wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 04:54 am BST

PDB ID	:	1PDW
Title	:	Crystal structure of human DJ-1, P 1 21 1 space group
Authors	:	Tao, X.; Tong, L.
Deposited on	:	2003-05-20
Resolution	:	2.20 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:

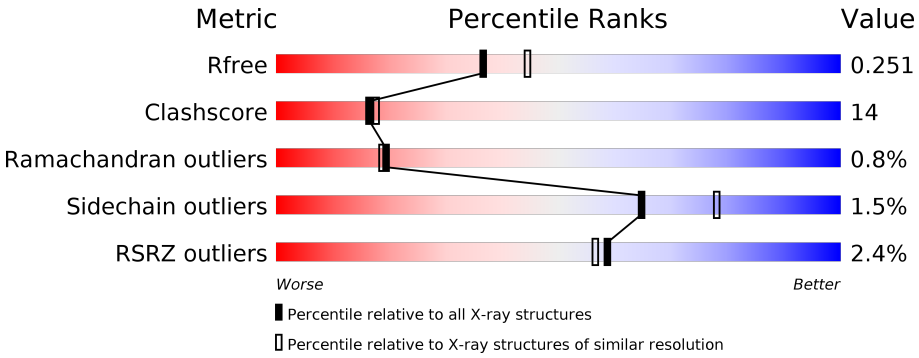
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.20 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.



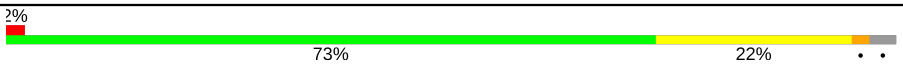
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	197	5%, 66%, 28%, • 5%
1	B	197	3%, 61%, 33%, • 5%
1	C	197	%, 64%, 29%, • 5%
1	D	197	2%, 68%, 25%, • 5%
1	E	197	4%, 67%, 27%, • •
1	F	197	%, 71%, 23%, • 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	197	
1	H	197	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 12021 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 DJ-1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	187	Total	C	N	O	S	Se	0	0	0
			1375	865	240	263	3	4			
1	B	187	Total	C	N	O	S	Se	0	0	0
			1375	865	240	263	3	4			
1	C	187	Total	C	N	O	S	Se	0	0	0
			1375	865	240	263	3	4			
1	D	187	Total	C	N	O	S	Se	0	0	0
			1375	865	240	263	3	4			
1	E	189	Total	C	N	O	S	Se	0	0	0
			1391	875	242	267	3	4			
1	F	187	Total	C	N	O	S	Se	0	0	0
			1375	865	240	263	3	4			
1	G	191	Total	C	N	O	S	Se	0	0	0
			1410	886	246	271	3	4			
1	H	187	Total	C	N	O	S	Se	0	0	0
			1375	865	240	263	3	4			

There are 96 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	17	MSE	MET	MODIFIED RESIDUE	UNP Q99497
A	26	MSE	MET	MODIFIED RESIDUE	UNP Q99497
A	133	MSE	MET	MODIFIED RESIDUE	UNP Q99497
A	134	MSE	MET	MODIFIED RESIDUE	UNP Q99497
A	190	LEU	-	EXPRESSION TAG	UNP Q99497
A	191	GLU	-	EXPRESSION TAG	UNP Q99497
A	192	HIS	-	EXPRESSION TAG	UNP Q99497
A	193	HIS	-	EXPRESSION TAG	UNP Q99497
A	194	HIS	-	EXPRESSION TAG	UNP Q99497
A	195	HIS	-	EXPRESSION TAG	UNP Q99497
A	196	HIS	-	EXPRESSION TAG	UNP Q99497
A	197	HIS	-	EXPRESSION TAG	UNP Q99497
B	217	MSE	MET	MODIFIED RESIDUE	UNP Q99497

