



wwPDB X-ray Structure Validation Summary Report

Aug 22, 2023 – 09:18 PM EDT

PDB ID : 3C9I
Title : Structure of P22 Tail-Needle GP26 Bound to Xenon Gas
Authors : Cingolani, G.; Olia, A.S.
Deposited on : 2008-02-15
Resolution : 1.95 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
Xtrriage (Phenix) : 1.13
EDS : 2.35
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

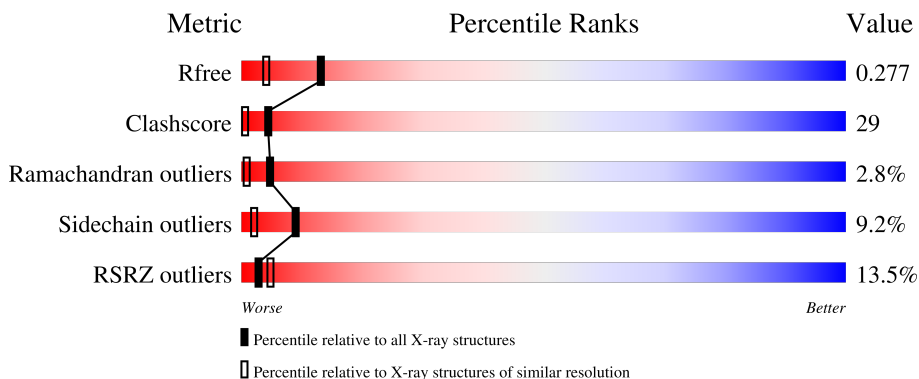
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.95 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	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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	242	
1	B	242	
1	C	242	
1	D	242	
1	E	242	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	242	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	XE	A	7	-	-	X	-
2	XE	D	11	-	-	X	-
2	XE	D	14	-	-	X	-

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 11851 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 Tail needle protein gp26.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	233	Total 1738	C 1065	N 311	O 360	S 2	0	0	0
1	B	242	Total 1800	C 1103	N 325	O 370	S 2	0	0	0
1	C	233	Total 1735	C 1063	N 311	O 359	S 2	0	0	0
1	D	221	Total 1645	C 1009	N 297	O 337	S 2	0	0	0
1	E	226	Total 1681	C 1029	N 307	O 343	S 2	0	0	0
1	F	221	Total 1644	C 1005	N 298	O 339	S 2	0	0	0

There are 66 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	992	GLY	-	expression tag	UNP P35837
A	993	ALA	-	expression tag	UNP P35837
A	994	LEU	-	expression tag	UNP P35837
A	995	VAL	-	expression tag	UNP P35837
A	996	PRO	-	expression tag	UNP P35837
A	997	ARG	-	expression tag	UNP P35837
A	998	GLY	-	expression tag	UNP P35837
A	999	SER	-	expression tag	UNP P35837
A	1000	HIS	-	expression tag	UNP P35837
A	1169	VAL	LEU	engineered mutation	UNP P35837
A	1222	MET	LEU	engineered mutation	UNP P35837
B	992	GLY	-	expression tag	UNP P35837
B	993	ALA	-	expression tag	UNP P35837
B	994	LEU	-	expression tag	UNP P35837
B	995	VAL	-	expression tag	UNP P35837
B	996	PRO	-	expression tag	UNP P35837
B	997	ARG	-	expression tag	UNP P35837

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	998	GLY	-	expression tag	UNP P35837
B	999	SER	-	expression tag	UNP P35837
B	1000	HIS	-	expression tag	UNP P35837
B	1169	VAL	LEU	engineered mutation	UNP P35837
B	1222	MET	LEU	engineered mutation	UNP P35837
C	992	GLY	-	expression tag	UNP P35837
C	993	ALA	-	expression tag	UNP P35837
C	994	LEU	-	expression tag	UNP P35837
C	995	VAL	-	expression tag	UNP P35837
C	996	PRO	-	expression tag	UNP P35837
C	997	ARG	-	expression tag	UNP P35837
C	998	GLY	-	expression tag	UNP P35837
C	999	SER	-	expression tag	UNP P35837
C	1000	HIS	-	expression tag	UNP P35837
C	1169	VAL	LEU	engineered mutation	UNP P35837
C	1222	MET	LEU	engineered mutation	UNP P35837
D	992	GLY	-	expression tag	UNP P35837
D	993	ALA	-	expression tag	UNP P35837
D	994	LEU	-	expression tag	UNP P35837
D	995	VAL	-	expression tag	UNP P35837
D	996	PRO	-	expression tag	UNP P35837
D	997	ARG	-	expression tag	UNP P35837
D	998	GLY	-	expression tag	UNP P35837
D	999	SER	-	expression tag	UNP P35837
D	1000	HIS	-	expression tag	UNP P35837
D	1169	VAL	LEU	engineered mutation	UNP P35837
D	1222	MET	LEU	engineered mutation	UNP P35837
E	992	GLY	-	expression tag	UNP P35837
E	993	ALA	-	expression tag	UNP P35837
E	994	LEU	-	expression tag	UNP P35837
E	995	VAL	-	expression tag	UNP P35837
E	996	PRO	-	expression tag	UNP P35837
E	997	ARG	-	expression tag	UNP P35837
E	998	GLY	-	expression tag	UNP P35837
E	999	SER	-	expression tag	UNP P35837
E	1000	HIS	-	expression tag	UNP P35837
E	1169	VAL	LEU	engineered mutation	UNP P35837
E	1222	MET	LEU	engineered mutation	UNP P35837
F	992	GLY	-	expression tag	UNP P35837
F	993	ALA	-	expression tag	UNP P35837
F	994	LEU	-	expression tag	UNP P35837
F	995	VAL	-	expression tag	UNP P35837

