wwPDB X-ray Structure Validation Summary Report (i)

Jun 15, 2024 – 03:14 PM EDT

PDB ID	:	2RFL
Title	:	Crystal structure of the putative phosphohistidine phosphatase SixA from Agrobacterium tumefaciens
Authors	:	Kim, Y.; Binkowski, T.; Xu, X.; Edwards, A.M.; Savchenko, A.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2007-10-01
Resolution	:	2.35 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
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:

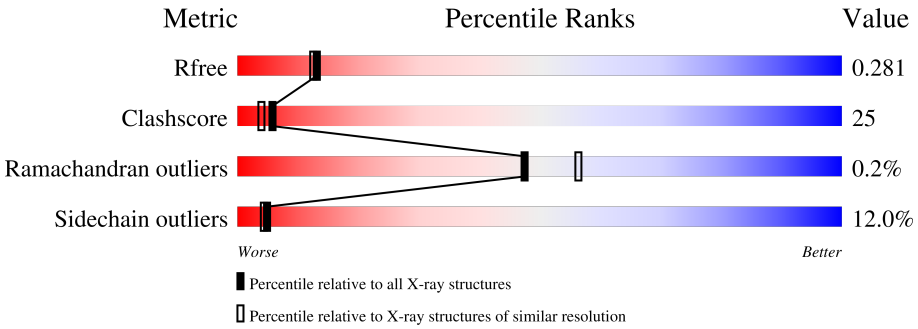
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.35 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)

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%

Mol	Chain	Length	Quality of chain
1	A	173	47% 35% • 14%
1	B	173	54% 29% 5% 11%
1	C	173	44% 36% 6% • 13%
1	D	173	47% 32% 5% 15%
1	E	173	53% 31% 6% 11%
1	F	173	41% 36% 8% 15%
1	G	173	54% 28% 6% • 11%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	H	173	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 10362 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 Putative phosphohistidine phosphatase SixA.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	149	Total	C	N	O	S	Se	0	1	0
			1186	749	210	222	1	4			
1	B	154	Total	C	N	O	S	Se	0	4	0
			1245	784	219	237	1	4			
1	C	151	Total	C	N	O	S	Se	0	4	0
			1226	774	217	230	1	4			
1	D	147	Total	C	N	O	S	Se	0	5	0
			1195	750	209	231	1	4			
1	E	154	Total	C	N	O	S	Se	0	6	0
			1258	792	224	237	1	4			
1	F	147	Total	C	N	O	S	Se	0	3	0
			1178	741	207	225	1	4			
1	G	154	Total	C	N	O	S	Se	0	1	0
			1217	768	217	227	1	4			
1	H	156	Total	C	N	O	S	Se	0	4	0
			1257	791	223	238	1	4			

There are 32 discrepancies between the modelled and reference sequences:

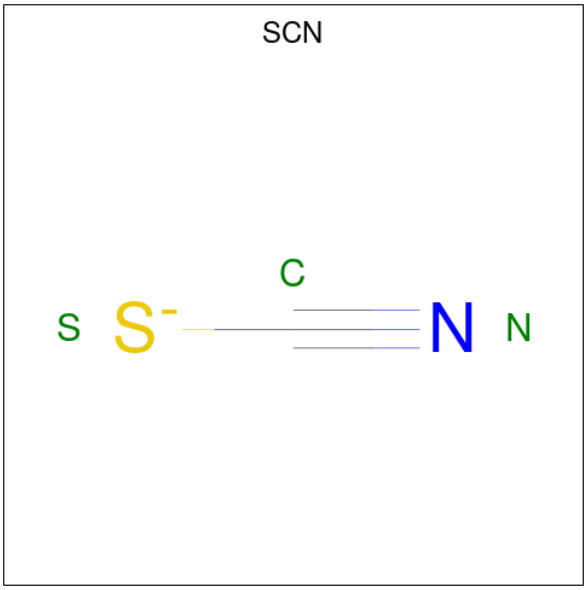
Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	EXPRESSION TAG	UNP Q8UFN9
A	0	HIS	-	EXPRESSION TAG	UNP Q8UFN9
A	170	GLY	-	EXPRESSION TAG	UNP Q8UFN9
A	171	SER	-	EXPRESSION TAG	UNP Q8UFN9
B	-1	GLY	-	EXPRESSION TAG	UNP Q8UFN9
B	0	HIS	-	EXPRESSION TAG	UNP Q8UFN9
B	170	GLY	-	EXPRESSION TAG	UNP Q8UFN9
B	171	SER	-	EXPRESSION TAG	UNP Q8UFN9
C	-1	GLY	-	EXPRESSION TAG	UNP Q8UFN9
C	0	HIS	-	EXPRESSION TAG	UNP Q8UFN9
C	170	GLY	-	EXPRESSION TAG	UNP Q8UFN9
C	171	SER	-	EXPRESSION TAG	UNP Q8UFN9
D	-1	GLY	-	EXPRESSION TAG	UNP Q8UFN9