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	226	MSE	MET	MODIFIED RESIDUE	UNP Q99497
B	333	MSE	MET	MODIFIED RESIDUE	UNP Q99497
B	334	MSE	MET	MODIFIED RESIDUE	UNP Q99497
B	390	LEU	-	EXPRESSION TAG	UNP Q99497
B	391	GLU	-	EXPRESSION TAG	UNP Q99497
B	392	HIS	-	EXPRESSION TAG	UNP Q99497
B	393	HIS	-	EXPRESSION TAG	UNP Q99497
B	394	HIS	-	EXPRESSION TAG	UNP Q99497
B	395	HIS	-	EXPRESSION TAG	UNP Q99497
B	396	HIS	-	EXPRESSION TAG	UNP Q99497
B	397	HIS	-	EXPRESSION TAG	UNP Q99497
C	1017	MSE	MET	MODIFIED RESIDUE	UNP Q99497
C	1026	MSE	MET	MODIFIED RESIDUE	UNP Q99497
C	1133	MSE	MET	MODIFIED RESIDUE	UNP Q99497
C	1134	MSE	MET	MODIFIED RESIDUE	UNP Q99497
C	1190	LEU	-	EXPRESSION TAG	UNP Q99497
C	1191	GLU	-	EXPRESSION TAG	UNP Q99497
C	1192	HIS	-	EXPRESSION TAG	UNP Q99497
C	1193	HIS	-	EXPRESSION TAG	UNP Q99497
C	1194	HIS	-	EXPRESSION TAG	UNP Q99497
C	1195	HIS	-	EXPRESSION TAG	UNP Q99497
C	1196	HIS	-	EXPRESSION TAG	UNP Q99497
C	1197	HIS	-	EXPRESSION TAG	UNP Q99497
D	1217	MSE	MET	MODIFIED RESIDUE	UNP Q99497
D	1226	MSE	MET	MODIFIED RESIDUE	UNP Q99497
D	1333	MSE	MET	MODIFIED RESIDUE	UNP Q99497
D	1334	MSE	MET	MODIFIED RESIDUE	UNP Q99497
D	1390	LEU	-	EXPRESSION TAG	UNP Q99497
D	1391	GLU	-	EXPRESSION TAG	UNP Q99497
D	1392	HIS	-	EXPRESSION TAG	UNP Q99497
D	1393	HIS	-	EXPRESSION TAG	UNP Q99497
D	1394	HIS	-	EXPRESSION TAG	UNP Q99497
D	1395	HIS	-	EXPRESSION TAG	UNP Q99497
D	1396	HIS	-	EXPRESSION TAG	UNP Q99497
D	1397	HIS	-	EXPRESSION TAG	UNP Q99497
E	2017	MSE	MET	MODIFIED RESIDUE	UNP Q99497
E	2026	MSE	MET	MODIFIED RESIDUE	UNP Q99497
E	2133	MSE	MET	MODIFIED RESIDUE	UNP Q99497
E	2134	MSE	MET	MODIFIED RESIDUE	UNP Q99497
E	2190	LEU	-	EXPRESSION TAG	UNP Q99497
E	2191	GLU	-	EXPRESSION TAG	UNP Q99497
E	2192	HIS	-	EXPRESSION TAG	UNP Q99497

