wwPDB X-ray Structure Validation Summary Report (i)

Oct 30, 2023 – 05:07 PM JST

PDB ID	:	8W86
Title	:	HLA-DQ2.5-B/C hordein peptide in complex with DQN0385AE02
Authors	:	Irie, M.; Tsushima, T.; Teranishi-Ikawa, Y.; Takahashi, N.; Ishii, S.; Okura, Y.; Fukami, T.A.; Torizawa, T.
Deposited on	:	2023-08-31
Resolution	:	2.24 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

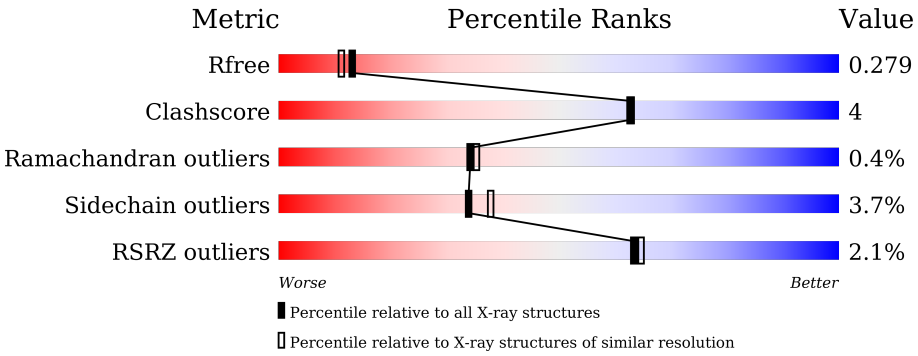
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.24 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.



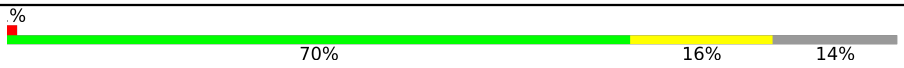
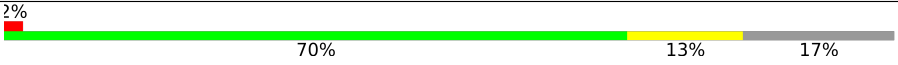
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2391 (2.26-2.22)
Clashscore	141614	2539 (2.26-2.22)
Ramachandran outliers	138981	2489 (2.26-2.22)
Sidechain outliers	138945	2490 (2.26-2.22)
RSRZ outliers	127900	2353 (2.26-2.22)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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	223	6% 79% 15% 5%
1	E	223	2% 46% 6% 48%
2	B	213	92% 7%
2	F	213	46% 50%
3	C	189	% 82% 13%
3	G	189	% 81% 14%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	D	230	
4	H	230	

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 11188 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 called DQN0385AE02 Fab heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	211	Total	C	N	O	S	0	0	0
			1538	976	250	306	6			
1	E	116	Total	C	N	O	S	0	0	0
			872	557	146	165	4			

- Molecule 2 is a protein called DQN0385AE02 Fab light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	211	Total	C	N	O	S	0	0	0
			1595	1002	261	327	5			
2	F	106	Total	C	N	O	S	0	0	0
			807	513	128	163	3			

- Molecule 3 is a protein called HLA class II histocompatibility antigen, DQ alpha 1 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	181	Total	C	N	O	S	0	0	0
			1440	927	235	276	2			
3	G	181	Total	C	N	O	S	0	0	0
			1422	916	234	270	2			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	47	SER	CYS	engineered mutation	UNP P01909
C	184	LEU	-	expression tag	UNP P01909
C	185	GLU	-	expression tag	UNP P01909
C	186	VAL	-	expression tag	UNP P01909
C	187	LEU	-	expression tag	UNP P01909
C	188	PHE	-	expression tag	UNP P01909
C	189	GLN	-	expression tag	UNP P01909
G	47	SER	CYS	engineered mutation	UNP P01909

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	184	LEU	-	expression tag	UNP P01909
G	185	GLU	-	expression tag	UNP P01909
G	186	VAL	-	expression tag	UNP P01909
G	187	LEU	-	expression tag	UNP P01909
G	188	PHE	-	expression tag	UNP P01909
G	189	GLN	-	expression tag	UNP P01909

- Molecule 4 is a protein called MHC class II HLA-DQ-beta-1 - B/C hordein peptide chimeric protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	197	Total	C	N	O	S	0	0	0
			1605	1015	283	300	7			
4	H	192	Total	C	N	O	S	0	0	0
			1529	975	265	282	7			

There are 46 discrepancies between the modelled and reference sequences:

