wwPDB X-ray Structure Validation Summary Report (i)

Jan 24, 2021 – 12:33 PM EST

PDB ID	:	2Q5E
Title	:	Crystal structure of human carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 2
Authors	:	Bonanno, J.B.; Dickey, M.; Bain, K.T.; Lau, C.; Romero, R.; Smith, D.; Wasserman, S.; Sauder, J.M.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2007-05-31
Resolution	:	2.51 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:

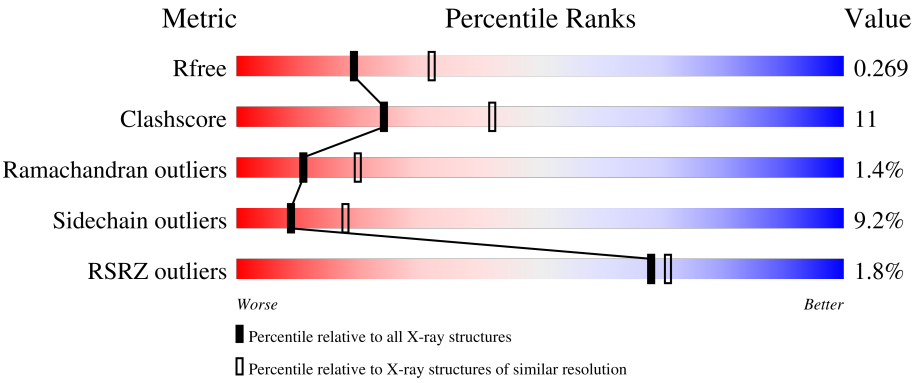
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.16
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.16

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 2.51 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.



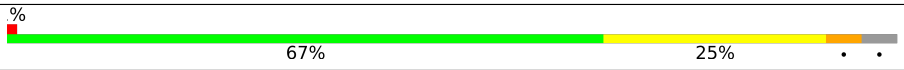
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	187	2% 71% 21%
1	B	187	2% 71% 20% 5%
1	C	187	2% 70% 20% 5%
1	D	187	2% 72% 18% 6%
1	E	187	% 62% 30%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	187	
1	G	187	
1	H	187	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 11566 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 Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	181	Total	C	N	O	S	0	0	0
			1454	930	241	276	7			
1	B	178	Total	C	N	O	S	0	0	0
			1433	919	236	271	7			
1	C	178	Total	C	N	O	S	0	0	0
			1433	919	236	271	7			
1	D	175	Total	C	N	O	S	0	0	0
			1405	902	229	267	7			
1	E	179	Total	C	N	O	S	0	0	0
			1436	918	236	274	8			
1	F	178	Total	C	N	O	S	0	0	0
			1428	916	233	271	8			
1	G	179	Total	C	N	O	S	0	0	0
			1439	922	237	272	8			
1	H	179	Total	C	N	O	S	0	0	0
			1439	922	237	272	8			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	85	SER	-	cloning artifact	UNP O14595
A	86	LEU	-	cloning artifact	UNP O14595
B	85	SER	-	cloning artifact	UNP O14595
B	86	LEU	-	cloning artifact	UNP O14595
C	85	SER	-	cloning artifact	UNP O14595
C	86	LEU	-	cloning artifact	UNP O14595
D	85	SER	-	cloning artifact	UNP O14595
D	86	LEU	-	cloning artifact	UNP O14595
E	85	SER	-	cloning artifact	UNP O14595
E	86	LEU	-	cloning artifact	UNP O14595
F	85	SER	-	cloning artifact	UNP O14595
F	86	LEU	-	cloning artifact	UNP O14595

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	85	SER	-	cloning artifact	UNP O14595
G	86	LEU	-	cloning artifact	UNP O14595
H	85	SER	-	cloning artifact	UNP O14595
H	86	LEU	-	cloning artifact	UNP O14595

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	G	1	Total Mg 1 1	0	0
2	D	1	Total Mg 1 1	0	0
2	E	1	Total Mg 1 1	0	0
2	H	1	Total Mg 1 1	0	0
2	B	1	Total Mg 1 1	0	0
2	C	1	Total Mg 1 1	0	0
2	A	1	Total Mg 1 1	0	0
2	F	1	Total Mg 1 1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	11	Total O 11 11	0	0
3	B	12	Total O 12 12	0	0
3	C	11	Total O 11 11	0	0
3	D	11	Total O 11 11	0	0
3	E	11	Total O 11 11	0	0
3	F	7	Total O 7 7	0	0
3	G	11	Total O 11 11	0	0

Continued on next page...

