wwPDB X-ray Structure Validation Summary Report (i)

May 25, 2020 – 08:55 am BST

PDB ID	:	5XAY
Title	:	Crystal structure of full length tylp, a tetr regulator from streptomyces fradiae
Authors	:	Ray, S.; Panjikar, S.; Anand, R.
Deposited on	:	2017-03-15
Resolution	:	2.58 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:

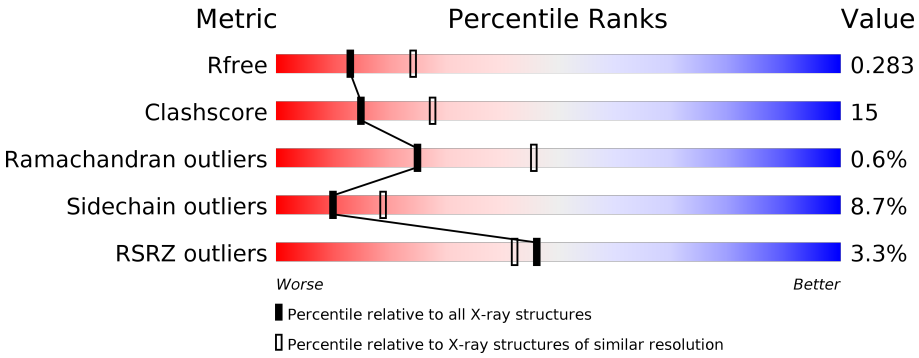
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.58 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.



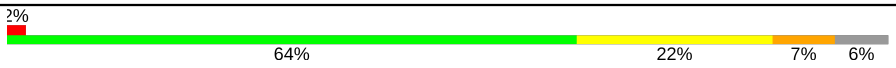
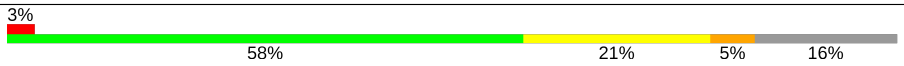
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3676 (2.60-2.56)
Clashscore	141614	4049 (2.60-2.56)
Ramachandran outliers	138981	3979 (2.60-2.56)
Sidechain outliers	138945	3979 (2.60-2.56)
RSRZ outliers	127900	3614 (2.60-2.56)

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	228	3% 71% 22% • 6%
1	B	228	2% 66% 25% • 6%
1	C	228	4% 69% 21% • • •
1	D	228	6% 51% 29% 6% 14%
1	E	228	% 60% 21% • 14%
1	F	228	2% 66% 18% • 13%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	228	
1	H	228	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 12871 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 Gamma-butyrolactone receptor protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	215	Total	C	N	O	Se	0	1	0
			1666	1042	309	310	5			
1	B	215	Total	C	N	O	Se	0	1	0
			1666	1043	309	308	6			
1	C	219	Total	C	N	O	Se	0	1	0
			1691	1057	316	312	6			
1	D	197	Total	C	N	O	Se	0	1	0
			1529	961	282	281	5			
1	E	197	Total	C	N	O	Se	0	0	0
			1525	958	280	282	5			
1	F	199	Total	C	N	O	Se	0	0	0
			1533	962	282	284	5			
1	G	215	Total	C	N	O	Se	0	0	0
			1661	1039	309	308	5			
1	H	192	Total	C	N	O	Se	0	1	0
			1486	937	268	276	5			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	227	LEU	-	expression tag	UNP Q9XCC7
A	228	GLU	-	expression tag	UNP Q9XCC7
A	229	HIS	-	expression tag	UNP Q9XCC7
B	227	LEU	-	expression tag	UNP Q9XCC7
B	228	GLU	-	expression tag	UNP Q9XCC7
B	229	HIS	-	expression tag	UNP Q9XCC7
C	227	LEU	-	expression tag	UNP Q9XCC7
C	228	GLU	-	expression tag	UNP Q9XCC7
C	229	HIS	-	expression tag	UNP Q9XCC7
D	227	LEU	-	expression tag	UNP Q9XCC7
D	228	GLU	-	expression tag	UNP Q9XCC7
D	229	HIS	-	expression tag	UNP Q9XCC7
E	227	LEU	-	expression tag	UNP Q9XCC7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	228	GLU	-	expression tag	UNP Q9XCC7
E	229	HIS	-	expression tag	UNP Q9XCC7
F	227	LEU	-	expression tag	UNP Q9XCC7
F	228	GLU	-	expression tag	UNP Q9XCC7
F	229	HIS	-	expression tag	UNP Q9XCC7
G	227	LEU	-	expression tag	UNP Q9XCC7
G	228	GLU	-	expression tag	UNP Q9XCC7
G	229	HIS	-	expression tag	UNP Q9XCC7
H	227	LEU	-	expression tag	UNP Q9XCC7
H	228	GLU	-	expression tag	UNP Q9XCC7
H	229	HIS	-	expression tag	UNP Q9XCC7