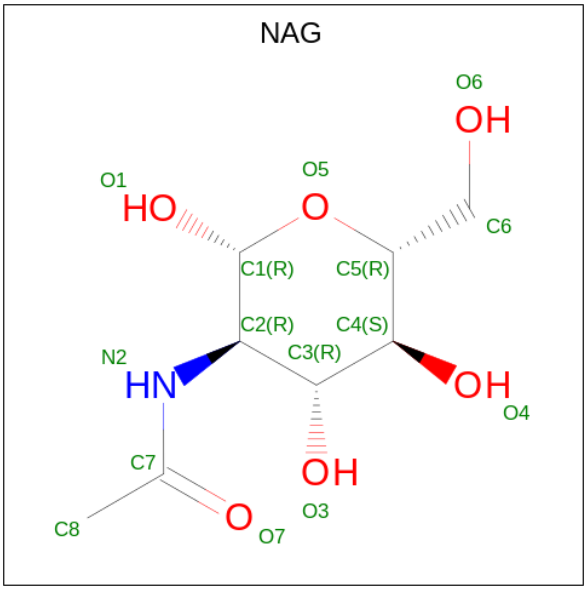
Chain	Residue	Modelled	Actual	Comment	Reference
D	984	GLY	-	linker	PDB ?
D	985	SER	-	linker	PDB ?
D	986	GLY	-	linker	PDB ?
D	987	GLY	-	linker	PDB ?
D	988	GLY	-	linker	PDB ?
D	989	GLY	-	linker	PDB ?
D	990	SER	-	linker	PDB ?
D	991	ILE	-	linker	PDB ?
D	992	GLU	-	linker	PDB ?
D	993	GLY	-	linker	PDB ?
D	994	ARG	-	linker	PDB ?
D	995	GLY	-	linker	PDB ?
D	996	SER	-	linker	PDB ?
D	997	GLY	-	linker	PDB ?
D	998	GLY	-	linker	PDB ?
D	999	GLY	-	linker	PDB ?
D	1000	SER	-	linker	PDB ?
D	1191	LEU	-	expression tag	UNP O19712
D	1192	GLU	-	expression tag	UNP O19712
D	1193	VAL	-	expression tag	UNP O19712
D	1194	LEU	-	expression tag	UNP O19712
D	1195	PHE	-	expression tag	UNP O19712
D	1196	GLN	-	expression tag	UNP O19712

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
H	984	GLY	-	linker	PDB ?
H	985	SER	-	linker	PDB ?
H	986	GLY	-	linker	PDB ?
H	987	GLY	-	linker	PDB ?
H	988	GLY	-	linker	PDB ?
H	989	GLY	-	linker	PDB ?
H	990	SER	-	linker	PDB ?
H	991	ILE	-	linker	PDB ?
H	992	GLU	-	linker	PDB ?
H	993	GLY	-	linker	PDB ?
H	994	ARG	-	linker	PDB ?
H	995	GLY	-	linker	PDB ?
H	996	SER	-	linker	PDB ?
H	997	GLY	-	linker	PDB ?
H	998	GLY	-	linker	PDB ?
H	999	GLY	-	linker	PDB ?
H	1000	SER	-	linker	PDB ?
H	1191	LEU	-	expression tag	UNP O19712
H	1192	GLU	-	expression tag	UNP O19712
H	1193	VAL	-	expression tag	UNP O19712
H	1194	LEU	-	expression tag	UNP O19712
H	1195	PHE	-	expression tag	UNP O19712
H	1196	GLN	-	expression tag	UNP O19712

- Molecule 5 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



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

- Molecule 6 is water.

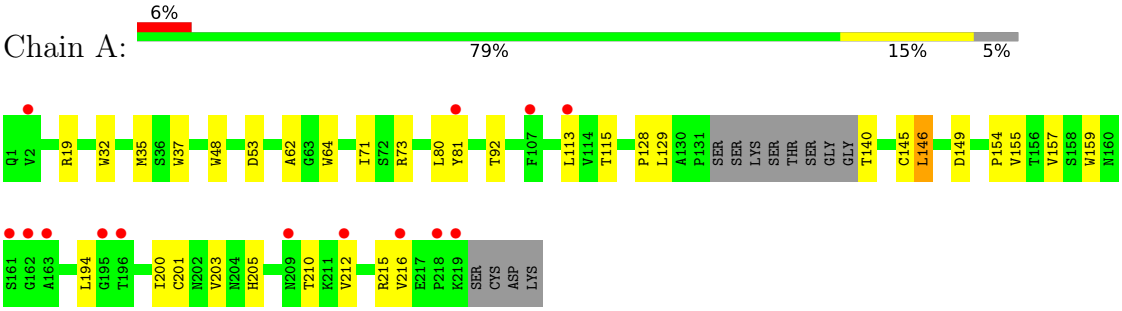
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	34	Total O 34 34	0	0
6	B	52	Total O 52 52	0	0
6	C	54	Total O 54 54	0	0
6	D	54	Total O 54 54	0	0
6	E	15	Total O 15 15	0	0
6	F	25	Total O 25 25	0	0
6	G	60	Total O 60 60	0	0
6	H	58	Total O 58 58	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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.

