Full wwPDB X-ray Structure Validation Report (i)

Oct 23, 2023 – 03:49 PM EDT

PDB ID	:	3AO4
Title	:	Fragment-based approach to the design of ligands targeting a novel site on HIV-1 integrase
Authors	:	Wielens, J.; Headey, S.J.; Parker, M.W.; Chalmers, D.K.; Scanlon, M.J.
Deposited on	:	2010-09-20
Resolution	:	1.95 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 (i)) were used in the production of this report:

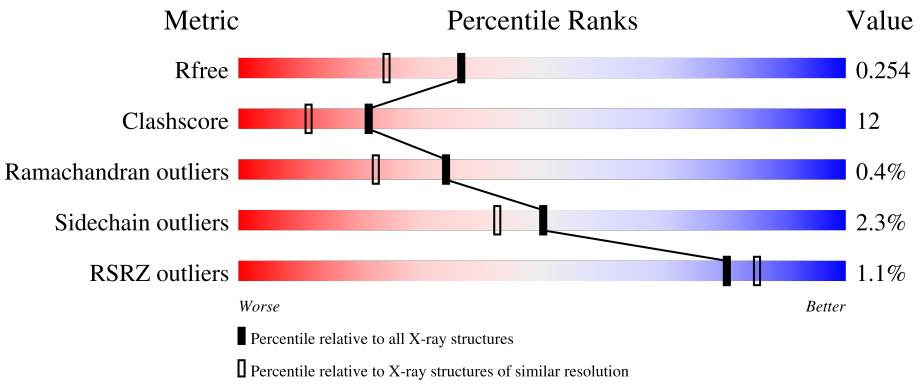
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 1.95 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	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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	163	2% 66% 19% • 15%
1	B	163	55% 25% • 18%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 2208 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 POL polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	139	1062	672	188	198	4	0	0	0
1	B	133	1033	654	183	192	4	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	56	SER	CYS	engineered mutation	UNP Q72498
A	123	GLY	SER	engineered mutation	UNP Q72498
A	124	ALA	THR	engineered mutation	UNP Q72498
A	127	ARG	LYS	engineered mutation	UNP Q72498
A	131	ASP	TRP	engineered mutation	UNP Q72498
A	139	ASP	PHE	engineered mutation	UNP Q72498
A	185	HIS	PHE	engineered mutation	UNP Q72498
B	56	SER	CYS	engineered mutation	UNP Q72498
B	123	GLY	SER	engineered mutation	UNP Q72498
B	124	ALA	THR	engineered mutation	UNP Q72498
B	127	ARG	LYS	engineered mutation	UNP Q72498
B	131	ASP	TRP	engineered mutation	UNP Q72498
B	139	ASP	PHE	engineered mutation	UNP Q72498
B	185	HIS	PHE	engineered mutation	UNP Q72498

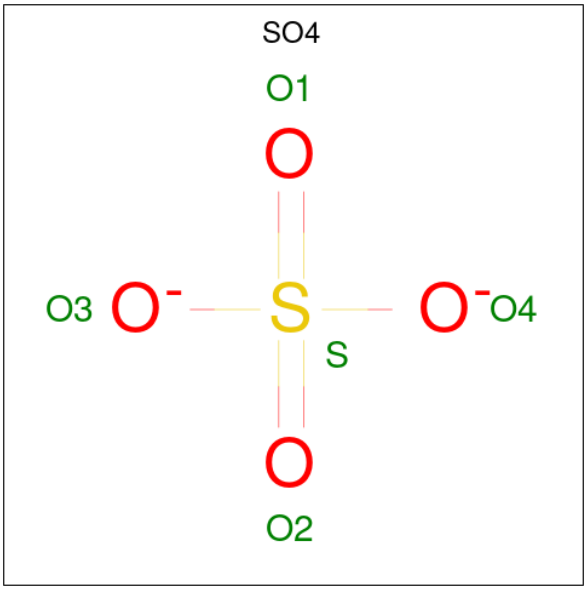
- Molecule 2 is CADMIUM ION (three-letter code: CD) (formula: Cd).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cd		
2	A	2	2	2	0	0
2	B	2	2	2	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

