wwPDB X-ray Structure Validation Summary Report (i)

Dec 21, 2024 – 04:03 pm GMT

PDB ID	:	8RJ8
Title	:	CytK nanopore mutant
Authors	:	Whittaker, J.J.; Sauciuc, A.; Guskov, A.
Deposited on	:	2023-12-20
Resolution	:	4.32 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:

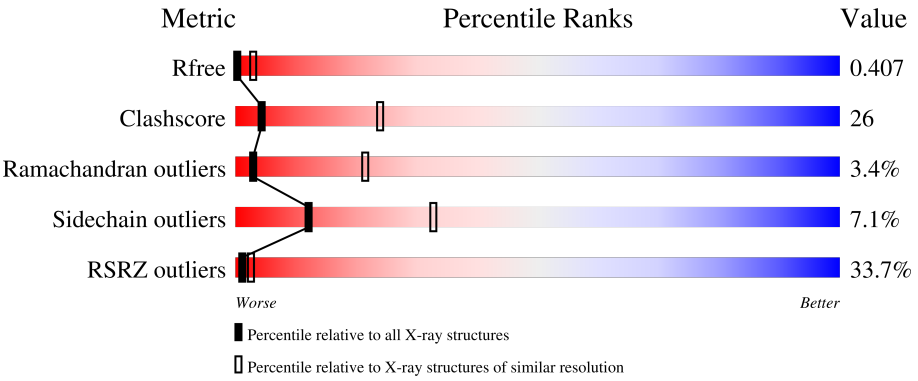
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 4.32 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.



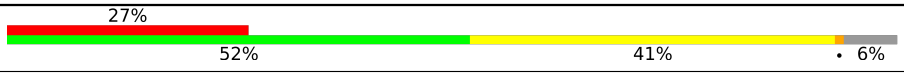
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	1002 (4.76-3.88)
Clashscore	180529	1034 (4.72-3.90)
Ramachandran outliers	177936	1027 (4.80-3.84)
Sidechain outliers	177891	1010 (4.80-3.84)
RSRZ outliers	164620	1000 (4.76-3.88)

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%

Mol	Chain	Length	Quality of chain
1	A	310	32% 41% 44% 7% • 6%
1	B	310	29% 45% 39% • 12%
1	C	310	33% 49% 38% • 8%
1	D	310	41% 38% 42% 8% • 10%
1	E	310	37% 51% 38% • • 6%
1	F	310	21% 45% 48% 5% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	310	

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 15699 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 Cytotoxin K.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	291	Total	C	N	O	S	0	0	0
			2267	1419	381	460	7			
1	B	272	Total	C	N	O	S	0	0	0
			2119	1324	357	431	7			
1	C	284	Total	C	N	O	S	0	0	0
			2227	1396	376	448	7			
1	D	279	Total	C	N	O	S	0	0	0
			2184	1371	368	438	7			
1	E	292	Total	C	N	O	S	0	0	0
			2267	1420	380	460	7			
1	F	302	Total	C	N	O	S	0	0	0
			2351	1472	393	478	8			
1	G	292	Total	C	N	O	S	0	0	0
			2284	1432	382	463	7			

There are 49 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP Q09KJ1
A	130	ASP	LYS	conflict	UNP Q09KJ1
A	149	ASP	THR	conflict	UNP Q09KJ1
A	157	ASP	LYS	conflict	UNP Q09KJ1
A	308	GLY	-	expression tag	UNP Q09KJ1
A	309	SER	-	expression tag	UNP Q09KJ1
A	310	ALA	-	expression tag	UNP Q09KJ1
B	1	MET	-	initiating methionine	UNP Q09KJ1
B	130	ASP	LYS	conflict	UNP Q09KJ1
B	149	ASP	THR	conflict	UNP Q09KJ1
B	157	ASP	LYS	conflict	UNP Q09KJ1
B	308	GLY	-	expression tag	UNP Q09KJ1
B	309	SER	-	expression tag	UNP Q09KJ1
B	310	ALA	-	expression tag	UNP Q09KJ1
C	1	MET	-	initiating methionine	UNP Q09KJ1

Continued on next page...

Continued from previous page...

