



wwPDB X-ray Structure Validation Summary Report

Nov 25, 2014 – 06:41 AM EST

PDB ID : 4PIP
Title : Engineered EgtD variant EgtD-M252V,E282A in complex with tryptophan and SAH
Authors : Vit, A.; Seebeck, F.P.; Blankenfeldt, W.
Deposited on : 2014-05-09
Resolution : 1.80 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

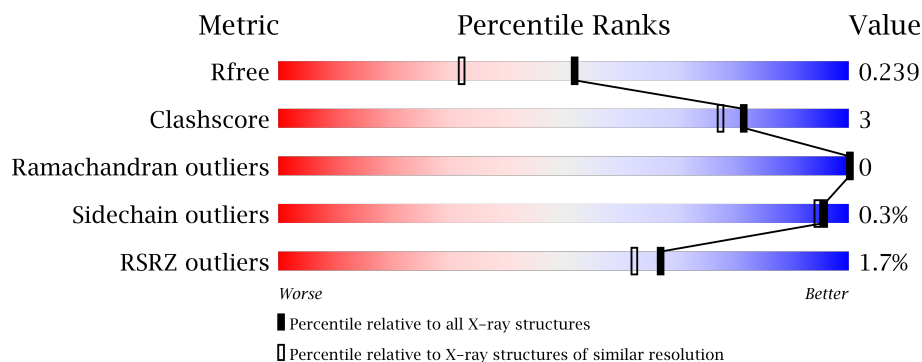
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.16 November 2013
Xtriage (Phenix) : dev-1439
EDS : stable24195
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.1.3
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable24195

1 Overall quality at a glance

The reported resolution of this entry is 1.80 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	3513 (1.80-1.80)
Clashscore	79885	4461 (1.80-1.80)
Ramachandran outliers	78287	4404 (1.80-1.80)
Sidechain outliers	78261	4403 (1.80-1.80)
RSRZ outliers	66119	3515 (1.80-1.80)

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	323	
1	B	323	
1	C	323	
1	D	323	

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 20818 atoms, of which 9603 are hydrogens 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 Histidine-specific methyltransferase EgtD.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	316	Total	C	H	N	O	S	0	6	0
			4839	1541	2395	427	471	5			
1	B	309	Total	C	H	N	O	S	0	1	0
			4733	1497	2352	421	458	5			
1	C	315	Total	C	H	N	O	S	0	2	0
			4727	1512	2330	416	463	6			
1	D	318	Total	C	H	N	O	S	0	7	0
			4868	1548	2411	429	475	5			

There are 32 discrepancies between the modelled and reference sequences:

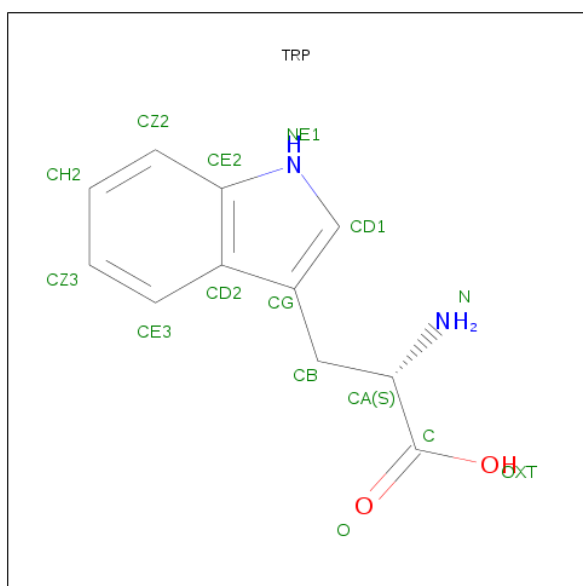
Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP A0R5M8
A	0	HIS	-	expression tag	UNP A0R5M8
A	2	ALA	THR	engineered mutation	UNP A0R5M8
A	29	THR	ALA	engineered mutation	UNP A0R5M8
A	30	GLN	PRO	engineered mutation	UNP A0R5M8
A	75	SER	ALA	engineered mutation	UNP A0R5M8
A	252	VAL	MET	engineered mutation	UNP A0R5M8
A	282	ALA	GLU	engineered mutation	UNP A0R5M8
B	-1	GLY	-	expression tag	UNP A0R5M8
B	0	HIS	-	expression tag	UNP A0R5M8
B	2	ALA	THR	engineered mutation	UNP A0R5M8
B	29	THR	ALA	engineered mutation	UNP A0R5M8
B	30	GLN	PRO	engineered mutation	UNP A0R5M8
B	75	SER	ALA	engineered mutation	UNP A0R5M8
B	252	VAL	MET	engineered mutation	UNP A0R5M8
B	282	ALA	GLU	engineered mutation	UNP A0R5M8
C	-1	GLY	-	expression tag	UNP A0R5M8
C	0	HIS	-	expression tag	UNP A0R5M8
C	2	ALA	THR	engineered mutation	UNP A0R5M8
C	29	THR	ALA	engineered mutation	UNP A0R5M8
C	30	GLN	PRO	engineered mutation	UNP A0R5M8

Continued on next page...

