wwPDB X-ray Structure Validation Summary Report (i)

Jul 26, 2016 – 07:43 PM EDT

PDB ID	:	5JMC
Title	:	Receptor binding domain of Botulinum neurotoxin A in complex with rat SV2C
Authors	:	Yao, G.; Zhang, S.; Mahrhold, S.; Lam, K.; Stern, D.; Bagramyan, K.; Perry, K.; Kalkum, M.; Rummel, A.; Dong, M.; Jin, R.
Deposited on	:	2016-04-28
Resolution	:	2.64 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
http://wwpdb.org/validation/2016/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:

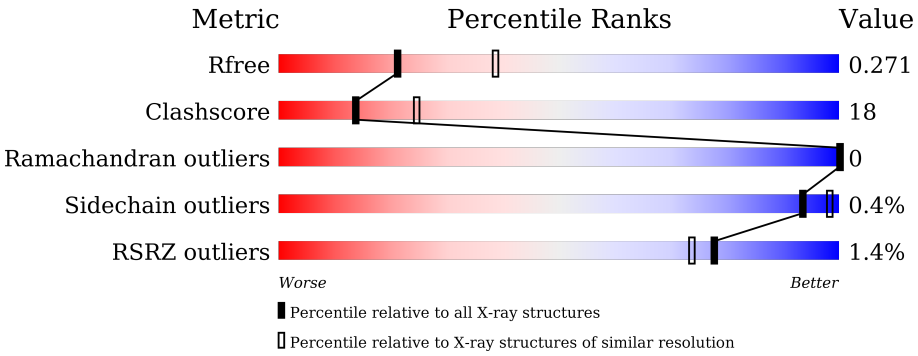
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 2.64 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}	91344	3377 (2.68-2.60)
Clashscore	102246	3781 (2.68-2.60)
Ramachandran outliers	100387	3722 (2.68-2.60)
Sidechain outliers	100360	3722 (2.68-2.60)
RSRZ outliers	91569	3388 (2.68-2.60)

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 <=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	433	70% 27%
1	C	433	74% 22%
1	E	433	75% 22%
1	G	433	73% 24%
2	B	124	4% 40% 34% 27%
2	D	124	50% 23% 26%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	124	
2	H	124	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 17000 atoms, of which 13 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 Botulinum neurotoxin type A.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
1	A	420	Total C H N O S 3471 2210 13 595 639 14	0	0	0
1	C	417	Total C N O S 3436 2195 589 638 14	0	0	0
1	E	419	Total C N O S 3455 2205 595 641 14	0	0	0
1	G	419	Total C N O S 3451 2204 592 641 14	0	0	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	868	GLY	-	expression tag	UNP P10845
A	869	SER	-	expression tag	UNP P10845
A	870	HIS	-	expression tag	UNP P10845
A	871	MET	-	expression tag	UNP P10845
A	1158	ALA	THR	engineered mutation	UNP P10845
A	1297	LEU	-	expression tag	UNP P10845
A	1298	VAL	-	expression tag	UNP P10845
A	1299	PRO	-	expression tag	UNP P10845
A	1300	ARG	-	expression tag	UNP P10845
C	868	GLY	-	expression tag	UNP P10845
C	869	SER	-	expression tag	UNP P10845
C	870	HIS	-	expression tag	UNP P10845
C	871	MET	-	expression tag	UNP P10845
C	1158	ALA	THR	engineered mutation	UNP P10845
C	1297	LEU	-	expression tag	UNP P10845
C	1298	VAL	-	expression tag	UNP P10845
C	1299	PRO	-	expression tag	UNP P10845
C	1300	ARG	-	expression tag	UNP P10845
E	868	GLY	-	expression tag	UNP P10845
E	869	SER	-	expression tag	UNP P10845
E	870	HIS	-	expression tag	UNP P10845

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	871	MET	-	expression tag	UNP P10845
E	1158	ALA	THR	engineered mutation	UNP P10845
E	1297	LEU	-	expression tag	UNP P10845
E	1298	VAL	-	expression tag	UNP P10845
E	1299	PRO	-	expression tag	UNP P10845
E	1300	ARG	-	expression tag	UNP P10845
G	868	GLY	-	expression tag	UNP P10845
G	869	SER	-	expression tag	UNP P10845
G	870	HIS	-	expression tag	UNP P10845
G	871	MET	-	expression tag	UNP P10845
G	1158	ALA	THR	engineered mutation	UNP P10845
G	1297	LEU	-	expression tag	UNP P10845
G	1298	VAL	-	expression tag	UNP P10845
G	1299	PRO	-	expression tag	UNP P10845
G	1300	ARG	-	expression tag	UNP P10845

