



wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 03:02 AM GMT

PDB ID : 4J5I
Title : Crystal structure of an alpha-ketoglutarate-dependent taurine dioxygenase from Mycobacterium smegmatis
Authors : Seattle Structural Genomics Center for Infectious Disease (SSGCID)
Deposited on : 2013-02-08
Resolution : 2.60 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>

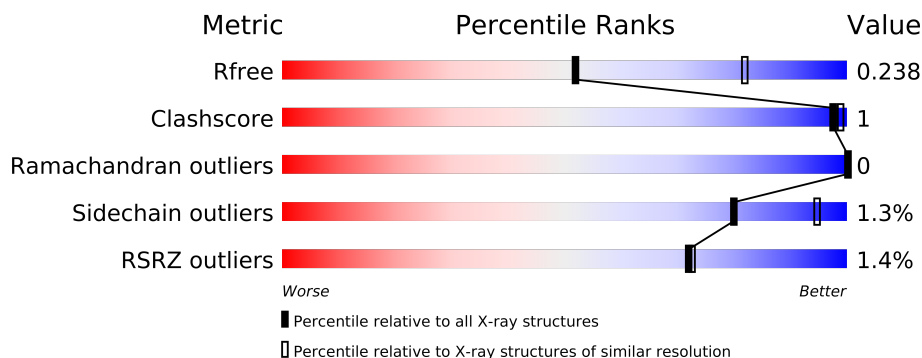
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 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 2.60 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	1718 (2.60-2.60)
Clashscore	79885	2154 (2.60-2.60)
Ramachandran outliers	78287	2113 (2.60-2.60)
Sidechain outliers	78261	2113 (2.60-2.60)
RSRZ outliers	66119	1718 (2.60-2.60)

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	304	
1	B	304	
1	C	304	
1	D	304	
1	E	304	
1	F	304	
1	G	304	
1	H	304	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 16323 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 Alpha-ketoglutarate-dependent taurine dioxygenase.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	248	Total	C	N	O	0	0	0
			1911	1217	339	355			
1	B	285	Total	C	N	O	0	0	0
			2186	1385	385	416			
1	C	271	Total	C	N	O	0	0	0
			2067	1311	363	393			
1	D	239	Total	C	N	O	0	0	0
			1839	1171	325	343			
1	E	250	Total	C	N	O	0	1	0
			1925	1225	338	362			
1	F	250	Total	C	N	O	0	0	0
			1929	1224	339	366			
1	G	264	Total	C	N	O	0	0	0
			2026	1289	356	381			
1	H	249	Total	C	N	O	0	0	0
			1918	1219	340	359			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	EXPRESSION TAG	UNP A0QZ23
A	-2	PRO	-	EXPRESSION TAG	UNP A0QZ23
A	-1	GLY	-	EXPRESSION TAG	UNP A0QZ23
A	0	SER	-	EXPRESSION TAG	UNP A0QZ23
A	1	MET	-	EXPRESSION TAG	UNP A0QZ23
A	2	VAL	-	EXPRESSION TAG	UNP A0QZ23
B	-3	GLY	-	EXPRESSION TAG	UNP A0QZ23
B	-2	PRO	-	EXPRESSION TAG	UNP A0QZ23
B	-1	GLY	-	EXPRESSION TAG	UNP A0QZ23
B	0	SER	-	EXPRESSION TAG	UNP A0QZ23
B	1	MET	-	EXPRESSION TAG	UNP A0QZ23
B	2	VAL	-	EXPRESSION TAG	UNP A0QZ23
C	-3	GLY	-	EXPRESSION TAG	UNP A0QZ23

