wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 03:10 PM GMT

PDB ID	:	4BQ6
Title	:	Crystal structure of the RGMB-NEO1 complex form 1
Authors	:	Bell, C.H.; Healey, E.; van Erp, S.; Bishop, B.; Tang, C.; Gilbert, R.J.C.; Aricescu, A.R.; Pasterkamp, R.J.; Siebold, C.
Deposited on	:	2013-05-30
Resolution	:	2.30 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:

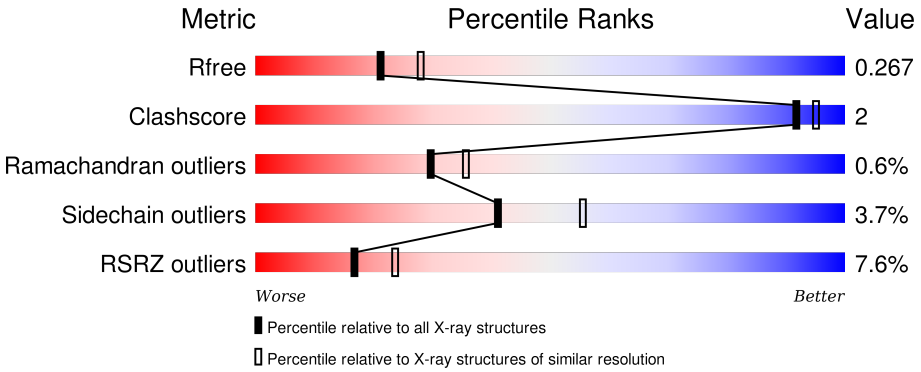
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.30 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	3852 (2.30-2.30)
Clashscore	102246	4452 (2.30-2.30)
Ramachandran outliers	100387	4410 (2.30-2.30)
Sidechain outliers	100360	4409 (2.30-2.30)
RSRZ outliers	91569	3857 (2.30-2.30)

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	A	264	7% 71% 6% 22%
1	B	264	4% 70% 6% 22%
2	C	122	6% 11% 88%
2	E	122	2% 12% 87%
3	D	251	6% 54% 5% 41%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	251	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6132 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 NEOGENIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	205	Total	C	N	O	S	0	0	0
			1611	1027	273	305	6			
1	B	205	Total	C	N	O	S	0	0	0
			1611	1027	273	305	6			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	880	GLU	-	EXPRESSION TAG	UNP P97798
A	881	THR	-	EXPRESSION TAG	UNP P97798
A	882	GLY	-	EXPRESSION TAG	UNP P97798
A	1134	ASN	-	EXPRESSION TAG	UNP P97798
A	1135	GLY	-	EXPRESSION TAG	UNP P97798
A	1136	THR	-	EXPRESSION TAG	UNP P97798
A	1137	LYS	-	EXPRESSION TAG	UNP P97798
A	1138	HIS	-	EXPRESSION TAG	UNP P97798
A	1139	HIS	-	EXPRESSION TAG	UNP P97798
A	1140	HIS	-	EXPRESSION TAG	UNP P97798
A	1141	HIS	-	EXPRESSION TAG	UNP P97798
A	1142	HIS	-	EXPRESSION TAG	UNP P97798
A	1143	HIS	-	EXPRESSION TAG	UNP P97798
B	880	GLU	-	EXPRESSION TAG	UNP P97798
B	881	THR	-	EXPRESSION TAG	UNP P97798
B	882	GLY	-	EXPRESSION TAG	UNP P97798
B	1134	ASN	-	EXPRESSION TAG	UNP P97798
B	1135	GLY	-	EXPRESSION TAG	UNP P97798
B	1136	THR	-	EXPRESSION TAG	UNP P97798
B	1137	LYS	-	EXPRESSION TAG	UNP P97798
B	1138	HIS	-	EXPRESSION TAG	UNP P97798
B	1139	HIS	-	EXPRESSION TAG	UNP P97798
B	1140	HIS	-	EXPRESSION TAG	UNP P97798
B	1141	HIS	-	EXPRESSION TAG	UNP P97798
B	1142	HIS	-	EXPRESSION TAG	UNP P97798

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1143	HIS	-	EXPRESSION TAG	UNP P97798

