wwPDB X-ray Structure Validation Summary Report (i)

Sep 25, 2023 – 04:18 AM EDT

PDB ID	:	5WAI
Title	:	Crystal Structure of a Suz12-Rbbp4-Jarid2-Aebp2 Heterotetrameric Complex
Authors	:	Chen, S.; Jiao, L.; Liu, X.
Deposited on	:	2017-06-26
Resolution	:	2.90 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:

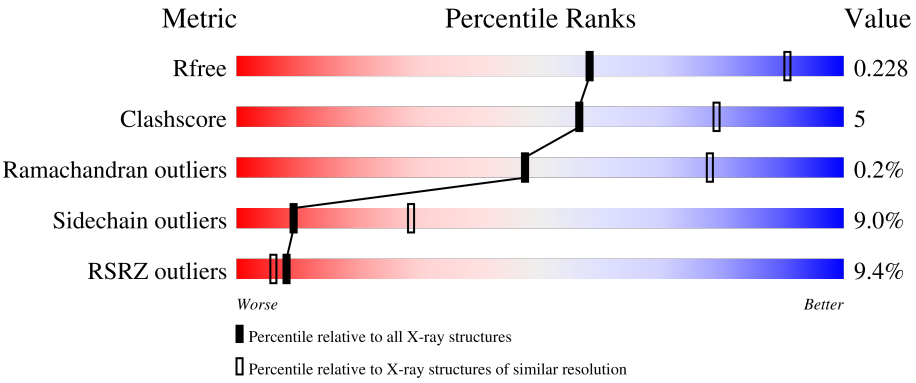
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.90 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.



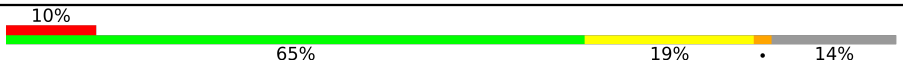
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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% 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	439	3% 74% 15% • 10%
1	E	439	3% 73% 14% • 11%
2	B	478	13% 50% 12% • 37%
2	F	478	9% 49% 11% • 40%
3	C	100	13% 74% 16% • 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	100	
4	D	19	
4	H	19	

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 12890 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 Histone-binding protein RBBP4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	397	Total	C	N	O	S	0	0	0
			3157	1992	536	619	10			
1	E	391	Total	C	N	O	S	0	0	0
			3112	1961	530	611	10			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	MET	-	initiating methionine	UNP Q09028
A	-12	SER	-	expression tag	UNP Q09028
A	-11	HIS	-	expression tag	UNP Q09028
A	-10	HIS	-	expression tag	UNP Q09028
A	-9	HIS	-	expression tag	UNP Q09028
A	-8	HIS	-	expression tag	UNP Q09028
A	-7	HIS	-	expression tag	UNP Q09028
A	-6	HIS	-	expression tag	UNP Q09028
A	-5	LEU	-	expression tag	UNP Q09028
A	-4	VAL	-	expression tag	UNP Q09028
A	-3	PRO	-	expression tag	UNP Q09028
A	-2	ARG	-	expression tag	UNP Q09028
A	-1	GLY	-	expression tag	UNP Q09028
A	0	SER	-	expression tag	UNP Q09028
E	-13	MET	-	initiating methionine	UNP Q09028
E	-12	SER	-	expression tag	UNP Q09028
E	-11	HIS	-	expression tag	UNP Q09028
E	-10	HIS	-	expression tag	UNP Q09028
E	-9	HIS	-	expression tag	UNP Q09028
E	-8	HIS	-	expression tag	UNP Q09028
E	-7	HIS	-	expression tag	UNP Q09028
E	-6	HIS	-	expression tag	UNP Q09028
E	-5	LEU	-	expression tag	UNP Q09028
E	-4	VAL	-	expression tag	UNP Q09028
E	-3	PRO	-	expression tag	UNP Q09028

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	-2	ARG	-	expression tag	UNP Q09028
E	-1	GLY	-	expression tag	UNP Q09028
E	0	SER	-	expression tag	UNP Q09028