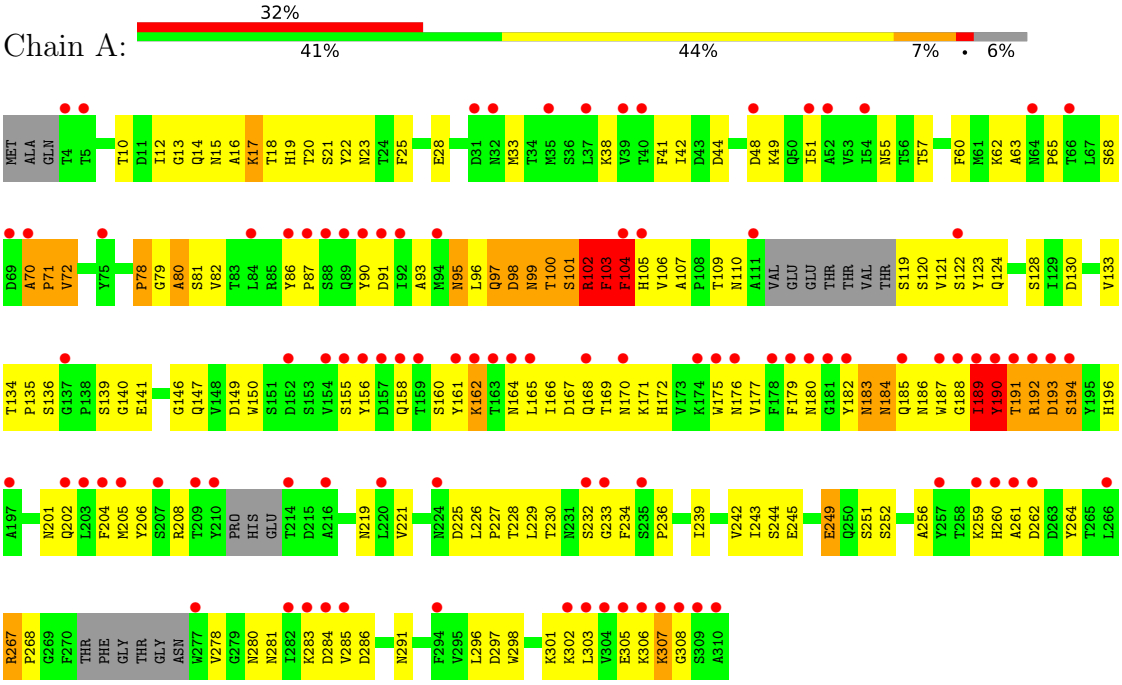
Chain	Residue	Modelled	Actual	Comment	Reference
C	130	ASP	LYS	conflict	UNP Q09KJ1
C	149	ASP	THR	conflict	UNP Q09KJ1
C	157	ASP	LYS	conflict	UNP Q09KJ1
C	308	GLY	-	expression tag	UNP Q09KJ1
C	309	SER	-	expression tag	UNP Q09KJ1
C	310	ALA	-	expression tag	UNP Q09KJ1
D	1	MET	-	initiating methionine	UNP Q09KJ1
D	130	ASP	LYS	conflict	UNP Q09KJ1
D	149	ASP	THR	conflict	UNP Q09KJ1
D	157	ASP	LYS	conflict	UNP Q09KJ1
D	308	GLY	-	expression tag	UNP Q09KJ1
D	309	SER	-	expression tag	UNP Q09KJ1
D	310	ALA	-	expression tag	UNP Q09KJ1
E	1	MET	-	initiating methionine	UNP Q09KJ1
E	130	ASP	LYS	conflict	UNP Q09KJ1
E	149	ASP	THR	conflict	UNP Q09KJ1
E	157	ASP	LYS	conflict	UNP Q09KJ1
E	308	GLY	-	expression tag	UNP Q09KJ1
E	309	SER	-	expression tag	UNP Q09KJ1
E	310	ALA	-	expression tag	UNP Q09KJ1
F	1	MET	-	initiating methionine	UNP Q09KJ1
F	130	ASP	LYS	conflict	UNP Q09KJ1
F	149	ASP	THR	conflict	UNP Q09KJ1
F	157	ASP	LYS	conflict	UNP Q09KJ1
F	308	GLY	-	expression tag	UNP Q09KJ1
F	309	SER	-	expression tag	UNP Q09KJ1
F	310	ALA	-	expression tag	UNP Q09KJ1
G	1	MET	-	initiating methionine	UNP Q09KJ1
G	130	ASP	LYS	conflict	UNP Q09KJ1
G	149	ASP	THR	conflict	UNP Q09KJ1
G	157	ASP	LYS	conflict	UNP Q09KJ1
G	308	GLY	-	expression tag	UNP Q09KJ1
G	309	SER	-	expression tag	UNP Q09KJ1
G	310	ALA	-	expression tag	UNP Q09KJ1

