



Full wwPDB X-ray Structure Validation Report

Feb 26, 2014 – 04:56 PM GMT

PDB ID : 3U3K
Title : Crystal structure of hSULT1A1 bound to PAP and 2-Naphtol
Authors : Guttman, C.; Berger, I.; Aharoni, A.; Zarivach, R.
Deposited on : 2011-10-06
Resolution : 2.36 Å(reported)

This is a full wwPDB validation 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/ValidationPDFNotes.html>

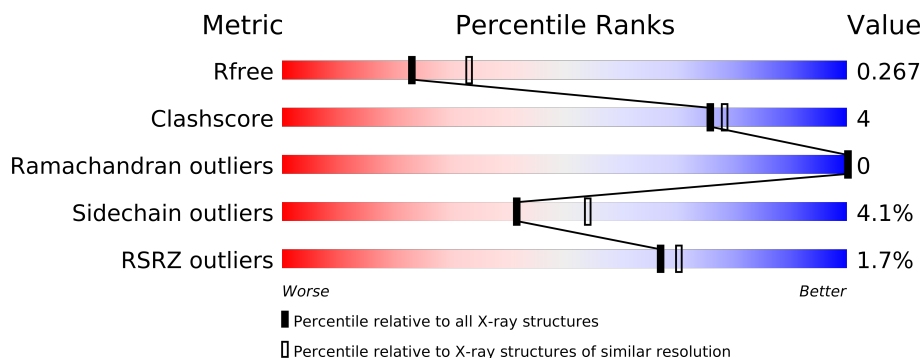
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : stable22639
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.36 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}	66092	3327 (2.40-2.32)
Clashscore	79885	1064 (2.38-2.34)
Ramachandran outliers	78287	1048 (2.38-2.34)
Sidechain outliers	78261	1049 (2.38-2.34)
RSRZ outliers	66119	3330 (2.40-2.32)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	315	
1	B	315	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4932 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Sulfotransferase 1A1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	289	Total	C	N	O	S	0	2	0
			2369	1539	386	428	16			
1	B	288	Total	C	N	O	S	0	4	0
			2374	1543	386	429	16			

There are 44 discrepancies between the modelled and reference sequences:

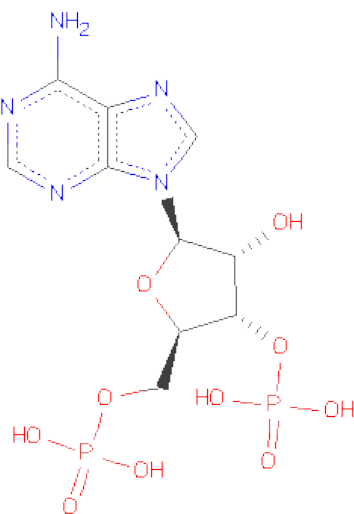
Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	EXPRESSION TAG	UNP P50225
A	-18	GLY	-	EXPRESSION TAG	UNP P50225
A	-17	SER	-	EXPRESSION TAG	UNP P50225
A	-16	SER	-	EXPRESSION TAG	UNP P50225
A	-15	HIS	-	EXPRESSION TAG	UNP P50225
A	-14	HIS	-	EXPRESSION TAG	UNP P50225
A	-13	HIS	-	EXPRESSION TAG	UNP P50225
A	-12	HIS	-	EXPRESSION TAG	UNP P50225
A	-11	HIS	-	EXPRESSION TAG	UNP P50225
A	-10	HIS	-	EXPRESSION TAG	UNP P50225
A	-9	SER	-	EXPRESSION TAG	UNP P50225
A	-8	SER	-	EXPRESSION TAG	UNP P50225
A	-7	GLY	-	EXPRESSION TAG	UNP P50225
A	-6	LEU	-	EXPRESSION TAG	UNP P50225
A	-5	VAL	-	EXPRESSION TAG	UNP P50225
A	-4	PRO	-	EXPRESSION TAG	UNP P50225
A	-3	ARG	-	EXPRESSION TAG	UNP P50225
A	-2	GLY	-	EXPRESSION TAG	UNP P50225
A	-1	SER	-	EXPRESSION TAG	UNP P50225
A	0	HIS	-	EXPRESSION TAG	UNP P50225
A	213	HIS	ARG	SEE REMARK 999	UNP P50225
A	223	MET	VAL	SEE REMARK 999	UNP P50225
B	-19	MET	-	EXPRESSION TAG	UNP P50225
B	-18	GLY	-	EXPRESSION TAG	UNP P50225
B	-17	SER	-	EXPRESSION TAG	UNP P50225