- Molecule 2 is a protein called Polycomb protein SUZ12.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	300	Total	C	N	O	S	0	1	0
			2504	1599	456	433	16			
2	F	289	Total	C	N	O	S	0	1	0
			2405	1535	440	415	15			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	546	TRP	-	expression tag	UNP Q15022
B	547	SER	-	expression tag	UNP Q15022
B	548	HIS	-	expression tag	UNP Q15022
B	549	PRO	-	expression tag	UNP Q15022
B	550	GLN	-	expression tag	UNP Q15022
B	551	PHE	-	expression tag	UNP Q15022
B	552	GLU	-	expression tag	UNP Q15022
B	553	LYS	-	expression tag	UNP Q15022
F	546	TRP	-	expression tag	UNP Q15022
F	547	SER	-	expression tag	UNP Q15022
F	548	HIS	-	expression tag	UNP Q15022
F	549	PRO	-	expression tag	UNP Q15022
F	550	GLN	-	expression tag	UNP Q15022
F	551	PHE	-	expression tag	UNP Q15022
F	552	GLU	-	expression tag	UNP Q15022
F	553	LYS	-	expression tag	UNP Q15022

- Molecule 3 is a protein called Zinc finger protein AEBP2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	91	Total	C	N	O	S	0	0	0
			744	478	136	128	2			
3	G	86	Total	C	N	O	S	0	0	0
			706	453	130	121	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	404	SER	-	expression tag	UNP Q6ZN18
C	405	ASN	-	expression tag	UNP Q6ZN18
C	406	ALA	-	expression tag	UNP Q6ZN18
G	404	SER	-	expression tag	UNP Q6ZN18
G	405	ASN	-	expression tag	UNP Q6ZN18
G	406	ALA	-	expression tag	UNP Q6ZN18

- Molecule 4 is a protein called Jumonji, AT-rich interactive domain 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	15	Total	C	N	O	S	0	0	0
			130	85	23	21	1			
4	H	15	Total	C	N	O	S	0	0	0
			130	85	23	21	1			

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

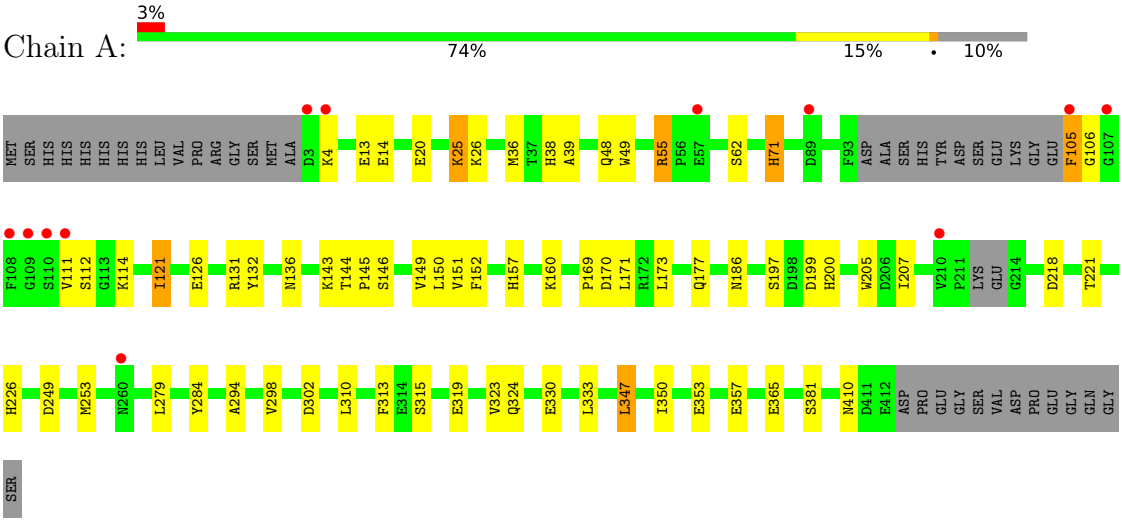
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Zn	0	0
			1	1		
5	F	1	Total	Zn	0	0
			1	1		

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 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.