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	996	PRO	-	expression tag	UNP P35837
F	997	ARG	-	expression tag	UNP P35837
F	998	GLY	-	expression tag	UNP P35837
F	999	SER	-	expression tag	UNP P35837
F	1000	HIS	-	expression tag	UNP P35837
F	1169	VAL	LEU	engineered mutation	UNP P35837
F	1222	MET	LEU	engineered mutation	UNP P35837

- Molecule 2 is XENON (three-letter code: XE) (formula: Xe).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	6	Total Xe 6 6	0	0
2	C	1	Total Xe 1 1	0	0
2	D	6	Total Xe 6 6	0	0
2	F	1	Total Xe 1 1	0	0

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

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

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total Cl 1 1	0	0
4	D	1	Total Cl 1 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	266	Total O 266 266	0	0

Continued on next page...

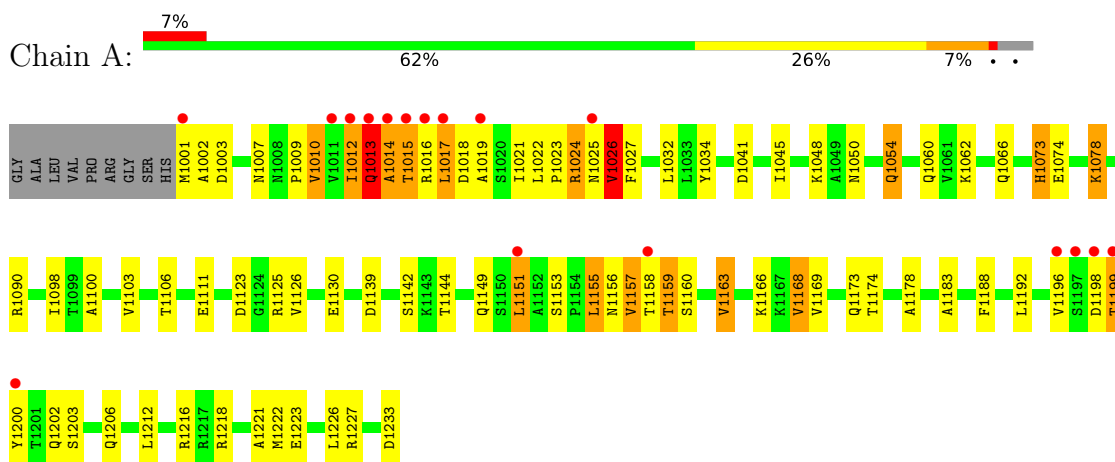
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	272	Total 272	O 272	0	0
5	C	277	Total 277	O 277	0	0
5	D	211	Total 211	O 211	0	0
5	E	275	Total 275	O 275	0	0
5	F	289	Total 289	O 289	0	0

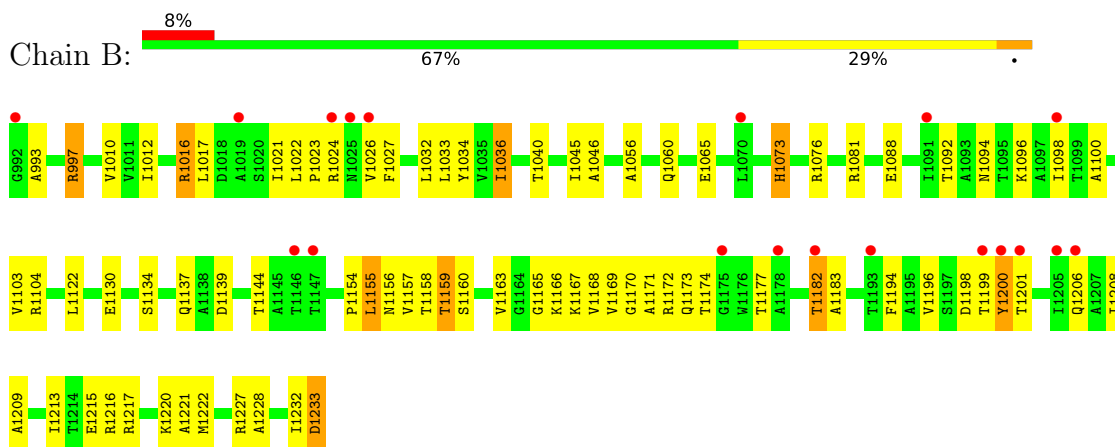
3 Residue-property plots [i](#)

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.