Continued on next page...

Continued from previous page...

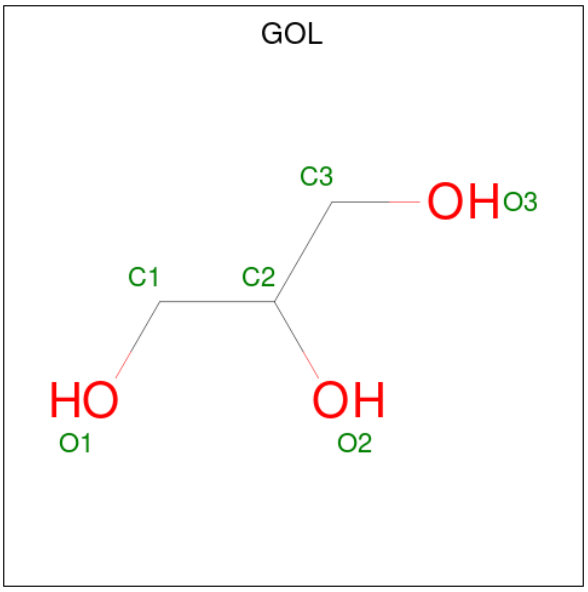
Chain	Residue	Modelled	Actual	Comment	Reference
D	0	HIS	-	EXPRESSION TAG	UNP Q8UFN9
D	170	GLY	-	EXPRESSION TAG	UNP Q8UFN9
D	171	SER	-	EXPRESSION TAG	UNP Q8UFN9
E	-1	GLY	-	EXPRESSION TAG	UNP Q8UFN9
E	0	HIS	-	EXPRESSION TAG	UNP Q8UFN9
E	170	GLY	-	EXPRESSION TAG	UNP Q8UFN9
E	171	SER	-	EXPRESSION TAG	UNP Q8UFN9
F	-1	GLY	-	EXPRESSION TAG	UNP Q8UFN9
F	0	HIS	-	EXPRESSION TAG	UNP Q8UFN9
F	170	GLY	-	EXPRESSION TAG	UNP Q8UFN9
F	171	SER	-	EXPRESSION TAG	UNP Q8UFN9
G	-1	GLY	-	EXPRESSION TAG	UNP Q8UFN9
G	0	HIS	-	EXPRESSION TAG	UNP Q8UFN9
G	170	GLY	-	EXPRESSION TAG	UNP Q8UFN9
G	171	SER	-	EXPRESSION TAG	UNP Q8UFN9
H	-1	GLY	-	EXPRESSION TAG	UNP Q8UFN9
H	0	HIS	-	EXPRESSION TAG	UNP Q8UFN9
H	170	GLY	-	EXPRESSION TAG	UNP Q8UFN9
H	171	SER	-	EXPRESSION TAG	UNP Q8UFN9

- Molecule 2 is THIOCYANATE ION (three-letter code: SCN) (formula: CNS).