- Molecule 2 is water.

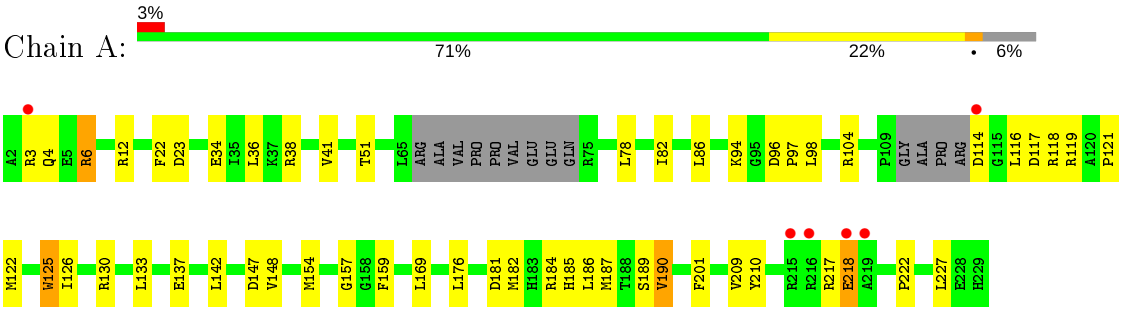
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	19	Total O 19 19	0	0
2	B	17	Total O 17 17	0	0
2	C	13	Total O 13 13	0	0
2	D	12	Total O 12 12	0	0
2	E	12	Total O 12 12	0	0
2	F	13	Total O 13 13	0	0
2	G	18	Total O 18 18	0	0
2	H	10	Total O 10 10	0	0

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.

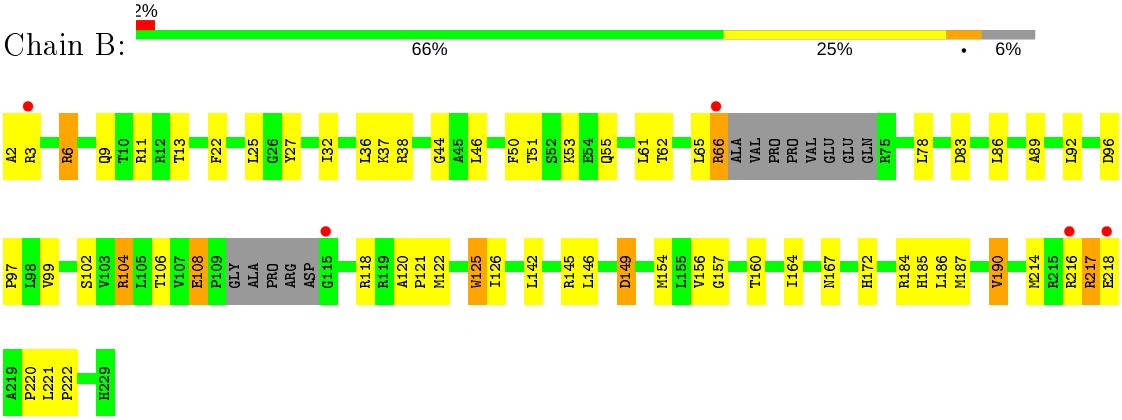
- Molecule 1: Gamma-butyrolactone receptor protein

