



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 25, 2017 – 12:22 AM EDT

PDB ID : 5WB8
Title : Crystal structure of the epidermal growth factor receptor extracellular region in complex with epigen
Authors : Bessman, N.J.; Freed, D.M.; Moore, J.O.; Ferguson, K.M.; Lemmon, M.A.
Deposited on : unknown
Resolution : 3.00 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 ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030345
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
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-20030345

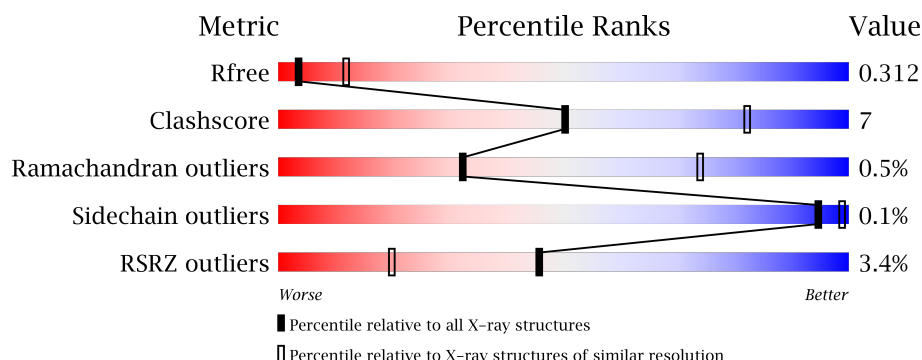
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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}	100719	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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 $\leq 5\%$. 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	507	
1	D	507	
2	B	61	
2	C	61	

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 8300 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 Epidermal growth factor receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	501	Total	C	N	O	S	0	0	0
			3740	2319	653	726	42			
1	D	504	Total	C	N	O	S	0	0	0
			3774	2343	663	726	42			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	502	HIS	-	expression tag	UNP P00533
A	503	HIS	-	expression tag	UNP P00533
A	504	HIS	-	expression tag	UNP P00533
A	505	HIS	-	expression tag	UNP P00533
A	506	HIS	-	expression tag	UNP P00533
A	507	HIS	-	expression tag	UNP P00533
D	502	HIS	-	expression tag	UNP P00533
D	503	HIS	-	expression tag	UNP P00533
D	504	HIS	-	expression tag	UNP P00533
D	505	HIS	-	expression tag	UNP P00533
D	506	HIS	-	expression tag	UNP P00533
D	507	HIS	-	expression tag	UNP P00533

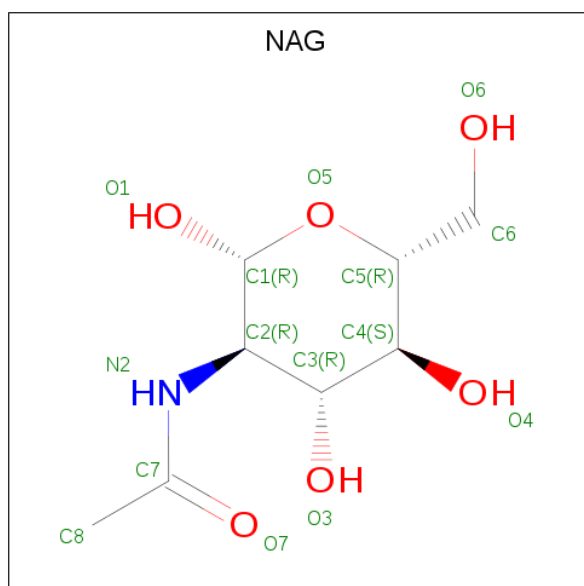
- Molecule 2 is a protein called Epigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	32	Total	C	N	O	S	0	0	0
			228	139	40	43	6			
2	C	36	Total	C	N	O	S	0	0	0
			278	173	48	52	5			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	26	SER	-	expression tag	UNP Q6UW88
C	26	SER	-	expression tag	UNP Q6UW88

