



wwPDB X-ray Structure Validation Summary Report

Sep 16, 2023 – 04:10 PM EDT

PDB ID : 4PEM
Title : Crystal Structure of S1G mutant of Penicillin G Acylase from *Kluyvera citrophila*
Authors : Ramasamy, S.; Varshney, N.K.; Brannigan, J.A.; Wilkinson, A.J.; Suresh, C.G.
Deposited on : 2014-04-24
Resolution : 2.50 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  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](#) ) 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

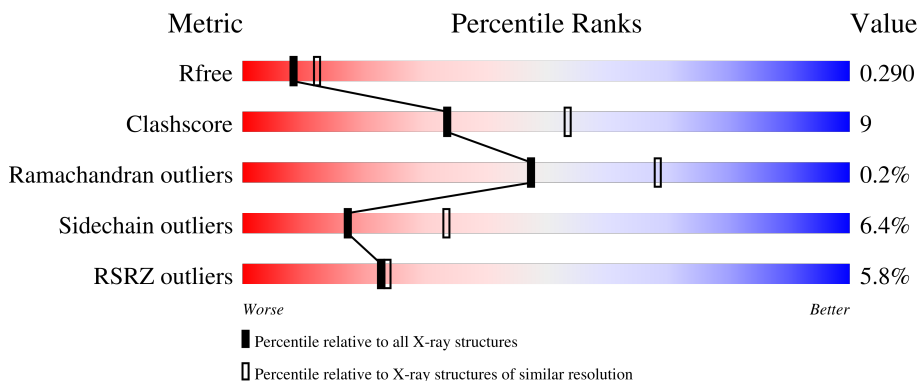
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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.



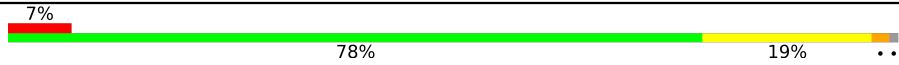
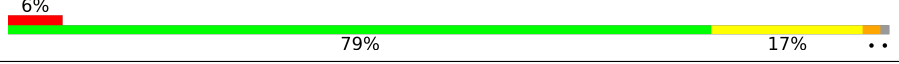

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 $\leq 5\%$. 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	286	<div style="display: flex; align-items: center;"> <div style="width: 3%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 66%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 13%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 10%; height: 10px; background-color: orange; margin-right: 5px;"></div> <div style="width: 8%; height: 10px; background-color: grey;"></div> </div> <p style="text-align: center;">3% 69% 13% • 10%</p>
1	C	286	<div style="display: flex; align-items: center;"> <div style="width: 7%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 66%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 16%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 10%; height: 10px; background-color: orange; margin-right: 5px;"></div> <div style="width: 1%; height: 10px; background-color: grey;"></div> </div> <p style="text-align: center;">7% 73% 16% • 10%</p>
1	E	286	<div style="display: flex; align-items: center;"> <div style="width: 5%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 70%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 14%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 10%; height: 10px; background-color: orange; margin-right: 5px;"></div> <div style="width: 1%; height: 10px; background-color: grey;"></div> </div> <p style="text-align: center;">5% 75% 14% • 10%</p>
1	G	286	<div style="display: flex; align-items: center;"> <div style="width: 4%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 71%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 14%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 10%; height: 10px; background-color: orange; margin-right: 5px;"></div> <div style="width: 1%; height: 10px; background-color: grey;"></div> </div> <p style="text-align: center;">4% 75% 14% • 10%</p>
2	B	568	<div style="display: flex; align-items: center;"> <div style="width: 5%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 73%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 17%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 5%; height: 10px; background-color: orange; margin-right: 5px;"></div> <div style="width: 0%; height: 10px; background-color: grey;"></div> </div> <p style="text-align: center;">5% 78% 17% • •</p>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	568	 7% 78% 19% ..
2	F	568	 6% 79% 17% ..
2	H	568	 7% 79% 17% ..