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-16	SER	-	EXPRESSION TAG	UNP P50225
B	-15	HIS	-	EXPRESSION TAG	UNP P50225
B	-14	HIS	-	EXPRESSION TAG	UNP P50225
B	-13	HIS	-	EXPRESSION TAG	UNP P50225
B	-12	HIS	-	EXPRESSION TAG	UNP P50225
B	-11	HIS	-	EXPRESSION TAG	UNP P50225
B	-10	HIS	-	EXPRESSION TAG	UNP P50225
B	-9	SER	-	EXPRESSION TAG	UNP P50225
B	-8	SER	-	EXPRESSION TAG	UNP P50225
B	-7	GLY	-	EXPRESSION TAG	UNP P50225
B	-6	LEU	-	EXPRESSION TAG	UNP P50225
B	-5	VAL	-	EXPRESSION TAG	UNP P50225
B	-4	PRO	-	EXPRESSION TAG	UNP P50225
B	-3	ARG	-	EXPRESSION TAG	UNP P50225
B	-2	GLY	-	EXPRESSION TAG	UNP P50225
B	-1	SER	-	EXPRESSION TAG	UNP P50225
B	0	HIS	-	EXPRESSION TAG	UNP P50225
B	213	HIS	ARG	SEE REMARK 999	UNP P50225
B	223	MET	VAL	SEE REMARK 999	UNP P50225

- Molecule 2 is ADENOSINE-3'-5'-DIPHOSPHATE (three-letter code: A3P) (formula: $C_{10}H_{15}N_5O_{10}P_2$).



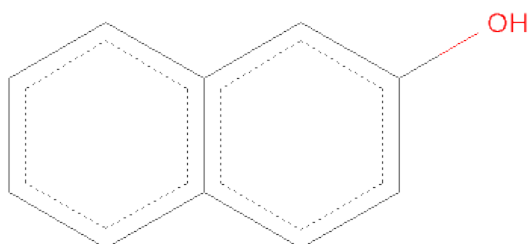
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	B	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

- Molecule 3 is NAPHTHALEN-2-OL (three-letter code: 03V) (formula: C₁₀H₈O).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			11	10	1		
3	B	1	Total	C	O	0	0
			11	10	1		

- Molecule 4 is water.

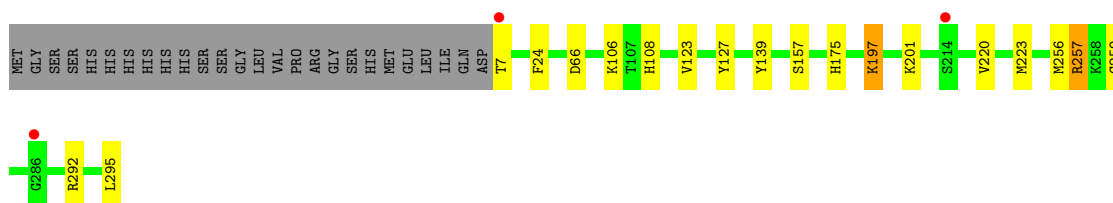
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	60	Total	O	0	0
			60	60		
4	B	53	Total	O	0	0
			53	53		

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 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.

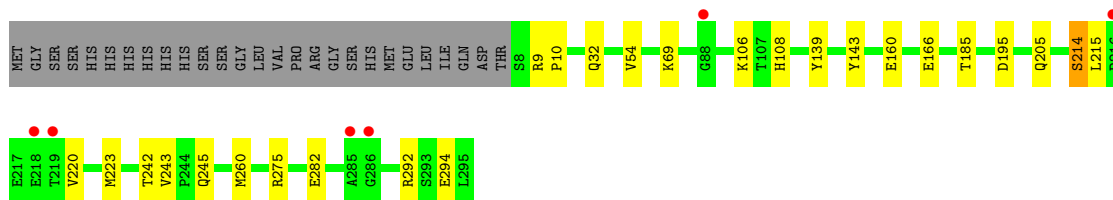
- Molecule 1: Sulfotransferase 1A1

Chain A: 