- Molecule 2 is a protein called Synaptic vesicle glycoprotein 2C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	91	Total	C	N	O	S	0	0	0
			757	487	117	148	5			
2	D	92	Total	C	N	O	S	0	0	0
			764	491	118	150	5			
2	F	92	Total	C	N	O	S	0	0	0
			764	491	118	150	5			
2	H	89	Total	C	N	O	S	0	0	0
			740	477	113	145	5			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	454	GLY	-	expression tag	UNP Q9Z2I6
D	454	GLY	-	expression tag	UNP Q9Z2I6
F	454	GLY	-	expression tag	UNP Q9Z2I6
H	454	GLY	-	expression tag	UNP Q9Z2I6

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	34	Total	O	0	0
			34	34		

Continued on next page...

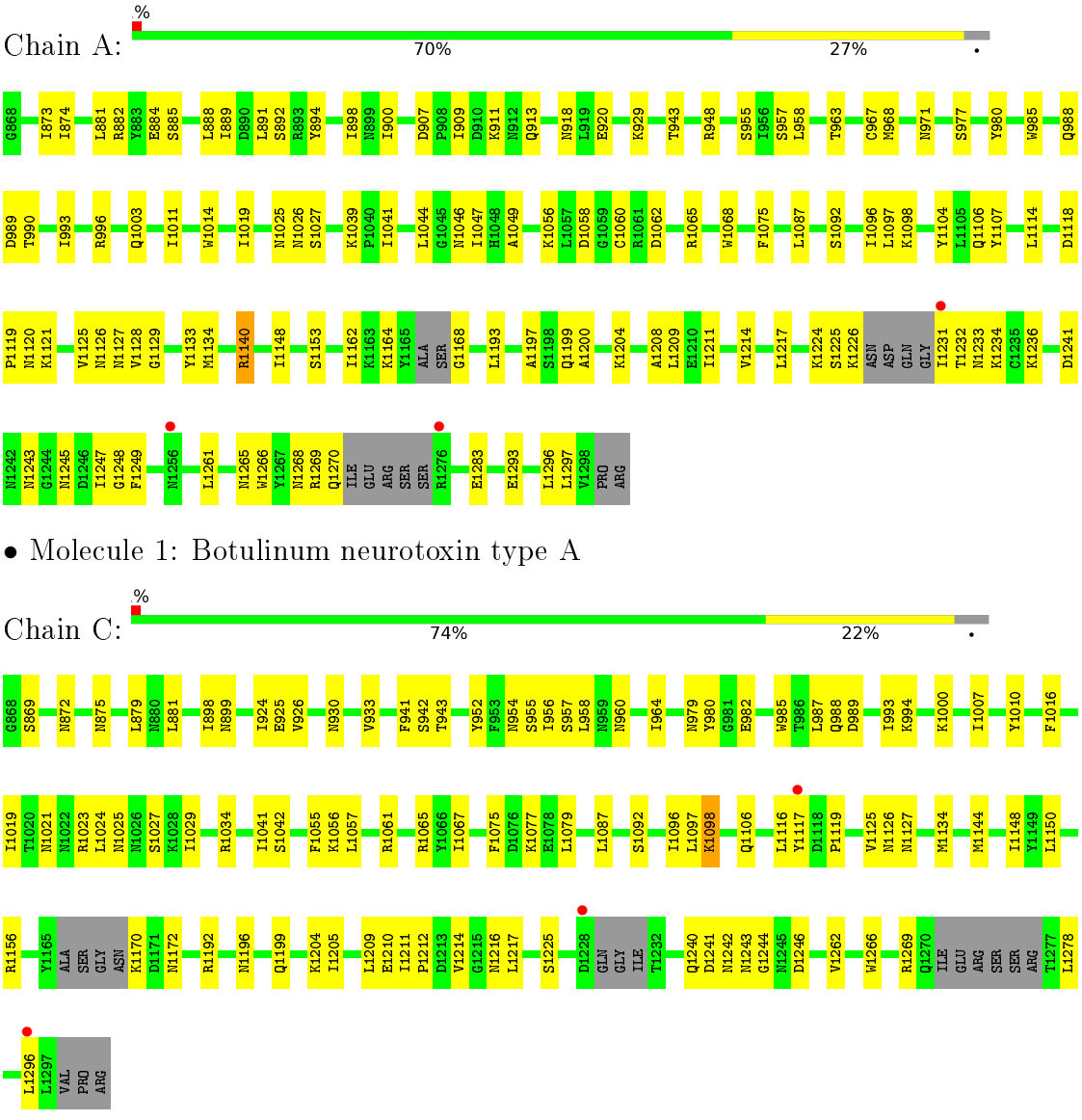
Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	8	Total O 8 8	0	0
3	C	32	Total O 32 32	0	0
3	D	8	Total O 8 8	0	0
3	E	33	Total O 33 33	0	0
3	F	7	Total O 7 7	0	0
3	G	31	Total O 31 31	0	0
3	H	9	Total O 9 9	0	0