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-2	PRO	-	EXPRESSION TAG	UNP A0QZ23
C	-1	GLY	-	EXPRESSION TAG	UNP A0QZ23
C	0	SER	-	EXPRESSION TAG	UNP A0QZ23
C	1	MET	-	EXPRESSION TAG	UNP A0QZ23
C	2	VAL	-	EXPRESSION TAG	UNP A0QZ23
D	-3	GLY	-	EXPRESSION TAG	UNP A0QZ23
D	-2	PRO	-	EXPRESSION TAG	UNP A0QZ23
D	-1	GLY	-	EXPRESSION TAG	UNP A0QZ23
D	0	SER	-	EXPRESSION TAG	UNP A0QZ23
D	1	MET	-	EXPRESSION TAG	UNP A0QZ23
D	2	VAL	-	EXPRESSION TAG	UNP A0QZ23
E	-3	GLY	-	EXPRESSION TAG	UNP A0QZ23
E	-2	PRO	-	EXPRESSION TAG	UNP A0QZ23
E	-1	GLY	-	EXPRESSION TAG	UNP A0QZ23
E	0	SER	-	EXPRESSION TAG	UNP A0QZ23
E	1	MET	-	EXPRESSION TAG	UNP A0QZ23
E	2	VAL	-	EXPRESSION TAG	UNP A0QZ23
F	-3	GLY	-	EXPRESSION TAG	UNP A0QZ23
F	-2	PRO	-	EXPRESSION TAG	UNP A0QZ23
F	-1	GLY	-	EXPRESSION TAG	UNP A0QZ23
F	0	SER	-	EXPRESSION TAG	UNP A0QZ23
F	1	MET	-	EXPRESSION TAG	UNP A0QZ23
F	2	VAL	-	EXPRESSION TAG	UNP A0QZ23
G	-3	GLY	-	EXPRESSION TAG	UNP A0QZ23
G	-2	PRO	-	EXPRESSION TAG	UNP A0QZ23
G	-1	GLY	-	EXPRESSION TAG	UNP A0QZ23
G	0	SER	-	EXPRESSION TAG	UNP A0QZ23
G	1	MET	-	EXPRESSION TAG	UNP A0QZ23
G	2	VAL	-	EXPRESSION TAG	UNP A0QZ23
H	-3	GLY	-	EXPRESSION TAG	UNP A0QZ23
H	-2	PRO	-	EXPRESSION TAG	UNP A0QZ23
H	-1	GLY	-	EXPRESSION TAG	UNP A0QZ23
H	0	SER	-	EXPRESSION TAG	UNP A0QZ23
H	1	MET	-	EXPRESSION TAG	UNP A0QZ23
H	2	VAL	-	EXPRESSION TAG	UNP A0QZ23

- Molecule 2 is FE (III) ION (three-letter code: FE) (formula: Fe).

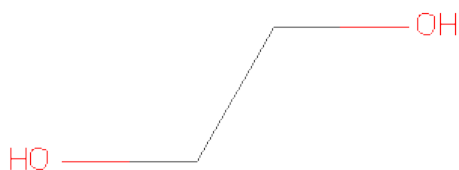
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	G	1	Total Fe 1 1	0	0
2	D	1	Total Fe 1 1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	E	1	Total 1	Fe 1	0	0
2	H	1	Total 1	Fe 1	0	0
2	B	1	Total 1	Fe 1	0	0
2	C	1	Total 1	Fe 1	0	0
2	A	1	Total 1	Fe 1	0	0
2	F	1	Total 1	Fe 1	0	0

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 4	C 2	O 2	0	0

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	F	1	Total 1	Mg 1	0	0

- Molecule 5 is water.

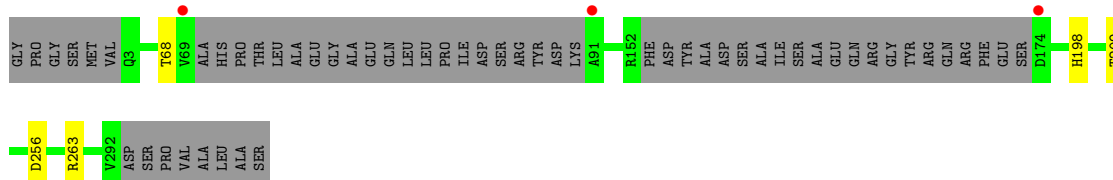
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	60	Total 60	O 60	0	0
5	B	88	Total 88	O 88	0	0
5	C	63	Total 63	O 63	0	0
5	D	44	Total 44	O 44	0	0
5	E	67	Total 67	O 67	0	0
5	F	67	Total 67	O 67	0	0
5	G	63	Total 63	O 63	0	0
5	H	57	Total 57	O 57	0	0

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.