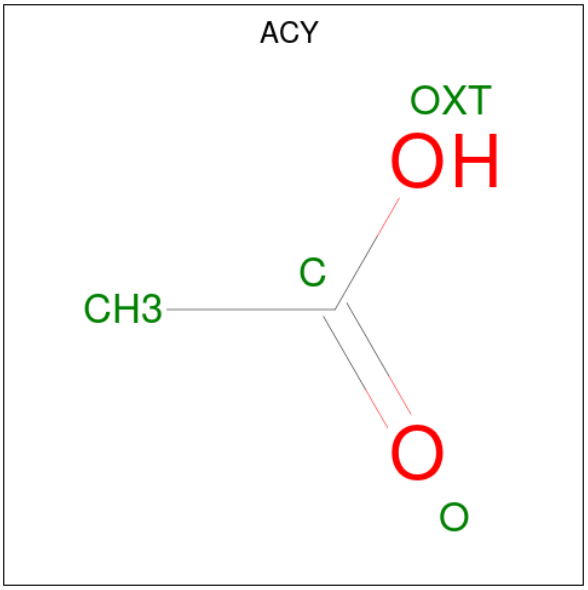
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	S	0	0
			3	1	1	1		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 6 3 3	0	0
3	G	1	Total C O 6 3 3	0	0

- Molecule 4 is ACETIC ACID (three-letter code: ACY) (formula: $C_2H_4O_2$).



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

- Molecule 5 is water.

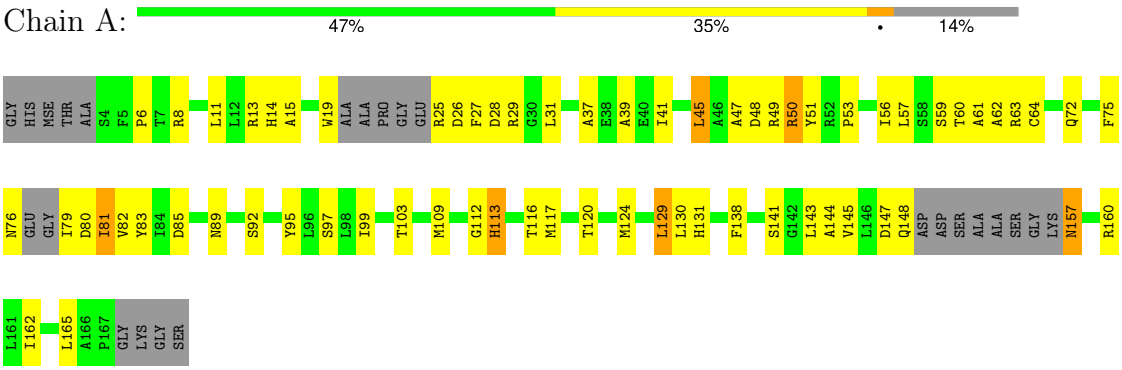
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	81	Total O 81 81	0	0
5	B	71	Total O 71 71	0	0
5	C	75	Total O 75 75	0	0
5	D	68	Total O 68 68	0	0
5	E	77	Total O 77 77	0	0
5	F	63	Total O 63 63	0	0
5	G	66	Total O 66 66	0	0
5	H	80	Total O 80 80	0	0

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. 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. 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.