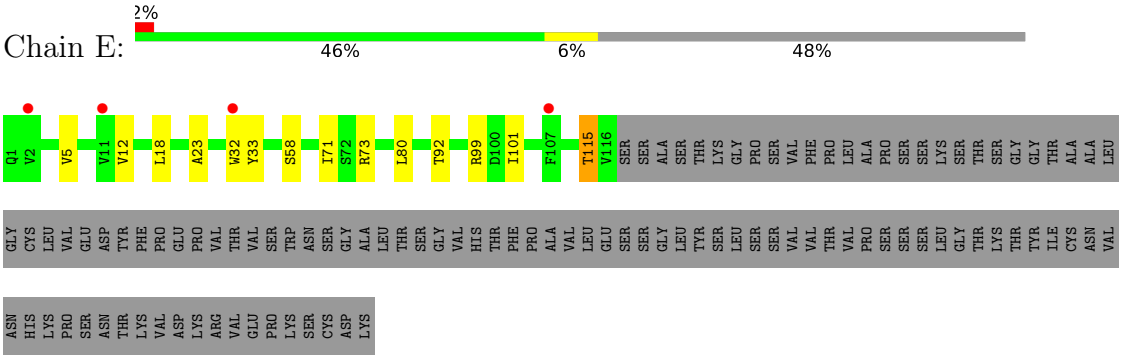
• Molecule 1: DQN0385AE02 Fab heavy chain

Chain A:



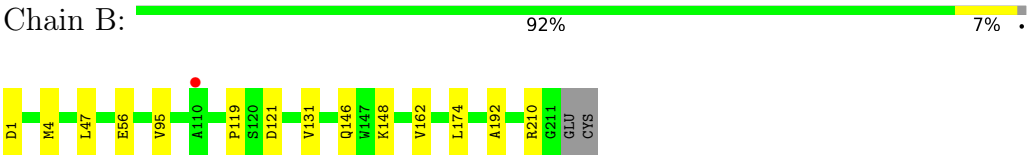
• Molecule 1: DQN0385AE02 Fab heavy chain

Chain E:



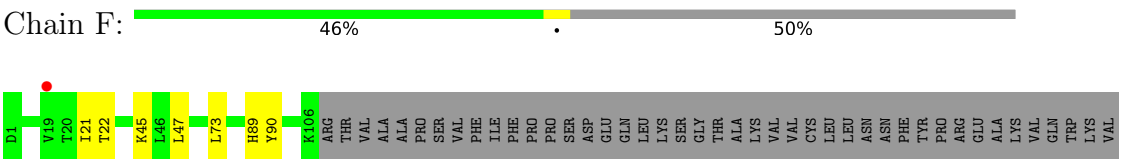
• Molecule 2: DQN0385AE02 Fab light chain

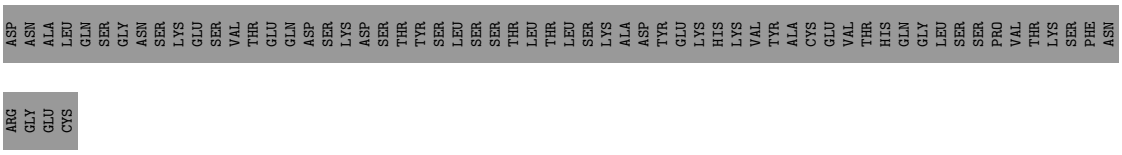
Chain B:



• Molecule 2: DQN0385AE02 Fab light chain

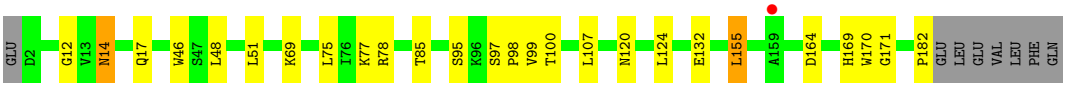
Chain F:





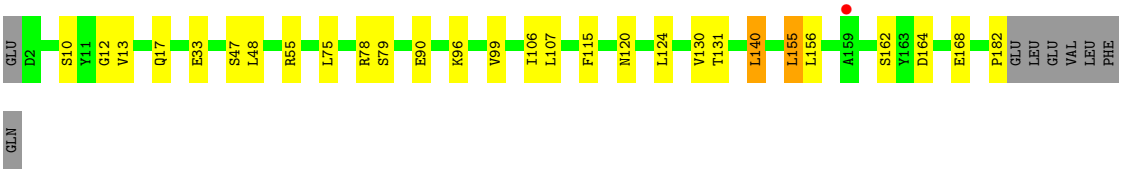
• Molecule 3: HLA class II histocompatibility antigen, DQ alpha 1 chain

Chain C:



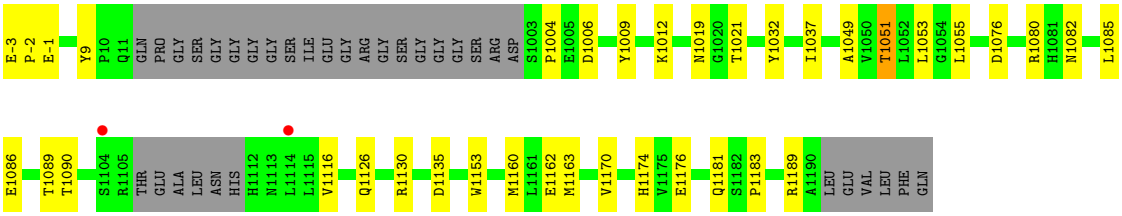
• Molecule 3: HLA class II histocompatibility antigen, DQ alpha 1 chain

Chain G:



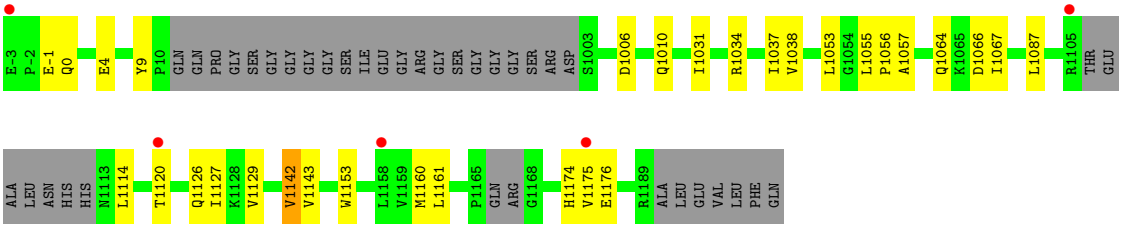
• Molecule 4: MHC class II HLA-DQ-beta-1 - B/C hordein peptide chimeric protein

Chain D:



• Molecule 4: MHC class II HLA-DQ-beta-1 - B/C hordein peptide chimeric protein

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	73.70Å 127.19Å 130.37Å 90.00° 104.45° 90.00°	Depositor
Resolution (Å)	126.22 – 2.24 126.24 – 2.24	Depositor EDS
% Data completeness (in resolution range)	69.1 (126.22-2.24) 69.1 (126.24-2.24)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.50 (at 2.22Å)	Xtriage
Refinement program	BUSTER 2.11.8	Depositor
R, R_{free}	0.241 , 0.280 0.234 , 0.279	Depositor DCC
R_{free} test set	4001 reflections (5.16%)	wwPDB-VP
Wilson B-factor (Å^2)	51.4	Xtriage
Anisotropy	0.021	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 37.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.014 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	11188	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.14% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

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

5 Model quality

5.1 Standard geometry

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	0.38	0/1578	0.63	0/2165
1	E	0.39	0/894	0.58	0/1218
2	B	0.43	1/1631 (0.1%)	0.61	0/2224
2	F	0.44	0/826	0.62	0/1122
3	C	0.45	0/1482	0.67	0/2026
3	G	0.44	0/1464	0.65	0/2004
4	D	0.42	0/1645	0.65	0/2242
4	H	0.44	0/1567	0.64	0/2139
All	All	0.42	1/11087 (0.0%)	0.63	0/15140

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	4	MET	SD-CE	-5.62	1.46	1.77

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	1538	0	1414	20	0
1	E	872	0	808	9	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	1595	0	1506	4	0
2	F	807	0	781	2	0
3	C	1440	0	1382	18	0
3	G	1422	0	1352	15	0
4	D	1605	0	1544	24	0
4	H	1529	0	1461	18	0
5	C	14	0	13	1	0
5	G	14	0	13	0	0
6	A	34	0	0	0	0
6	B	52	0	0	0	0
6	C	54	0	0	0	0
6	D	54	0	0	0	0
6	E	15	0	0	0	0
6	F	25	0	0	0	0
6	G	60	0	0	1	0
6	H	58	0	0	0	0
All	All	11188	0	10274	90	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 4.