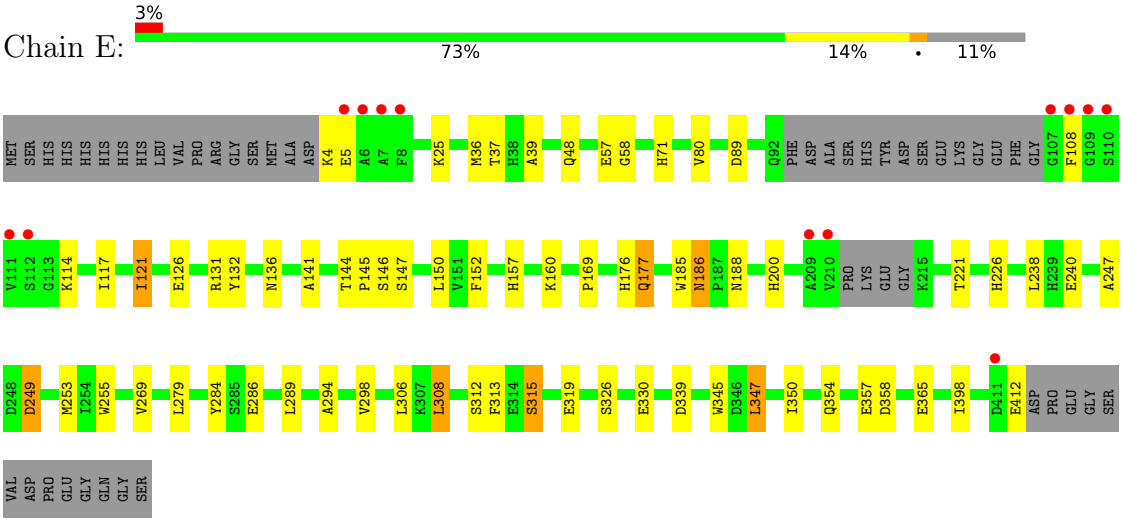
• Molecule 1: Histone-binding protein RBBP4

Chain A:



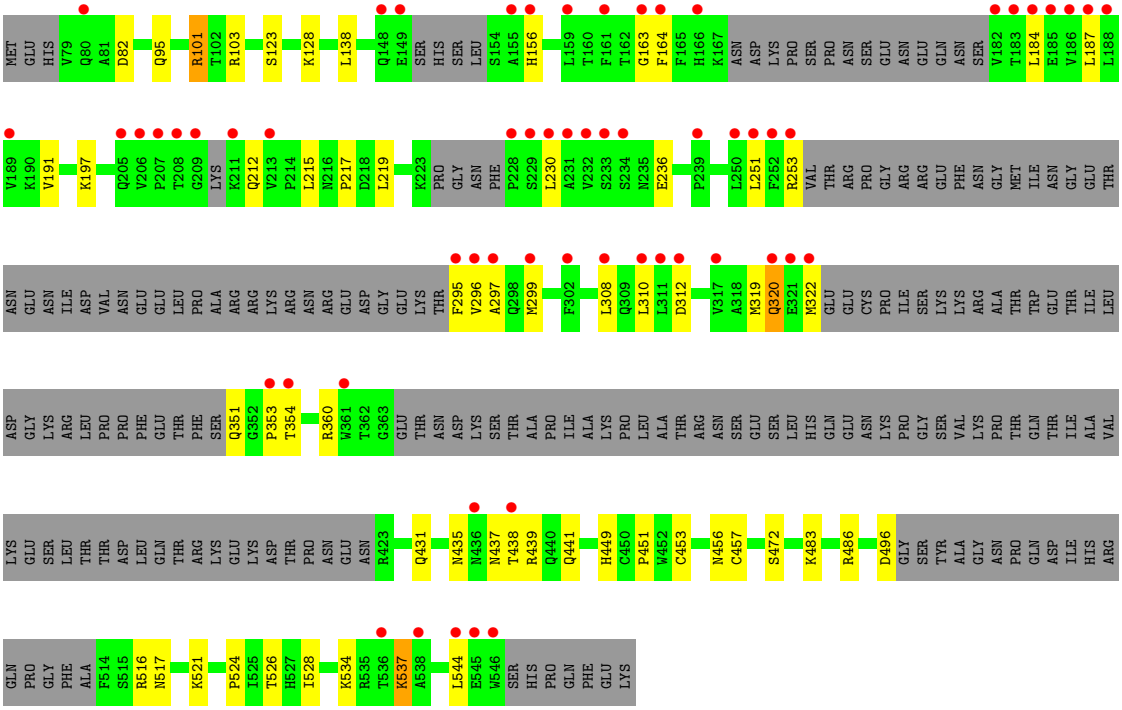
• Molecule 1: Histone-binding protein RBBP4

Chain E:



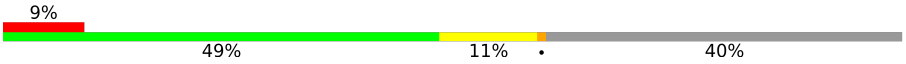
• Molecule 2: Polycomb protein SUZ12

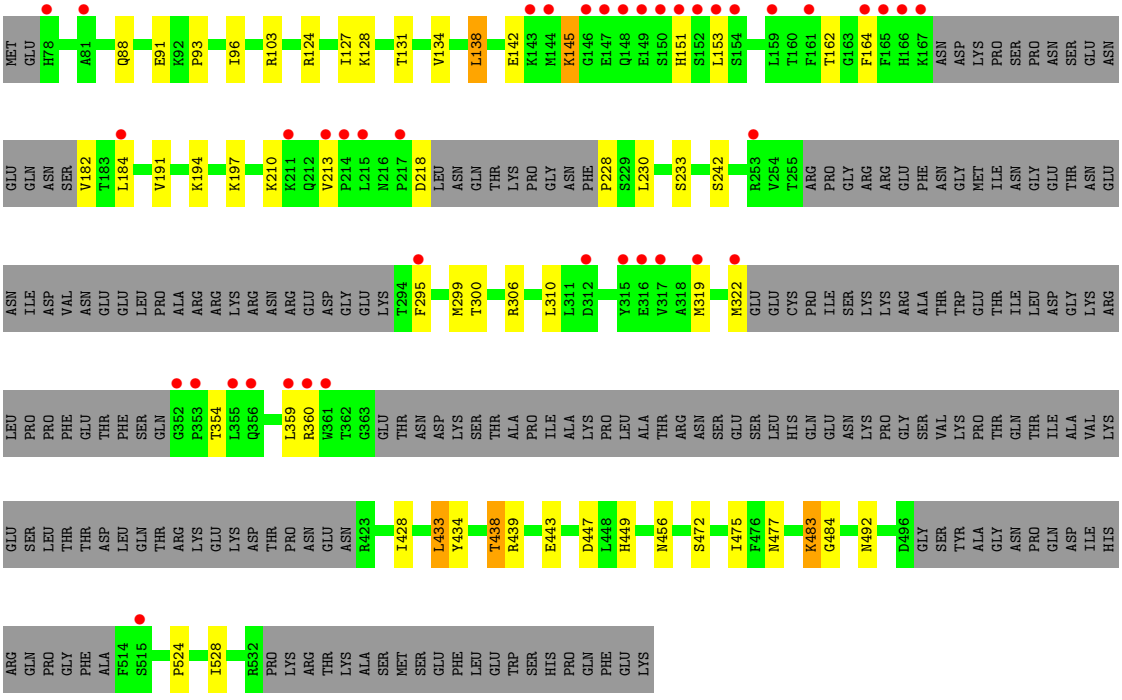
Chain B:



• Molecule 2: Polycomb protein SUZ12

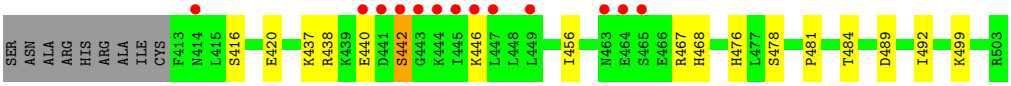
Chain F:





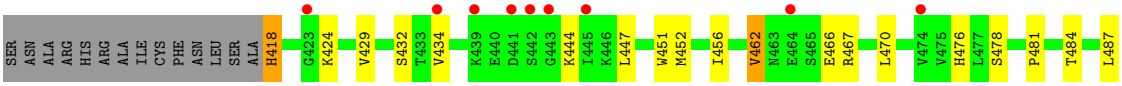
• Molecule 3: Zinc finger protein AEBP2

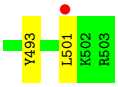
Chain C:



• Molecule 3: Zinc finger protein AEBP2

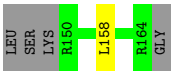
Chain G:





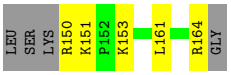
• Molecule 4: Jumonji, AT-rich interactive domain 2

Chain D:



• Molecule 4: Jumonji, AT-rich interactive domain 2

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	95.30Å 111.43Å 253.25Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.10 – 2.90 48.10 – 2.90	Depositor EDS
% Data completeness (in resolution range)	94.1 (48.10-2.90) 94.4 (48.10-2.90)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.30 (at 2.91Å)	Xtriage
Refinement program	BUSTER 2.10.2	Depositor
R, R_{free}	0.171 , 0.216 0.181 , 0.228	Depositor DCC
R_{free} test set	2867 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	52.5	Xtriage
Anisotropy	0.104	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 69.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	12890	wwPDB-VP
Average B, all atoms (Å^2)	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.89% 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: ZN