- Molecule 3 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



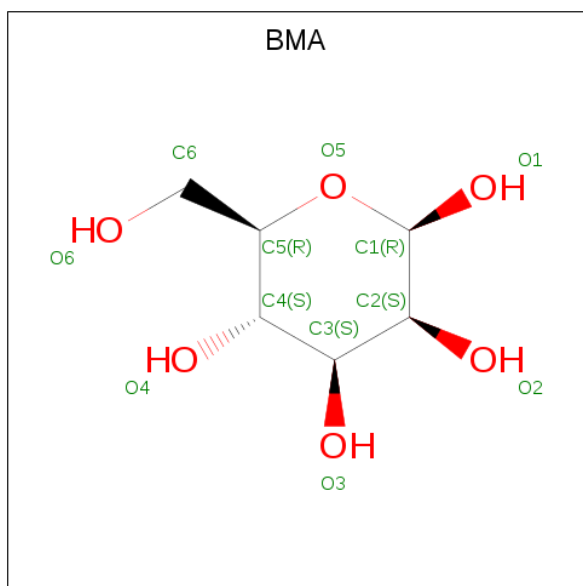
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	D	1	Total	C	N	O	0	0
			14	8	1	5		
3	D	1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

Continued on next page...

Continued from previous page...

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

- Molecule 4 is BETA-D-MANNOSE (three-letter code: BMA) (formula: $C_6H_{12}O_6$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			11	6	5		
4	A	1	Total	C	O	0	0
			11	6	5		
4	D	1	Total	C	O	0	0
			11	6	5		

- Molecule 5 is ALPHA-D-MANNOSE (three-letter code: MAN) (formula: $C_6H_{12}O_6$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	D	1	Total	C	O	0	0
			11	6	5		

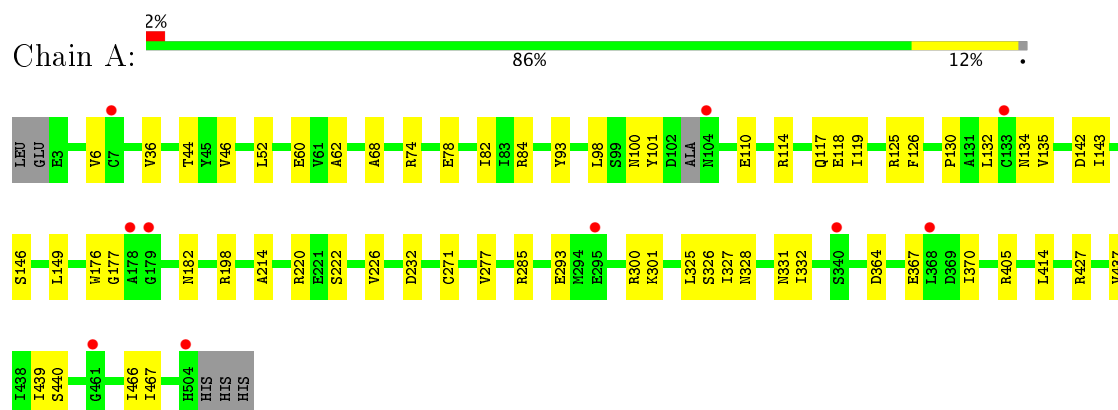
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	33	Total	O	0	0
			33	33		
6	C	4	Total	O	0	0
			4	4		
6	D	31	Total	O	0	0
			31	31		