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 25838 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 Penicillin G acylase Alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	235	Total 1831	C 1171	N 308	O 344	S 8	0	0	0
1	C	258	Total 2026	C 1290	N 343	O 385	S 8	0	0	0
1	E	258	Total 2026	C 1290	N 343	O 385	S 8	0	0	0
1	G	258	Total 2026	C 1290	N 343	O 385	S 8	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-25	MET	-	initiating methionine	UNP A0A068F6N5
A	-24	LYS	-	expression tag	UNP A0A068F6N5
A	-23	ASN	-	expression tag	UNP A0A068F6N5
A	-22	ARG	-	expression tag	UNP A0A068F6N5
A	-21	ASN	-	expression tag	UNP A0A068F6N5
A	-20	ARG	-	expression tag	UNP A0A068F6N5
C	-25	MET	-	initiating methionine	UNP A0A068F6N5
C	-24	LYS	-	expression tag	UNP A0A068F6N5
C	-23	ASN	-	expression tag	UNP A0A068F6N5
C	-22	ARG	-	expression tag	UNP A0A068F6N5
C	-21	ASN	-	expression tag	UNP A0A068F6N5
C	-20	ARG	-	expression tag	UNP A0A068F6N5
E	-25	MET	-	initiating methionine	UNP A0A068F6N5
E	-24	LYS	-	expression tag	UNP A0A068F6N5
E	-23	ASN	-	expression tag	UNP A0A068F6N5
E	-22	ARG	-	expression tag	UNP A0A068F6N5
E	-21	ASN	-	expression tag	UNP A0A068F6N5
E	-20	ARG	-	expression tag	UNP A0A068F6N5
G	-25	MET	-	initiating methionine	UNP A0A068F6N5
G	-24	LYS	-	expression tag	UNP A0A068F6N5
G	-23	ASN	-	expression tag	UNP A0A068F6N5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	-22	ARG	-	expression tag	UNP A0A068F6N5
G	-21	ASN	-	expression tag	UNP A0A068F6N5
G	-20	ARG	-	expression tag	UNP A0A068F6N5

- Molecule 2 is a protein called Penicillin G acylase Beta.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
2	B	549	Total	C	N	O	S	0	0	0
			4304	2743	744	809	8			
2	D	560	Total	C	N	O	S	0	0	0
			4399	2800	762	829	8			
2	F	560	Total	C	N	O	S	0	0	0
			4398	2800	762	828	8			
2	H	560	Total	C	N	O	S	0	0	0
			4399	2800	762	829	8			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	264	GLY	SER	engineered mutation	UNP A0A068F6N5
D	264	GLY	SER	engineered mutation	UNP A0A068F6N5
F	264	GLY	SER	engineered mutation	UNP A0A068F6N5
H	264	GLY	SER	engineered mutation	UNP A0A068F6N5

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Ca	0	0
			1	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	30	Total	O	0	0
			30	30		
4	B	81	Total	O	0	0
			81	81		
4	C	42	Total	O	0	0
			42	42		
4	D	69	Total	O	0	0
			69	69		

Continued on next page...

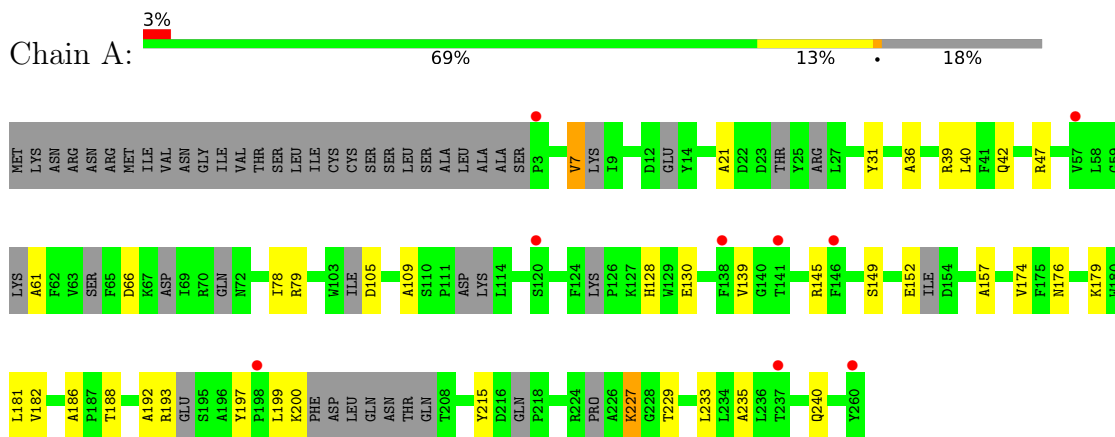
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	38	Total 38	O 38	0	0
4	F	63	Total 63	O 63	0	0
4	G	31	Total 31	O 31	0	0
4	H	74	Total 74	O 74	0	0