Continued from previous page...

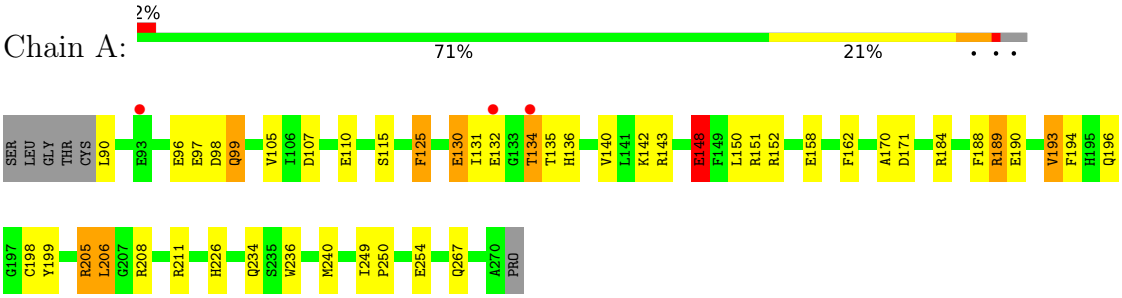
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	H	17	Total	O	0	0
			17	17		

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.

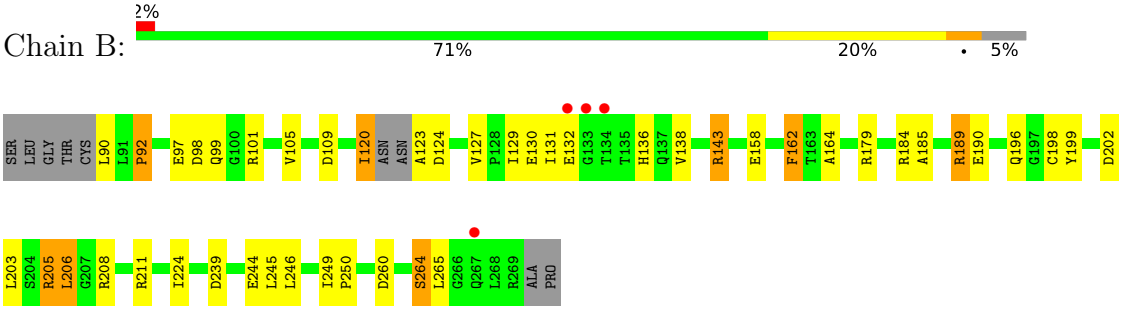
• Molecule 1: Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 2

Chain A:



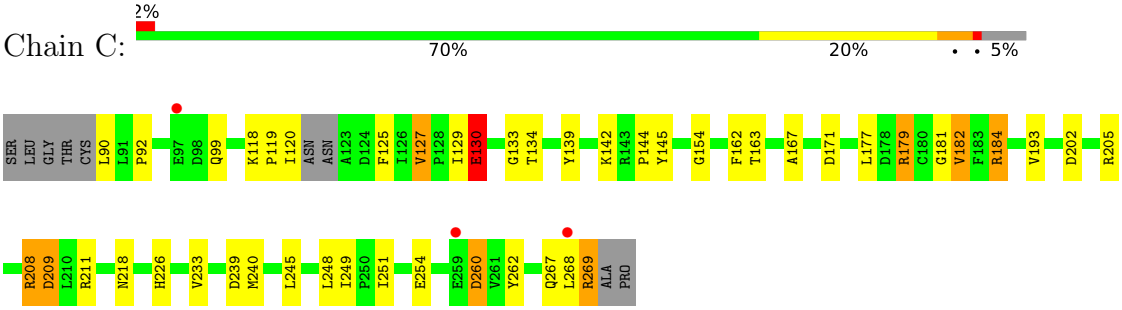
• Molecule 1: Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 2

Chain B:



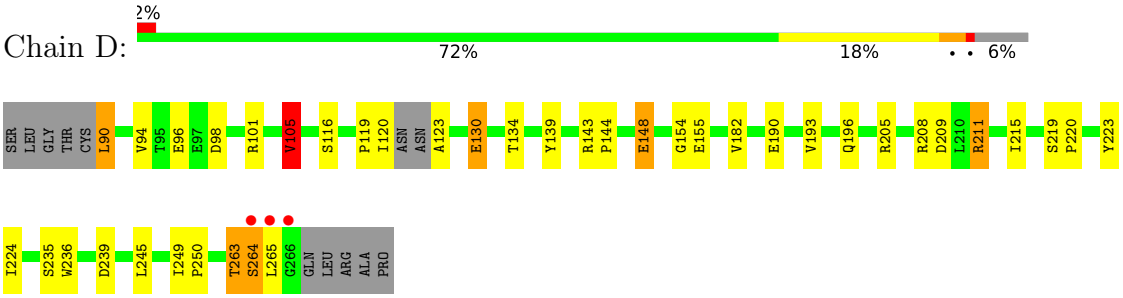
• Molecule 1: Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 2

Chain C:



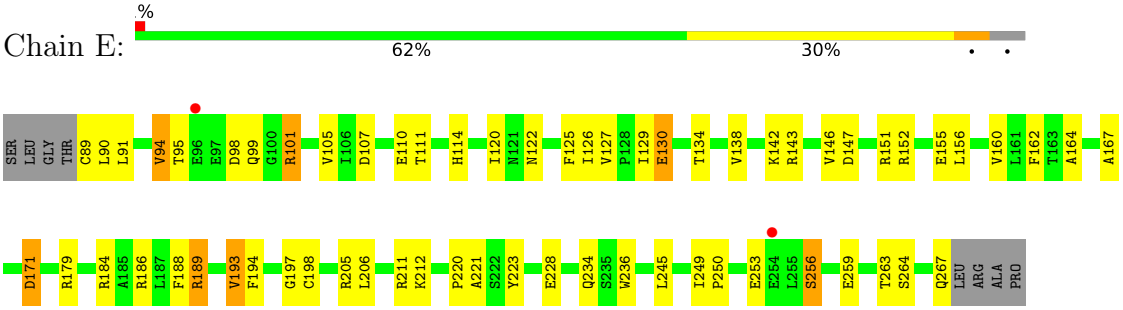
• Molecule 1: Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 2

Chain D:



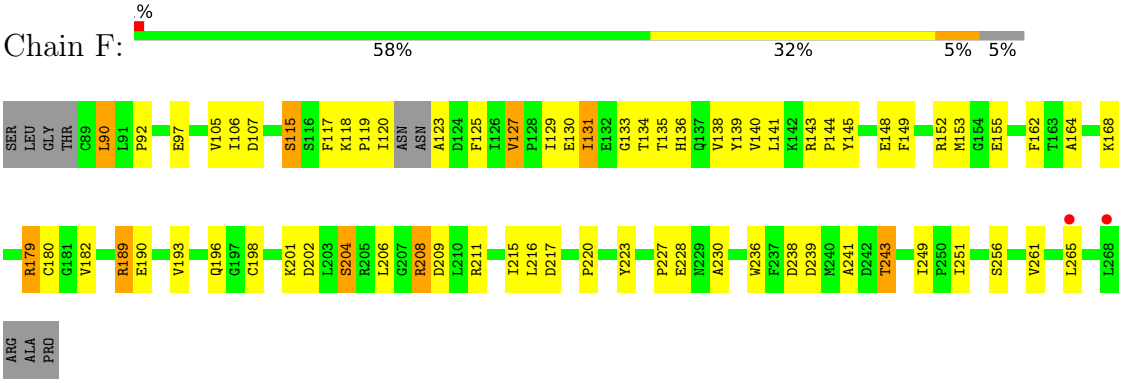
• Molecule 1: Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 2

Chain E:



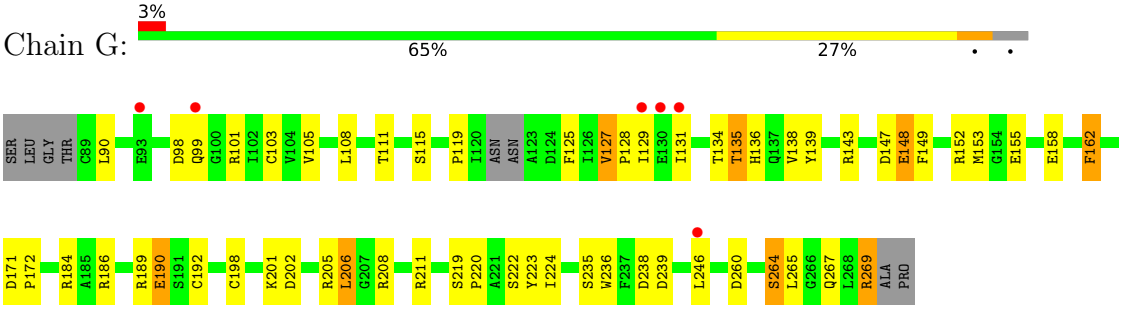
• Molecule 1: Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 2