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	0.54	0/3243	0.76	0/4418
1	E	0.49	0/3195	0.75	0/4353
2	B	0.52	0/2559	0.74	1/3435 (0.0%)
2	F	0.50	0/2460	0.71	0/3307
3	C	0.47	0/761	0.72	0/1026
3	G	0.45	0/722	0.71	0/973
4	D	0.52	0/132	0.64	0/175
4	H	0.49	0/132	0.55	0/175
All	All	0.51	0/13204	0.73	1/17862 (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})
2	B	457	CYS	C-N-CA	-5.42	108.16	121.70

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	3157	0	2989	38	0
1	E	3112	0	2954	35	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	2504	0	2533	29	0
2	F	2405	0	2434	20	0
3	C	744	0	781	8	0
3	G	706	0	745	10	0
4	D	130	0	138	1	0
4	H	130	0	138	1	0
5	B	1	0	0	0	0
5	F	1	0	0	0	0
All	All	12890	0	12712	128	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 5.

The worst 5 of 128 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)
2:B:295:PHE:HE2	2:B:353:PRO:CG	1.52	1.23
2:B:295:PHE:CE2	2:B:353:PRO:HG2	1.79	1.16
2:B:295:PHE:HE2	2:B:353:PRO:HG2	1.14	1.02
2:B:295:PHE:CE2	2:B:353:PRO:CG	2.40	0.99
2:B:295:PHE:HA	2:B:320:GLN:O	1.68	0.92

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	391/439 (89%)	372 (95%)	18 (5%)	1 (0%)	41 71
1	E	385/439 (88%)	372 (97%)	12 (3%)	1 (0%)	41 71
2	B	283/478 (59%)	263 (93%)	20 (7%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	F	276/478 (58%)	259 (94%)	16 (6%)	1 (0%)	34 66
3	C	89/100 (89%)	85 (96%)	4 (4%)	0	100 100
3	G	84/100 (84%)	78 (93%)	6 (7%)	0	100 100
4	D	13/19 (68%)	11 (85%)	2 (15%)	0	100 100
4	H	13/19 (68%)	13 (100%)	0	0	100 100
All	All	1534/2072 (74%)	1453 (95%)	78 (5%)	3 (0%)	47 78

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	315	SER
2	F	153	LEU
1	A	315	SER

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	353/388 (91%)	328 (93%)	25 (7%)	14 40
1	E	349/388 (90%)	324 (93%)	25 (7%)	14 39
2	B	283/442 (64%)	251 (89%)	32 (11%)	6 18
2	F	273/442 (62%)	243 (89%)	30 (11%)	6 19
3	C	85/92 (92%)	78 (92%)	7 (8%)	11 32
3	G	81/92 (88%)	74 (91%)	7 (9%)	10 30
4	D	15/18 (83%)	15 (100%)	0	100 100
4	H	15/18 (83%)	11 (73%)	4 (27%)	0 1
All	All	1454/1880 (77%)	1324 (91%)	130 (9%)	9 29

5 of 130 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	438	THR
3	G	418	HIS
2	B	438	THR
2	B	437	ASN
3	G	444	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	83	HIS
2	F	216	ASN
2	F	309	GLN
1	A	354	GLN
1	A	328	HIS

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)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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.

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

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	397/439 (90%)	-0.04	12 (3%) 50 45	16, 44, 104, 135	0
1	E	391/439 (89%)	-0.01	13 (3%) 46 41	22, 56, 112, 146	0
2	B	300/478 (62%)	0.83	60 (20%) 1 0	18, 77, 135, 160	0
2	F	289/478 (60%)	0.52	41 (14%) 2 2	27, 73, 142, 162	0
3	C	91/100 (91%)	0.55	13 (14%) 2 2	37, 84, 137, 145	0
3	G	86/100 (86%)	0.55	10 (11%) 4 3	41, 88, 139, 147	0
4	D	15/19 (78%)	-0.12	0 100 100	33, 41, 78, 90	0
4	H	15/19 (78%)	-0.14	0 100 100	37, 48, 77, 109	0
All	All	1584/2072 (76%)	0.30	149 (9%) 8 6	16, 60, 127, 162	0

The worst 5 of 149 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	208	THR	6.2
1	A	110	SER	6.1
2	B	186	VAL	6.1
1	A	109	GLY	6.0
2	B	183	THR	5.9

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 monosaccharides 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. The B-factors column lists the minimum, median, 95^{th} 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	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	ZN	B	1001	1/1	0.99	0.14	30,30,30,30	0
5	ZN	F	1001	1/1	1.00	0.12	39,39,39,39	0

6.5 Other polymers

There are no such residues in this entry.