wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 10:20 AM GMT

PDB ID	:	3LQA
Title	:	Crystal structure of clade C gp120 in complex with sCD4 and 21c Fab
Authors	:	Diskin, R.; Marcovecchio, P.M.; Bjorkman, P.J.
Deposited on	:	2010-02-08
Resolution	:	3.40 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
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:

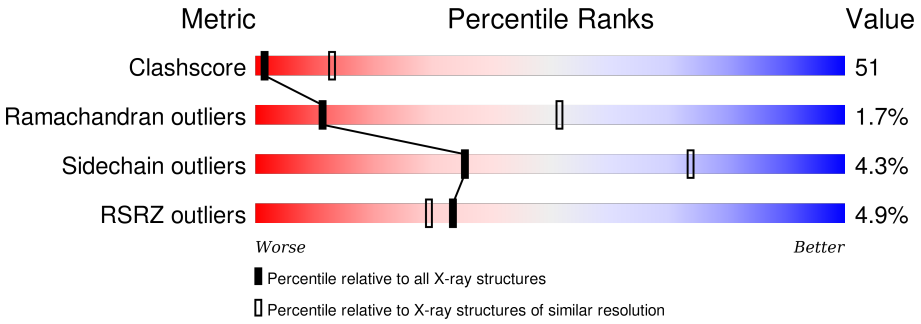
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 3.40 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))
Clashscore	102246	1611 (3.50-3.30)
Ramachandran outliers	100387	1571 (3.50-3.30)
Sidechain outliers	100360	1571 (3.50-3.30)
RSRZ outliers	91569	1485 (3.50-3.30)

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	C	192	3%, 30%, 55%, 6%, 9%
2	G	332	10%, 28%, 48%, 10%, 13%
3	H	231	2%, 35%, 51%, 10%
4	L	217	47%, 42%, 7%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	NAG	G	2500	X	-	-	-

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6945 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 T-cell surface glycoprotein CD4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	175	Total	C	N	O	S	0	0	0
			1361	850	238	269	4			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	183	ILE	-	EXPRESSION TAG	UNP P01730
C	184	ASP	-	EXPRESSION TAG	UNP P01730
C	185	GLY	-	EXPRESSION TAG	UNP P01730
C	186	ARG	-	EXPRESSION TAG	UNP P01730
C	187	HIS	-	EXPRESSION TAG	UNP P01730
C	188	HIS	-	EXPRESSION TAG	UNP P01730
C	189	HIS	-	EXPRESSION TAG	UNP P01730
C	190	HIS	-	EXPRESSION TAG	UNP P01730
C	191	HIS	-	EXPRESSION TAG	UNP P01730
C	192	HIS	-	EXPRESSION TAG	UNP P01730

- Molecule 2 is a protein called Envelope glycoprotein gp160.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	290	Total	C	N	O	S	0	0	0
			2271	1434	386	433	18			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	294	GLY	-	LINKER	UNP Q1PHM6
G	323	ALA	-	LINKER	UNP Q1PHM6
G	324	GLY	-	LINKER	UNP Q1PHM6
G	89	ILE	THR	ENGINEERED	UNP Q1PHM6
G	226	ASP	ASN	ENGINEERED	UNP Q1PHM6
G	232	ILE	THR	ENGINEERED	UNP Q1PHM6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	285	THR	ASN	ENGINEERED	UNP Q1PHM6
G	329	ASN	SER	ENGINEERED	UNP Q1PHM6
G	388	ILE	THR	ENGINEERED	UNP Q1PHM6
G	447	ASP	ASN	ENGINEERED	UNP Q1PHM6
G	495	SER	-	EXPRESSION TAG	UNP Q1PHM6
G	496	GLY	-	EXPRESSION TAG	UNP Q1PHM6
G	497	HIS	-	EXPRESSION TAG	UNP Q1PHM6
G	498	HIS	-	EXPRESSION TAG	UNP Q1PHM6
G	499	HIS	-	EXPRESSION TAG	UNP Q1PHM6
G	500	HIS	-	EXPRESSION TAG	UNP Q1PHM6
G	501	HIS	-	EXPRESSION TAG	UNP Q1PHM6
G	502	HIS	-	EXPRESSION TAG	UNP Q1PHM6