Chain F:



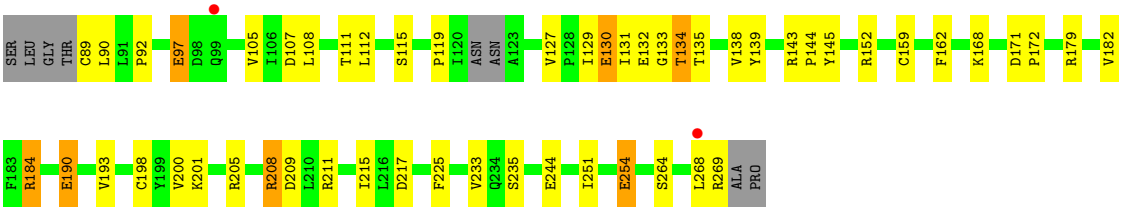
• Molecule 1: Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 2

Chain G:



• Molecule 1: Carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 2

Chain H:


SER
LEU
GLY
THR
C89
L90
L91
P92
E97
D98
Q99
V105
I106
D107
L108
T111
L112
S115
P119
I120
ASN
ASN
A123
V127
P128
I129
E130
I131
E132
G133
T134
T135
V138
Y139
R143
P144
Y145
R152
C159
F162
K168
D171
P172
R179
V182
F183
R184
E190
V193
C198
Y199
V200
K201
R205
R208
D209
L210
R211
I215
L216
D217
F225
V233
Q234
S235
E244
I251
E254
S264
L268
R269
ALA
PRO

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	108.64Å 117.64Å 170.13Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.51 31.30 – 2.51	Depositor EDS
% Data completeness (in resolution range)	96.4 (20.00-2.51) 96.4 (31.30-2.51)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	2.53 (at 2.51Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.216 , 0.278 0.211 , 0.269	Depositor DCC
R_{free} test set	3635 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	46.2	Xtriage
Anisotropy	0.214	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 34.0	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	0.94	EDS
Total number of atoms	11566	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 2.92% 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: MG

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	1.16	3/1485 (0.2%)	1.16	7/2018 (0.3%)
1	B	1.08	2/1463 (0.1%)	1.06	6/1986 (0.3%)
1	C	1.09	1/1463 (0.1%)	1.10	4/1986 (0.2%)
1	D	1.05	2/1435 (0.1%)	1.02	2/1949 (0.1%)
1	E	1.09	3/1467 (0.2%)	1.00	2/1994 (0.1%)
1	F	1.05	0/1458	1.03	6/1980 (0.3%)
1	G	1.07	3/1469 (0.2%)	0.96	4/1994 (0.2%)
1	H	1.17	3/1469 (0.2%)	1.11	4/1994 (0.2%)
All	All	1.10	17/11709 (0.1%)	1.06	35/15901 (0.2%)

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	D	0	1
1	E	0	1
All	All	0	2

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	198	CYS	CB-SG	-13.77	1.58	1.82
1	B	198	CYS	CB-SG	-10.49	1.64	1.82
1	E	198	CYS	CB-SG	-10.37	1.64	1.82
1	A	254	GLU	CG-CD	7.82	1.63	1.51
1	H	198	CYS	CB-SG	-7.23	1.70	1.82

The worst 5 of 35 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	208	ARG	NE-CZ-NH2	-10.89	114.85	120.30
1	H	208	ARG	NE-CZ-NH2	-10.56	115.02	120.30
1	D	208	ARG	NE-CZ-NH1	9.23	124.92	120.30
1	A	208	ARG	NE-CZ-NH2	-8.62	115.99	120.30
1	C	208	ARG	NE-CZ-NH1	8.39	124.50	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	263	THR	Peptide
1	E	89	CYS	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	1454	0	1431	32	0
1	B	1433	0	1413	28	0
1	C	1433	0	1413	30	1
1	D	1405	0	1381	24	0
1	E	1436	0	1407	35	0
1	F	1428	0	1405	42	0
1	G	1439	0	1418	37	0
1	H	1439	0	1418	29	1
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
2	G	1	0	0	0	0
2	H	1	0	0	0	0
3	A	11	0	0	3	0
3	B	12	0	0	0	0
3	C	11	0	0	3	0
3	D	11	0	0	0	0
3	E	11	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	F	7	0	0	0	0
3	G	11	0	0	3	0
3	H	17	0	0	4	0
All	All	11566	0	11286	251	1

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 11.