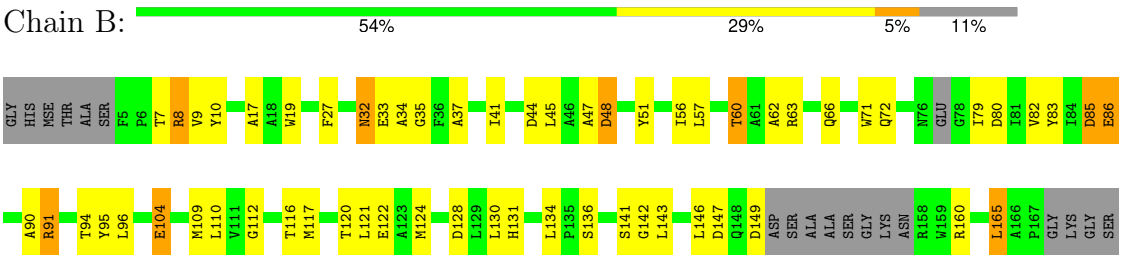
• Molecule 1: Putative phosphohistidine phosphatase SixA

Chain A:



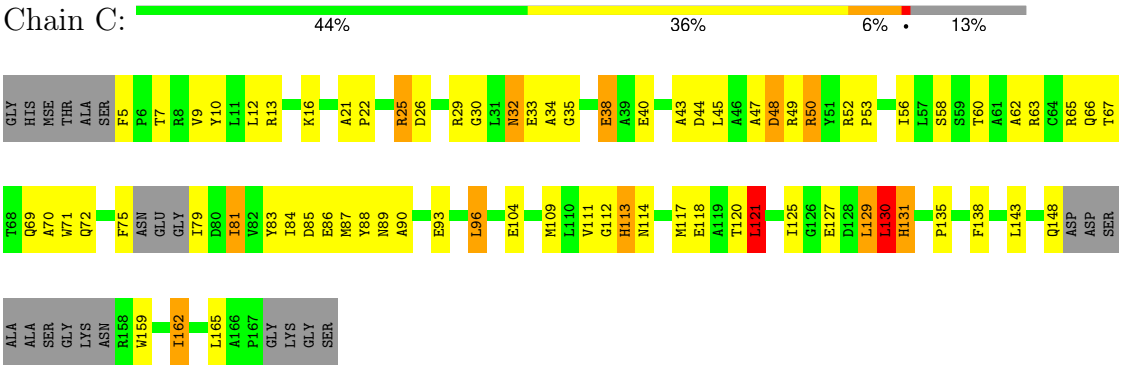
• Molecule 1: Putative phosphohistidine phosphatase SixA

Chain B:



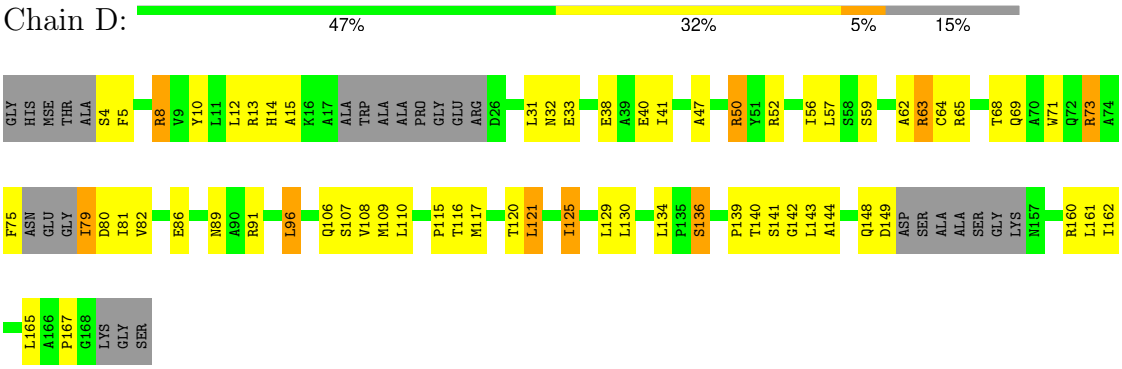
• Molecule 1: Putative phosphohistidine phosphatase SixA

Chain C:



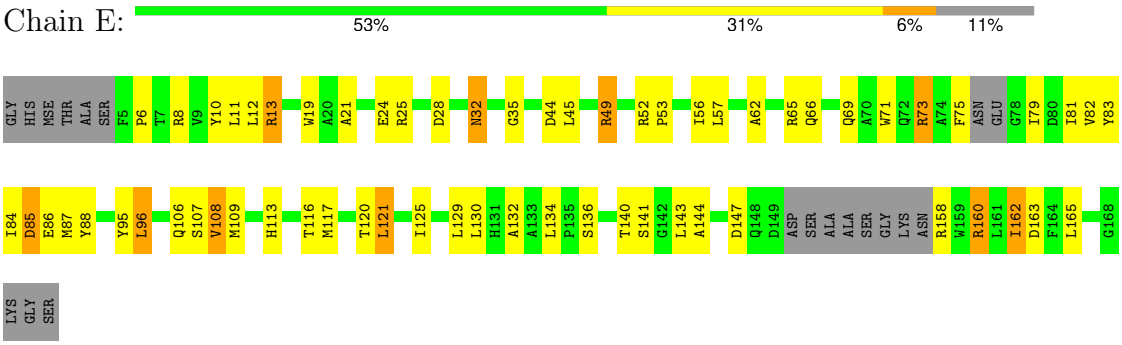
• Molecule 1: Putative phosphohistidine phosphatase SixA

Chain D:



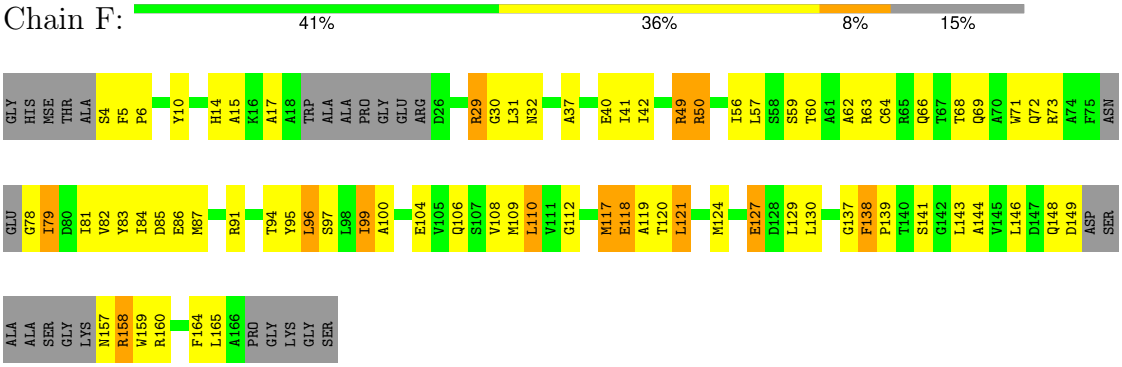
• Molecule 1: Putative phosphohistidine phosphatase SixA

Chain E:



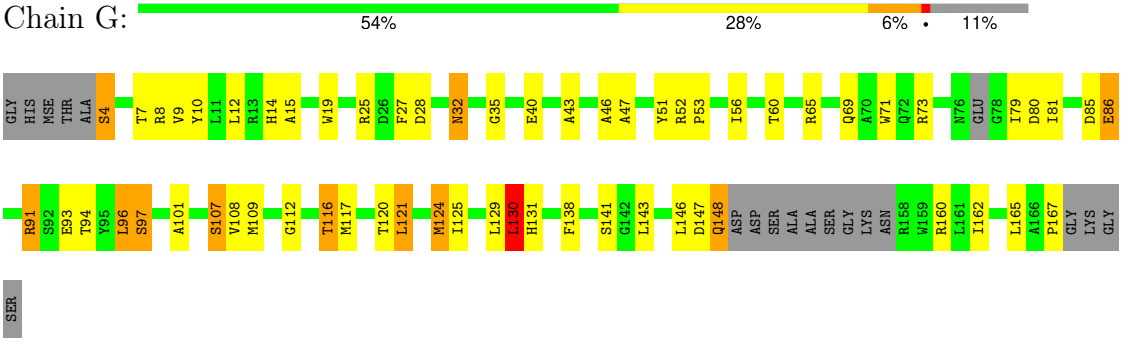
• Molecule 1: Putative phosphohistidine phosphatase SixA