Continued from previous page...

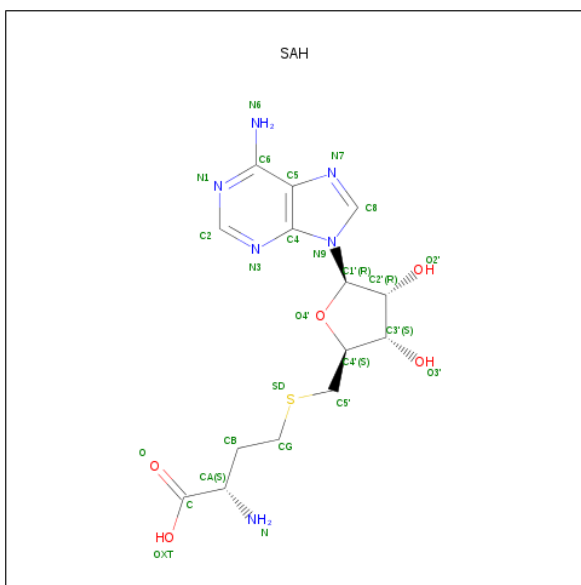
Chain	Residue	Modelled	Actual	Comment	Reference
C	75	SER	ALA	engineered mutation	UNP A0R5M8
C	252	VAL	MET	engineered mutation	UNP A0R5M8
C	282	ALA	GLU	engineered mutation	UNP A0R5M8
D	-1	GLY	-	expression tag	UNP A0R5M8
D	0	HIS	-	expression tag	UNP A0R5M8
D	2	ALA	THR	engineered mutation	UNP A0R5M8
D	29	THR	ALA	engineered mutation	UNP A0R5M8
D	30	GLN	PRO	engineered mutation	UNP A0R5M8
D	75	SER	ALA	engineered mutation	UNP A0R5M8
D	252	VAL	MET	engineered mutation	UNP A0R5M8
D	282	ALA	GLU	engineered mutation	UNP A0R5M8

- Molecule 2 is TRYPTOPHAN (three-letter code: TRP) (formula: $C_{11}H_{12}N_2O_2$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	H	N	O	0	0
			27	11	12	2	2		
2	B	1	Total	C	H	N	O	0	0
			24	11	9	2	2		
2	C	1	Total	C	H	N	O	0	0
			24	11	9	2	2		
2	D	1	Total	C	H	N	O	0	0
			24	11	9	2	2		

- Molecule 3 is S-ADENOSYL-L-HOMOCYSTEINE (three-letter code: SAH) (formula: $C_{14}H_{20}N_6O_5S$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total 45	C 14	H 19	N 6	O 5	S 1	0	0
3	B	1	Total 45	C 14	H 19	N 6	O 5	S 1	0	0
3	C	1	Total 45	C 14	H 19	N 6	O 5	S 1	0	0
3	D	1	Total 45	C 14	H 19	N 6	O 5	S 1	0	0

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mg	0	0
			1	1		
4	C	1	Total	Mg	0	0
			1	1		

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	1	Total	Cl	0	0
			1	1		

- Molecule 6 is water.

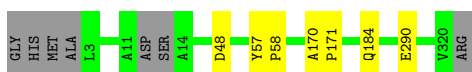
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	343	Total 343	O 343	0	0
6	B	325	Total 325	O 325	0	0
6	C	360	Total 360	O 360	0	0
6	D	341	Total 341	O 341	0	0

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: Histidine-specific methyltransferase EgtD

Chain A: 



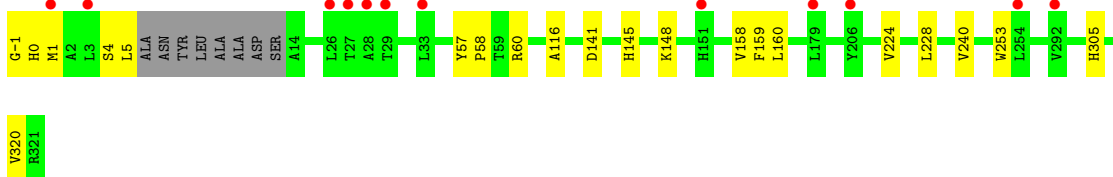
- Molecule 1: Histidine-specific methyltransferase EgtD