Chain A:



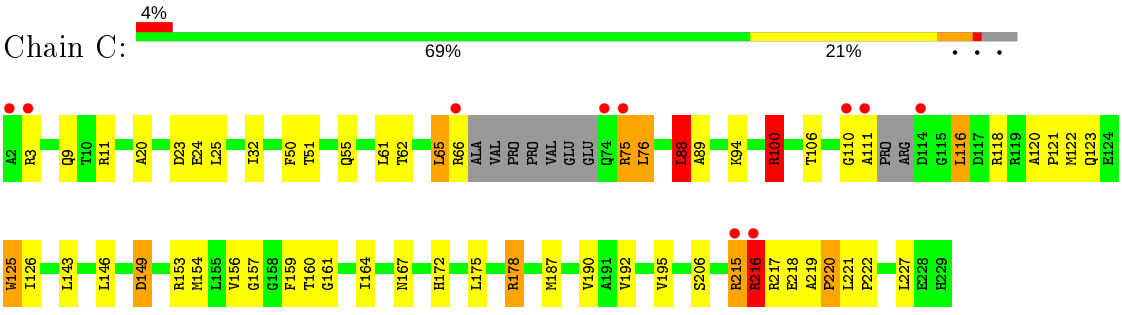
- Molecule 1: Gamma-butyrolactone receptor protein

Chain B:



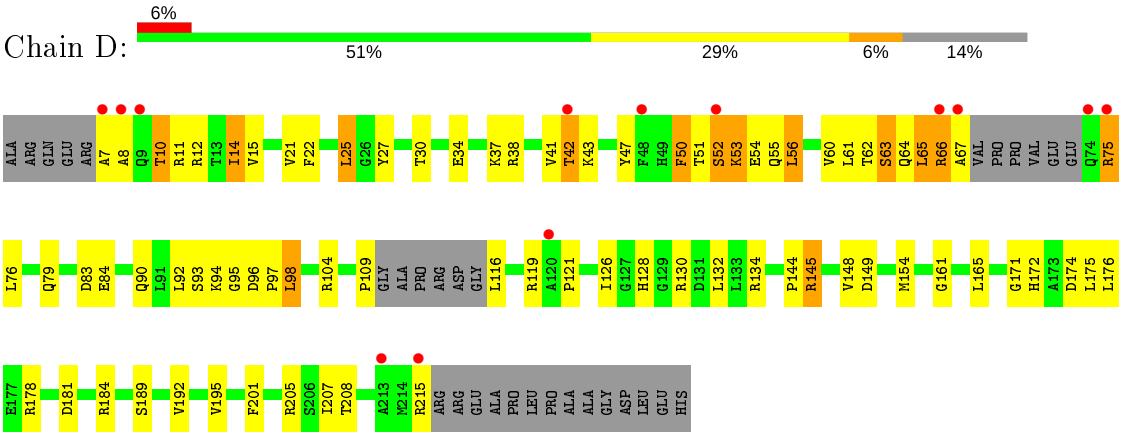
- Molecule 1: Gamma-butyrolactone receptor protein

Chain C:



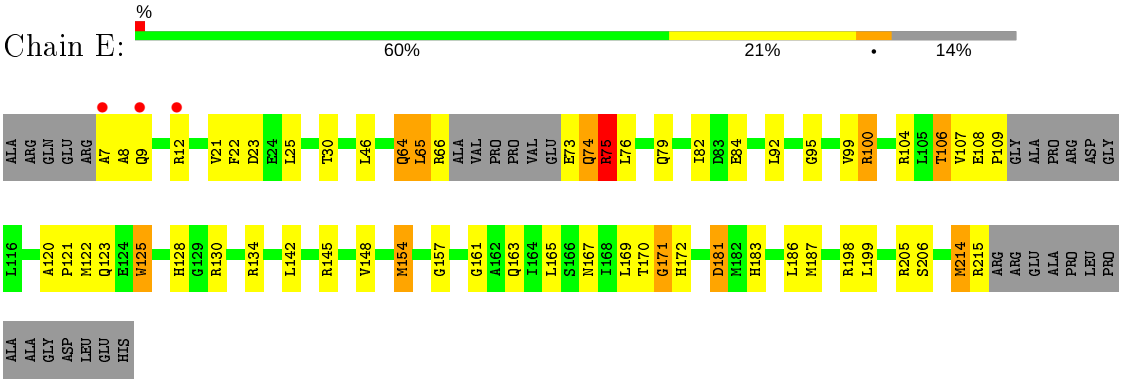
- Molecule 1: Gamma-butyrolactone receptor protein