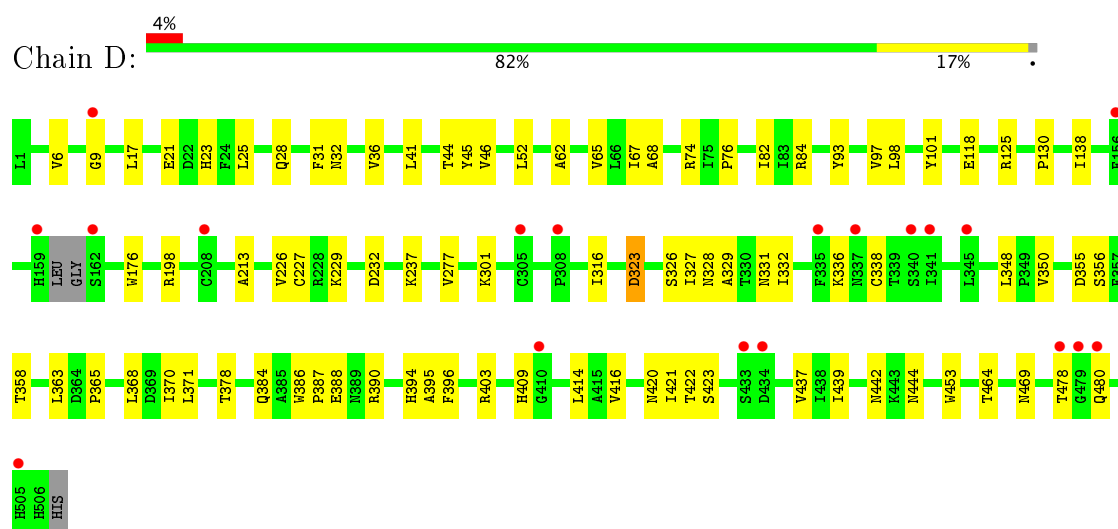
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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 ($\text{RSRZ} > 2$). 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.

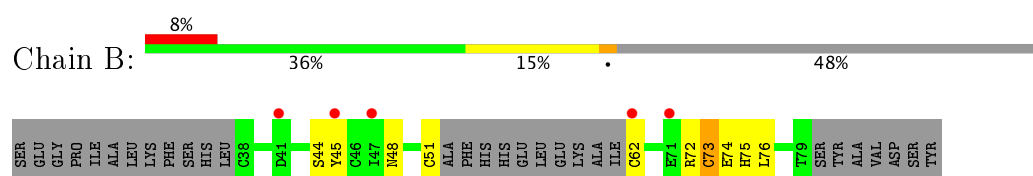
• Molecule 1: Epidermal growth factor receptor



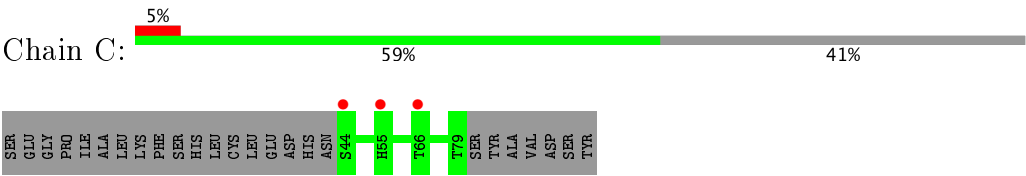
• Molecule 1: Epidermal growth factor receptor



• Molecule 2: Epigen



● Molecule 2: Epigen



4 Data and refinement statistics

Property	Value	Source
Space group	P 43 2 2	Depositor
Cell constants a, b, c, α , β , γ	104.05Å 104.05Å 285.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.89 – 3.00 48.89 – 3.00	Depositor EDS
% Data completeness (in resolution range)	96.9 (48.89-3.00) 96.9 (48.89-3.00)	Depositor EDS
R_{merge}	0.28	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.18 (at 3.01Å)	Xtriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, R_{free}	0.265 , 0.312 0.265 , 0.312	Depositor DCC
R_{free} test set	1992 reflections (6.34%)	DCC
Wilson B-factor (Å ²)	25.3	Xtriage
Anisotropy	0.006	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 42.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.74	EDS
Total number of atoms	8300	wwPDB-VP
Average B, all atoms (Å ²)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.81% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 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: BMA, NAG, MAN

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	0.25	0/3809	0.44	0/5171
1	D	0.25	0/3848	0.46	0/5225
2	B	0.25	0/231	0.48	0/311
2	C	0.26	0/283	0.44	0/380
All	All	0.25	0/8171	0.45	0/11087

There are no bond length outliers.