Chain F:



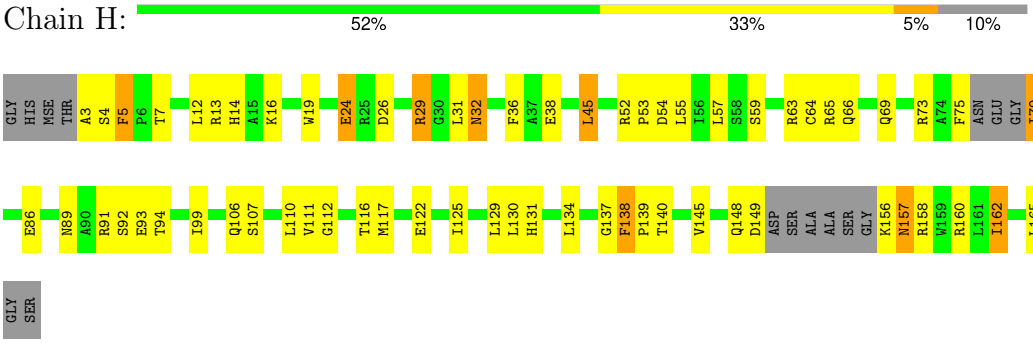
• Molecule 1: Putative phosphohistidine phosphatase SixA

Chain G:



• Molecule 1: Putative phosphohistidine phosphatase SixA

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	58.97Å 201.50Å 72.72Å 90.00° 113.98° 90.00°	Depositor
Resolution (Å)	47.25 – 2.35 47.24 – 2.34	Depositor EDS
% Data completeness (in resolution range)	97.6 (47.25-2.35) 97.3 (47.24-2.34)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	2.02 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.230 , 0.283 0.230 , 0.281	Depositor DCC
R_{free} test set	6354 reflections (10.09%)	wwPDB-VP
Wilson B-factor (Å^2)	40.8	Xtriage
Anisotropy	0.442	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 38.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	0.458 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	10362	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.50% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: SCN, GOL, ACY

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	0.80	0/1204	0.89	1/1626 (0.1%)
1	B	0.75	1/1266 (0.1%)	0.90	1/1712 (0.1%)
1	C	0.78	0/1246	0.88	3/1684 (0.2%)
1	D	0.77	0/1211	0.95	2/1634 (0.1%)
1	E	0.89	0/1277	0.97	3/1725 (0.2%)
1	F	0.76	2/1194 (0.2%)	0.87	2/1611 (0.1%)
1	G	0.86	2/1237 (0.2%)	0.92	1/1672 (0.1%)
1	H	0.94	4/1277 (0.3%)	0.91	2/1726 (0.1%)
All	All	0.82	9/9912 (0.1%)	0.91	15/13390 (0.1%)

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	H	24[A]	GLU	CD-OE1	9.50	1.36	1.25
1	H	24[B]	GLU	CD-OE1	9.50	1.36	1.25
1	H	24[A]	GLU	CD-OE2	8.19	1.34	1.25
1	H	24[B]	GLU	CD-OE2	8.19	1.34	1.25
1	G	86	GLU	CG-CD	6.97	1.62	1.51

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	50	ARG	NE-CZ-NH2	-11.48	114.56	120.30
1	G	130	LEU	CA-CB-CG	8.40	134.62	115.30
1	C	13	ARG	NE-CZ-NH2	-7.13	116.74	120.30
1	F	130	LEU	CA-CB-CG	6.74	130.80	115.30
1	C	121	LEU	CA-CB-CG	6.46	130.16	115.30