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	2193	HIS	-	EXPRESSION TAG	UNP Q99497
E	2194	HIS	-	EXPRESSION TAG	UNP Q99497
E	2195	HIS	-	EXPRESSION TAG	UNP Q99497
E	2196	HIS	-	EXPRESSION TAG	UNP Q99497
E	2197	HIS	-	EXPRESSION TAG	UNP Q99497
F	2217	MSE	MET	MODIFIED RESIDUE	UNP Q99497
F	2226	MSE	MET	MODIFIED RESIDUE	UNP Q99497
F	2333	MSE	MET	MODIFIED RESIDUE	UNP Q99497
F	2334	MSE	MET	MODIFIED RESIDUE	UNP Q99497
F	2390	LEU	-	EXPRESSION TAG	UNP Q99497
F	2391	GLU	-	EXPRESSION TAG	UNP Q99497
F	2392	HIS	-	EXPRESSION TAG	UNP Q99497
F	2393	HIS	-	EXPRESSION TAG	UNP Q99497
F	2394	HIS	-	EXPRESSION TAG	UNP Q99497
F	2395	HIS	-	EXPRESSION TAG	UNP Q99497
F	2396	HIS	-	EXPRESSION TAG	UNP Q99497
F	2397	HIS	-	EXPRESSION TAG	UNP Q99497
G	3017	MSE	MET	MODIFIED RESIDUE	UNP Q99497
G	3026	MSE	MET	MODIFIED RESIDUE	UNP Q99497
G	3133	MSE	MET	MODIFIED RESIDUE	UNP Q99497
G	3134	MSE	MET	MODIFIED RESIDUE	UNP Q99497
G	3190	LEU	-	EXPRESSION TAG	UNP Q99497
G	3191	GLU	-	EXPRESSION TAG	UNP Q99497
G	3192	HIS	-	EXPRESSION TAG	UNP Q99497
G	3193	HIS	-	EXPRESSION TAG	UNP Q99497
G	3194	HIS	-	EXPRESSION TAG	UNP Q99497
G	3195	HIS	-	EXPRESSION TAG	UNP Q99497
G	3196	HIS	-	EXPRESSION TAG	UNP Q99497
G	3197	HIS	-	EXPRESSION TAG	UNP Q99497
H	3217	MSE	MET	MODIFIED RESIDUE	UNP Q99497
H	3226	MSE	MET	MODIFIED RESIDUE	UNP Q99497
H	3333	MSE	MET	MODIFIED RESIDUE	UNP Q99497
H	3334	MSE	MET	MODIFIED RESIDUE	UNP Q99497
H	3390	LEU	-	EXPRESSION TAG	UNP Q99497
H	3391	GLU	-	EXPRESSION TAG	UNP Q99497
H	3392	HIS	-	EXPRESSION TAG	UNP Q99497
H	3393	HIS	-	EXPRESSION TAG	UNP Q99497
H	3394	HIS	-	EXPRESSION TAG	UNP Q99497
H	3395	HIS	-	EXPRESSION TAG	UNP Q99497
H	3396	HIS	-	EXPRESSION TAG	UNP Q99497
H	3397	HIS	-	EXPRESSION TAG	UNP Q99497

- Molecule 2 is water.

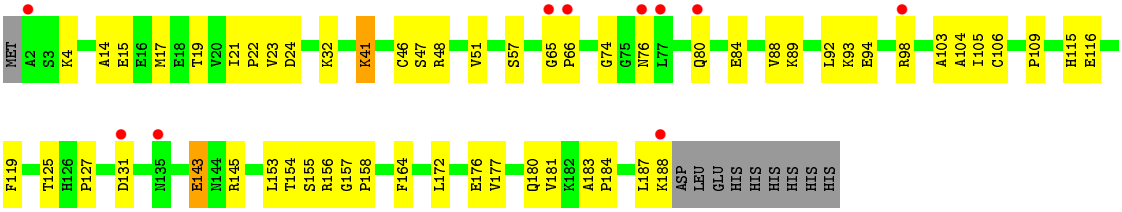
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	76	Total O 76 76	0	0
2	B	95	Total O 95 95	0	0
2	C	124	Total O 124 124	0	0
2	D	107	Total O 107 107	0	0
2	E	104	Total O 104 104	0	0
2	F	141	Total O 141 141	0	0
2	G	151	Total O 151 151	0	0
2	H	172	Total O 172 172	0	0

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.

