



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2018 – 03:34 am GMT

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 ⓘ symbol.

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

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : trunk30967
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk30967

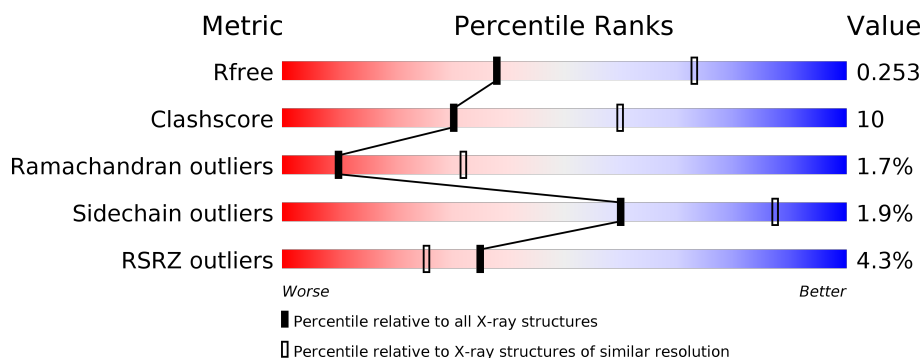
1 Overall quality at a glance

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}	111664	2792 (2.80-2.80)
Clashscore	122126	3209 (2.80-2.80)
Ramachandran outliers	120053	3158 (2.80-2.80)
Sidechain outliers	120020	3160 (2.80-2.80)
RSRZ outliers	108989	2726 (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. 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 $\leq 5\%$. 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	<div> <div>0%</div> <div> <div>75%</div> <div>15%</div> <div>9%</div> </div> </div>
1	B	601	<div> <div>2%</div> <div> <div>75%</div> <div>15%</div> <div>9%</div> </div> </div>
1	C	601	<div> <div>2%</div> <div> <div>70%</div> <div>19%</div> <div>9%</div> </div> </div>
1	D	601	<div> <div>4%</div> <div> <div>72%</div> <div>16%</div> <div>9%</div> </div> </div>
1	E	601	<div> <div>4%</div> <div> <div>69%</div> <div>19%</div> <div>9%</div> </div> </div>
1	F	601	<div> <div>5%</div> <div> <div>67%</div> <div>20%</div> <div>9%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	601	<div><div></div><div>3%</div><div>71%</div><div>18%</div><div>•</div><div>9%</div></div>
1	H	601	<div><div></div><div>10%</div><div>61%</div><div>24%</div><div>• •</div><div>11%</div></div>

2 Entry composition

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.

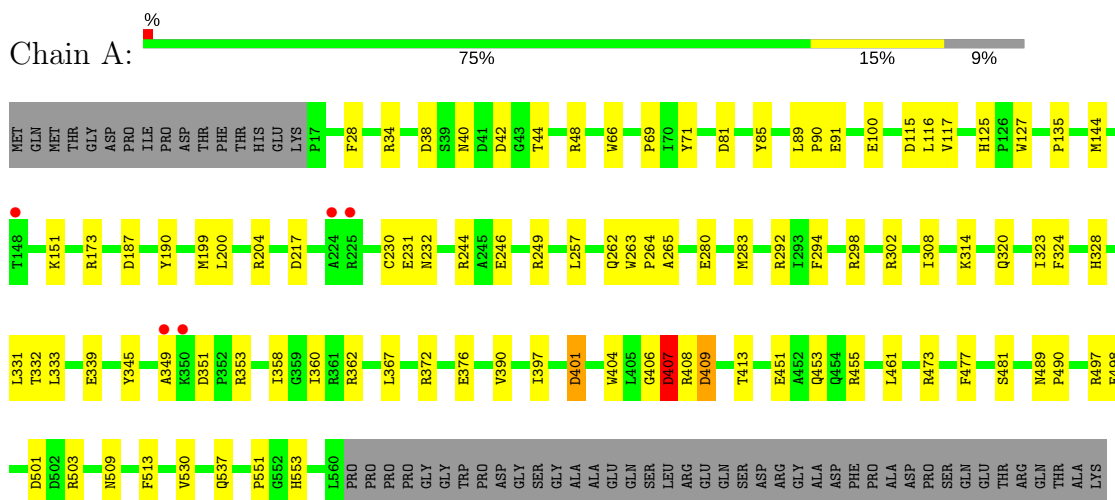
- Molecule 1 is a protein called Trehalose synthase.