- Molecule 1: Sulfotransferase 1A1

Chain B: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	72.18Å 81.27Å 121.76Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	22.00 – 2.36 21.67 – 2.36	Depositor EDS
% Data completeness (in resolution range)	95.3 (22.00-2.36) 95.4 (21.67-2.36)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.85 (at 2.36Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.214 , 0.271 0.214 , 0.267	Depositor DCC
R_{free} test set	1456 reflections (5.35%)	DCC
Wilson B-factor (Å ²)	47.1	Xtriage
Anisotropy	0.079	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 19.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.35$	Xtriage
Outliers	2 of 28688 reflections (0.007%)	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	4932	wwPDB-VP
Average B, all atoms (Å ²)	44.0	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 46.36 % 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 1.1574e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: A3P, 03V

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	0.94	0/2449	0.81	1/3322 (0.0%)
1	B	0.88	0/2460	0.79	0/3336
All	All	0.91	0/4909	0.80	1/6658 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	257	ARG	NE-CZ-NH1	5.20	122.90	120.30

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2369	0	0	10	0
1	B	2374	0	0	9	0
2	A	27	0	11	1	0
2	B	27	0	11	0	0
3	A	11	0	7	2	0
3	B	11	0	7	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	60	0	0	3	0
4	B	53	0	0	1	0
All	All	4932	0	36	19	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 4.

All (19) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:108:HIS:NE2	3:B:2002:03V:O1	2.24	0.71
1:A:108:HIS:NE2	3:A:2002:03V:O1	2.30	0.64
1:A:106:LYS:NZ	4:A:306:HOH:O	2.33	0.61
1:A:197:LYS:NZ	1:A:259:GLY:O	2.36	0.58
1:B:220:VAL:O	1:B:223:MET:N	2.37	0.57
1:A:220:VAL:O	1:A:223:MET:N	2.45	0.50
1:B:106:LYS:NZ	3:B:2002:03V:H1	2.26	0.49
1:B:160:GLU:OE1	1:B:275:ARG:NH2	2.46	0.48
1:B:214:SER:O	1:B:215:LEU:CD1	2.62	0.48
1:A:66:ASP:OD1	4:A:339:HOH:O	2.20	0.47
1:B:166:GLU:OE2	4:B:318:HOH:O	2.21	0.47
1:A:127:TYR:OH	1:A:175:HIS:ND1	2.50	0.45
1:B:9:ARG:N	1:B:10:PRO:CD	2.80	0.44
1:B:205[B]:GLN:NE2	1:B:220:VAL:CG2	2.81	0.44
1:A:256:MET:SD	2:A:2001:A3P:C8	3.06	0.44
1:A:106:LYS:NZ	3:A:2002:03V:H1	2.32	0.44
1:B:195:ASP:OD1	1:B:292:ARG:NH1	2.51	0.43
1:A:257:ARG:NH2	4:A:329:HOH:O	2.53	0.42
1:A:24:PHE:N	1:A:24:PHE:CD1	2.89	0.41

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone ⓘ

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	289/315 (92%)	277 (96%)	12 (4%)	0	100	100
1	B	290/315 (92%)	280 (97%)	10 (3%)	0	100	100
All	All	579/630 (92%)	557 (96%)	22 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	261/282 (93%)	253 (97%)	8 (3%)	52	68
1	B	262/282 (93%)	248 (95%)	14 (5%)	32	39
All	All	523/564 (93%)	501 (96%)	22 (4%)	41	52

All (22) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	7	THR
1	A	123	VAL
1	A	139	TYR
1	A	157	SER
1	A	197	LYS
1	A	201	LYS
1	A	292	ARG
1	A	295	LEU
1	B	32	GLN
1	B	54	VAL
1	B	69	LYS
1	B	139	TYR
1	B	143	TYR
1	B	185	THR
1	B	214	SER
1	B	242	THR
1	B	243	VAL
1	B	245	GLN
1	B	260[A]	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	260[B]	MET
1	B	282	GLU
1	B	294	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

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	$\# Z > 2$
2	A3P	A	2001	-	29,29,29	1.21	1 (3%)	45,45,45	2.50	14 (31%)
3	03V	A	2002	-	12,12,12	1.55	3 (25%)	16,16,16	1.30	1 (6%)
2	A3P	B	2001	-	29,29,29	0.98	1 (3%)	45,45,45	2.29	15 (33%)
3	03V	B	2002	-	12,12,12	1.58	2 (16%)	16,16,16	1.32	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
2	A3P	A	2001	-	-	0/15/31/31	0/1/3/3
3	03V	A	2002	-	-	0/0/0/0	0/0/2/2
2	A3P	B	2001	-	-	0/15/31/31	0/1/3/3
3	03V	B	2002	-	-	0/0/0/0	0/0/2/2

