wwPDB X-ray Structure Validation Summary Report (i)

May 23, 2020 – 06:04 pm BST

PDB ID	:	5H2T
Title	:	Structure of trehalose synthase
Authors	:	Wang, D.; Huang, H.; Zhou, J.; Jiang, L.
Deposited on	:	2016-10-18
Resolution	:	2.80 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 following versions of software and data (see references (1)) were used in the production of this report:

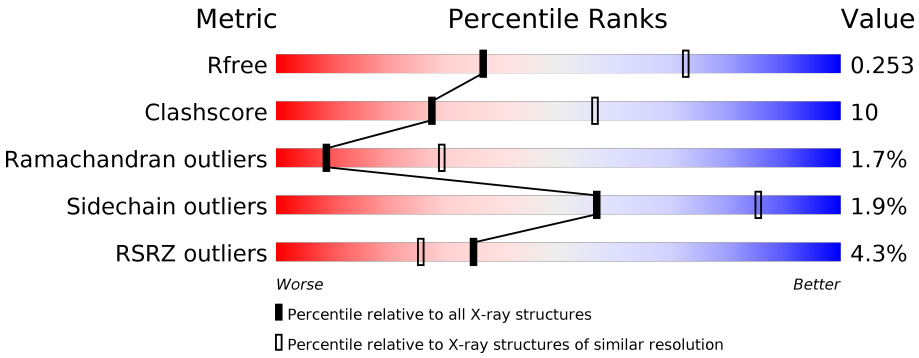
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.80 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.



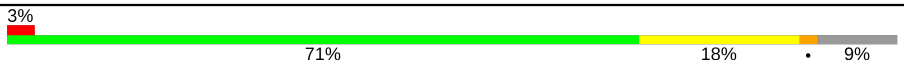
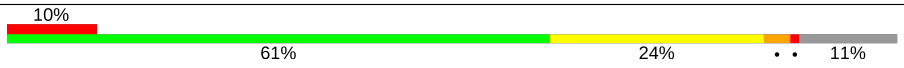
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 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	601	75% 15% 9%
1	B	601	2% 75% 15% 9%
1	C	601	2% 70% 19% 9%
1	D	601	4% 72% 16% 9%
1	E	601	4% 69% 19% 9%
1	F	601	5% 67% 20% 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	601	
1	H	601	

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 35620 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 Trehalose synthase.

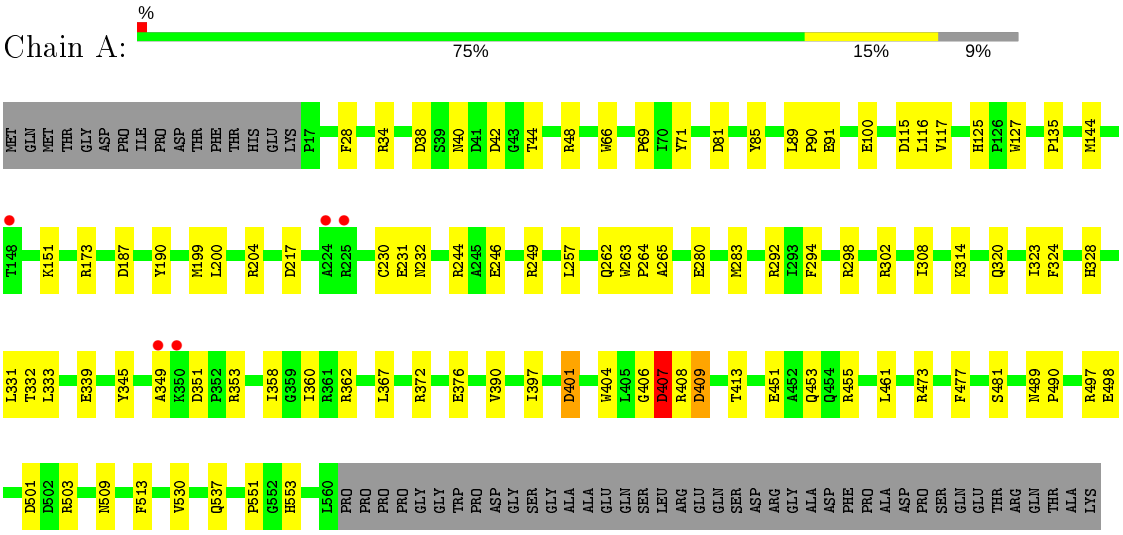
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	544	Total	C	N	O	S	0	0	0
			4462	2858	763	818	23			
1	B	544	Total	C	N	O	S	0	0	0
			4462	2858	763	818	23			
1	C	544	Total	C	N	O	S	0	0	0
			4462	2858	763	818	23			
1	D	544	Total	C	N	O	S	0	0	0
			4462	2858	763	818	23			
1	E	544	Total	C	N	O	S	0	0	0
			4462	2858	763	818	23			
1	F	544	Total	C	N	O	S	0	0	0
			4462	2858	763	818	23			
1	G	544	Total	C	N	O	S	0	0	0
			4462	2858	763	818	23			
1	H	536	Total	C	N	O	S	0	0	0
			4386	2810	748	805	23			