Chain B: 



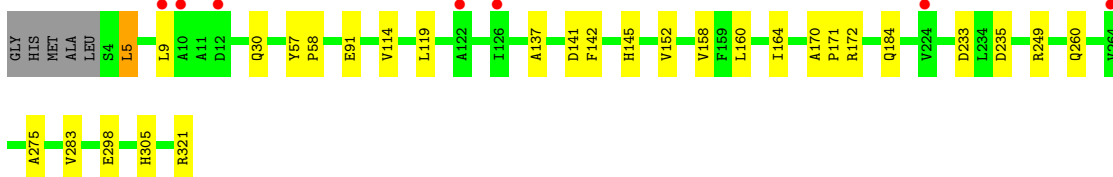
- Molecule 1: Histidine-specific methyltransferase EgtD

Chain C: 



- Molecule 1: Histidine-specific methyltransferase EgtD

Chain D: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	79.05Å 67.63Å 112.12Å 90.00° 109.24° 90.00°	Depositor
Resolution (Å)	41.86 – 1.80 41.86 – 1.80	Depositor EDS
% Data completeness (in resolution range)	99.9 (41.86-1.80) 99.9 (41.86-1.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.46 (at 1.79Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, R_{free}	0.194 , 0.239 0.195 , 0.239	Depositor DCC
R_{free} test set	5270 reflections (5.09%)	DCC
Wilson B-factor (Å ²)	16.2	Xtriage
Anisotropy	0.286	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 39.4	EDS
Estimated twinning fraction	0.000 for h,-k,-h-l	Xtriage
L-test for twinning	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	3 of 103581 reflections (0.003%)	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	20818	wwPDB-VP
Average B, all atoms (Å ²)	24.0	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 69.76 % 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 3.5033e-06. 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: MG, SAH, CL

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	0.44	2/2500 (0.1%)	0.61	0/3400
1	B	0.44	0/2428	0.57	1/3300 (0.0%)
1	C	0.39	0/2445	0.54	0/3325
1	D	0.37	0/2523	0.56	1/3435 (0.0%)
All	All	0.41	2/9896 (0.0%)	0.57	2/13460 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	48[A]	ASP	CA-C	5.25	1.66	1.52
1	A	48[B]	ASP	CA-C	5.25	1.66	1.52

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	5	LEU	CA-CB-CG	5.35	127.61	115.30
1	B	90	SER	CB-CA-C	5.06	119.71	110.10

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	2444	2395	2408	6	0
1	B	2381	2352	2349	8	0
1	C	2397	2330	2362	22	0
1	D	2457	2411	2414	23	1
2	A	15	12	9	1	0
2	B	15	9	9	1	0
2	C	15	9	9	0	0
2	D	15	9	9	0	0
3	A	26	19	19	0	0
3	B	26	19	19	0	0
3	C	26	19	19	0	0
3	D	26	19	19	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
5	C	1	0	0	5	1
6	A	343	0	0	2	4
6	B	325	0	0	6	4
6	C	360	0	0	5	3
6	D	341	0	0	10	2
All	All	11215	9603	9645	62	8

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 3.

The worst 5 of 62 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:C:1:MET:CE	1:C:4:SER:OG	1.92	1.17
1:C:0:HIS:ND1	5:C:401:CL:CL	2.17	1.13
1:C:-1:GLY:N	5:C:401:CL:CL	2.23	1.08
1:C:1:MET:HE1	1:C:4:SER:OG	1.51	1.03
1:C:0:HIS:N	5:C:401:CL:CL	2.28	1.02

The worst 5 of 8 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	Distance(Å)	Clash(Å)
1:D:305:HIS:HE2	5:C:401:CL:CL[1_445]	1.42	0.18
6:A:558:HOH:O	6:D:511:HOH:O[1_655]	2.08	0.12
6:A:618:HOH:O	6:C:607:HOH:O[1_545]	2.09	0.11
6:B:544:HOH:O	6:D:610:HOH:O[1_565]	2.11	0.09
6:A:596:HOH:O	6:A:624:HOH:O[2_544]	2.11	0.09