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	2001	A3P	C4-N9	-4.40	1.31	1.37
2	B	2001	A3P	C4-N9	-2.81	1.33	1.37
3	B	2002	03V	C5-C4	2.63	1.48	1.42
3	B	2002	03V	C11-C4	-2.60	1.35	1.41
3	A	2002	03V	C7-C5	-2.56	1.35	1.41
3	A	2002	03V	C5-C4	2.50	1.48	1.42
3	A	2002	03V	C11-C4	-2.02	1.36	1.41

All (31) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	2001	A3P	N3-C2-N1	-8.37	121.71	128.71
2	B	2001	A3P	N3-C2-N1	-8.28	121.79	128.71
2	A	2001	A3P	C8-N9-C4	5.94	111.43	106.90
2	A	2001	A3P	C4'-O4'-C1'	-5.77	103.48	109.75
2	B	2001	A3P	O6P-P2-O5'	-4.32	94.72	106.65
2	A	2001	A3P	O4'-C1'-N9	4.05	112.21	108.44
2	B	2001	A3P	N3-C4-N9	3.85	132.38	125.43
2	B	2001	A3P	C4-C5-N7	-3.84	106.24	109.52
2	A	2001	A3P	N3-C4-N9	3.69	132.10	125.43
2	B	2001	A3P	C8-N9-C4	3.68	109.71	106.90
2	A	2001	A3P	C1'-N9-C4	-3.57	120.47	126.64
2	A	2001	A3P	O2P-P1-O3'	-3.37	97.38	107.09
2	A	2001	A3P	O3P-P1-O2P	3.35	120.64	107.61
2	B	2001	A3P	O2P-P1-O3'	-3.26	97.69	107.09
2	A	2001	A3P	O6P-P2-O5'	-3.24	97.69	106.65
2	B	2001	A3P	P1-O3'-C3'	3.22	128.73	121.96
2	A	2001	A3P	P1-O3'-C3'	3.07	128.41	121.96
2	B	2001	A3P	O3'-P1-O1P	-3.01	98.39	106.79
2	B	2001	A3P	O6P-P2-O5P	2.96	119.13	107.61
2	A	2001	A3P	O6P-P2-O4P	2.83	119.68	110.44

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	2001	A3P	C5-C4-N3	-2.79	119.62	125.70
3	A	2002	03V	C2-C3-C4	-2.66	117.14	120.49
2	A	2001	A3P	N7-C8-N9	-2.64	106.88	114.36
2	B	2001	A3P	C1'-N9-C4	-2.61	122.13	126.64
3	B	2002	03V	C9-C10-C11	2.58	124.24	120.47
2	B	2001	A3P	N7-C8-N9	-2.57	107.09	114.36
2	A	2001	A3P	O3'-P1-O1P	-2.52	99.76	106.79
2	B	2001	A3P	C2-N3-C4	2.15	120.12	114.01
2	A	2001	A3P	O5P-P2-O5'	-2.14	100.73	106.65
2	B	2001	A3P	O5P-P2-O5'	-2.08	100.90	106.65
2	B	2001	A3P	O3P-P1-O1P	2.00	116.99	110.44

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues

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, 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	289/315 (91%)	-0.15	3 (1%) 79 83	24, 43, 67, 81	3 (1%)
1	B	288/315 (91%)	-0.14	6 (2%) 60 63	27, 42, 66, 76	3 (1%)
All	All	577/630 (91%)	-0.15	9 (1%) 67 72	24, 43, 67, 81	6 (1%)

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	216	PRO	3.2
1	B	285	ALA	3.1
1	B	286	GLY	3.0
1	A	7	THR	2.6
1	A	286	GLY	2.5
1	B	218	GLU	2.2
1	B	219	THR	2.2
1	B	88	GLY	2.1
1	A	214	SER	2.1

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, 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	RSR	LLDF	B-factors(Å ²)	Q<0.9
3	03V	A	2002	11/11	0.13	0.46	39,41,43,43	0
3	03V	B	2002	11/11	0.14	0.37	40,43,44,45	0
2	A3P	A	2001	27/27	0.08	-1.38	28,33,37,38	0
2	A3P	B	2001	27/27	0.09	-1.54	23,31,33,34	0

6.5 Other polymers ⓘ

There are no such residues in this entry.