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

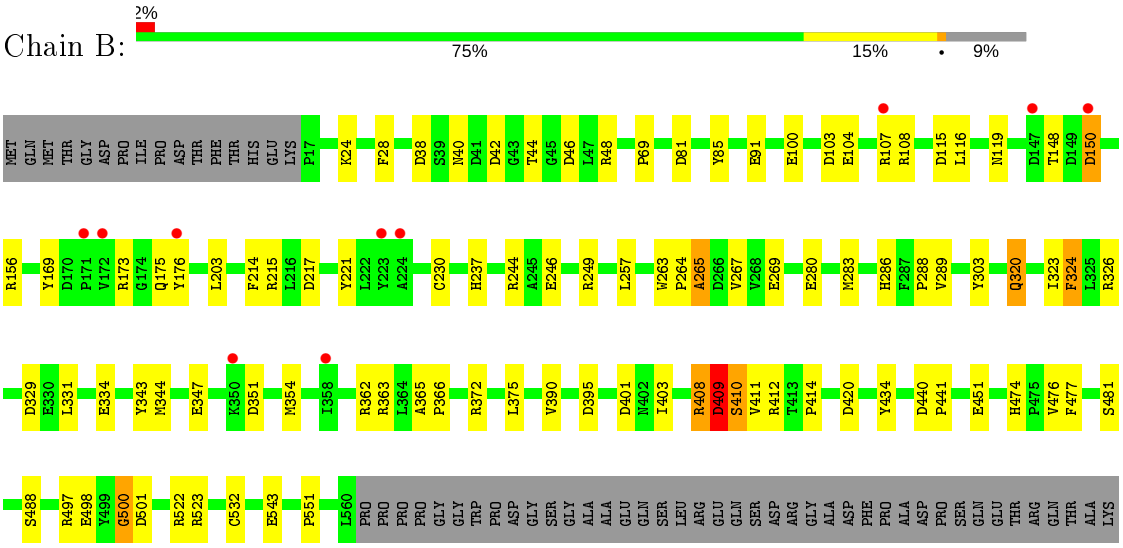
- Molecule 1: Trehalose synthase

Chain A:



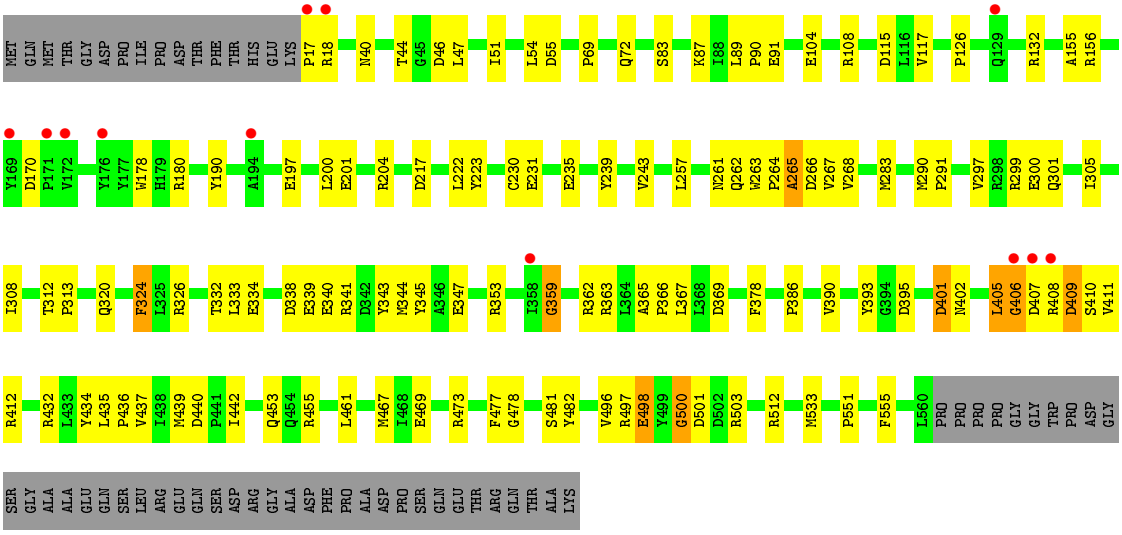
- Molecule 1: Trehalose synthase

Chain B:



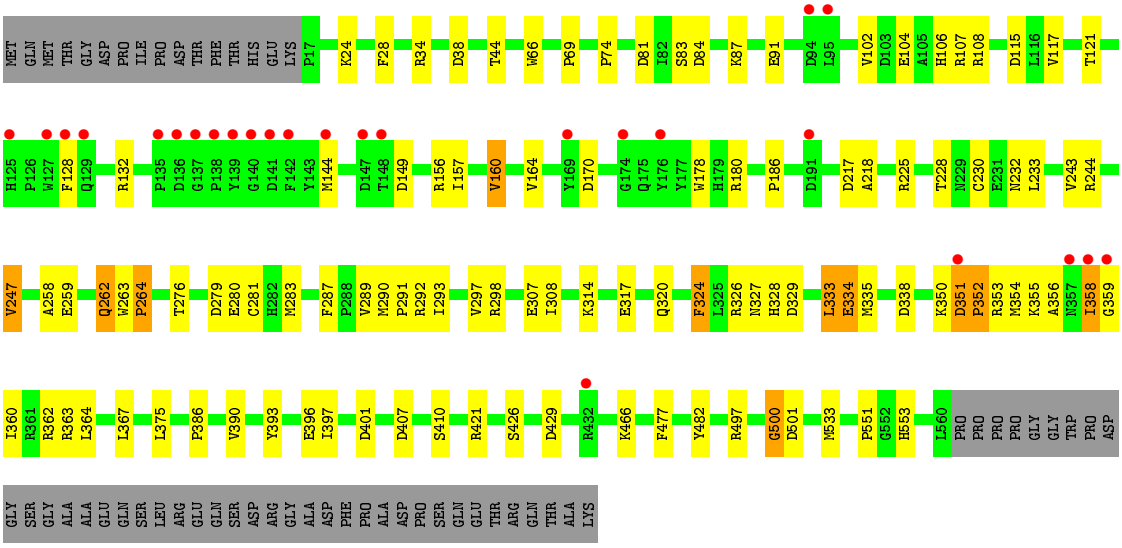
- Molecule 1: Trehalose synthase

Chain C:



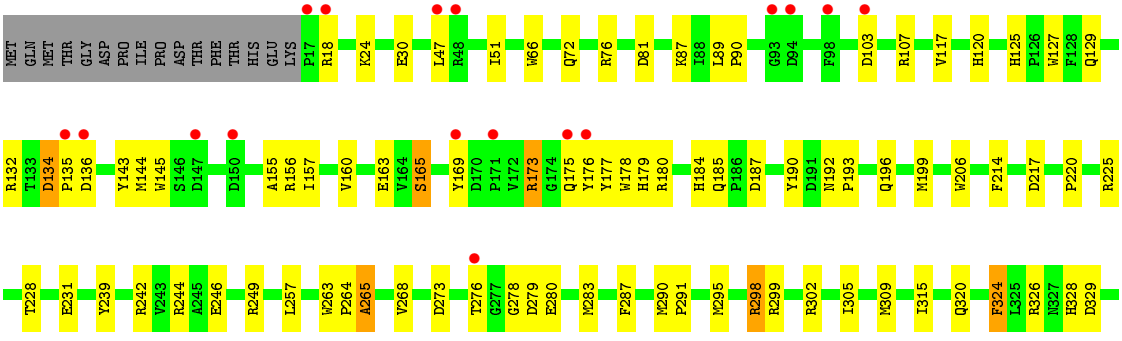
- Molecule 1: Trehalose synthase

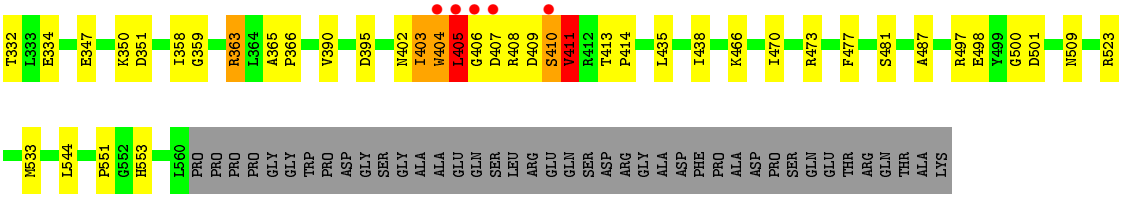
Chain D:



- Molecule 1: Trehalose synthase

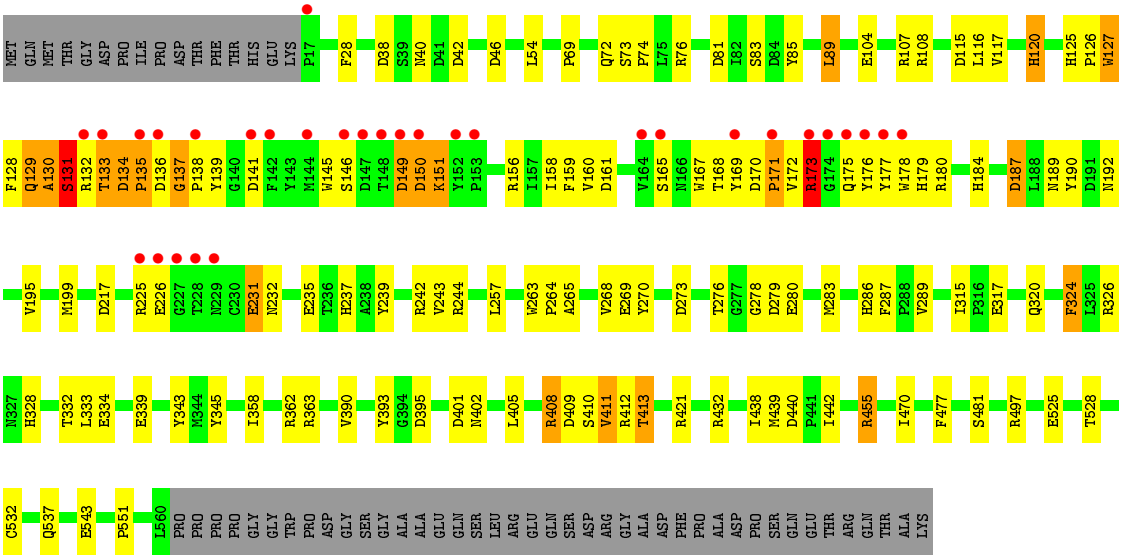
Chain E:





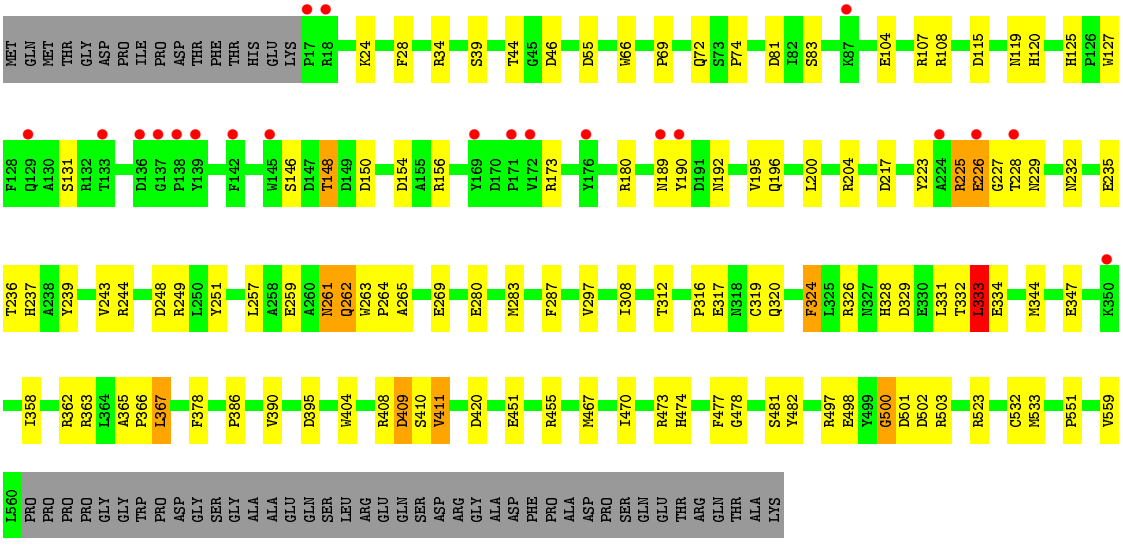
• Molecule 1: Trehalose synthase

Chain F:



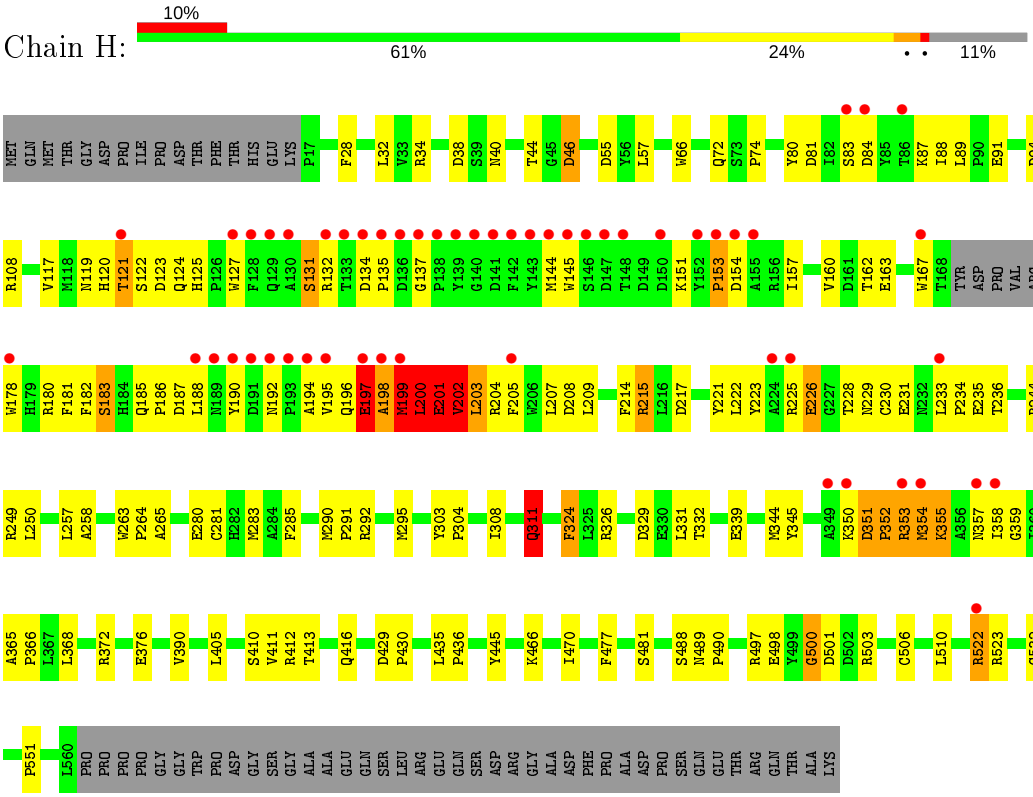
• Molecule 1: Trehalose synthase

Chain G:



• Molecule 1: Trehalose synthase

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	248.76Å 133.12Å 211.78Å 90.00° 112.54° 90.00°	Depositor
Resolution (Å)	43.57 – 2.80 49.18 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.0 (43.57-2.80) 99.2 (49.18-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.73 (at 2.81Å)	Xtriage
Refinement program	PHENIX 1.10.1-2155	Depositor
R, R_{free}	0.205 , 0.252 0.207 , 0.253	Depositor DCC
R_{free} test set	7745 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	46.0	Xtriage
Anisotropy	0.048	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 32.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	35620	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.17% 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

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	0.65	0/4599	0.69	1/6269 (0.0%)
1	B	0.63	1/4599 (0.0%)	0.68	2/6269 (0.0%)
1	C	0.59	0/4599	0.67	1/6269 (0.0%)
1	D	0.59	0/4599	0.69	1/6269 (0.0%)
1	E	0.59	0/4599	0.69	4/6269 (0.1%)
1	F	0.61	0/4599	0.76	6/6269 (0.1%)
1	G	0.59	1/4599 (0.0%)	0.70	4/6269 (0.1%)
1	H	0.61	1/4519 (0.0%)	0.78	6/6159 (0.1%)
All	All	0.61	3/36712 (0.0%)	0.71	25/50042 (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
1	D	0	2
1	E	0	2
1	F	0	2
1	G	0	2
1	H	0	9
All	All	0	18

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	404	TRP	CD2-CE2	7.18	1.50	1.41
1	H	506	CYS	CB-SG	-5.55	1.72	1.81
1	B	532	CYS	CB-SG	-5.33	1.73	1.81

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	H	200	LEU	CA-CB-CG	11.07	140.76	115.30
1	G	333	LEU	CA-CB-CG	9.66	137.52	115.30
1	G	404	TRP	CG-CD2-CE3	7.93	141.03	133.90
1	H	311	GLN	CA-CB-CG	6.89	128.57	113.40
1	F	89	LEU	CA-CB-CG	6.70	130.70	115.30

There are no chirality outliers.

5 of 18 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	407	ASP	Peptide
1	D	351	ASP	Peptide
1	D	352	PRO	Peptide
1	E	165	SER	Peptide
1	E	410	SER	Peptide

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	4462	0	4252	59	0
1	B	4462	0	4252	64	0
1	C	4462	0	4252	81	0
1	D	4462	0	4252	78	0
1	E	4462	0	4252	90	0
1	F	4462	0	4252	114	0
1	G	4462	0	4252	79	0
1	H	4386	0	4178	141	0
All	All	35620	0	33942	692	0

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 10.