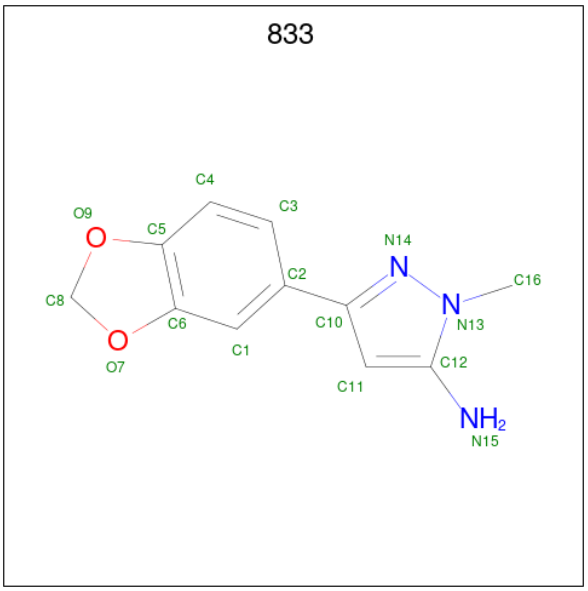
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Cl 1 1	0	0
3	B	1	Total Cl 1 1	0	0

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0
4	B	1	Total O S 5 4 1	0	0

- Molecule 5 is 3-(1,3-benzodioxol-5-yl)-1-methyl-1H-pyrazol-5-amine (three-letter code: 833) (formula: $C_{11}H_{11}N_3O_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
5	A	1	16	11	3	2	0	0
			Total	C	N	O		
5	B	1	16	11	3	2	0	0

- Molecule 6 is water.

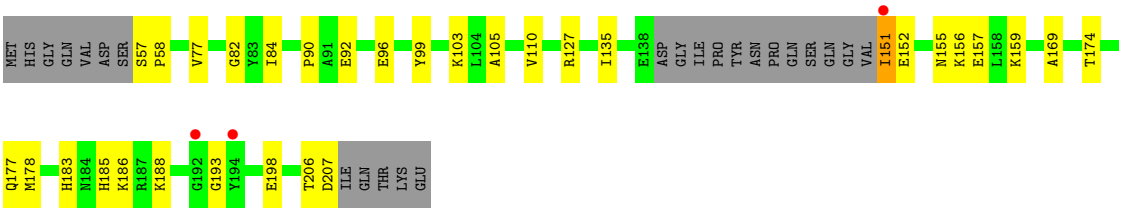
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	31	31	31	0	0
			Total	O		
6	B	29	29	29	0	0

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.

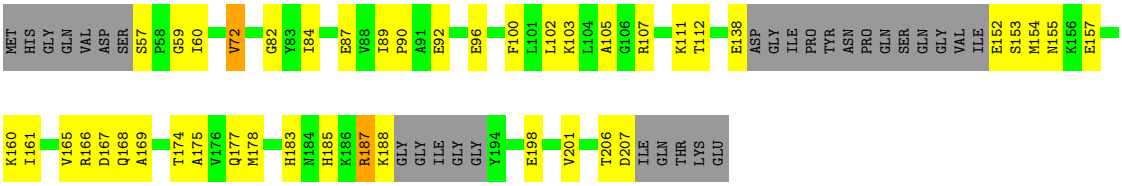
• Molecule 1: POL polyprotein

Chain A:



• Molecule 1: POL polyprotein

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α, β, γ	48.20Å 48.20Å 103.60Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	38.72 – 1.95 38.72 – 1.95	Depositor EDS
% Data completeness (in resolution range)	100.0 (38.72-1.95) 99.3 (38.72-1.95)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	0.04	Depositor
$< I/\sigma(I) >$ [1]	3.38 (at 1.95Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.203 , 0.253 0.205 , 0.254	Depositor DCC
R_{free} test set	975 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	32.3	Xtriage
Anisotropy	0.024	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 29.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.026 for -h,-k,l 0.488 for h,-h-k,-l 0.028 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	2208	wwPDB-VP
Average B, all atoms (Å^2)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.11% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: 833, SO4, CL, CD

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	1.08	3/1081 (0.3%)	0.91	2/1462 (0.1%)
1	B	1.13	4/1051 (0.4%)	0.96	2/1420 (0.1%)
All	All	1.11	7/2132 (0.3%)	0.94	4/2882 (0.1%)

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	92	GLU	CG-CD	6.67	1.61	1.51
1	A	92	GLU	CG-CD	5.76	1.60	1.51
1	A	110	VAL	CB-CG1	5.68	1.64	1.52
1	B	92	GLU	CD-OE1	5.62	1.31	1.25
1	B	72	VAL	CB-CG1	-5.51	1.41	1.52
1	A	92	GLU	CD-OE1	5.16	1.31	1.25
1	B	107	ARG	CZ-NH1	5.03	1.39	1.33

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	187	ARG	NE-CZ-NH1	5.32	122.96	120.30
1	A	92	GLU	OE1-CD-OE2	-5.14	117.14	123.30
1	A	92	GLU	CG-CD-OE1	5.03	128.35	118.30
1	B	187	ARG	NE-CZ-NH2	-5.01	117.80	120.30

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	1062	0	1058	23	2
1	B	1033	0	1032	36	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	10	0	0	0	0
4	B	5	0	0	0	0
5	A	16	0	11	1	0
5	B	16	0	11	2	0
6	A	31	0	0	0	0
6	B	29	0	0	3	0
All	All	2208	0	2112	51	2

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 12.