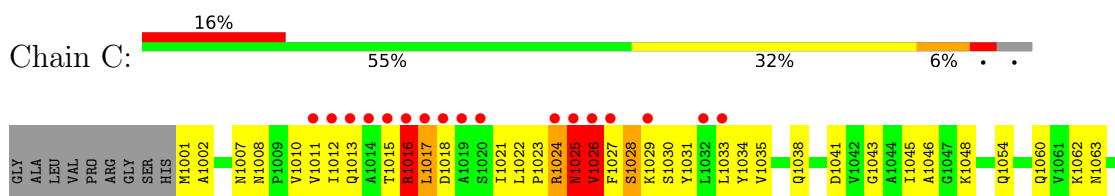
- Molecule 1: Tail needle protein gp26

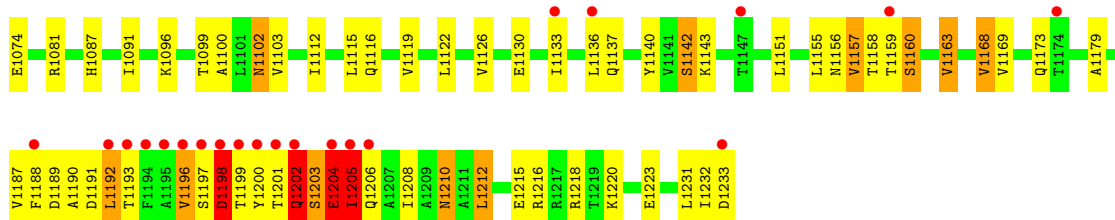


- Molecule 1: Tail needle protein gp26

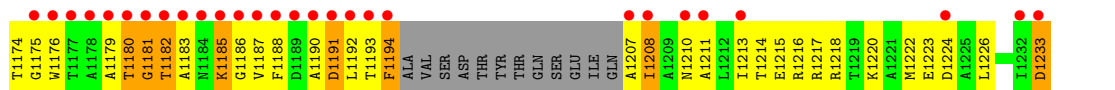
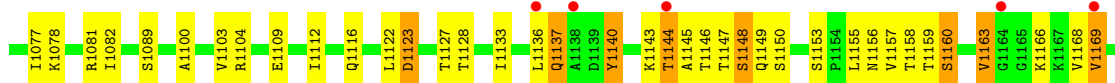
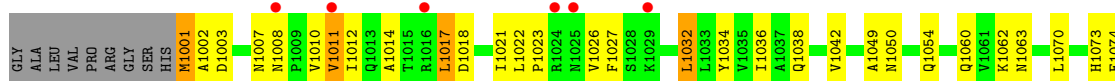