3 Residue-property plots (i)

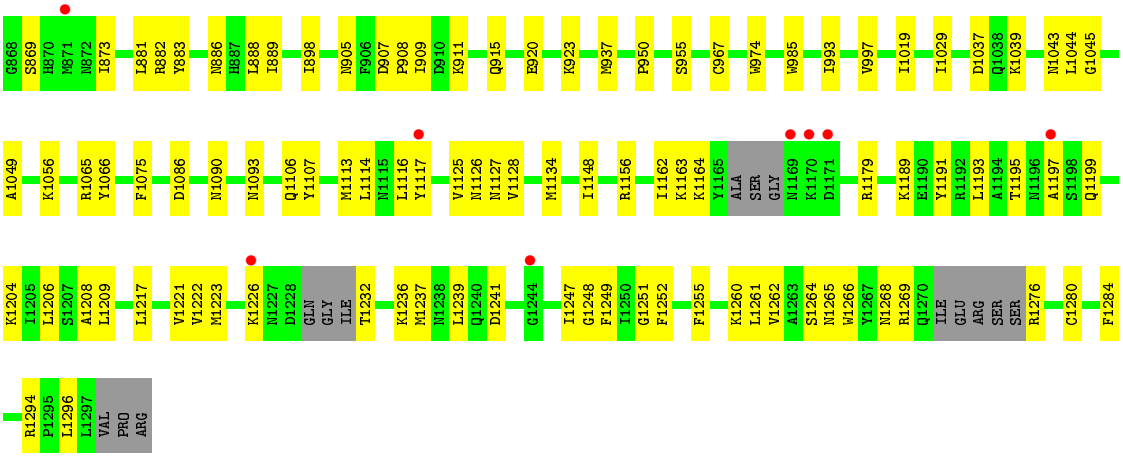
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Botulinum neurotoxin type A



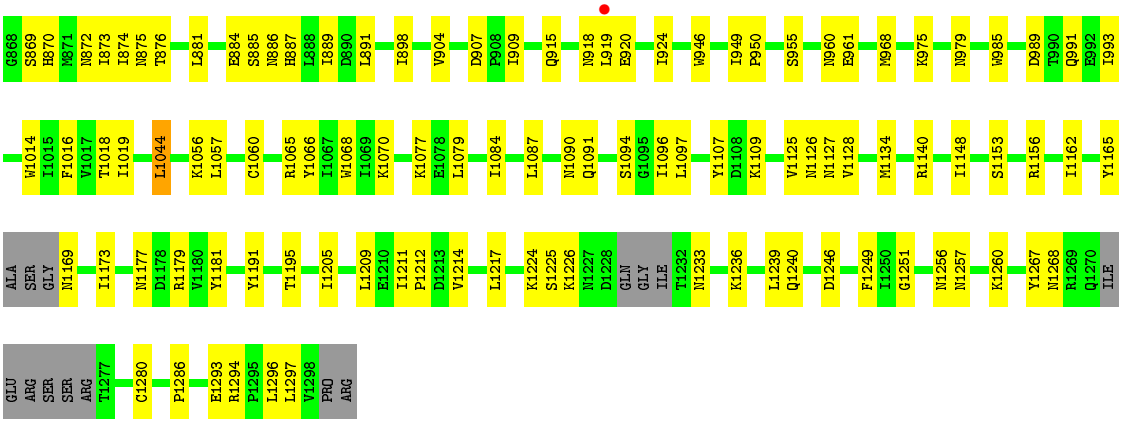
- Molecule 1: Botulinum neurotoxin type A

- Molecule 1: Botulinum neurotoxin type A



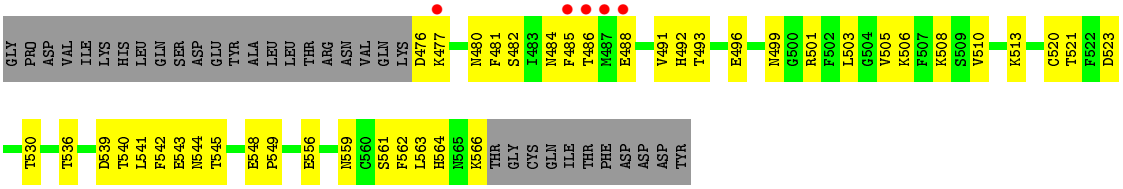
• Molecule 1: Botulinum neurotoxin type A