All (51) 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:B:154:MET:SD	1:B:183:HIS:NE2	2.52	0.83
1:A:57:SER:N	1:A:58:PRO:HD2	1.98	0.78
1:A:186:LYS:HG2	1:A:188:LYS:HG3	1.69	0.75
1:A:155:ASN:O	1:A:159:LYS:HG3	1.86	0.75
1:B:165:VAL:HB	1:B:178:MET:HE3	1.69	0.72
1:B:206:THR:O	1:B:207:ASP:HB2	1.88	0.72
1:B:82:GLY:O	1:B:183:HIS:HE1	1.71	0.72
1:B:112:THR:HG23	1:B:138:GLU:HG2	1.73	0.71
1:B:72:VAL:HG13	1:B:89:ILE:HG13	1.73	0.70
1:A:82:GLY:O	1:A:183:HIS:HE1	1.75	0.69
1:B:165:VAL:HB	1:B:178:MET:CE	2.22	0.69
1:B:72:VAL:CG1	1:B:89:ILE:HG13	2.22	0.69
1:A:90:PRO:HD2	1:A:96:GLU:OE1	2.00	0.61
1:A:99:TYR:CE2	1:A:103:LYS:HD2	2.37	0.60
1:B:187:ARG:O	1:B:188:LYS:HB2	2.01	0.59

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:112:THR:CG2	1:B:138:GLU:HG2	2.33	0.59
1:B:157:GLU:OE1	1:B:183:HIS:HD2	1.87	0.58
1:B:206:THR:O	1:B:207:ASP:CB	2.51	0.58
1:A:198:GLU:HG2	5:A:1:833:C12	2.35	0.56
1:B:57:SER:HB3	1:B:60:ILE:HD12	1.88	0.56
1:B:167:ASP:OD2	1:B:168:GLN:HG3	2.07	0.55
1:B:72:VAL:HG13	1:B:89:ILE:CG1	2.36	0.53
1:A:178:MET:HG3	1:B:102:LEU:HD21	1.91	0.53
1:B:198:GLU:HG2	5:B:1:833:C12	2.40	0.52
1:B:111:LYS:HD2	6:B:217:HOH:O	2.10	0.52
1:A:206:THR:O	1:A:207:ASP:HB2	2.11	0.51
1:A:103:LYS:CG	1:B:177:GLN:HG3	2.41	0.51
1:A:157:GLU:OE1	1:A:183:HIS:HD2	1.95	0.50
1:A:103:LYS:HG3	1:B:177:GLN:HG3	1.92	0.50
1:B:111:LYS:HE3	6:B:216:HOH:O	2.12	0.50
1:A:77:VAL:HG22	1:A:84:ILE:HG12	1.94	0.49
1:A:57:SER:N	1:A:58:PRO:CD	2.73	0.48
1:B:152:GLU:O	1:B:155:ASN:HB2	2.15	0.47
1:A:185:HIS:HE1	1:B:105:ALA:O	1.97	0.47
1:B:84:ILE:HD11	1:B:154:MET:HB3	1.97	0.47
1:A:151:ILE:HG12	1:A:152:GLU:N	2.29	0.47
1:B:90:PRO:HD2	1:B:96:GLU:OE1	2.15	0.47
1:A:105:ALA:HB2	1:A:135:ILE:HD11	1.98	0.46
1:A:177:GLN:HG3	1:B:103:LYS:HG2	1.97	0.46
1:B:59:GLY:HA3	1:B:111:LYS:HD2	1.98	0.46
1:A:169:ALA:HB1	1:A:174:THR:HB	1.97	0.45
1:B:201:VAL:HG21	5:B:1:833:H11	1.99	0.45
1:B:138:GLU:O	6:B:47:HOH:O	2.21	0.45
1:B:161:ILE:O	1:B:165:VAL:HG22	2.18	0.43
1:B:169:ALA:HB1	1:B:174:THR:HB	2.00	0.43
1:A:105:ALA:O	1:B:185:HIS:HE1	2.03	0.42
1:A:99:TYR:CZ	1:A:103:LYS:HD2	2.55	0.42
1:A:177:GLN:HB3	1:B:102:LEU:HG	2.02	0.41
1:B:87:GLU:HG3	1:B:100:PHE:CE2	2.55	0.41
1:A:177:GLN:HG3	1:B:103:LYS:CG	2.51	0.40
1:B:175:ALA:HA	1:B:178:MET:HE2	2.02	0.40