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			

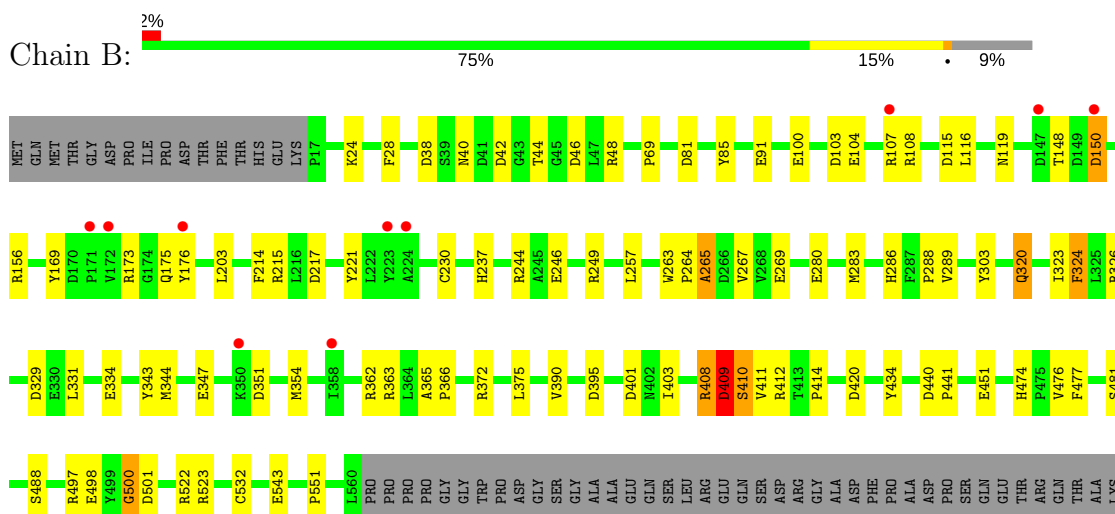
3 Residue-property plots

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

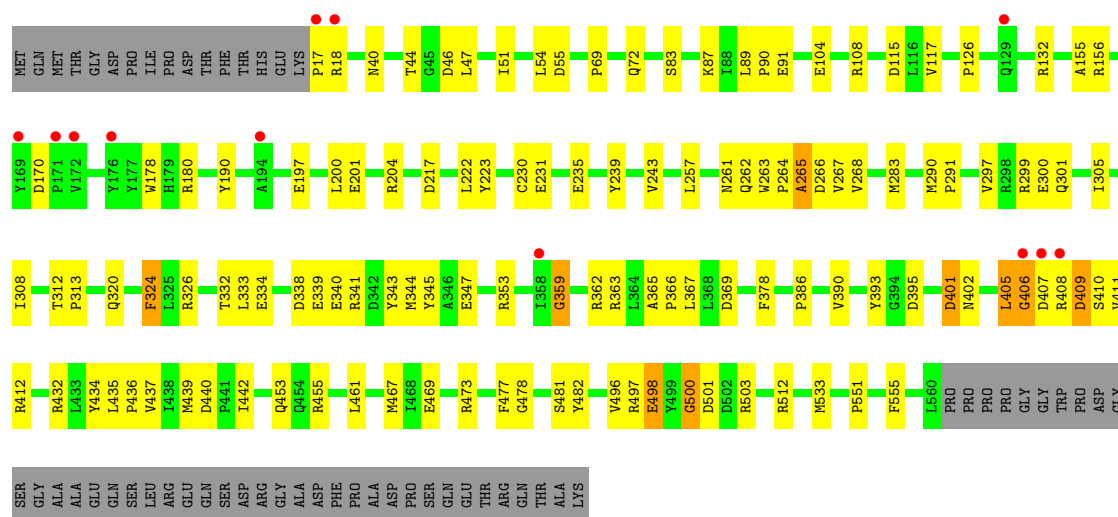


• Molecule 1: Trehalose synthase



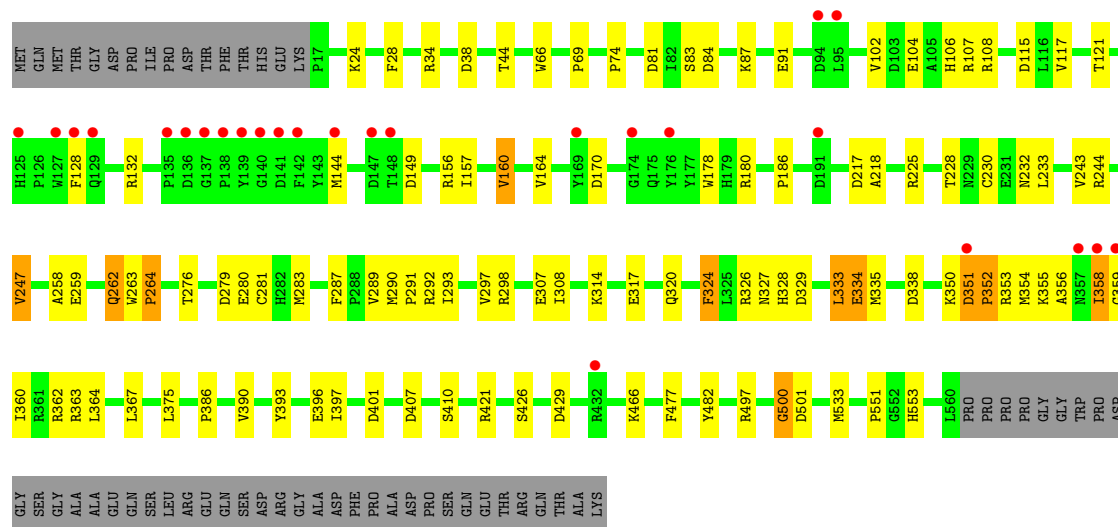
• Molecule 1: Trehalose synthase





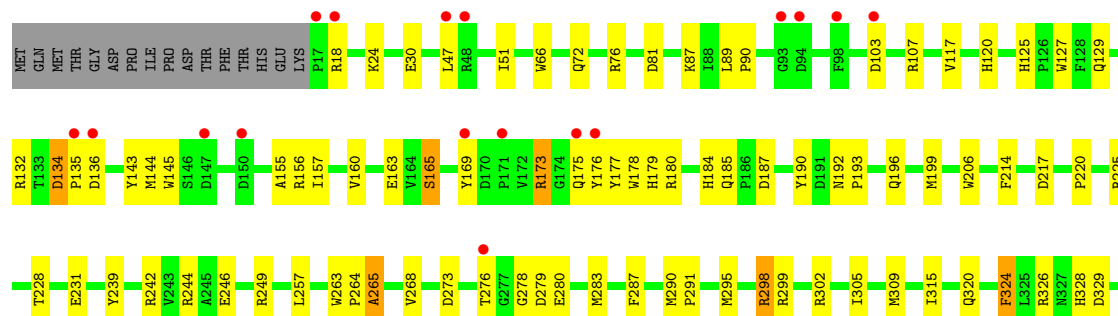
• Molecule 1: Trehalose synthase

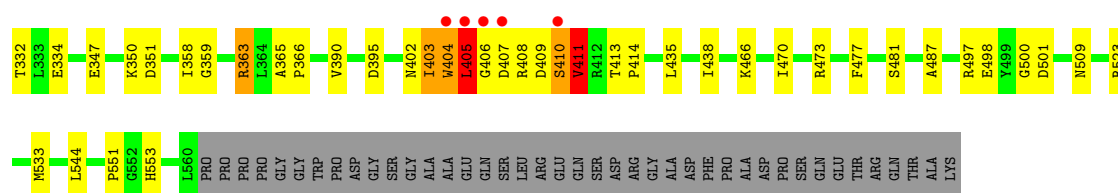
Chain D: 72% 16% 9% 4%



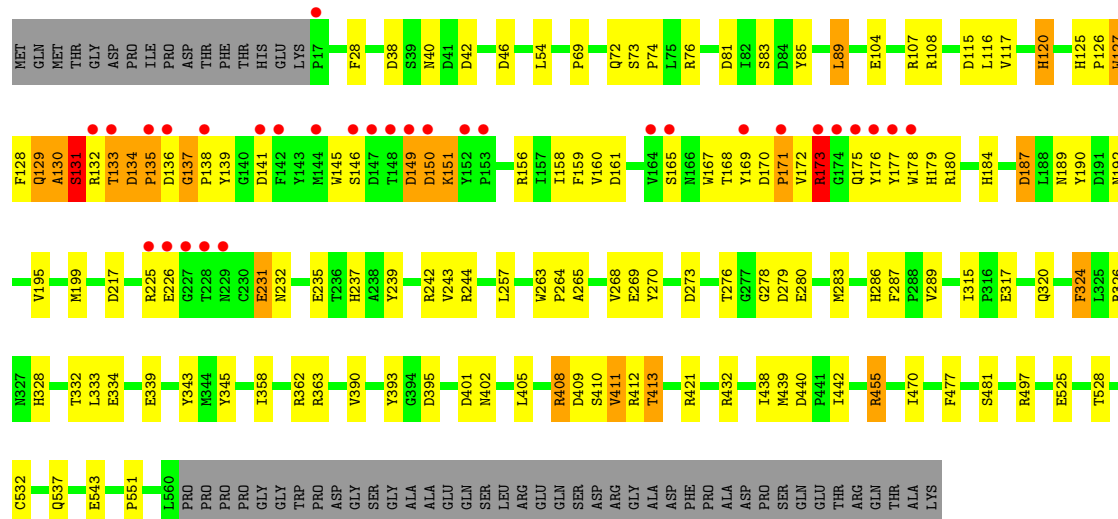
• Molecule 1: Trehalose synthase

Chain E: 69% 19% 9% 4%

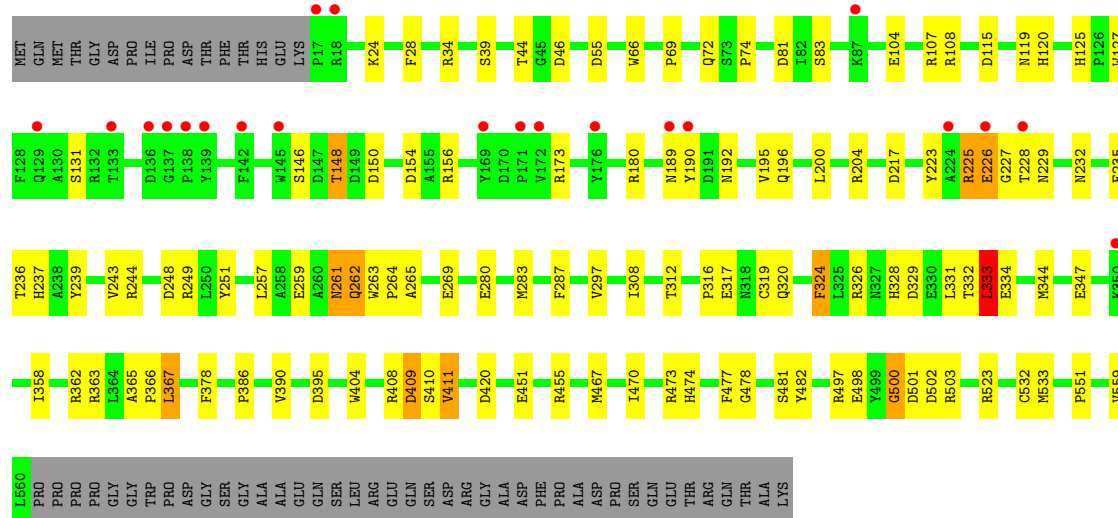




• Molecule 1: Trehalose synthase



• Molecule 1: Trehalose synthase



• Molecule 1: Trehalose synthase



P551	A365	R249	W178	R108	MET
L560	P366	L250	H179	V117	GLN
PRO	L367	L257	R180	M118	MET
PRO	L368	A258	F181	N119	THR
PRO	R372	W263	F182	H120	GLY
PRO	E376	P264	S183	T121	ASP
GLY	V390	A265	H184	S122	PRO
TRP	L405	E280	Q185	D123	ILE
ASP	S410	C281	P186	Q124	PRO