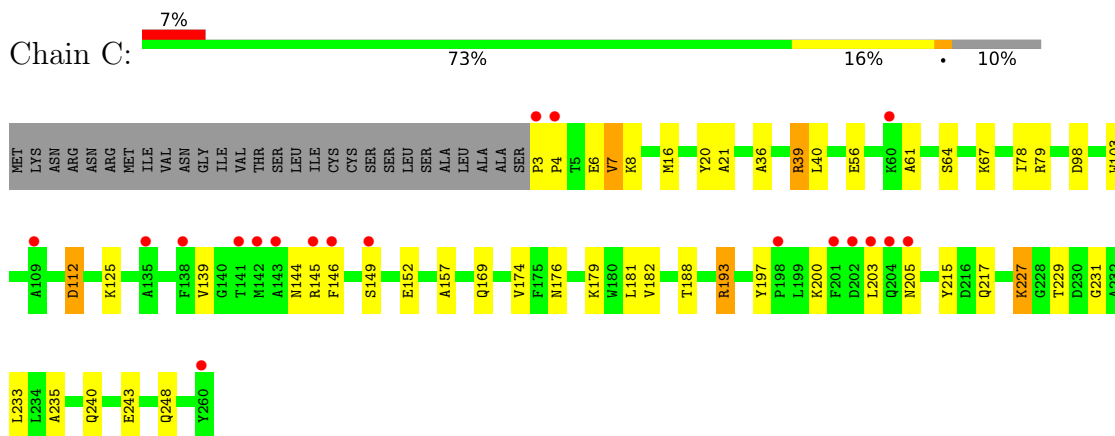
3 Residue-property plots [i](#)

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.

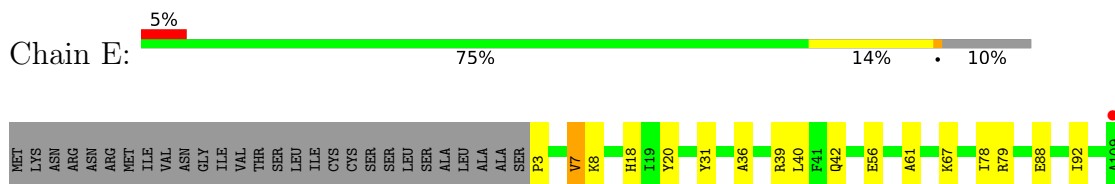
- Molecule 1: Penicillin G acylase Alpha

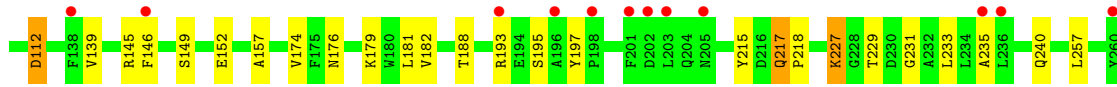


- Molecule 1: Penicillin G acylase Alpha

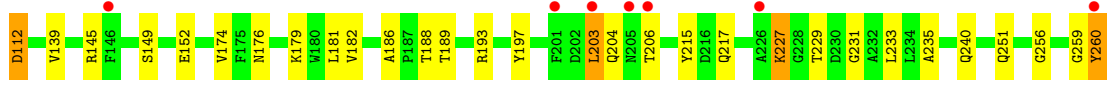
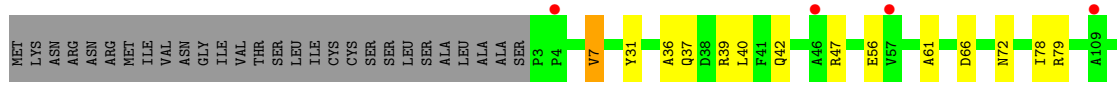
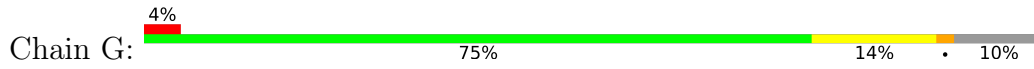