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	317/323 (98%)	307 (97%)	10 (3%)	0	100	100
1	B	308/323 (95%)	300 (97%)	8 (3%)	0	100	100
1	C	313/323 (97%)	309 (99%)	4 (1%)	0	100	100
1	D	323/323 (100%)	318 (98%)	5 (2%)	0	100	100
All	All	1261/1292 (98%)	1234 (98%)	27 (2%)	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	251/253 (99%)	251 (100%)	0	100	100
1	B	244/253 (96%)	243 (100%)	1 (0%)	95	93
1	C	244/253 (96%)	243 (100%)	1 (0%)	95	93
1	D	251/253 (99%)	250 (100%)	1 (0%)	95	93
All	All	990/1012 (98%)	987 (100%)	3 (0%)	96	94

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	142	PHE
1	C	159	PHE
1	D	142	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	145	HIS
1	D	145	HIS
1	D	184	GLN

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 ⓘ

Of 11 ligands modelled in this entry, 3 are monoatomic - leaving 8 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$
2	TRP	A	401	-	16,16,16	0.68	0	22,22,22	0.88	2 (9%)
3	SAH	A	402	-	28,28,28	1.00	2 (7%)	40,40,40	2.14	6 (15%)
2	TRP	B	401	-	16,16,16	0.68	0	22,22,22	0.99	2 (9%)
3	SAH	B	402	-	28,28,28	0.98	1 (3%)	40,40,40	2.31	8 (20%)
2	TRP	C	402	-	16,16,16	0.72	0	22,22,22	0.91	1 (4%)
3	SAH	C	403	-	28,28,28	1.00	2 (7%)	40,40,40	2.12	6 (15%)
2	TRP	D	401	-	16,16,16	0.70	0	22,22,22	0.85	0
3	SAH	D	402	-	28,28,28	1.05	2 (7%)	40,40,40	2.40	6 (15%)

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	TRP	A	401	-	-	0/8/8/8	0/2/2/2
3	SAH	A	402	-	-	0/15/31/31	0/3/3/3
2	TRP	B	401	-	-	0/8/8/8	0/2/2/2
3	SAH	B	402	-	-	0/15/31/31	0/3/3/3
2	TRP	C	402	-	-	0/8/8/8	0/2/2/2
3	SAH	C	403	-	-	0/15/31/31	0/3/3/3
2	TRP	D	401	-	-	0/8/8/8	0/2/2/2
3	SAH	D	402	-	-	0/15/31/31	0/3/3/3

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	402	SAH	C2-N3	3.28	1.37	1.32
3	D	402	SAH	C2-N3	3.25	1.37	1.32
3	C	403	SAH	C2-N3	2.99	1.37	1.32
3	B	402	SAH	C2-N3	2.74	1.36	1.32
3	D	402	SAH	C2-N1	2.37	1.38	1.33

The worst 5 of 31 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	402	SAH	N3-C2-N1	-11.42	118.84	128.89
3	B	402	SAH	N3-C2-N1	-11.23	119.01	128.89
3	A	402	SAH	N3-C2-N1	-10.36	119.77	128.89
3	C	403	SAH	N3-C2-N1	-10.14	119.97	128.89
3	D	402	SAH	C5-C4-N3	-5.11	121.00	125.98

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	316/323 (97%)	-0.29	0 100 100	8, 18, 34, 42	0
1	B	309/323 (95%)	-0.24	2 (0%) 86 85	9, 19, 34, 60	0
1	C	315/323 (97%)	0.18	12 (3%) 38 31	13, 22, 37, 57	0
1	D	318/323 (98%)	-0.03	7 (2%) 59 53	9, 21, 45, 53	0
All	All	1258/1292 (97%)	-0.09	21 (1%) 67 62	8, 20, 39, 60	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	26	LEU	5.6
1	C	33	LEU	3.3
1	D	126	ILE	3.3
1	C	27	THR	3.0
1	C	292	VAL	2.8

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, 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(\AA^2)	Q<0.9
3	SAH	B	402	26/26	0.11	1.46	12,16,29,30	0
5	CL	C	401	1/1	0.22	1.36	22,22,22,22	0
2	TRP	C	402	15/15	0.14	0.80	26,34,42,43	0
2	TRP	B	401	15/15	0.10	0.65	12,17,21,25	0
3	SAH	A	402	26/26	0.10	0.47	11,15,20,20	0
2	TRP	D	401	15/15	0.09	-0.07	14,21,30,35	0
4	MG	C	404	1/1	0.08	-0.37	34,34,34,34	0
3	SAH	C	403	26/26	0.09	-0.70	11,16,25,27	0
3	SAH	D	402	26/26	0.08	-0.75	11,20,26,30	0
2	TRP	A	401	15/15	0.07	-1.00	7,10,14,17	0
4	MG	B	403	1/1	0.04	-4.40	12,12,12,12	0

6.5 Other polymers ⓘ

There are no such residues in this entry.