Chain D:



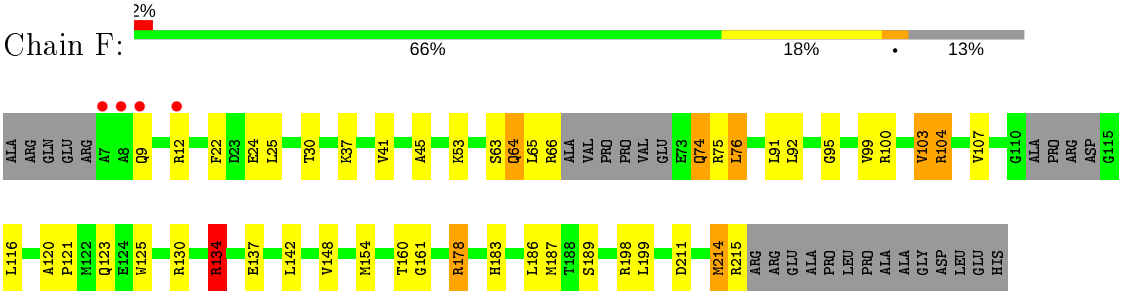
• Molecule 1: Gamma-butyrolactone receptor protein

Chain E:



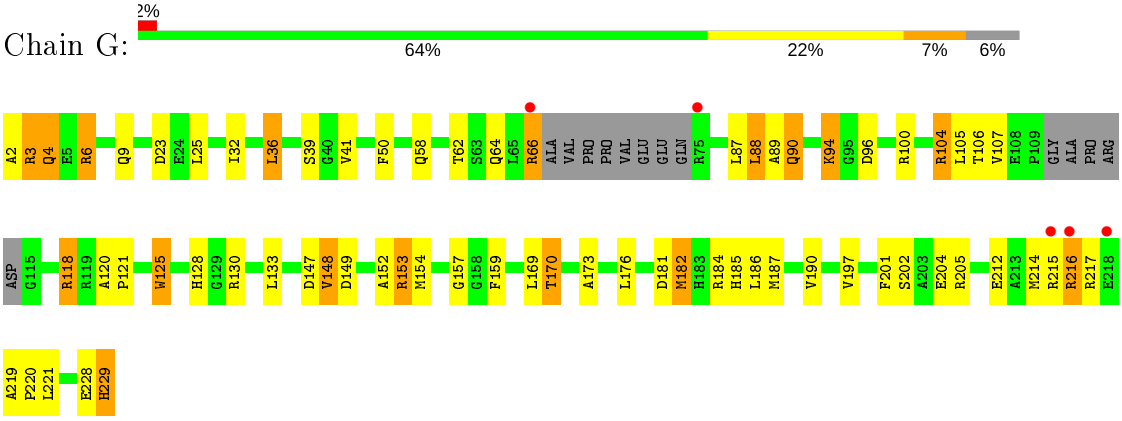
• Molecule 1: Gamma-butyrolactone receptor protein

Chain F:



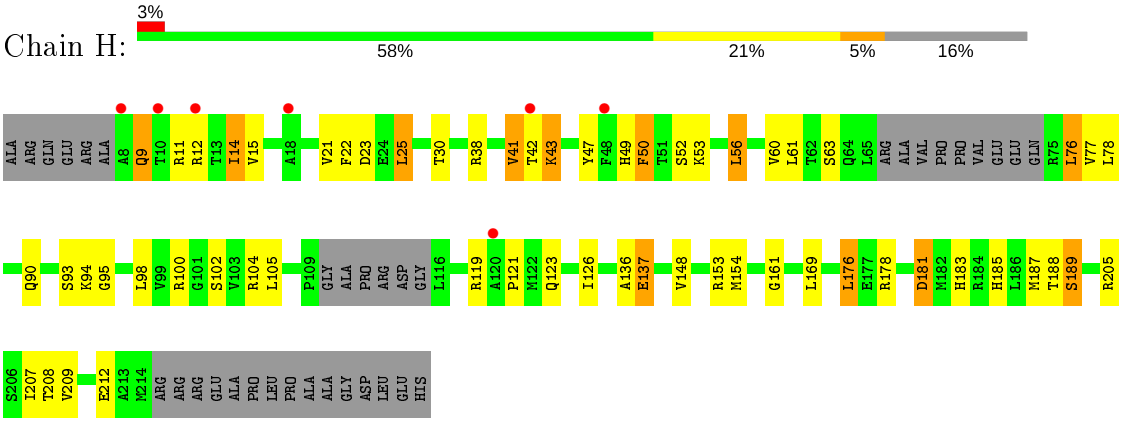
• Molecule 1: Gamma-butyrolactone receptor protein

Chain G:



• Molecule 1: Gamma-butyrolactone receptor protein

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	101.33Å 71.89Å 160.07Å 90.00° 102.54° 90.00°	Depositor
Resolution (Å)	19.92 – 2.58 19.92 – 2.58	Depositor EDS
% Data completeness (in resolution range)	99.6 (19.92-2.58) 99.8 (19.92-2.58)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.55 (at 2.59Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.227 , 0.284 0.228 , 0.283	Depositor DCC
R_{free} test set	994 reflections (1.40%)	wwPDB-VP
Wilson B-factor (Å^2)	30.8	Xtriage
Anisotropy	0.117	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 33.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	12871	wwPDB-VP
Average B, all atoms (Å^2)	34.0	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 65.46 % 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 7.0608e-06. 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	0.92	1/1684 (0.1%)	1.09	6/2268 (0.3%)
1	B	0.94	2/1684 (0.1%)	1.05	7/2267 (0.3%)
1	C	0.89	2/1709 (0.1%)	1.05	5/2300 (0.2%)
1	D	0.79	0/1544	0.98	3/2078 (0.1%)
1	E	0.77	1/1537 (0.1%)	1.00	7/2069 (0.3%)
1	F	0.79	1/1545 (0.1%)	0.95	2/2079 (0.1%)
1	G	0.94	1/1676 (0.1%)	1.15	11/2257 (0.5%)
1	H	0.77	0/1501	0.92	1/2022 (0.0%)
All	All	0.86	8/12880 (0.1%)	1.03	42/17340 (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	E	0	1
1	G	0	1
All	All	0	2

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	125	TRP	CD2-CE2	6.39	1.49	1.41
1	C	125	TRP	CD2-CE2	5.63	1.48	1.41
1	A	125	TRP	CD2-CE2	5.53	1.48	1.41
1	B	125	TRP	CD2-CE2	5.51	1.48	1.41
1	B	149	ASP	CB-CG	5.47	1.63	1.51

The worst 5 of 42 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	G	6	ARG	NE-CZ-NH2	-12.26	114.17	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	G	6	ARG	NE-CZ-NH1	10.43	125.51	120.30
1	A	12	ARG	NE-CZ-NH2	-9.36	115.62	120.30
1	G	130	ARG	NE-CZ-NH2	-8.26	116.17	120.30
1	A	6	ARG	NE-CZ-NH1	8.10	124.35	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	E	171	GLY	Peptide
1	G	106	THR	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	1666	0	1705	54	0
1	B	1666	0	1710	50	1
1	C	1691	0	1733	52	1
1	D	1529	0	1581	80	0
1	E	1525	0	1571	50	0
1	F	1533	0	1577	32	0
1	G	1661	0	1701	64	0
1	H	1486	0	1532	44	0
2	A	19	0	0	2	0
2	B	17	0	0	1	0
2	C	13	0	0	1	0
2	D	12	0	0	0	0
2	E	12	0	0	5	0
2	F	13	0	0	1	0
2	G	18	0	0	3	0
2	H	10	0	0	0	0
All	All	12871	0	13110	390	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 15.