There are no bond angle outliers.

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	3740	0	3491	42	0
1	D	3774	0	3522	55	0
2	B	228	0	186	7	0
2	C	278	0	255	0	0
3	A	84	0	73	2	0
3	D	84	0	74	6	0
4	A	22	0	20	1	0
4	D	11	0	9	0	0
5	D	11	0	10	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	A	33	0	0	2	0
6	C	4	0	0	0	0
6	D	31	0	0	0	0
All	All	8300	0	7640	105	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 7.

The worst 5 of 105 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:328:ASN:H	1:A:331:ASN:HB2	1.48	0.77
2:B:51:CYS:HA	2:B:62:CYS:HB3	1.69	0.75
1:A:332:ILE:HG13	1:A:370:ILE:HD12	1.77	0.66
1:D:323:ASP:OD2	5:D:3206:MAN:O6	2.13	0.66
1:D:388:GLU:HB3	1:D:420:ASN:HD22	1.60	0.66

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	497/507 (98%)	464 (93%)	32 (6%)	1 (0%)	51	86
1	D	500/507 (99%)	465 (93%)	32 (6%)	3 (1%)	28	70
2	B	28/61 (46%)	23 (82%)	4 (14%)	1 (4%)	4	22
2	C	34/61 (56%)	31 (91%)	3 (9%)	0	100	100
All	All	1059/1136 (93%)	983 (93%)	71 (7%)	5 (0%)	32	74

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	323	ASP
1	D	478	THR
1	D	480	GLN
1	A	232	ASP
2	B	73	CYS

5.3.2 Protein sidechains ⓘ

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	405/446 (91%)	405 (100%)	0	100	100
1	D	408/446 (92%)	407 (100%)	1 (0%)	94	98
2	B	23/52 (44%)	23 (100%)	0	100	100
2	C	29/52 (56%)	29 (100%)	0	100	100
All	All	865/996 (87%)	864 (100%)	1 (0%)	94	99

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	338	CYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	117	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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](#)

16 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
3	NAG	A	3201	1,3	14,14,15	0.26	0	15,19,21	0.64	0
3	NAG	A	3202	3	14,14,15	0.43	0	15,19,21	1.23	2 (13%)
3	NAG	A	3203	1,3	14,14,15	0.20	0	15,19,21	0.44	0
3	NAG	A	3204	3,4	14,14,15	0.23	0	15,19,21	0.60	0
4	BMA	A	3205	3	11,11,12	0.54	0	13,15,17	1.19	2 (15%)
3	NAG	A	3206	1,3	14,14,15	0.32	0	15,19,21	0.52	0
3	NAG	A	3207	3,4	14,14,15	0.14	0	15,19,21	0.50	0
4	BMA	A	3208	3	11,11,12	0.44	0	13,15,17	0.90	0
3	NAG	D	3201	1,3	14,14,15	0.44	0	15,19,21	0.57	0
3	NAG	D	3202	3	14,14,15	1.17	1 (7%)	15,19,21	0.71	0
3	NAG	D	3203	1,3	14,14,15	0.46	0	15,19,21	0.51	0
3	NAG	D	3204	3,4	14,14,15	0.87	1 (7%)	15,19,21	0.66	0
4	BMA	D	3205	3,5	11,11,12	0.55	0	13,15,17	0.93	0
5	MAN	D	3206	4	11,11,12	0.87	1 (9%)	13,15,17	0.91	1 (7%)
3	NAG	D	3207	1,3	14,14,15	0.80	1 (7%)	15,19,21	0.95	0
3	NAG	D	3208	3	14,14,15	0.28	0	15,19,21	0.62	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
3	NAG	A	3201	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	3202	3	-	0/6/23/26	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	A	3203	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	3204	3,4	-	0/6/23/26	0/1/1/1
4	BMA	A	3205	3	-	0/2/19/22	0/1/1/1
3	NAG	A	3206	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	3207	3,4	-	0/6/23/26	0/1/1/1
4	BMA	A	3208	3	-	0/2/19/22	0/1/1/1
3	NAG	D	3201	1,3	-	0/6/23/26	0/1/1/1
3	NAG	D	3202	3	-	0/6/23/26	0/1/1/1
3	NAG	D	3203	1,3	-	0/6/23/26	0/1/1/1
3	NAG	D	3204	3,4	-	0/6/23/26	0/1/1/1
4	BMA	D	3205	3,5	-	0/2/19/22	0/1/1/1
5	MAN	D	3206	4	-	0/2/19/22	0/1/1/1
3	NAG	D	3207	1,3	-	0/6/23/26	0/1/1/1
3	NAG	D	3208	3	-	0/6/23/26	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	3207	NAG	O5-C1	-2.24	1.40	1.43
5	D	3206	MAN	O5-C1	-2.08	1.40	1.43
3	D	3204	NAG	C1-C2	2.82	1.56	1.52
3	D	3202	NAG	C1-C2	4.19	1.58	1.52