- Molecule 1: Penicillin G acylase Alpha

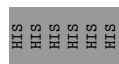
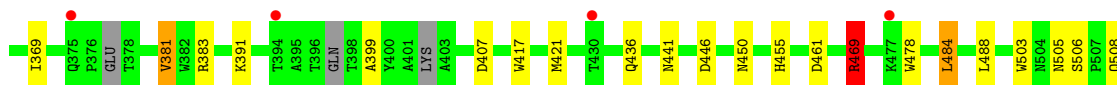
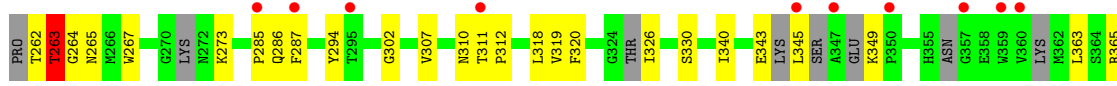
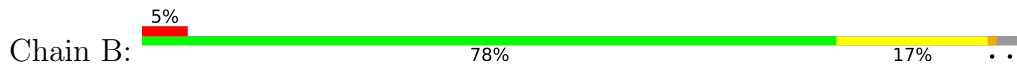




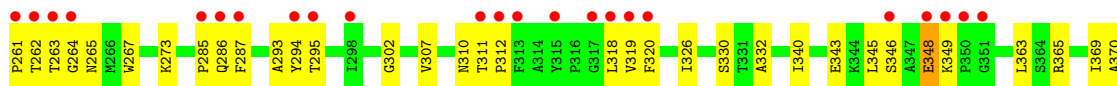
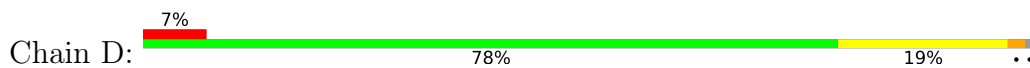
• Molecule 1: Penicillin G acylase Alpha



• Molecule 2: Penicillin G acylase Beta

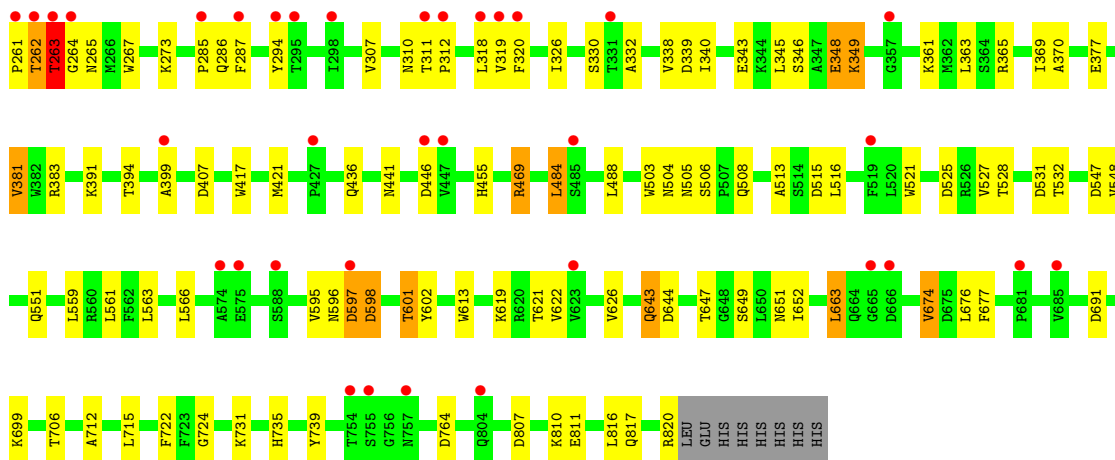
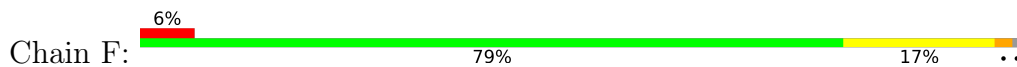