The worst 5 of 390 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:214:MSE:HE1	1:B:217:ARG:NH1	1.40	1.37
1:A:122:MSE:HE2	2:A:318:HOH:O	1.17	1.28
1:D:65:LEU:HD22	1:D:65:LEU:O	1.46	1.14
1:H:9:GLN:HE21	1:H:9:GLN:HA	0.96	1.11
1:E:171:GLY:CA	2:E:301:HOH:O	1.96	1.10

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:B:6:ARG:NH2	1:C:50:PHE:O[1_565]	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	210/228 (92%)	201 (96%)	9 (4%)	0	100	100
1	B	210/228 (92%)	199 (95%)	10 (5%)	1 (0%)	29	50
1	C	214/228 (94%)	202 (94%)	10 (5%)	2 (1%)	17	34
1	D	192/228 (84%)	173 (90%)	18 (9%)	1 (0%)	29	50
1	E	191/228 (84%)	185 (97%)	5 (3%)	1 (0%)	29	50
1	F	193/228 (85%)	181 (94%)	11 (6%)	1 (0%)	29	50
1	G	209/228 (92%)	194 (93%)	14 (7%)	1 (0%)	29	50
1	H	187/228 (82%)	167 (89%)	18 (10%)	2 (1%)	14	28
All	All	1606/1824 (88%)	1502 (94%)	95 (6%)	9 (1%)	25	45

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	216	ARG
1	D	52	SER
1	F	64	GLN
1	G	88	LEU
1	B	220	PRO

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	173/178 (97%)	165 (95%)	8 (5%)	27 49
1	B	173/178 (97%)	163 (94%)	10 (6%)	20 38
1	C	174/178 (98%)	160 (92%)	14 (8%)	12 22
1	D	160/178 (90%)	140 (88%)	20 (12%)	4 8
1	E	160/178 (90%)	150 (94%)	10 (6%)	18 35
1	F	160/178 (90%)	147 (92%)	13 (8%)	11 22
1	G	172/178 (97%)	154 (90%)	18 (10%)	7 12
1	H	157/178 (88%)	135 (86%)	22 (14%)	3 5
All	All	1329/1424 (93%)	1214 (91%)	115 (9%)	10 19

5 of 115 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	64	GLN
1	F	74	GLN
1	H	98	LEU
1	E	74	GLN
1	E	198	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	185	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	74	GLN
1	G	64	GLN
1	D	128	HIS
1	G	229	HIS

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

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

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

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, 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	210/228 (92%)	-0.38	6 (2%)	51	47	13, 26, 51, 74	0
1	B	210/228 (92%)	-0.40	5 (2%)	59	55	13, 25, 56, 93	0
1	C	214/228 (93%)	-0.29	10 (4%)	31	28	13, 25, 54, 84	0
1	D	192/228 (84%)	0.04	13 (6%)	17	14	18, 39, 85, 106	0
1	E	192/228 (84%)	-0.31	3 (1%)	72	69	16, 31, 62, 85	0
1	F	194/228 (85%)	-0.29	4 (2%)	63	60	15, 32, 66, 84	0
1	G	210/228 (92%)	-0.38	5 (2%)	59	55	12, 27, 49, 79	0
1	H	187/228 (82%)	-0.14	7 (3%)	41	37	16, 39, 85, 103	0
All	All	1609/1824 (88%)	-0.27	53 (3%)	46	42	12, 29, 69, 106	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	7	ALA	8.4
1	C	111	ALA	6.9
1	D	67	ALA	6.5
1	D	8	ALA	5.9
1	C	110	GLY	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 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.