The worst 5 of 251 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:E:98:ASP:O	1:E:101:ARG:HG3	1.59	1.02
1:G:147:ASP:HA	3:G:282:HOH:O	1.57	1.02
1:F:227:PRO:HD2	1:F:228:GLU:OE2	1.63	0.96
1:A:131:ILE:N	1:A:134:THR:O	1.98	0.95
1:H:264:SER:O	1:H:268:LEU:HD13	1.72	0.90

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:179:ARG:NH2	1:H:145:TYR:OH[4_455]	2.16	0.04

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	179/187 (96%)	166 (93%)	12 (7%)	1 (1%)	25 43
1	B	174/187 (93%)	161 (92%)	11 (6%)	2 (1%)	14 26
1	C	174/187 (93%)	158 (91%)	13 (8%)	3 (2%)	9 16

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	D	171/187 (91%)	156 (91%)	12 (7%)	3 (2%)	8 14
1	E	177/187 (95%)	157 (89%)	18 (10%)	2 (1%)	14 26
1	F	174/187 (93%)	158 (91%)	14 (8%)	2 (1%)	14 26
1	G	175/187 (94%)	159 (91%)	11 (6%)	5 (3%)	4 6
1	H	175/187 (94%)	169 (97%)	5 (3%)	1 (1%)	25 43
All	All	1399/1496 (94%)	1284 (92%)	96 (7%)	19 (1%)	11 20

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	264	SER
1	E	99	GLN
1	G	99	GLN
1	B	185	ALA
1	A	196	GLN

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	163/168 (97%)	150 (92%)	13 (8%)	12 23
1	B	161/168 (96%)	148 (92%)	13 (8%)	11 23
1	C	161/168 (96%)	148 (92%)	13 (8%)	11 23
1	D	158/168 (94%)	146 (92%)	12 (8%)	13 25
1	E	162/168 (96%)	145 (90%)	17 (10%)	7 13
1	F	161/168 (96%)	139 (86%)	22 (14%)	3 7
1	G	162/168 (96%)	148 (91%)	14 (9%)	10 20
1	H	162/168 (96%)	147 (91%)	15 (9%)	9 17
All	All	1290/1344 (96%)	1171 (91%)	119 (9%)	9 18

5 of 119 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	122	ASN
1	E	256	SER
1	H	135	THR
1	E	130	GLU
1	E	189	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	99	GLN
1	C	226	HIS
1	G	136	HIS
1	B	226	HIS
1	E	226	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 8 ligands modelled in this entry, 8 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	181/187 (96%)	-0.35	3 (1%)	70	72	24, 34, 63, 81	0
1	B	178/187 (95%)	-0.29	4 (2%)	62	65	26, 39, 65, 74	0
1	C	178/187 (95%)	-0.23	3 (1%)	70	72	27, 42, 63, 81	0
1	D	175/187 (93%)	-0.23	3 (1%)	70	72	27, 42, 60, 82	0
1	E	179/187 (95%)	-0.10	2 (1%)	80	82	30, 45, 73, 84	0
1	F	178/187 (95%)	-0.14	2 (1%)	80	82	27, 47, 67, 81	0
1	G	179/187 (95%)	0.00	6 (3%)	45	48	29, 54, 78, 89	0
1	H	179/187 (95%)	-0.32	2 (1%)	80	82	22, 38, 56, 74	0
All	All	1427/1496 (95%)	-0.21	25 (1%)	68	71	22, 42, 70, 89	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	265	LEU	6.7
1	A	132	GLU	4.3
1	C	259	GLU	3.3
1	B	133	GLY	3.1
1	G	130	GLU	3.1

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 (i)

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
2	MG	F	1	1/1	0.74	0.12	59,59,59,59	0
2	MG	H	1	1/1	0.75	0.13	47,47,47,47	0
2	MG	D	1	1/1	0.81	0.17	58,58,58,58	0
2	MG	C	1	1/1	0.83	0.32	60,60,60,60	0
2	MG	G	1	1/1	0.84	0.34	50,50,50,50	0
2	MG	A	1	1/1	0.84	0.20	38,38,38,38	0
2	MG	E	1	1/1	0.90	0.25	47,47,47,47	0
2	MG	B	1	1/1	0.91	0.09	40,40,40,40	0

6.5 Other polymers (i)

There are no such residues in this entry.