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. 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. 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.

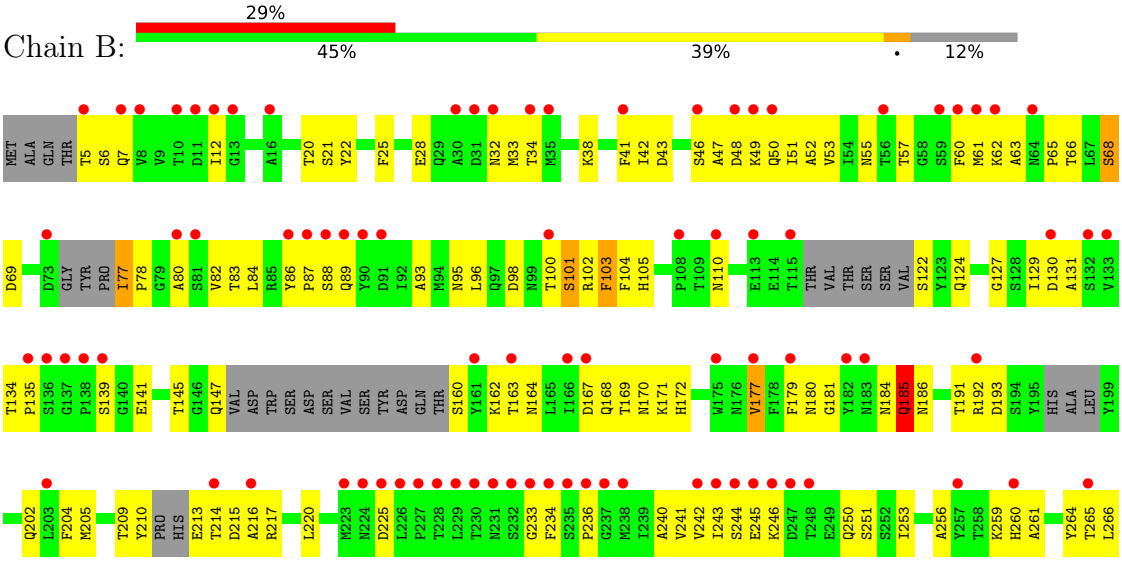
• Molecule 1: Cytotoxin K

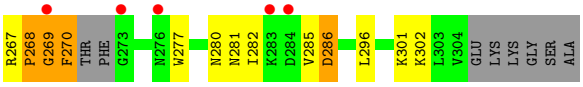
Chain A:



• Molecule 1: Cytotoxin K

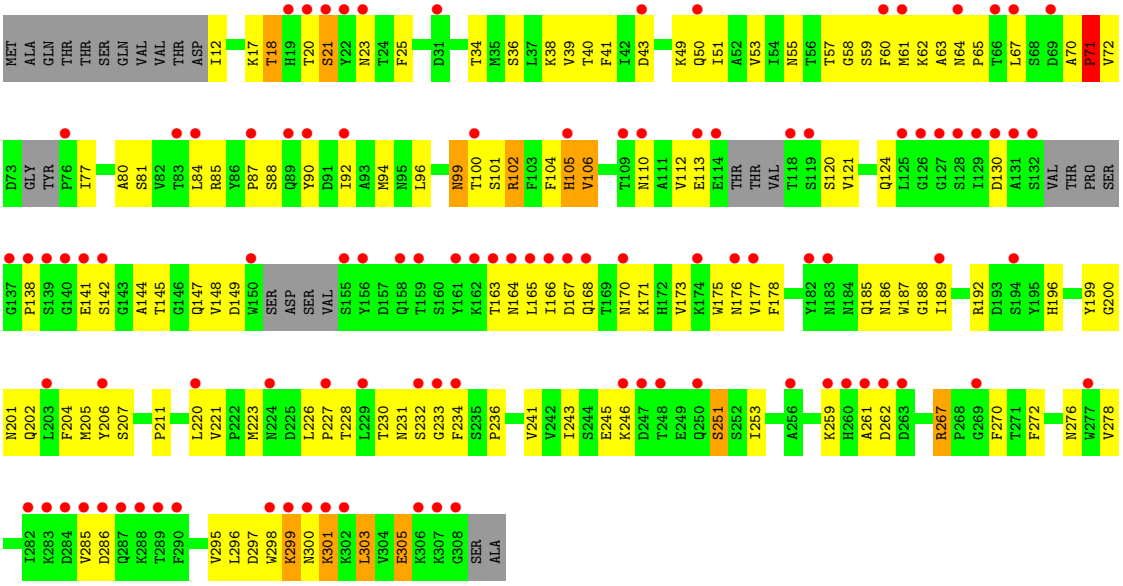
Chain B:





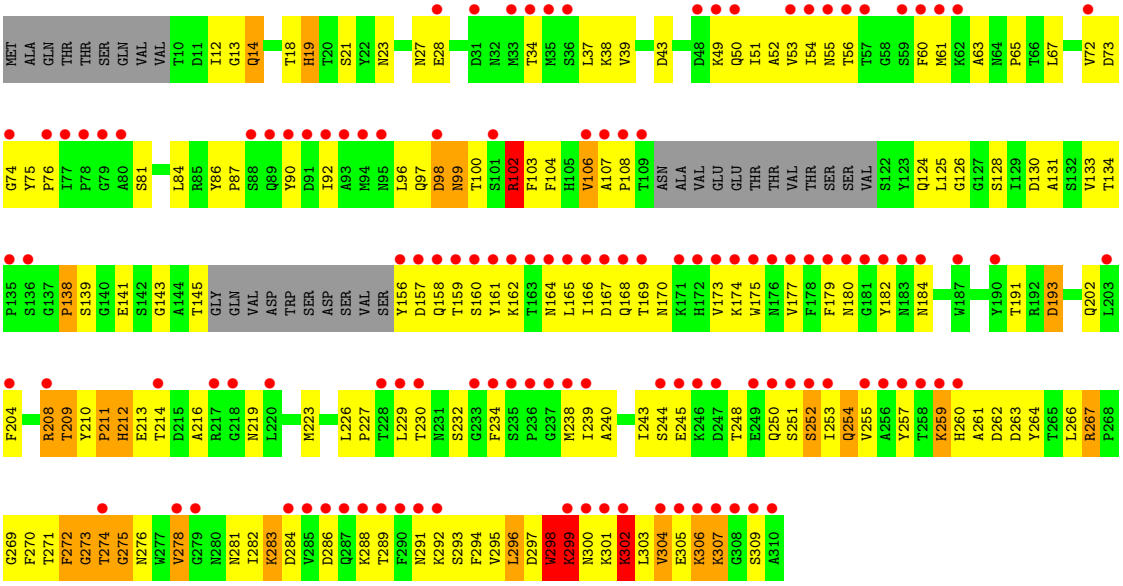
• Molecule 1: Cytotoxin K

Chain C:



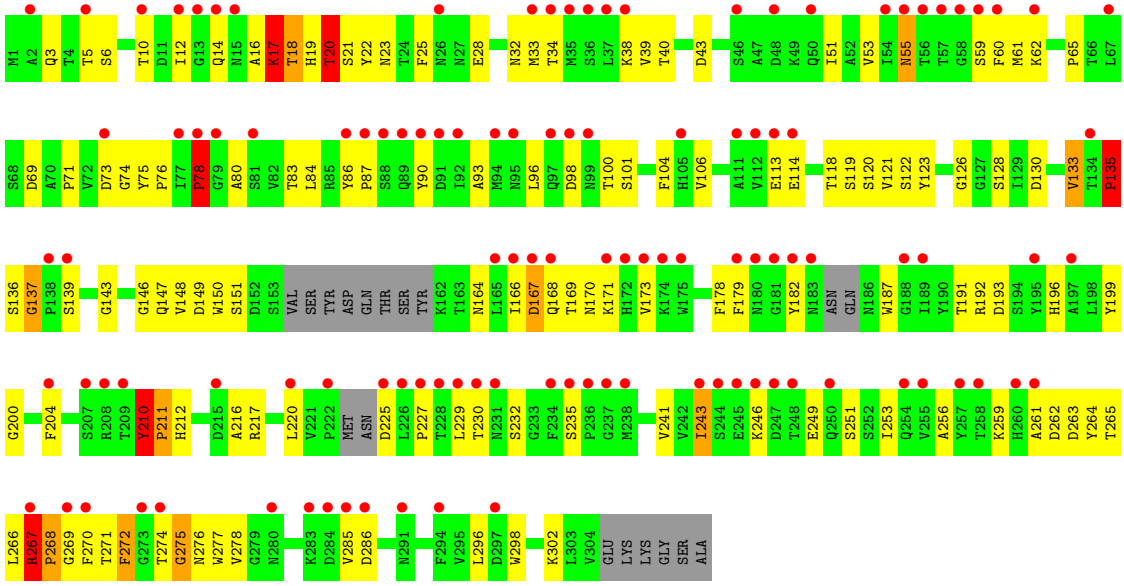
• Molecule 1: Cytotoxin K

Chain D:

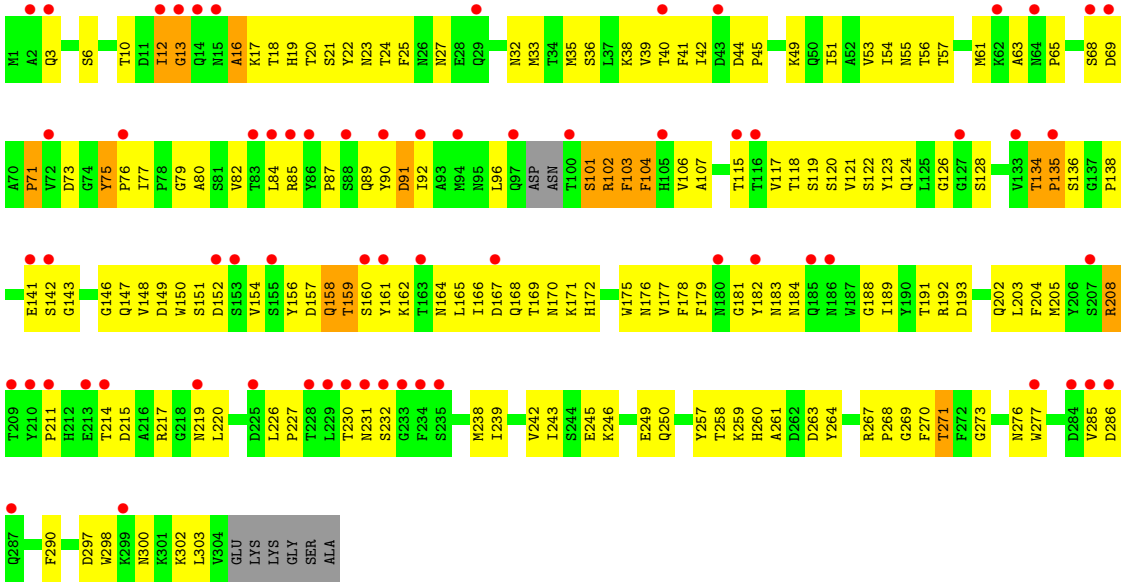


• Molecule 1: Cytotoxin K

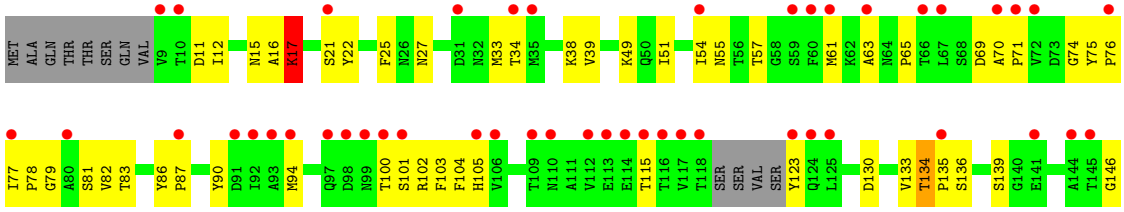
Chain E:

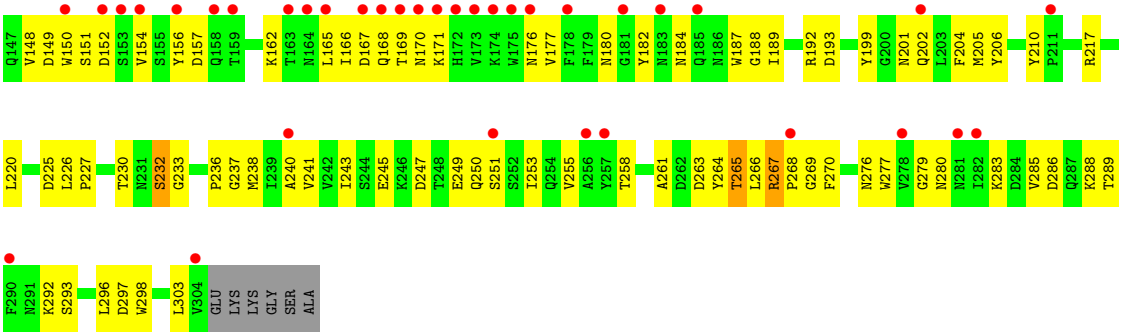


• Molecule 1: Cytotoxin K



• Molecule 1: Cytotoxin K




Q147 V148 D149 W150 S151 D152 S153 V154 S155 Y156 D157 Q158 T159 K162 T163 N164 L165 I166 D167 Q168 T169 N170 K171 H172 V173 K174 W175 N176 V177 F178 F179 N180 G181 Y182 N183 N184 Q185 N186 W187 G188 I189 R192 D193 Y199 G200 N201 Q202 L203 F204 M205 Y206 Y210 P211 R217
L220 D225 L226 P227 T230 N231 S232 G233 P236 G237 M238 I239 A240 V241 V242 I243 S244 E245 K246 D247 T248 E249 Q250 S251 S252 I253 Q254 V255 A256 Y257 T258 A261 D262 D263 Y264 T265 L266 R267 P268 G269 F270 N276 W277 V278 G279 N280 N281 I282 K283 D284 V285 D286 Q287 K288 T289
P290 N291 K292 S293 L296 D297 W298 L303 V304 GLU LYS LYS GLY SER ALA

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	103.23Å 186.22Å 104.53Å 90.00° 115.85° 90.00°	Depositor
Resolution (Å)	94.07 – 4.32 94.07 – 4.32	Depositor EDS
% Data completeness (in resolution range)	95.5 (94.07-4.32) 97.0 (94.07-4.32)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.42 (at 4.30Å)	Xtriage
Refinement program	PHENIX (1.20.1_4487: ???)	Depositor
R, R_{free}	0.322 , 0.407 0.322 , 0.407	Depositor DCC
R_{free} test set	1942 reflections (5.49%)	wwPDB-VP
Wilson B-factor (Å^2)	84.7	Xtriage
Anisotropy	0.135	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.22 , 178.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.37$, $< L^2 > = 0.19$	Xtriage
Estimated twinning fraction	0.059 for l,-k,h	Xtriage
F_o,F_c correlation	0.65	EDS
Total number of atoms	15699	wwPDB-VP
Average B, all atoms (Å^2)	89.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.91% 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

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	0.43	0/2318	0.68	1/3153 (0.0%)
1	B	0.33	0/2161	0.63	0/2933
1	C	0.37	0/2278	0.65	1/3093 (0.0%)
1	D	0.45	0/2237	0.69	0/3042
1	E	0.38	0/2319	0.67	2/3159 (0.1%)
1	F	0.32	0/2407	0.65	0/3282
1	G	0.33	0/2340	0.63	0/3191
All	All	0.38	0/16060	0.66	4/21853 (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
1	A	0	2
1	B	0	1
1	C	0	1
1	D	0	3
1	E	0	1
1	F	0	1
1	G	0	1
All	All	0	10