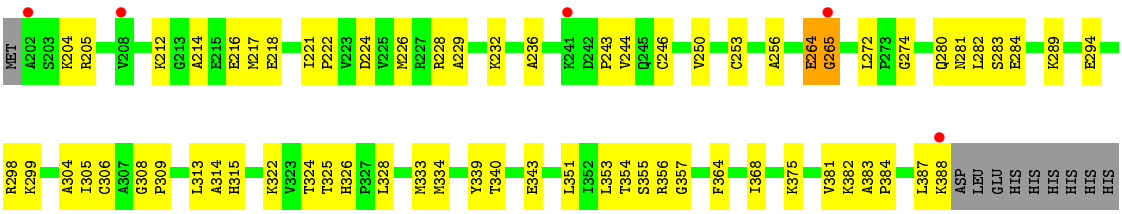
- Molecule 1: DJ-1

Chain A:



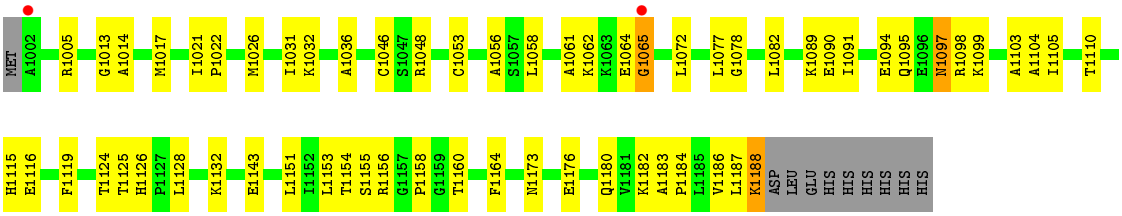
- Molecule 1: DJ-1

Chain B:



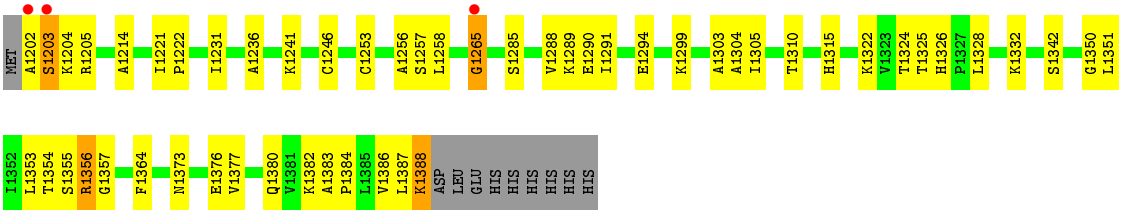
- Molecule 1: DJ-1

Chain C:



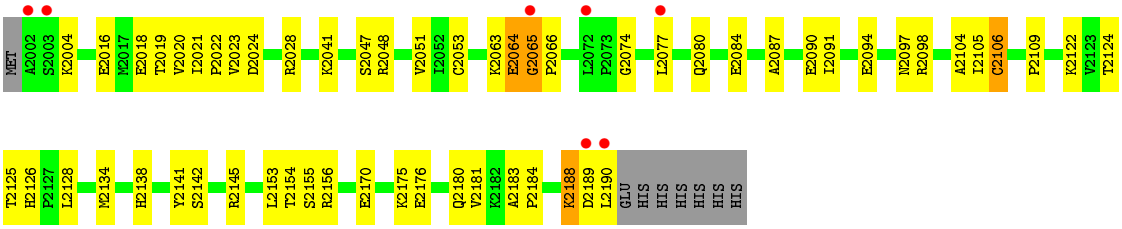
- Molecule 1: DJ-1

Chain D:



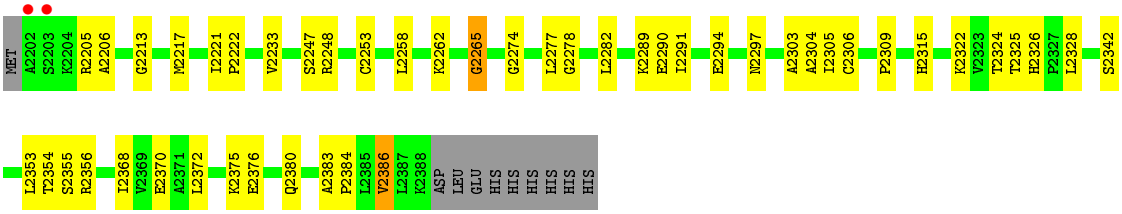
- Molecule 1: DJ-1

Chain E:



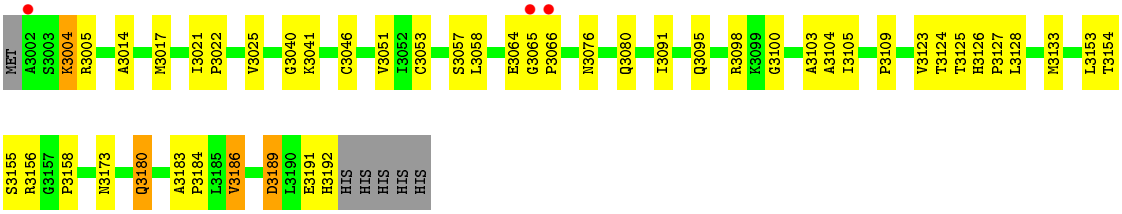
- Molecule 1: DJ-1

Chain F:



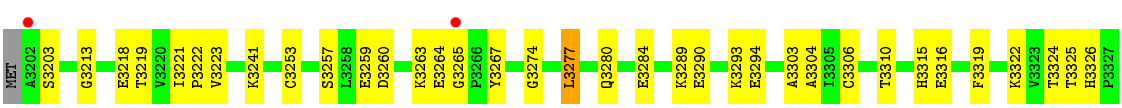
- Molecule 1: DJ-1

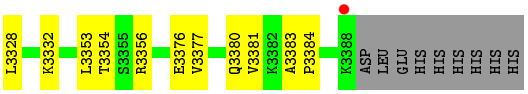
Chain G:



- Molecule 1: DJ-1

Chain H:




L3328
K3332
L3353
T3354
S3355
R3356
E3376
V3377
Q3380
V3381
K3382
A3383
P3384
K3388
ASP
LEU
GLU
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	71.26Å 83.66Å 114.16Å 90.00° 100.56° 90.00°	Depositor
Resolution (Å)	19.91 – 2.20 24.16 – 2.19	Depositor EDS
% Data completeness (in resolution range)	91.1 (19.91-2.20) 95.7 (24.16-2.19)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	5.53 (at 2.19Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.180 , 0.243 0.192 , 0.251	Depositor DCC
R_{free} test set	4986 reflections (7.65%)	wwPDB-VP
Wilson B-factor (Å^2)	14.8	Xtriage
Anisotropy	0.491	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 51.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	12021	wwPDB-VP
Average B, all atoms (Å^2)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.91% 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	0.27	0/1388	0.58	0/1868
1	B	0.27	0/1388	0.58	0/1868
1	C	0.30	0/1388	0.62	0/1868
1	D	0.31	0/1388	0.61	1/1868 (0.1%)
1	E	0.28	0/1404	0.58	0/1890
1	F	0.28	0/1388	0.61	0/1868
1	G	0.30	0/1424	0.60	0/1917
1	H	0.32	0/1388	0.63	0/1868
All	All	0.29	0/11156	0.60	1/15015 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	1356	ARG	N-CA-C	5.09	124.73	111.00

There are no chirality outliers.

There are no planarity outliers.