GLY	V411	H282	D187	H125	THR
SER	R412	M283	L188	H126	THR
ALA	T413	A284	M189	F126	PHE
ALA	Q416	F285	Y190	W127	THR
GLU	D429	M290	D191	F128	HIS
GLN	P430	P291	N192	Q129	GLU
SER	L435	R292	P193	A130	LYS
ASP	P436	M296	A194	S131	P17
ARG	Y445	Y303	V195	R132	
GLY	K466	P304	Q196	T133	F28
ALA	I470	I308	E197	D134	L32
ASP	P477	F324	A198	P135	L32
PRO	S481	L325	M199	D136	V33
ALA	S488	R326	L200	G137	R34
SER	M489	E330	E201	F138	D38
GLN	P490	L331	W202	G140	S39
THR	R497	T332	L203	D141	N40
ARG	E498	E339	R204	F142	T44
GLN	Y499	M344	W206	Y143	G45
ALA	G500	Y345	D208	W145	D46
ALA	D501		L209	S146	D55
THR	R503		F214	D147	Y56
THR	C506		R215	T148	L57
LYS	L510		L216	D150	W66
	R522		D217	Y152	Q72
	R523		Y221	P153	S73
	C532		L222	D154	P74
	R542		Y223	A155	Y80
	E543		A224	R156	D81
	L544		R225	I157	I82
			E226	V160	S83
			G227	D161	D84
			T228	T162	Y85
			N229	E163	T86
			C230		K87
			E231	W167	I88
			W322	T168	L89
			L233	TYR	
			P234	ASP	
			E235	PRO	
			T236	VAL	
				ARG	
			R244	GLY	