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	E	78	PRO	N-CA-CB	-7.71	94.05	103.30
1	E	135	PRO	N-CA-CB	-7.28	94.56	103.30
1	C	71	PRO	N-CA-CB	-6.28	95.69	102.60
1	A	104	PHE	N-CA-CB	5.22	119.99	110.60

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	102	ARG	Sidechain
1	A	190	TYR	Mainchain
1	B	267	ARG	Sidechain
1	C	102	ARG	Sidechain
1	D	102	ARG	Sidechain

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	2267	0	2148	149	0
1	B	2119	0	2006	121	0
1	C	2227	0	2102	141	0
1	D	2184	0	2067	113	0
1	E	2267	0	2149	118	0
1	F	2351	0	2226	149	0
1	G	2284	0	2151	103	0
All	All	15699	0	14849	805	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 26.

The worst 5 of 805 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:C:296:LEU:HA	1:C:303:LEU:HD23	1.28	1.12
1:D:87:PRO:HA	1:D:261:ALA:HA	1.12	1.11
1:A:72:VAL:HA	1:A:79:GLY:HA2	1.12	1.11
1:A:96:LEU:HD21	1:A:101:SER:H	1.21	1.05
1:G:265:THR:OG1	1:G:267:ARG:HG2	1.58	1.01

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	283/310 (91%)	222 (78%)	42 (15%)	19 (7%)	1 12
1	B	258/310 (83%)	223 (86%)	31 (12%)	4 (2%)	8 37
1	C	274/310 (88%)	231 (84%)	37 (14%)	6 (2%)	5 30
1	D	273/310 (88%)	206 (76%)	52 (19%)	15 (6%)	1 15
1	E	284/310 (92%)	235 (83%)	38 (13%)	11 (4%)	2 19
1	F	298/310 (96%)	257 (86%)	34 (11%)	7 (2%)	5 29
1	G	288/310 (93%)	247 (86%)	36 (12%)	5 (2%)	7 36
All	All	1958/2170 (90%)	1621 (83%)	270 (14%)	67 (3%)	3 21

5 of 67 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	78	PRO
1	A	98	ASP
1	A	103	PHE
1	A	104	PHE
1	A	135	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	254/270 (94%)	230 (91%)	24 (9%)	7 23
1	B	237/270 (88%)	225 (95%)	12 (5%)	20 42

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	C	247/270 (92%)	236 (96%)	11 (4%)	23 46
1	D	242/270 (90%)	203 (84%)	39 (16%)	2 12
1	E	254/270 (94%)	238 (94%)	16 (6%)	15 36
1	F	264/270 (98%)	254 (96%)	10 (4%)	28 50
1	G	255/270 (94%)	242 (95%)	13 (5%)	20 42
All	All	1753/1890 (93%)	1628 (93%)	125 (7%)	12 33

5 of 125 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	177	VAL
1	F	159	THR
1	D	282	ILE
1	F	158	GLN
1	G	180	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	14	GLN
1	D	276	ASN
1	F	168	GLN
1	D	300	ASN
1	A	201	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 oligosaccharides 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

Warning: The R factor obtained from EDS is 0.376, which does not match the depositor's R factor of 0.322. Please interpret the results in this section carefully.

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	291/310 (93%)	1.79	98 (33%)	1	2	30, 84, 154, 262	0
1	B	272/310 (87%)	1.96	90 (33%)	1	2	39, 89, 150, 256	0
1	C	284/310 (91%)	2.20	102 (35%)	1	2	30, 93, 166, 237	0
1	D	279/310 (90%)	2.74	126 (45%)	1	1	30, 82, 141, 167	0
1	E	292/310 (94%)	2.19	114 (39%)	1	2	30, 80, 134, 204	0
1	F	302/310 (97%)	1.39	66 (21%)	3	5	30, 85, 137, 197	0
1	G	292/310 (94%)	1.81	83 (28%)	1	3	43, 80, 137, 203	0
All	All	2012/2170 (92%)	2.00	679 (33%)	1	2	30, 85, 145, 262	0

The worst 5 of 679 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	128	SER	23.3
1	C	140	GLY	19.4
1	C	130	ASP	18.4
1	C	141	GLU	18.4
1	D	165	LEU	16.8

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.