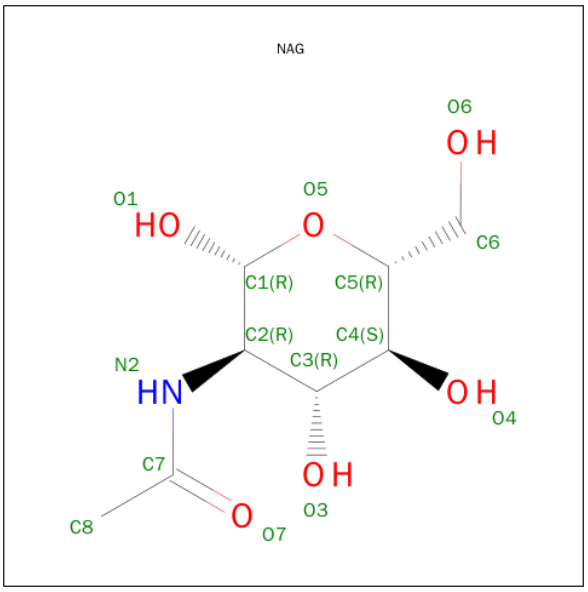
- Molecule 3 is a protein called Heavy chain of anti HIV Fab from human 21c antibody.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	H	225	Total	C	N	O	S	0	0	0
			1685	1068	277	333	7			

- Molecule 4 is a protein called Light chain of anti HIV Fab from human 21c antibody.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	L	213	Total	C	N	O	S	0	0	0
			1572	984	262	321	5			

- Molecule 5 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



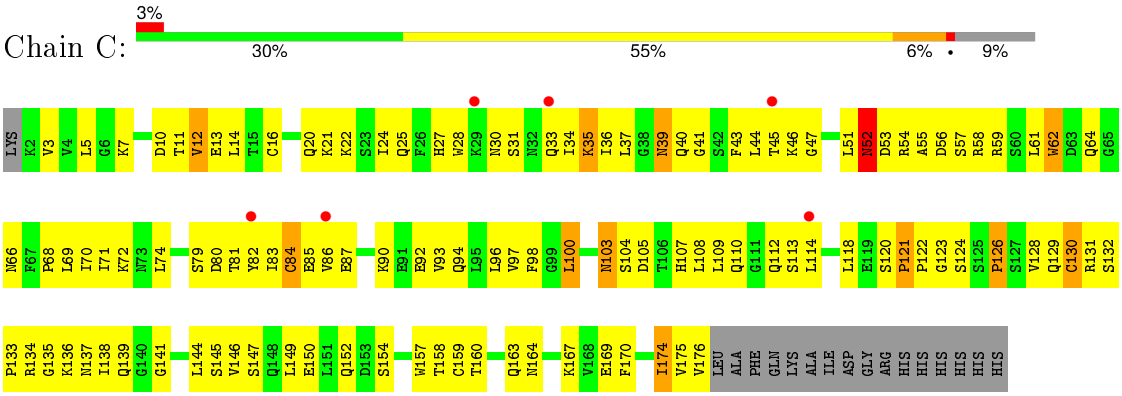
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	G	1	Total C N O 14 8 1 5	0	0
5	G	1	Total C N O 14 8 1 5	0	0
5	G	1	Total C N O 14 8 1 5	0	0
5	G	1	Total C N O 14 8 1 5	0	0

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.