- Molecule 2 is a protein called RGM DOMAIN FAMILY MEMBER B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	15	Total	C	N	O	S	0	0	0
			118	79	16	21	2			
2	E	16	Total	C	N	O	S	0	0	0
			128	85	19	22	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	47	GLU	-	EXPRESSION TAG	UNP Q6NW40
C	48	THR	-	EXPRESSION TAG	UNP Q6NW40
C	49	GLY	-	EXPRESSION TAG	UNP Q6NW40
E	47	GLU	-	EXPRESSION TAG	UNP Q6NW40
E	48	THR	-	EXPRESSION TAG	UNP Q6NW40
E	49	GLY	-	EXPRESSION TAG	UNP Q6NW40

- Molecule 3 is a protein called RGM DOMAIN FAMILY MEMBER B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	148	Total	C	N	O	S	0	0	0
			1148	721	197	223	7			
3	F	149	Total	C	N	O	S	0	0	0
			1159	727	201	224	7			

There are 20 discrepancies between the modelled and reference sequences:

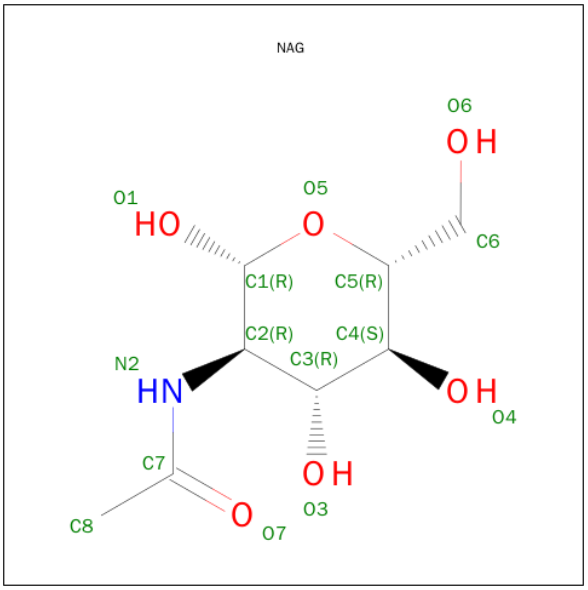
Chain	Residue	Modelled	Actual	Comment	Reference
D	411	GLY	-	EXPRESSION TAG	UNP Q6NW40
D	412	THR	-	EXPRESSION TAG	UNP Q6NW40
D	413	LYS	-	EXPRESSION TAG	UNP Q6NW40
D	414	HIS	-	EXPRESSION TAG	UNP Q6NW40
D	415	HIS	-	EXPRESSION TAG	UNP Q6NW40
D	416	HIS	-	EXPRESSION TAG	UNP Q6NW40
D	417	HIS	-	EXPRESSION TAG	UNP Q6NW40
D	418	HIS	-	EXPRESSION TAG	UNP Q6NW40
D	419	HIS	-	EXPRESSION TAG	UNP Q6NW40
D	225	GLY	GLU	CONFLICT	UNP Q6NW40
F	411	GLY	-	EXPRESSION TAG	UNP Q6NW40

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	412	THR	-	EXPRESSION TAG	UNP Q6NW40
F	413	LYS	-	EXPRESSION TAG	UNP Q6NW40
F	414	HIS	-	EXPRESSION TAG	UNP Q6NW40
F	415	HIS	-	EXPRESSION TAG	UNP Q6NW40
F	416	HIS	-	EXPRESSION TAG	UNP Q6NW40
F	417	HIS	-	EXPRESSION TAG	UNP Q6NW40
F	418	HIS	-	EXPRESSION TAG	UNP Q6NW40
F	419	HIS	-	EXPRESSION TAG	UNP Q6NW40
F	225	GLY	GLU	CONFLICT	UNP Q6NW40

- Molecule 4 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	119	Total	O	0	0
			119	119		
5	B	98	Total	O	0	0
			98	98		

Continued on next page...

Continued from previous page...