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	1375	0	1431	44	0
1	B	1375	0	1431	57	0
1	C	1375	0	1431	52	0
1	D	1375	0	1431	38	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	1391	0	1446	49	0
1	F	1375	0	1431	31	0
1	G	1410	0	1459	33	0
1	H	1375	0	1431	34	0
2	A	76	0	0	3	0
2	B	95	0	0	6	0
2	C	124	0	0	3	0
2	D	107	0	0	4	0
2	E	104	0	0	4	0
2	F	141	0	0	2	0
2	G	151	0	0	3	0
2	H	172	0	0	8	0
All	All	12021	0	11491	322	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 14.

The worst 5 of 322 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:313:LEU:HB2	1:B:333:MSE:HE1	1.13	1.07
1:B:313:LEU:HB2	1:B:333:MSE:CE	1.96	0.94
1:H:3324:THR:OG1	1:H:3356:ARG:HD3	1.73	0.88
1:C:1183:ALA:HB3	1:C:1184:PRO:HD3	1.56	0.88
1:C:1124:THR:OG1	1:C:1156:ARG:HD3	1.75	0.87

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	185/197 (94%)	171 (92%)	13 (7%)	1 (0%)	29	31
1	B	185/197 (94%)	177 (96%)	7 (4%)	1 (0%)	29	31
1	C	185/197 (94%)	178 (96%)	6 (3%)	1 (0%)	29	31
1	D	185/197 (94%)	176 (95%)	7 (4%)	2 (1%)	14	12
1	E	187/197 (95%)	180 (96%)	4 (2%)	3 (2%)	9	7
1	F	185/197 (94%)	180 (97%)	3 (2%)	2 (1%)	14	12
1	G	189/197 (96%)	183 (97%)	4 (2%)	2 (1%)	14	12
1	H	185/197 (94%)	178 (96%)	7 (4%)	0	100	100
All	All	1486/1576 (94%)	1423 (96%)	51 (3%)	12 (1%)	19	19

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	1203	SER
1	A	65	GLY
1	C	1065	GLY
1	E	2064	GLU
1	F	2265	GLY

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	145/151 (96%)	142 (98%)	3 (2%)	53	67
1	B	145/151 (96%)	143 (99%)	2 (1%)	67	80
1	C	145/151 (96%)	142 (98%)	3 (2%)	53	67
1	D	145/151 (96%)	144 (99%)	1 (1%)	84	91
1	E	147/151 (97%)	144 (98%)	3 (2%)	55	69
1	F	145/151 (96%)	144 (99%)	1 (1%)	84	91
1	G	149/151 (99%)	146 (98%)	3 (2%)	55	69
1	H	145/151 (96%)	144 (99%)	1 (1%)	84	91

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1166/1208 (96%)	1149 (98%)	17 (2%)	65 78

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	1188	LYS
1	D	1388	LYS
1	G	3004	LYS
1	C	1143	GLU
1	G	3180	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 49 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	1297	ASN
1	D	1380	GLN
1	H	3315	HIS
1	D	1326	HIS
1	E	2081	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

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, 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	183/197 (92%)	0.31	10 (5%)	25 24	10, 24, 43, 53	0
1	B	183/197 (92%)	0.13	5 (2%)	54 52	12, 21, 34, 43	0
1	C	183/197 (92%)	-0.22	2 (1%)	80 79	6, 14, 28, 43	0
1	D	183/197 (92%)	-0.10	3 (1%)	72 70	6, 16, 32, 47	0
1	E	185/197 (93%)	0.15	7 (3%)	40 38	8, 21, 39, 50	0
1	F	183/197 (92%)	-0.16	2 (1%)	80 79	7, 15, 30, 51	0
1	G	187/197 (94%)	-0.22	3 (1%)	72 70	4, 13, 32, 46	0
1	H	183/197 (92%)	-0.29	3 (1%)	72 70	3, 11, 25, 42	0
All	All	1470/1576 (93%)	-0.05	35 (2%)	59 56	3, 17, 35, 53	0

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	2002	ALA	7.1
1	F	2202	ALA	6.9
1	D	1202	ALA	6.4
1	H	3202	ALA	6.4
1	A	2	ALA	5.7

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.