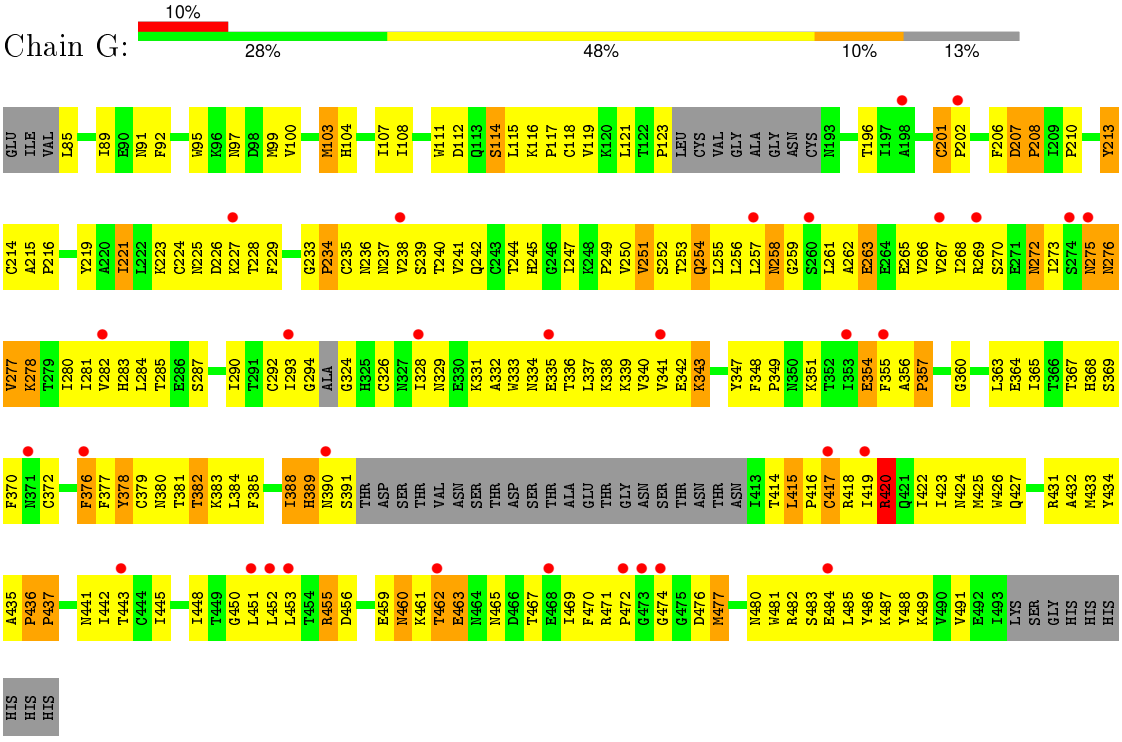
- Molecule 1: T-cell surface glycoprotein CD4

Chain C:



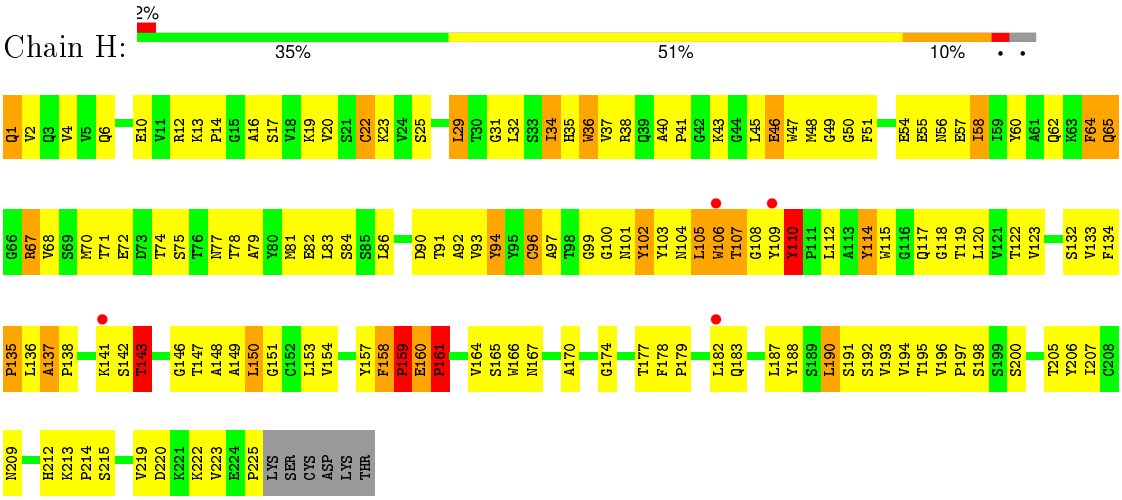
- Molecule 2: Envelope glycoprotein gp160

Chain G:



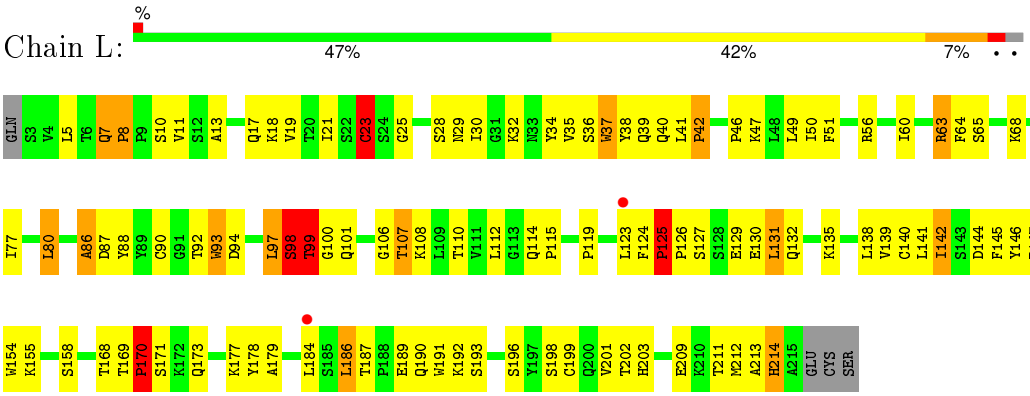
- Molecule 3: Heavy chain of anti HIV Fab from human 21c antibody

Chain H:



• Molecule 4: Light chain of anti HIV Fab from human 21c antibody

Chain L:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	93.73Å 187.93Å 151.67Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	84.00 – 3.40 83.88 – 3.39	Depositor EDS
% Data completeness (in resolution range)	79.6 (84.00-3.40) 78.9 (83.88-3.39)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	3.04 (at 3.41Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.234 , 0.322 0.228 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	90.5	Xtriage
Anisotropy	0.033	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 83.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.26$	Xtriage
Outliers	1 of 14987 reflections (0.007%)	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	6945	wwPDB-VP
Average B, all atoms (Å^2)	102.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.76% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	C	0.89	1/1380 (0.1%)	1.48	15/1862 (0.8%)
2	G	1.04	7/2316 (0.3%)	1.57	41/3142 (1.3%)
3	H	1.18	4/1728 (0.2%)	1.66	33/2358 (1.4%)
4	L	1.08	2/1610 (0.1%)	1.46	16/2197 (0.7%)
All	All	1.06	14/7034 (0.2%)	1.55	105/9559 (1.1%)

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
3	H	0	4
4	L	0	3
All	All	0	7

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	G	258	ASN	CB-CG	12.20	1.79	1.51
2	G	272	ASN	CB-CG	7.98	1.69	1.51
4	L	37	TRP	CB-CG	-7.20	1.37	1.50
2	G	441	ASN	CB-CG	6.95	1.67	1.51
2	G	258	ASN	CA-CB	6.68	1.70	1.53

The worst 5 of 105 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	H	160	GLU	C-N-CD	-19.55	77.58	120.60
3	H	58	ILE	N-CA-C	-14.99	70.53	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	H	158	PHE	C-N-CD	-14.53	88.63	120.60
2	G	276	ASN	N-CA-CB	-10.18	92.27	110.60
3	H	105	LEU	CA-CB-CG	9.28	136.64	115.30

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	H	102	TYR	Sidechain
3	H	110	TYR	Sidechain
3	H	114	TYR	Sidechain
3	H	94	TYR	Sidechain
4	L	34	TYR	Sidechain

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	C	1361	0	1382	159	0
2	G	2271	0	2232	278	0
3	H	1685	0	1645	202	0
4	L	1572	0	1530	126	0
5	G	56	0	51	11	0
All	All	6945	0	6840	700	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 51.

The worst 5 of 700 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)
5:G:2500:NAG:C1	5:G:2500:NAG:C2	1.75	1.56
2:G:258:ASN:CB	2:G:258:ASN:CG	1.79	1.51
3:H:158:PHE:HB3	3:H:159:PRO:HD2	1.22	1.20
2:G:420:ARG:HH12	3:H:55:GLU:HB2	1.06	1.13
2:G:355:PHE:HB3	2:G:385:PHE:HB3	1.23	1.12