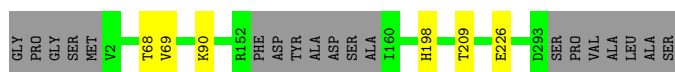
- Molecule 1: Alpha-ketoglutarate-dependent taurine dioxygenase

Chain A: 



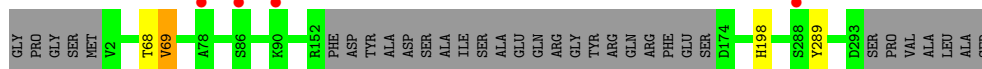
- Molecule 1: Alpha-ketoglutarate-dependent taurine dioxygenase

Chain B: 



- Molecule 1: Alpha-ketoglutarate-dependent taurine dioxygenase

Chain C: 



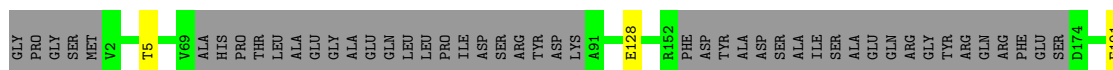
- Molecule 1: Alpha-ketoglutarate-dependent taurine dioxygenase

Chain D: 



- Molecule 1: Alpha-ketoglutarate-dependent taurine dioxygenase

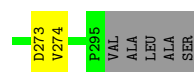
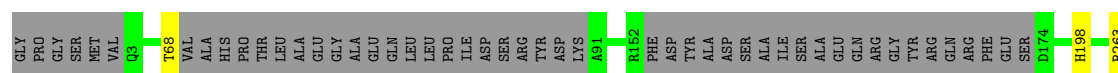
Chain E: 





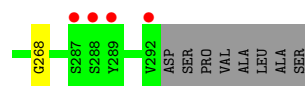
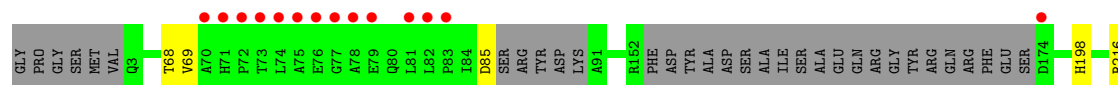
- Molecule 1: Alpha-ketoglutarate-dependent taurine dioxygenase

Chain F:



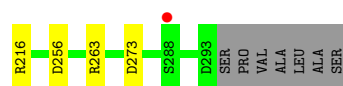
- Molecule 1: Alpha-ketoglutarate-dependent taurine dioxygenase

Chain G:



- Molecule 1: Alpha-ketoglutarate-dependent taurine dioxygenase

Chain H:



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	77.01Å 166.68Å 99.72Å 90.00° 90.57° 90.00°	Depositor
Resolution (Å)	49.40 – 2.60 49.35 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.0 (49.40-2.60) 99.0 (49.35-2.60)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.81 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.211 , 0.234 0.215 , 0.238	Depositor DCC
R_{free} test set	3833 reflections (5.28%)	DCC
Wilson B-factor (Å ²)	26.6	Xtriage
Anisotropy	0.043	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 2.1	EDS
Estimated twinning fraction	0.045 for h,-k,-l	Xtriage
L-test for twinning	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 76379 reflections	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	16323	wwPDB-VP
Average B, all atoms (Å ²)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.17% of the height of the origin peak. No significant pseudotranslation is detected.*

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, FE, EDO

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	0.50	0/1957	0.74	3/2675 (0.1%)
1	B	0.54	1/2239 (0.0%)	0.67	0/3063
1	C	0.47	0/2117	0.63	0/2901
1	D	0.48	0/1883	0.66	1/2576 (0.0%)
1	E	0.47	0/1974	0.67	2/2702 (0.1%)
1	F	0.49	0/1976	0.64	0/2704
1	G	0.47	0/2075	0.64	0/2840
1	H	0.48	0/1964	0.67	2/2688 (0.1%)
All	All	0.49	1/16185 (0.0%)	0.67	8/22149 (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
1	A	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	226	GLU	CD-OE1	-6.51	1.18	1.25

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	263	ARG	NE-CZ-NH2	13.14	126.87	120.30
1	A	263	ARG	NE-CZ-NH1	-11.16	114.72	120.30
1	H	256	ASP	CB-CG-OD1	8.15	125.63	118.30
1	E	234	LYS	CD-CE-NZ	7.96	130.02	111.70

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	234	LYS	CD-CE-NZ	7.83	129.71	111.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	256	ASP	Peptide

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	1911	0	0	0	0
1	B	2186	0	0	1	0
1	C	2067	0	0	2	0
1	D	1839	0	0	1	0
1	E	1925	0	0	4	0
1	F	1929	0	0	2	0
1	G	2026	0	0	3	0
1	H	1918	0	0	1	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
2	G	1	0	0	0	0
2	H	1	0	0	0	0
3	A	4	0	6	0	0
4	F	1	0	0	0	0
5	A	60	0	0	0	0
5	B	88	0	0	0	0
5	C	63	0	0	0	0
5	D	44	0	0	1	0
5	E	67	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	F	67	0	0	1	0
5	G	63	0	0	1	0
5	H	57	0	0	1	0
All	All	16323	0	6	13	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 1.