• Molecule 2: Penicillin G acylase Beta

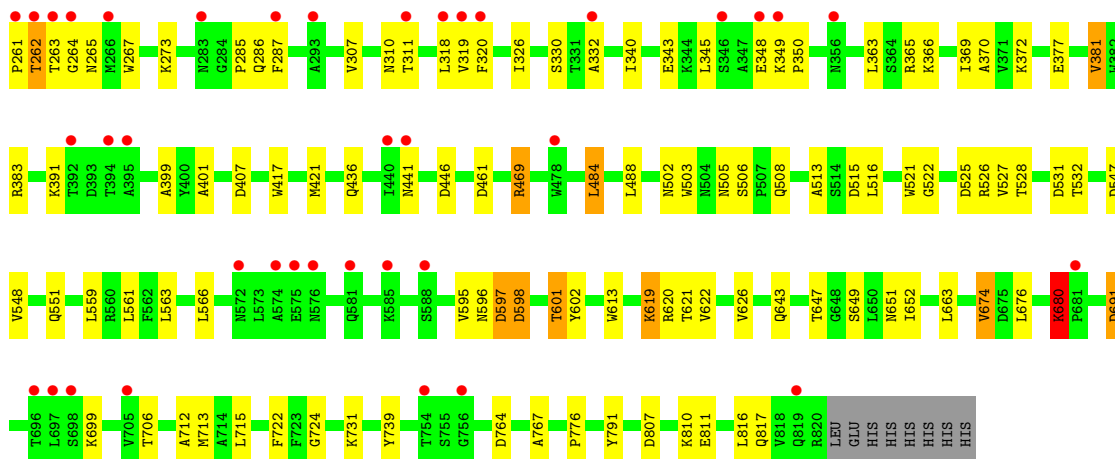
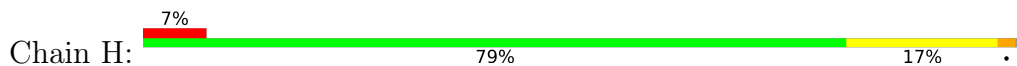




• Molecule 2: Penicillin G acylase Beta



• Molecule 2: Penicillin G acylase Beta



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	54.00Å 124.56Å 135.14Å 104.05° 101.37° 96.51°	Depositor
Resolution (Å)	38.00 – 2.50 37.85 – 2.50	Depositor EDS
% Data completeness (in resolution range)	76.3 (38.00-2.50) 76.4 (37.85-2.50)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.59 (at 2.51Å)	Xtrriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.248 , 0.292 0.251 , 0.290	Depositor DCC
R_{free} test set	4354 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å ²)	42.9	Xtrriage
Anisotropy	0.343	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 41.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	25838	wwPDB-VP
Average B, all atoms (Å ²)	44.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.49% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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:
CA

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	0.57	0/1862	0.64	0/2503
1	C	0.57	0/2075	0.66	1/2816 (0.0%)
1	E	0.53	0/2075	0.62	0/2816
1	G	0.54	0/2075	0.63	0/2816
2	B	0.54	0/4422	0.63	1/6019 (0.0%)
2	D	0.54	0/4528	0.62	0/6175
2	F	0.52	0/4527	0.62	1/6174 (0.0%)
2	H	0.53	0/4528	0.62	0/6175
All	All	0.54	0/26092	0.63	3/35494 (0.0%)

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
2	F	0	1
2	H	0	1
All	All	0	2

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	469	ARG	NE-CZ-NH1	6.08	123.34	120.30
2	F	263	THR	CA-CB-CG2	-5.12	105.23	112.40
1	C	39	ARG	NE-CZ-NH2	5.10	122.85	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	F	263	THR	Peptide
2	H	680	LYS	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	1831	0	1769	39	0
1	C	2026	0	1978	48	1
1	E	2026	0	1978	38	0
1	G	2026	0	1978	33	0
2	B	4304	0	4108	69	0
2	D	4399	0	4221	83	0
2	F	4398	0	4218	106	1
2	H	4399	0	4221	93	0
3	A	1	0	0	0	0
4	A	30	0	0	3	0
4	B	81	0	0	7	0
4	C	42	0	0	11	0
4	D	69	0	0	12	0
4	E	38	0	0	10	0
4	F	63	0	0	9	0
4	G	31	0	0	6	0
4	H	74	0	0	8	0
All	All	25838	0	24471	426	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 9.