Chain G:



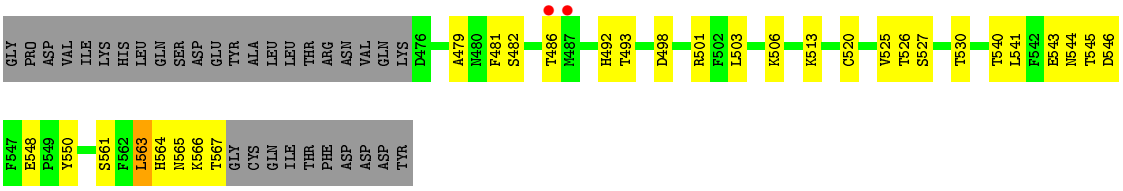
• Molecule 2: Synaptic vesicle glycoprotein 2C

Chain B:



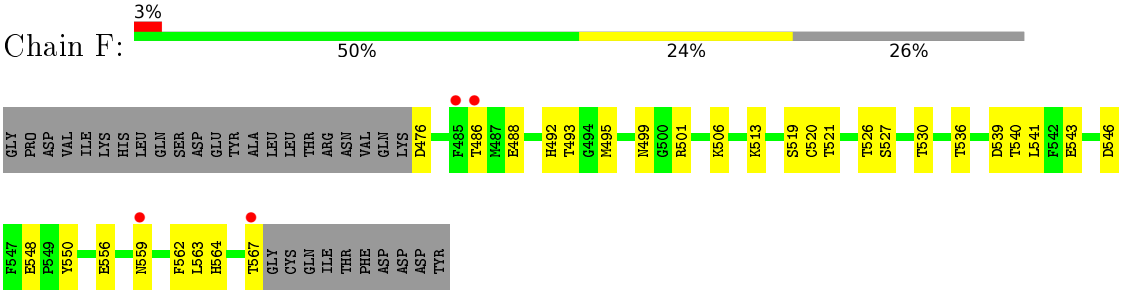
• Molecule 2: Synaptic vesicle glycoprotein 2C

Chain D:



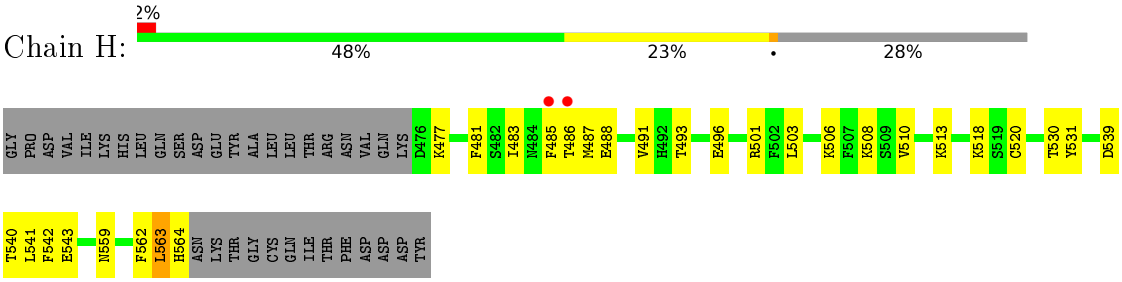
• Molecule 2: Synaptic vesicle glycoprotein 2C

Chain F:



• Molecule 2: Synaptic vesicle glycoprotein 2C

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	88.66Å 143.99Å 110.92Å 90.00° 93.62° 90.00°	Depositor
Resolution (Å)	87.76 – 2.64 87.76 – 2.64	Depositor EDS
% Data completeness (in resolution range)	99.3 (87.76-2.64) 88.3 (87.76-2.64)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.09 (at 2.65Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.239 , 0.275 0.235 , 0.271	Depositor DCC
R_{free} test set	3617 reflections (5.03%)	DCC
Wilson B-factor (Å^2)	36.6	Xtriage
Anisotropy	0.838	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 32.3	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	17000	wwPDB-VP
Average B, all atoms (Å^2)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 38.26 % of the origin peak, indicating pseudo translational symmetry.The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 3.7888e-04.The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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	0.67	0/3529	0.64	0/4768
1	C	0.62	0/3507	0.61	2/4739 (0.0%)
1	E	0.60	0/3526	0.62	0/4764
1	G	0.61	0/3522	0.63	1/4760 (0.0%)
2	B	0.67	0/777	0.66	0/1046
2	D	0.68	0/784	0.61	0/1056
2	F	0.58	0/784	0.58	0/1056
2	H	0.55	0/760	0.63	1/1024 (0.1%)
All	All	0.63	0/17189	0.62	4/23213 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	1023	ARG	NE-CZ-NH1	-5.99	117.30	120.30
1	C	1023	ARG	NE-CZ-NH2	5.35	122.97	120.30
1	G	1044	LEU	CA-CB-CG	5.10	127.02	115.30
2	H	563	LEU	CA-CB-CG	5.04	126.90	115.30