The worst 5 of 13 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:H:263:ARG:NH1	5:H:553:HOH:O	2.21	0.73
1:E:263:ARG:NH1	5:E:564:HOH:O	2.24	0.70
1:C:68:THR:CG2	1:C:69:VAL:O	2.41	0.68
1:F:263:ARG:NH1	5:F:561:HOH:O	2.28	0.66
1:D:133:GLN:NE2	5:D:525:HOH:O	2.41	0.53

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	242/304 (80%)	235 (97%)	7 (3%)	0	100	100
1	B	281/304 (92%)	271 (96%)	10 (4%)	0	100	100
1	C	267/304 (88%)	257 (96%)	10 (4%)	0	100	100
1	D	233/304 (77%)	225 (97%)	8 (3%)	0	100	100
1	E	245/304 (81%)	238 (97%)	7 (3%)	0	100	100
1	F	244/304 (80%)	238 (98%)	6 (2%)	0	100	100
1	G	258/304 (85%)	252 (98%)	6 (2%)	0	100	100
1	H	243/304 (80%)	236 (97%)	7 (3%)	0	100	100
All	All	2013/2432 (83%)	1952 (97%)	61 (3%)	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	195/248 (79%)	192 (98%)	3 (2%)	76	93
1	B	221/248 (89%)	218 (99%)	3 (1%)	78	94
1	C	210/248 (85%)	208 (99%)	2 (1%)	85	96
1	D	186/248 (75%)	184 (99%)	2 (1%)	84	96
1	E	195/248 (79%)	192 (98%)	3 (2%)	76	93
1	F	199/248 (80%)	196 (98%)	3 (2%)	76	93
1	G	207/248 (84%)	205 (99%)	2 (1%)	85	96
1	H	195/248 (79%)	192 (98%)	3 (2%)	76	93
All	All	1608/1984 (81%)	1587 (99%)	21 (1%)	80	95

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	209	THR
1	E	198	HIS
1	G	198	HIS
1	D	198	HIS
1	H	132	GLN

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 ⓘ

Of 10 ligands modelled in this entry, 9 are monoatomic - leaving 1 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	$\# Z > 2$
3	EDO	A	402	-	3,3,3	0.50	0	2,2,2	0.42	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	EDO	A	402	-	-	0/1/1/1	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

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	248/304 (81%)	-0.34	3 (1%) 75 77	15, 23, 45, 58	0
1	B	285/304 (93%)	-0.39	0 100 100	11, 20, 35, 44	0
1	C	271/304 (89%)	-0.19	4 (1%) 70 71	16, 29, 56, 72	0
1	D	239/304 (78%)	-0.31	1 (0%) 90 91	14, 27, 44, 56	0
1	E	250/304 (82%)	-0.30	2 (0%) 83 85	13, 24, 45, 68	0
1	F	250/304 (82%)	-0.43	0 100 100	14, 22, 38, 62	0
1	G	264/304 (86%)	0.07	18 (6%) 17 14	15, 31, 64, 79	0
1	H	249/304 (81%)	-0.40	1 (0%) 90 91	14, 27, 43, 58	0
All	All	2056/2432 (84%)	-0.28	29 (1%) 72 72	11, 25, 47, 79	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	81	LEU	4.8
1	G	75	ALA	4.5
1	G	78	ALA	3.9
1	G	72	PRO	3.6
1	A	91	ALA	3.5

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	EDO	A	402	4/4	0.14	-0.10	27,29,29,30	0
4	MG	F	402	1/1	0.11	-0.90	3,3,3,3	0
2	FE	D	401	1/1	0.07	-1.11	35,35,35,35	0
2	FE	F	401	1/1	0.08	-1.48	38,38,38,38	0
2	FE	C	401	1/1	0.07	-1.50	46,46,46,46	0
2	FE	H	401	1/1	0.06	-1.65	43,43,43,43	0
2	FE	G	401	1/1	0.06	-1.71	56,56,56,56	0
2	FE	E	401	1/1	0.05	-2.33	41,41,41,41	0
2	FE	B	401	1/1	0.08	-2.73	54,54,54,54	0
2	FE	A	401	1/1	0.06	-8.50	33,33,33,33	0

6.5 Other polymers

There are no such residues in this entry.