All (2) 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)
1:A:127:ARG:NH2	1:A:186:LYS:CD[3_565]	2.12	0.08
1:A:127:ARG:NH2	1:A:186:LYS:NZ[3_565]	2.18	0.02

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	135/163 (83%)	131 (97%)	3 (2%)	1 (1%)	22	11
1	B	127/163 (78%)	125 (98%)	2 (2%)	0	100	100
All	All	262/326 (80%)	256 (98%)	5 (2%)	1 (0%)	34	22

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	193	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	109/131 (83%)	107 (98%)	2 (2%)	59	53
1	B	108/131 (82%)	105 (97%)	3 (3%)	43	33
All	All	217/262 (83%)	212 (98%)	5 (2%)	50	42

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	151	ILE
1	A	156	LYS
1	B	153	SER
1	B	160	LYS
1	B	166	ARG

Sometimes 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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 11 ligands modelled in this entry, 6 are monoatomic - leaving 5 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	$\#\|Z\| > 2$
4	SO4	A	753	-	4,4,4	0.10	0	6,6,6	0.32	0
5	833	A	1	-	15,18,18	1.35	1 (6%)	16,26,26	2.17	6 (37%)
4	SO4	B	751	-	4,4,4	0.08	0	6,6,6	0.61	0
5	833	B	1	-	15,18,18	1.21	1 (6%)	16,26,26	2.06	5 (31%)
4	SO4	A	752	-	4,4,4	0.23	0	6,6,6	0.68	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
5	833	B	1	-	-	0/0/10/10	0/3/3/3
5	833	A	1	-	-	0/0/10/10	0/3/3/3

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	1	833	C6-C5	3.59	1.48	1.39
5	B	1	833	C6-C5	3.32	1.47	1.39

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	1	833	C16-N13-C12	4.70	134.50	128.86
5	B	1	833	C16-N13-C12	4.38	134.11	128.86
5	A	1	833	O7-C6-C1	4.01	131.32	128.11
5	B	1	833	O7-C6-C1	3.52	130.92	128.11
5	B	1	833	C1-C2-C3	3.11	122.03	118.15
5	B	1	833	C8-O9-C5	2.72	108.88	105.34
5	A	1	833	O7-C6-C5	-2.66	106.75	109.78
5	A	1	833	C8-O7-C6	2.57	108.68	105.34
5	A	1	833	C1-C2-C3	2.26	120.97	118.15
5	A	1	833	C6-C1-C2	-2.21	116.27	120.06
5	B	1	833	C6-C1-C2	-2.01	116.62	120.06

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	1	833	1	0
5	B	1	833	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	139/163 (85%)	0.02	3 (2%) 62 70	20, 31, 63, 77	0
1	B	133/163 (81%)	-0.02	0 100 100	19, 30, 56, 80	0
All	All	272/326 (83%)	0.00	3 (1%) 80 85	19, 31, 62, 80	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	151	ILE	3.2
1	A	192	GLY	2.6
1	A	194	TYR	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 monosaccharides 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. 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	B-factors(Å^2)	Q<0.9
4	SO4	A	752	5/5	0.89	0.15	52,58,60,62	0
4	SO4	B	751	5/5	0.91	0.17	54,56,60,63	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	833	B	1	16/16	0.91	0.17	34,46,53,53	0
5	833	A	1	16/16	0.92	0.20	34,44,54,54	0
4	SO4	A	753	5/5	0.94	0.14	81,82,84,85	0
3	CL	B	502	1/1	0.97	0.04	29,29,29,29	0
3	CL	A	501	1/1	0.98	0.04	30,30,30,30	0
2	CD	B	1004	1/1	1.00	0.05	32,32,32,32	0
2	CD	A	1001	1/1	1.00	0.09	30,30,30,30	0
2	CD	A	1002	1/1	1.00	0.08	32,32,32,32	0
2	CD	B	1003	1/1	1.00	0.07	30,30,30,30	0

6.5 Other polymers (i)

There are no such residues in this entry.