175	LEU	3.4
1	A	148	ILE	3.3
1	C	139	SER	3.1
3	D	103	PRO	3.1
3	D	2	ASP	3.1
1	A	91	VAL	3.0
1	C	162	ASP	3.0
6	G	202	ALA	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
3	D	165	PRO	2.9
3	D	166	GLN	2.9
1	A	86	VAL	2.9
1	A	92	PHE	2.9
1	C	89	VAL	2.8
1	C	175	LEU	2.8
1	C	138	LEU	2.8
1	A	116	VAL	2.7
1	A	12	LEU	2.7
3	D	148	ILE	2.7
1	A	89	VAL	2.6
1	A	110	ASP	2.6
1	C	145	PHE	2.6
6	G	140	LEU	2.6
3	D	158	LEU	2.6
1	C	174	LEU	2.5
1	C	109	VAL	2.5
1	A	90	THR	2.5
1	C	177	HIS	2.4
1	C	91	VAL	2.4
3	D	171	TYR	2.4
1	A	16	TYR	2.3
1	A	108	LEU	2.3
5	F	239	ASP	2.3
4	E	152	LEU	2.3
1	A	157	ALA	2.3
1	C	137	PHE	2.2
6	G	212	GLU	2.2
6	G	151	CYS	2.2
3	D	160	MET	2.2
5	H	237	THR	2.2
1	A	105	LEU	2.1
1	A	106	ILE	2.1
4	E	137	VAL	2.1
5	F	130	LEU	2.0
5	H	252	ALA	2.0
1	C	178	TRP	2.0
1	C	165	VAL	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
8	NAG	A	1000	14/?	0.61	0.24	-	93,99,107,108	0
8	NAG	A	1002	14/?	0.83	0.29	-	98,100,106,106	0
8	NAG	C	1001	14/?	0.90	0.25	-	80,86,89,92	0
8	NAG	A	1001	14/?	0.92	0.21	-	78,84,90,93	0
8	NAG	C	1002	14/?	0.87	0.27	-	95,99,101,102	0
8	NAG	B	1000	14/?	0.61	0.21	-	109,111,116,117	0
8	NAG	C	1000	14/?	0.64	0.26	-	116,121,123,123	0
8	NAG	D	1000	14/?	0.59	0.25	-	115,118,122,123	0

6.5 Other polymers (i)

There are no such residues in this entry.