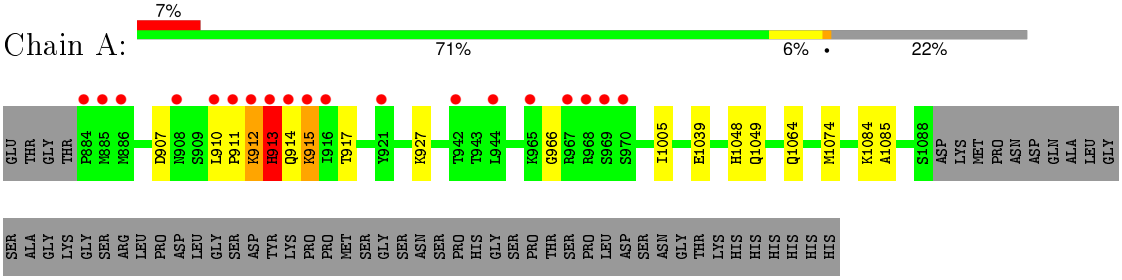
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	C	45	Total O 45 45	0	0
5	D	67	Total O 67 67	0	0

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.

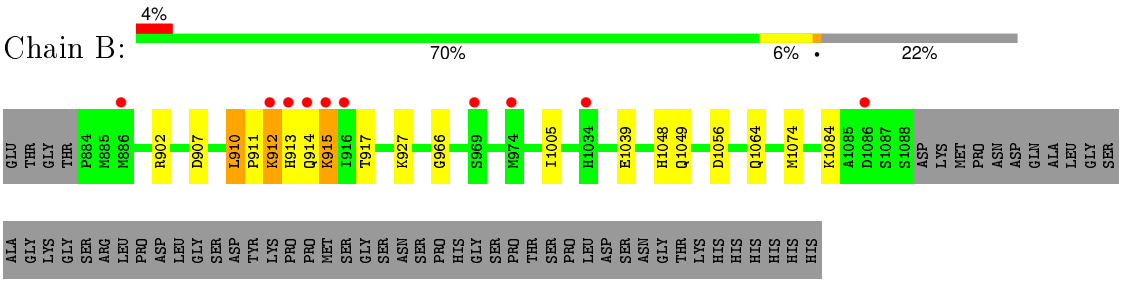
- Molecule 1: NEOGENIN

Chain A:



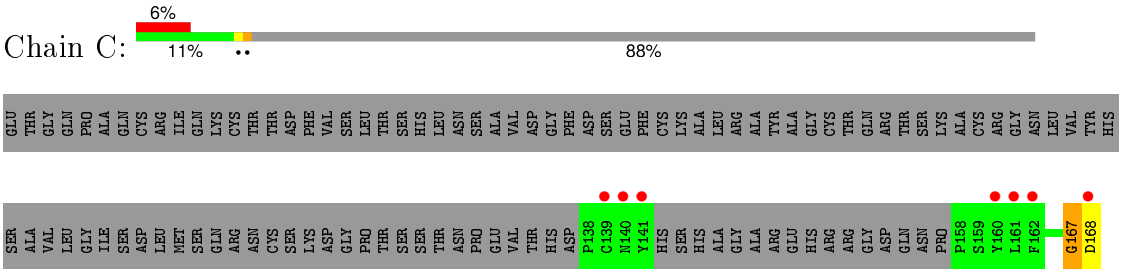
- Molecule 1: NEOGENIN

Chain B:



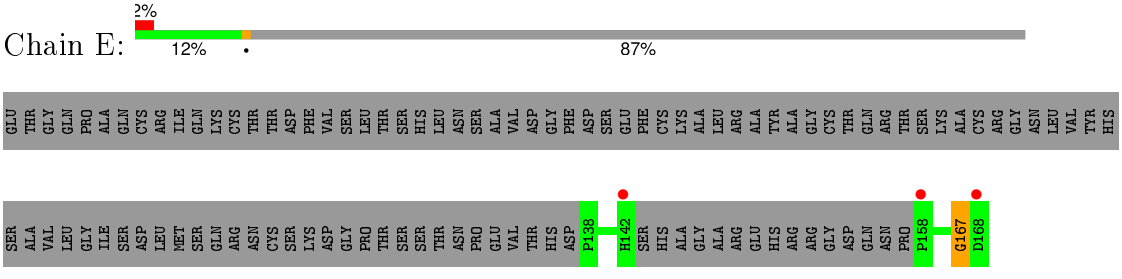
- Molecule 2: RGM DOMAIN FAMILY MEMBER B

Chain C:



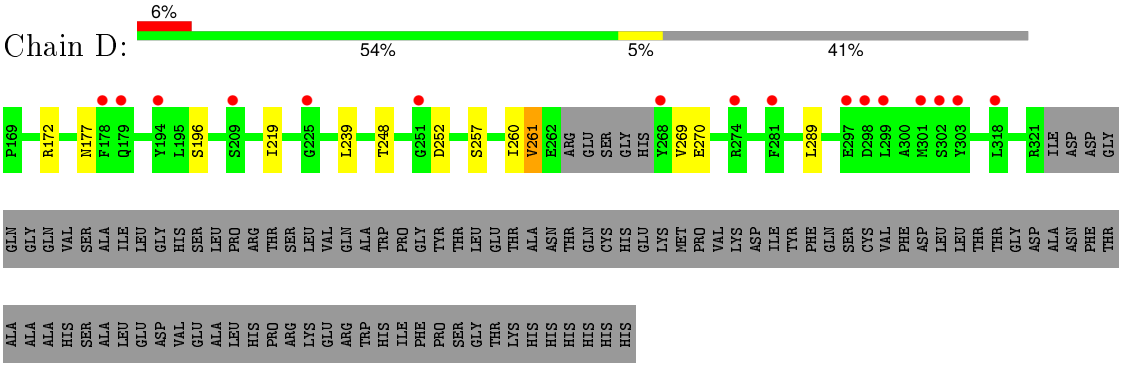
- Molecule 2: RGM DOMAIN FAMILY MEMBER B

Chain E:



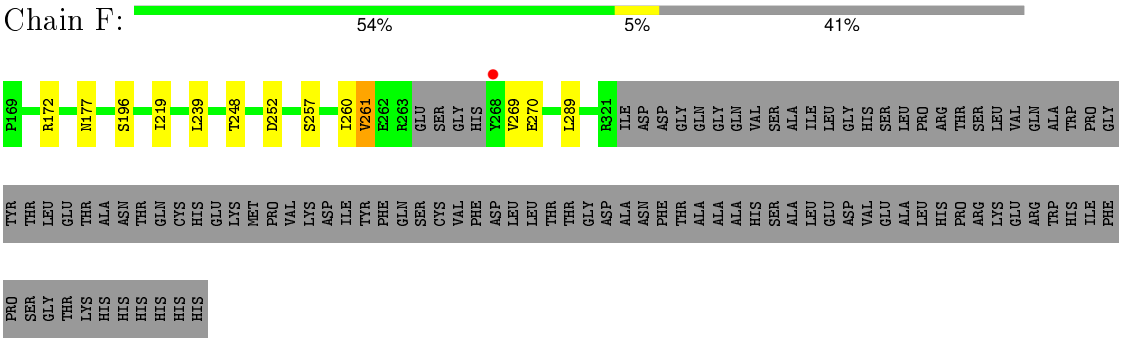
• Molecule 3: RGM DOMAIN FAMILY MEMBER B

Chain D:



• Molecule 3: RGM DOMAIN FAMILY MEMBER B

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	91.32Å 100.39Å 103.69Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.30 56.60 – 2.25	Depositor EDS
% Data completeness (in resolution range)	99.4 (30.00-2.30) 98.9 (56.60-2.25)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.50 (at 2.25Å)	Xtriage
Refinement program	BUSTER 2.11.2	Depositor
R, R_{free}	0.218 , 0.267 0.220 , 0.267	Depositor DCC
R_{free} test set	2169 reflections (5.34%)	DCC
Wilson B-factor (Å^2)	43.6	Xtriage
Anisotropy	0.030	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 51.1	EDS
Estimated twinning fraction	0.024 for -h,l,k	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Outliers	6 of 45400 reflections (0.013%)	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6132	wwPDB-VP
Average B, all atoms (Å^2)	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 54.22 % of the origin peak, indicating pseudo translational symmetry.The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 3.7690e-05.The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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	A	0.51	0/1655	0.72	1/2261 (0.0%)
1	B	0.49	0/1655	0.71	1/2261 (0.0%)
2	C	0.46	0/122	1.01	1/162 (0.6%)
2	E	0.52	0/133	1.06	1/177 (0.6%)
3	D	0.52	0/1170	0.74	0/1590
3	F	0.53	0/1181	0.75	0/1604
All	All	0.51	0/5916	0.74	4/8055 (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	C	0	1
2	E	0	2
All	All	0	3