The worst 5 of 692 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:D:356:ALA:HB1	1:D:360:ILE:H	0.99	1.13

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:334:GLU:N	1:D:358:ILE:O	1.99	0.94
1:D:356:ALA:HB1	1:D:360:ILE:N	1.83	0.94
1:H:201:GLU:O	1:H:204:ARG:N	2.04	0.91
1:A:125:HIS:HD2	1:A:127:TRP:H	1.20	0.90

There are no symmetry-related clashes.

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	542/601 (90%)	521 (96%)	17 (3%)	4 (1%)	22	53
1	B	542/601 (90%)	522 (96%)	15 (3%)	5 (1%)	17	46
1	C	542/601 (90%)	513 (95%)	21 (4%)	8 (2%)	10	33
1	D	542/601 (90%)	516 (95%)	21 (4%)	5 (1%)	17	46
1	E	542/601 (90%)	515 (95%)	21 (4%)	6 (1%)	14	41
1	F	542/601 (90%)	496 (92%)	26 (5%)	20 (4%)	3	11
1	G	542/601 (90%)	518 (96%)	16 (3%)	8 (2%)	10	33
1	H	532/601 (88%)	491 (92%)	24 (4%)	17 (3%)	4	13
All	All	4326/4808 (90%)	4092 (95%)	161 (4%)	73 (2%)	9	29

5 of 73 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	265	ALA
1	A	407	ASP
1	B	150	ASP
1	D	358	ILE
1	E	403	ILE

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	476/522 (91%)	468 (98%)	8 (2%)	60	87
1	B	476/522 (91%)	472 (99%)	4 (1%)	81	94
1	C	476/522 (91%)	468 (98%)	8 (2%)	60	87
1	D	476/522 (91%)	467 (98%)	9 (2%)	57	85
1	E	476/522 (91%)	469 (98%)	7 (2%)	65	89
1	F	476/522 (91%)	462 (97%)	14 (3%)	42	76
1	G	476/522 (91%)	467 (98%)	9 (2%)	57	85
1	H	468/522 (90%)	455 (97%)	13 (3%)	43	77
All	All	3800/4176 (91%)	3728 (98%)	72 (2%)	57	85

5 of 72 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	405	LEU
1	F	231	GLU
1	H	248	ASP
1	E	411	VAL
1	F	107	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	129	GLN
1	F	166	ASN
1	H	120	HIS
1	E	184	HIS
1	E	416	GLN

5.3.3 RNA

There are no RNA molecules 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

There are no ligands in this entry.

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, 95^{th} 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	544/601 (90%)	-0.31	5 (0%) 84 80	13, 24, 43, 71	0
1	B	544/601 (90%)	-0.25	10 (1%) 68 61	12, 25, 48, 72	0
1	C	544/601 (90%)	-0.22	12 (2%) 62 52	14, 31, 51, 102	0
1	D	544/601 (90%)	0.03	26 (4%) 30 21	15, 34, 54, 71	0
1	E	544/601 (90%)	-0.01	22 (4%) 38 28	15, 34, 68, 94	0
1	F	544/601 (90%)	0.07	31 (5%) 23 15	14, 31, 87, 103	0
1	G	544/601 (90%)	-0.07	21 (3%) 39 29	15, 36, 63, 87	0
1	H	536/601 (89%)	0.36	58 (10%) 5 3	16, 38, 80, 95	0
All	All	4344/4808 (90%)	-0.05	185 (4%) 35 25	12, 30, 65, 103	0

The worst 5 of 185 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	H	349	ALA	14.4
1	F	169	TYR	10.5
1	H	138	PRO	8.3
1	F	228	THR	8.2
1	D	358	ILE	7.8

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.