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	1186	0	1158	57	0
1	B	1245	0	1198	76	0
1	C	1226	0	1193	69	1
1	D	1195	0	1153	53	0
1	E	1258	0	1223	65	0
1	F	1178	0	1143	72	1
1	G	1217	0	1190	63	0
1	H	1257	0	1224	55	0
2	A	3	0	0	0	0
3	A	6	0	8	0	0
3	G	6	0	8	0	0
4	D	4	0	3	0	0
5	A	81	0	0	7	1
5	B	71	0	0	16	0
5	C	75	0	0	22	0
5	D	68	0	0	15	0
5	E	77	0	0	9	0
5	F	63	0	0	10	1
5	G	66	0	0	14	0
5	H	80	0	0	11	0
All	All	10362	0	9501	490	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 25.

The worst 5 of 490 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:60:THR:HA	5:B:234:HOH:O	1.31	1.24
1:C:7:THR:HG22	1:C:148:GLN:OE1	1.45	1.16
1:B:90:ALA:CB	1:B:94:THR:HG21	1.76	1.15
1:B:90:ALA:HB1	1:B:94:THR:CG2	1.77	1.14
1:D:149:ASP:HB2	5:D:235:HOH:O	1.50	1.11

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-

metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:162:ILE:O	1:F:32:ASN:ND2[2_556]	2.02	0.18
5:A:253:HOH:O	5:F:185:HOH:O[1_454]	2.19	0.01

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	142/173 (82%)	134 (94%)	8 (6%)	0	100	100
1	B	152/173 (88%)	142 (93%)	10 (7%)	0	100	100
1	C	149/173 (86%)	145 (97%)	4 (3%)	0	100	100
1	D	144/173 (83%)	136 (94%)	8 (6%)	0	100	100
1	E	154/173 (89%)	146 (95%)	7 (4%)	1 (1%)	25	27
1	F	142/173 (82%)	139 (98%)	3 (2%)	0	100	100
1	G	149/173 (86%)	140 (94%)	9 (6%)	0	100	100
1	H	154/173 (89%)	147 (96%)	6 (4%)	1 (1%)	25	27
All	All	1186/1384 (86%)	1129 (95%)	55 (5%)	2 (0%)	47	56

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	86	GLU
1	E	132	ALA

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	122/129 (95%)	109 (89%)	13 (11%)	6	6
1	B	126/129 (98%)	113 (90%)	13 (10%)	7	6
1	C	124/129 (96%)	108 (87%)	16 (13%)	4	4
1	D	124/129 (96%)	103 (83%)	21 (17%)	2	1
1	E	127/129 (98%)	115 (91%)	12 (9%)	8	8
1	F	121/129 (94%)	105 (87%)	16 (13%)	4	4
1	G	123/129 (95%)	104 (85%)	19 (15%)	2	3
1	H	128/129 (99%)	115 (90%)	13 (10%)	7	6
All	All	995/1032 (96%)	872 (88%)	123 (12%)	5	4

5 of 123 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	116	THR
1	H	4[A]	SER
1	E	129	LEU
1	G	148	GLN
1	H	116	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	H	131	HIS
1	H	157	ASN
1	F	14	HIS
1	G	14	HIS
1	G	32	ASN

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)

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	$\#\|Z\| > 2$
3	GOL	G	172	-	5,5,5	0.38	0	5,5,5	0.52	0
2	SCN	A	172	-	1,2,2	1.36	0	0,1,1	-	-
3	GOL	A	173	-	5,5,5	0.42	0	5,5,5	0.57	0
4	ACY	D	172	-	3,3,3	0.85	0	3,3,3	1.11	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	GOL	G	172	-	-	2/4/4/4	-
3	GOL	A	173	-	-	4/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	173	GOL	O1-C1-C2-O2
3	A	173	GOL	O1-C1-C2-C3
3	A	173	GOL	C1-C2-C3-O3
3	G	172	GOL	O1-C1-C2-C3
3	G	172	GOL	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

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

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.