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	C	173/192 (90%)	168 (97%)	3 (2%)	2 (1%)	16 59
2	G	282/332 (85%)	273 (97%)	3 (1%)	6 (2%)	9 47
3	H	223/231 (96%)	216 (97%)	4 (2%)	3 (1%)	15 57
4	L	211/217 (97%)	204 (97%)	3 (1%)	4 (2%)	10 49
All	All	889/972 (92%)	861 (97%)	13 (2%)	15 (2%)	11 51

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	52	ASN
1	C	126	PRO
2	G	208	PRO
2	G	216	PRO
2	G	254	GLN

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	C	159/173 (92%)	152 (96%)	7 (4%)	35 73
2	G	261/296 (88%)	250 (96%)	11 (4%)	36 74
3	H	188/194 (97%)	181 (96%)	7 (4%)	41 77

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
4	L	176/180 (98%)	167 (95%)	9 (5%)	29 69
All	All	784/843 (93%)	750 (96%)	34 (4%)	35 74

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	G	436	PRO
3	H	74	THR
4	L	170	PRO
2	G	445	ILE
1	C	130	CYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
2	G	102	GLN
2	G	275	ASN
4	L	7	GLN
2	G	105	GLN
2	G	236	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)

4 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	G	1000	2	14,14,15	2.36	3 (21%)	15,19,21	1.81	2 (13%)
5	NAG	G	1500	2	14,14,15	2.89	3 (21%)	15,19,21	1.41	3 (20%)
5	NAG	G	2000	2	14,14,15	2.46	4 (28%)	15,19,21	1.43	2 (13%)
5	NAG	G	2500	2	14,14,15	4.61	3 (21%)	15,19,21	1.07	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
5	NAG	G	1000	2	-	0/6/23/26	0/1/1/1
5	NAG	G	1500	2	-	2/6/23/26	0/1/1/1
5	NAG	G	2000	2	-	0/6/23/26	0/1/1/1
5	NAG	G	2500	2	1/1/5/7	0/6/23/26	0/1/1/1

The worst 5 of 13 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	G	2500	NAG	C8-C7	2.08	1.54	1.50
5	G	2000	NAG	C3-C2	2.08	1.57	1.52
5	G	1000	NAG	C3-C2	2.13	1.57	1.52
5	G	1500	NAG	C2-N2	2.13	1.50	1.46
5	G	2000	NAG	O5-C5	2.22	1.48	1.43

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	G	1000	NAG	C2-N2-C7	-4.01	117.89	123.04
5	G	1500	NAG	O7-C7-C8	-2.90	116.74	122.06
5	G	2000	NAG	C3-C2-N2	-2.45	104.69	110.56
5	G	2500	NAG	O7-C7-C8	-2.44	117.59	122.06
5	G	1500	NAG	C3-C2-N2	-2.39	104.84	110.56

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
5	G	2500	NAG	C1

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	G	1500	NAG	C8-C7-N2-C2
5	G	1500	NAG	O7-C7-N2-C2

There are no ring outliers.

4 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	G	1000	NAG	2	0
5	G	1500	NAG	4	0
5	G	2000	NAG	3	0
5	G	2500	NAG	2	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	C	175/192 (91%)	0.35	6 (3%) 49 44	51, 103, 139, 185	0
2	G	290/332 (87%)	0.69	32 (11%) 7 7	52, 118, 174, 225	0
3	H	225/231 (97%)	0.36	4 (1%) 71 65	36, 84, 136, 185	0
4	L	213/217 (98%)	0.26	2 (0%) 85 81	39, 83, 133, 162	0
All	All	903/972 (92%)	0.44	44 (4%) 33 29	36, 97, 162, 225	0

The worst 5 of 44 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	G	390	ASN	5.5
3	H	109	TYR	4.8
2	G	443	THR	4.3
2	G	257	LEU	4.2
2	G	353	ILE	3.8

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 carbohydrates 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. 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, 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	LLDF	B-factors(Å^2)	Q<0.9
5	NAG	G	2000	14/15	0.74	0.31	-0.28	125,125,125,125	0
5	NAG	G	1000	14/15	0.87	0.17	-1.21	105,105,105,105	0
5	NAG	G	2500	14/15	0.82	0.17	-	136,136,136,136	0
5	NAG	G	1500	14/15	0.90	0.29	-	122,122,122,122	0

6.5 Other polymers (i)

There are no such residues in this entry.