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	3205	BMA	O2-C2-C3	-2.25	105.76	110.17
5	D	3206	MAN	O2-C2-C3	-2.11	106.04	110.17
3	A	3202	NAG	C1-C2-N2	2.35	114.51	110.49
4	A	3205	BMA	C1-O5-C5	2.87	116.12	112.17
3	A	3202	NAG	C1-O5-C5	2.97	116.26	112.17

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

9 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	3203	NAG	1	0
3	A	3204	NAG	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	3205	BMA	1	0
3	D	3201	NAG	1	0
3	D	3202	NAG	1	0
3	D	3203	NAG	1	0
3	D	3204	NAG	1	0
5	D	3206	MAN	1	0
3	D	3207	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 [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	501/507 (98%)	0.52	10 (1%) 65 36	20, 49, 75, 102	0
1	D	504/507 (99%)	0.55	19 (3%) 41 17	25, 53, 84, 120	0
2	B	32/61 (52%)	1.31	5 (15%) 2 1	44, 80, 110, 130	0
2	C	36/61 (59%)	0.67	3 (8%) 12 4	30, 54, 78, 96	0
All	All	1073/1136 (94%)	0.56	37 (3%) 46 20	20, 51, 84, 130	0

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	504	HIS	4.6
1	D	479	GLY	4.3
1	A	104	ASN	3.1
2	B	45	TYR	3.0
1	D	305	CYS	2.9

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(\AA^2)	Q<0.9
3	NAG	D	3203	14/15	0.71	0.30	0.12	66,74,103,111	0
3	NAG	A	3206	14/15	0.85	0.26	0.08	17,39,65,76	0
3	NAG	A	3203	14/15	0.68	0.28	-0.11	34,50,55,67	0
3	NAG	D	3207	14/15	0.74	0.26	-0.79	71,84,99,101	0
4	BMA	A	3205	11/12	0.77	0.22	-	57,79,102,117	0
4	BMA	D	3205	11/12	0.57	0.40	-	89,108,121,129	0
3	NAG	D	3208	14/15	0.65	0.28	-	81,97,105,105	0
3	NAG	A	3202	14/15	0.53	0.48	-	73,100,107,113	0
3	NAG	D	3204	14/15	0.64	0.32	-	84,90,95,102	0
3	NAG	A	3204	14/15	0.73	0.24	-	45,76,90,90	0
5	MAN	D	3206	11/12	0.30	0.24	-	53,83,104,107	0
3	NAG	A	3207	14/15	0.75	0.28	-	65,78,90,103	0
3	NAG	A	3201	14/15	0.66	0.43	-	48,81,94,95	0
3	NAG	D	3201	14/15	0.76	0.27	-	47,70,77,79	0
4	BMA	A	3208	11/12	0.55	0.41	-	77,89,103,107	0
3	NAG	D	3202	14/15	0.62	0.47	-	81,94,104,109	0

6.5 Other polymers [i](#)

There are no such residues in this entry.