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	3458	13	3428	140	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	3436	0	3396	143	1
1	E	3455	0	3415	107	0
1	G	3451	0	3409	98	0
2	B	757	0	696	48	0
2	D	764	0	703	37	0
2	F	764	0	703	35	1
2	H	740	0	677	36	0
3	A	34	0	0	8	0
3	B	8	0	0	1	0
3	C	32	0	0	4	0
3	D	8	0	0	0	0
3	E	33	0	0	8	0
3	F	7	0	0	2	0
3	G	31	0	0	2	0
3	H	9	0	0	1	0
All	All	16987	13	16427	614	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 18.

The worst 5 of 614 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:G:1249:PHE:CD2	1:G:1268:ASN:ND2	1.70	1.56
1:C:1125:VAL:CG1	1:C:1134:MET:HG2	1.42	1.48
1:E:920:GLU:OE1	1:E:1056:LYS:NZ	1.63	1.31
1:C:1125:VAL:CG1	1:C:1134:MET:CG	2.12	1.27
1:A:1249:PHE:CD2	1:A:1268:ASN:ND2	2.05	1.25

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:869:SER:OG	2:F:492:HIS:ND1[1_556]	2.18	0.02

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	412/433 (95%)	394 (96%)	18 (4%)	0	100	100
1	C	409/433 (94%)	393 (96%)	16 (4%)	0	100	100
1	E	411/433 (95%)	391 (95%)	20 (5%)	0	100	100
1	G	411/433 (95%)	387 (94%)	24 (6%)	0	100	100
2	B	89/124 (72%)	88 (99%)	1 (1%)	0	100	100
2	D	90/124 (73%)	88 (98%)	2 (2%)	0	100	100
2	F	90/124 (73%)	90 (100%)	0	0	100	100
2	H	87/124 (70%)	81 (93%)	6 (7%)	0	100	100
All	All	1999/2228 (90%)	1912 (96%)	87 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	387/398 (97%)	384 (99%)	3 (1%)	86	95
1	C	385/398 (97%)	383 (100%)	2 (0%)	92	98
1	E	387/398 (97%)	386 (100%)	1 (0%)	94	99
1	G	387/398 (97%)	386 (100%)	1 (0%)	94	99
2	B	87/117 (74%)	87 (100%)	0	100	100
2	D	88/117 (75%)	87 (99%)	1 (1%)	80	92

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	F	88/117 (75%)	88 (100%)	0	100 100
2	H	85/117 (73%)	85 (100%)	0	100 100
All	All	1894/2060 (92%)	1886 (100%)	8 (0%)	93 98

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	1065	ARG
1	G	1065	ARG
2	D	563	LEU
1	A	1140	ARG
1	C	1098	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	960	ASN
1	C	1021	ASN
1	E	1238	ASN
1	C	954	ASN
1	E	1172	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 (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	420/433 (96%)	-0.17	3 (0%)	89	87	32, 54, 90, 107	6 (1%)
1	C	417/433 (96%)	-0.18	3 (0%)	89	87	36, 58, 85, 99	5 (1%)
1	E	419/433 (96%)	-0.02	8 (1%)	70	65	36, 56, 103, 123	12 (2%)
1	G	419/433 (96%)	-0.25	1 (0%)	95	95	34, 56, 73, 81	7 (1%)
2	B	91/124 (73%)	0.05	5 (5%)	29	22	48, 56, 78, 117	9 (9%)
2	D	92/124 (74%)	-0.12	2 (2%)	65	59	44, 53, 69, 88	9 (9%)
2	F	92/124 (74%)	0.07	4 (4%)	39	32	45, 54, 71, 94	8 (8%)
2	H	89/124 (71%)	-0.07	2 (2%)	65	59	46, 54, 71, 78	6 (6%)
All	All	2039/2228 (91%)	-0.13	28 (1%)	78	74	32, 55, 89, 123	62 (3%)

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	1169	ASN	5.3
2	F	567	THR	3.8
2	F	485	PHE	3.8
1	E	1244	GLY	3.3
1	E	1117	TYR	3.2

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.