The worst 5 of 90 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:A:200:ILE:HA	1:A:215:ARG:HA	1.62	0.79
3:C:75:LEU:HD23	4:D:1032:TYR:HB2	1.65	0.79
4:D:1085:LEU:O	4:D:1089:THR:HG22	1.85	0.76
4:D:1116:VAL:HG22	4:D:1160:MET:HG2	1.74	0.70
1:A:113:LEU:HD23	1:A:154:PRO:HD3	1.75	0.69

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	207/223 (93%)	196 (95%)	9 (4%)	2 (1%)	15 11
1	E	114/223 (51%)	108 (95%)	5 (4%)	1 (1%)	17 13
2	B	209/213 (98%)	198 (95%)	10 (5%)	1 (0%)	29 28
2	F	104/213 (49%)	101 (97%)	3 (3%)	0	100 100
3	C	179/189 (95%)	176 (98%)	3 (2%)	0	100 100
3	G	179/189 (95%)	170 (95%)	9 (5%)	0	100 100
4	D	191/230 (83%)	180 (94%)	10 (5%)	1 (0%)	29 28
4	H	184/230 (80%)	177 (96%)	7 (4%)	0	100 100
All	All	1367/1710 (80%)	1306 (96%)	56 (4%)	5 (0%)	34 35

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	149	ASP
1	A	194	LEU
4	D	1126	GLN
2	B	210	ARG
1	E	32	TRP

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	161/187 (86%)	157 (98%)	4 (2%)	47 54
1	E	86/187 (46%)	83 (96%)	3 (4%)	36 40
2	B	177/188 (94%)	172 (97%)	5 (3%)	43 49
2	F	90/188 (48%)	86 (96%)	4 (4%)	28 30
3	C	164/173 (95%)	157 (96%)	7 (4%)	29 31
3	G	159/173 (92%)	151 (95%)	8 (5%)	24 24

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
4	D	178/204 (87%)	172 (97%)	6 (3%)	37 42
4	H	166/204 (81%)	159 (96%)	7 (4%)	30 32
All	All	1181/1504 (78%)	1137 (96%)	44 (4%)	34 38

5 of 44 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	90	TYR
3	G	156	LEU
3	G	47	SER
3	G	90	GLU
4	H	4	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
3	C	120	ASN
3	C	169	HIS
4	D	1174	HIS
3	G	179	HIS
4	H	1174	HIS

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

2 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	$\#\|Z\| > 2$
5	NAG	C	1000	-	14,14,15	0.35	0	17,19,21	1.18	1 (5%)
5	NAG	G	1000	-	14,14,15	0.33	0	17,19,21	1.39	1 (5%)

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
5	NAG	C	1000	-	-	0/6/23/26	0/1/1/1
5	NAG	G	1000	-	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	G	1000	NAG	O5-C1-C2	-5.41	102.74	111.29
5	C	1000	NAG	O5-C1-C2	-4.59	104.04	111.29

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	C	1000	NAG	1	0

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

6.1 Protein, DNA and RNA chains

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, 95^{th} 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	211/223 (94%)	0.65	14 (6%)	18	18	33, 69, 103, 115	0
1	E	116/223 (52%)	0.41	4 (3%)	45	45	43, 63, 75, 84	0
2	B	211/213 (99%)	0.26	1 (0%)	91	91	30, 55, 80, 88	0
2	F	106/213 (49%)	0.22	1 (0%)	84	84	39, 54, 69, 82	0
3	C	181/189 (95%)	0.29	1 (0%)	89	89	30, 45, 69, 80	0
3	G	181/189 (95%)	0.23	1 (0%)	89	89	30, 44, 70, 82	0
4	D	197/230 (85%)	0.24	2 (1%)	82	83	34, 51, 70, 83	0
4	H	192/230 (83%)	0.41	5 (2%)	56	57	33, 51, 68, 84	0
All	All	1395/1710 (81%)	0.35	29 (2%)	63	65	30, 54, 82, 115	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	161	SER	3.6
1	A	2	VAL	3.4
1	A	212	VAL	3.4
1	E	2	VAL	3.3
1	A	219	LYS	3.2

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

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. The B-factors column lists the minimum, median, 95^{th} 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	B-factors(Å^2)	Q<0.9
5	NAG	C	1000	14/15	0.66	0.22	81,83,84,84	0
5	NAG	G	1000	14/15	0.76	0.22	87,89,90,90	0

6.5 Other polymers

There are no such residues in this entry.