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	167	GLY	C-N-CA	9.17	144.63	121.70
2	C	167	GLY	C-N-CA	8.67	143.37	121.70
1	A	912	LYS	C-N-CA	6.33	137.53	121.70
1	B	912	LYS	C-N-CA	5.63	135.77	121.70

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	167	GLY	Peptide
2	E	167	GLY	Mainchain,Peptide

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	1611	0	1599	7	0
1	B	1611	0	1599	9	0
2	C	118	0	101	0	0
2	E	128	0	108	0	0
3	D	1148	0	1122	4	0
3	F	1159	0	1135	4	0
4	A	14	0	13	0	0
4	B	14	0	13	0	0
5	A	119	0	0	0	0
5	B	98	0	0	0	0
5	C	45	0	0	0	0
5	D	67	0	0	0	0
All	All	6132	0	5690	24	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.

The worst 5 of 24 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:913:HIS:HA	1:A:914:GLN:HG2	1.72	0.71
3:F:261:VAL:HG23	3:F:270:GLU:HG3	1.73	0.71
3:D:261:VAL:HG23	3:D:270:GLU:HG3	1.73	0.68
1:B:1039:GLU:HG2	1:B:1048:HIS:NE2	2.15	0.62
1:A:1039:GLU:HG2	1:A:1048:HIS:NE2	2.18	0.58

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	203/264 (77%)	190 (94%)	10 (5%)	3 (2%)	13 12
1	B	203/264 (77%)	192 (95%)	10 (5%)	1 (0%)	34 41
2	C	11/122 (9%)	11 (100%)	0	0	100 100
2	E	12/122 (10%)	12 (100%)	0	0	100 100
3	D	144/251 (57%)	133 (92%)	11 (8%)	0	100 100
3	F	145/251 (58%)	134 (92%)	11 (8%)	0	100 100
All	All	718/1274 (56%)	672 (94%)	42 (6%)	4 (1%)	30 36

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	912	LYS
1	A	913	HIS
1	A	966	GLY
1	B	966	GLY

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	181/232 (78%)	176 (97%)	5 (3%)	51 68
1	B	181/232 (78%)	175 (97%)	6 (3%)	45 61
2	C	13/103 (13%)	12 (92%)	1 (8%)	16 20

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	E	14/103 (14%)	14 (100%)	0	100 100
3	D	128/215 (60%)	122 (95%)	6 (5%)	32 43
3	F	129/215 (60%)	123 (95%)	6 (5%)	32 43
All	All	646/1100 (59%)	622 (96%)	24 (4%)	41 55

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	927	LYS
3	D	177	ASN
3	F	261	VAL
2	C	168	ASP
3	D	172	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	913	HIS
1	B	1079	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)

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	$\#\|Z\| > 2$
4	NAG	A	2089	1	14,14,15	0.27	0	15,19,21	0.80	1 (6%)
4	NAG	B	2089	1	14,14,15	0.26	0	15,19,21	0.68	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
4	NAG	A	2089	1	-	0/6/23/26	0/1/1/1
4	NAG	B	2089	1	-	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})
4	B	2089	NAG	C1-O5-C5	2.33	115.21	112.25
4	A	2089	NAG	C1-O5-C5	2.86	115.87	112.25

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

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	205/264 (77%)	0.38	19 (9%)	11	16	28, 45, 92, 136	0
1	B	205/264 (77%)	0.23	10 (4%)	33	42	28, 46, 94, 130	0
2	C	15/122 (12%)	1.72	7 (46%)	0	0	39, 66, 104, 111	0
2	E	16/122 (13%)	0.94	3 (18%)	2	2	35, 55, 92, 108	0
3	D	148/251 (58%)	0.52	16 (10%)	8	11	28, 48, 90, 110	0
3	F	149/251 (59%)	0.32	1 (0%)	89	92	28, 42, 75, 105	0
All	All	738/1274 (57%)	0.40	56 (7%)	17	24	28, 46, 93, 136	0

The worst 5 of 56 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	913	HIS	8.9
1	B	913	HIS	7.0
1	A	912	LYS	6.6
2	E	168	ASP	5.3
3	D	268	TYR	4.4

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
4	NAG	B	2089	14/15	0.89	0.14	-	76,79,82,83	0
4	NAG	A	2089	14/15	0.86	0.22	-	77,81,84,85	0

6.5 Other polymers

There are no such residues in this entry.