			A245	GLN	
			E246	TYR	
			V247	E104	
			D248	Y177	

4 Data and refinement statistics

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
$\langle I/\sigma(I) \rangle$ 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	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	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.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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(°)	Ideal(°)
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

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

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	542/601 (90%)	521 (96%)	17 (3%)	4 (1%)	24	57
1	B	542/601 (90%)	522 (96%)	15 (3%)	5 (1%)	19	50
1	C	542/601 (90%)	513 (95%)	21 (4%)	8 (2%)	11	36
1	D	542/601 (90%)	516 (95%)	21 (4%)	5 (1%)	19	50
1	E	542/601 (90%)	515 (95%)	21 (4%)	6 (1%)	16	45
1	F	542/601 (90%)	496 (92%)	26 (5%)	20 (4%)	4	12
1	G	542/601 (90%)	518 (96%)	16 (3%)	8 (2%)	11	36
1	H	532/601 (88%)	491 (92%)	24 (4%)	17 (3%)	4	15
All	All	4326/4808 (90%)	4092 (95%)	161 (4%)	73 (2%)	10	32

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%)	63	89
1	B	476/522 (91%)	472 (99%)	4 (1%)	83	95
1	C	476/522 (91%)	468 (98%)	8 (2%)	63	89
1	D	476/522 (91%)	467 (98%)	9 (2%)	60	87
1	E	476/522 (91%)	469 (98%)	7 (2%)	67	91
1	F	476/522 (91%)	462 (97%)	14 (3%)	45	79
1	G	476/522 (91%)	467 (98%)	9 (2%)	60	87
1	H	468/522 (90%)	455 (97%)	13 (3%)	47	80
All	All	3800/4176 (91%)	3728 (98%)	72 (2%)	60	87

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 [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

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 ⓘ

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	544/601 (90%)	-0.31	5 (0%) 84 79	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%) 24 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%) 6 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 [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.