The worst 5 of 426 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:F:263:THR:HG21	2:F:504:ASN:ND2	1.59	1.14
2:D:263:THR:HB	2:D:264:GLY:HA3	1.31	1.11
2:F:263:THR:CG2	2:F:504:ASN:ND2	2.13	1.11
2:F:262:THR:HG21	2:F:649:SER:HB3	1.28	1.07
2:F:820:ARG:NH2	4:F:907:HOH:O	1.85	1.07

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:98:ASP:OD2	2:F:361:LYS:NZ[1_565]	2.11	0.09

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	202/286 (71%)	196 (97%)	6 (3%)	0	100	100
1	C	256/286 (90%)	249 (97%)	6 (2%)	1 (0%)	34	54
1	E	256/286 (90%)	252 (98%)	4 (2%)	0	100	100
1	G	256/286 (90%)	249 (97%)	7 (3%)	0	100	100
2	B	529/568 (93%)	509 (96%)	18 (3%)	2 (0%)	34	54
2	D	558/568 (98%)	533 (96%)	24 (4%)	1 (0%)	47	68
2	F	558/568 (98%)	537 (96%)	19 (3%)	2 (0%)	34	54
2	H	558/568 (98%)	534 (96%)	23 (4%)	1 (0%)	47	68
All	All	3173/3416 (93%)	3059 (96%)	107 (3%)	7 (0%)	47	68

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	F	262	THR
1	C	4	PRO
2	B	263	THR
2	D	455	HIS
2	F	455	HIS

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	188/235 (80%)	179 (95%)	9 (5%)	25	48
1	C	211/235 (90%)	200 (95%)	11 (5%)	23	44
1	E	211/235 (90%)	200 (95%)	11 (5%)	23	44
1	G	211/235 (90%)	198 (94%)	13 (6%)	18	35
2	B	439/459 (96%)	409 (93%)	30 (7%)	16	30
2	D	451/459 (98%)	419 (93%)	32 (7%)	14	28
2	F	450/459 (98%)	420 (93%)	30 (7%)	16	31
2	H	451/459 (98%)	420 (93%)	31 (7%)	15	30
All	All	2612/2776 (94%)	2445 (94%)	167 (6%)	17	33

5 of 167 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	663	LEU
2	H	377	GLU
2	F	706	THR
1	G	193	ARG
2	H	531	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	508	GLN
1	E	217	GLN
1	G	217	GLN
1	C	205	ASN
2	B	508	GLN

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 1 ligands modelled in this entry, 1 is 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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	235/286 (82%)	0.34	9 (3%) 40 43	29, 39, 52, 75	0
1	C	258/286 (90%)	0.60	19 (7%) 14 15	27, 42, 63, 110	0
1	E	258/286 (90%)	0.53	13 (5%) 28 30	31, 44, 65, 105	0
1	G	258/286 (90%)	0.41	11 (4%) 35 38	27, 43, 63, 103	0
2	B	549/568 (96%)	0.51	27 (4%) 29 31	25, 40, 62, 80	0
2	D	560/568 (98%)	0.53	37 (6%) 18 19	27, 42, 64, 99	0
2	F	560/568 (98%)	0.56	35 (6%) 20 21	29, 45, 67, 94	0
2	H	560/568 (98%)	0.58	38 (6%) 17 17	29, 43, 65, 106	0
All	All	3238/3416 (94%)	0.52	189 (5%) 23 24	25, 43, 64, 110	0

The worst 5 of 189 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	261	PRO	10.3
2	D	262	THR	10.1
2	D	261	PRO	6.3
1	G	201	PHE	5.3
1	C	205	ASN	5.1

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

6.3 Carbohydrates [i](#)

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, 95th 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
3	CA	A	901	1/1	0.95	0.04	52,52,52,52	0

6.5 Other polymers [i](#)

There are no such residues in this entry.