- Molecule 1: Tail needle protein gp26

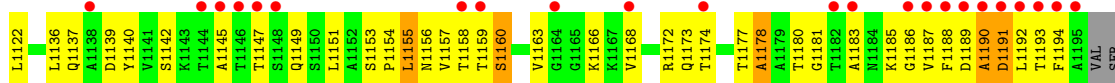




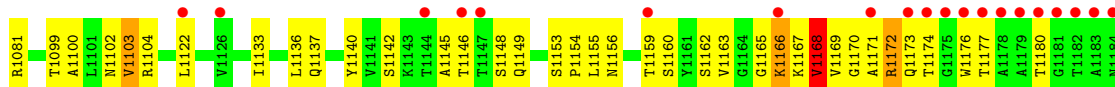
• Molecule 1: Tail needle protein gp26

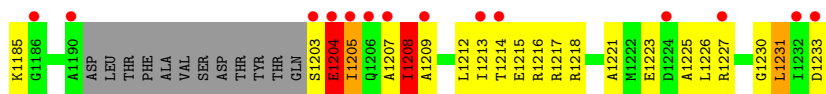


• Molecule 1: Tail needle protein gp26



• Molecule 1: Tail needle protein gp26





4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	40.54Å 114.53Å 172.04Å 90.00° 91.01° 90.00°	Depositor
Resolution (Å)	20.00 – 1.95 14.99 – 1.95	Depositor EDS
% Data completeness (in resolution range)	92.5 (20.00-1.95) 91.5 (14.99-1.95)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.17 (at 1.95Å)	Xtrriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.234 , 0.269 0.244 , 0.277	Depositor DCC
R_{free} test set	5233 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	21.8	Xtrriage
Anisotropy	0.828	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 67.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.028 for h,-k,-l	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	11851	wwPDB-VP
Average B, all atoms (Å ²)	40.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.95% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CL, XE, 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	0.34	0/1755	0.57	0/2387
1	B	0.34	0/1819	0.61	0/2474
1	C	0.33	0/1752	0.69	2/2382 (0.1%)
1	D	0.43	2/1660 (0.1%)	0.62	1/2255 (0.0%)
1	E	0.31	0/1697	0.59	0/2304
1	F	0.32	0/1658	0.58	0/2250
All	All	0.35	2/10341 (0.0%)	0.61	3/14052 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	1169	VAL	CA-CB	8.12	1.71	1.54
1	D	1169	VAL	C-O	5.01	1.32	1.23

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	1205	ILE	N-CA-C	-7.39	91.05	111.00
1	C	1198	ASP	CB-CA-C	6.78	123.95	110.40
1	D	1180	THR	N-CA-C	5.60	126.12	111.00

There are no chirality outliers.

There are no planarity outliers.

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	1738	0	1728	122	0
1	B	1800	0	1791	102	0
1	C	1735	0	1722	145	0
1	D	1645	0	1643	145	0
1	E	1681	0	1676	108	0
1	F	1644	0	1642	124	0
2	A	6	0	0	2	0
2	C	1	0	0	1	0
2	D	6	0	0	4	0
2	F	1	0	0	0	0
3	A	1	0	0	0	0
3	D	1	0	0	0	0
4	A	1	0	0	0	0
4	D	1	0	0	0	0
5	A	266	0	0	13	0
5	B	272	0	0	7	0
5	C	277	0	0	16	0
5	D	211	0	0	12	0
5	E	275	0	0	15	0
5	F	289	0	0	19	0
All	All	11851	0	10202	590	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 29.

The worst 5 of 590 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:D:1193:THR:HG21	1:F:1216:ARG:CD	1.60	1.32
1:A:1199:THR:HB	5:A:1472:HOH:O	1.45	1.16
1:C:1024:ARG:CD	1:C:1024:ARG:H	1.64	1.07
1:D:1193:THR:CG2	1:F:1216:ARG:HD2	1.86	1.06
1:C:1024:ARG:H	1:C:1024:ARG:HD2	1.00	1.05

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	231/242 (96%)	219 (95%)	7 (3%)	5 (2%)	6 1
1	B	240/242 (99%)	231 (96%)	5 (2%)	4 (2%)	9 2
1	C	231/242 (96%)	206 (89%)	13 (6%)	12 (5%)	2 0
1	D	217/242 (90%)	197 (91%)	14 (6%)	6 (3%)	5 1
1	E	222/242 (92%)	208 (94%)	9 (4%)	5 (2%)	6 1
1	F	217/242 (90%)	205 (94%)	6 (3%)	6 (3%)	5 1
All	All	1358/1452 (94%)	1266 (93%)	54 (4%)	38 (3%)	5 1

5 of 38 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1014	ALA
1	A	1159	THR
1	A	1199	THR
1	B	1200	TYR
1	C	1016	ARG

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	184/190 (97%)	164 (89%)	20 (11%)	6 1
1	B	190/190 (100%)	180 (95%)	10 (5%)	22 10
1	C	183/190 (96%)	163 (89%)	20 (11%)	6 1

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	173/190 (91%)	155 (90%)	18 (10%)	7	1
1	E	176/190 (93%)	160 (91%)	16 (9%)	9	2
1	F	173/190 (91%)	158 (91%)	15 (9%)	10	2
All	All	1079/1140 (95%)	980 (91%)	99 (9%)	9	2

5 of 99 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	1149	GLN
1	E	1015	THR
1	D	1163	VAL
1	D	1223	GLU
1	E	1035	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 51 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	1060	GLN
1	E	1000	HIS
1	F	1149	GLN
1	D	1073	HIS
1	D	1132	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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 18 ligands modelled in this entry, 18 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 [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	233/242 (96%)	0.59	17 (7%) 15 23	15, 33, 63, 99	0
1	B	242/242 (100%)	0.64	19 (7%) 12 19	10, 34, 60, 88	0
1	C	233/242 (96%)	0.93	38 (16%) 1 2	13, 33, 86, 112	0
1	D	221/242 (91%)	1.12	39 (17%) 1 1	18, 38, 80, 97	0
1	E	226/242 (93%)	1.00	38 (16%) 1 2	10, 37, 78, 98	0
1	F	221/242 (91%)	0.89	35 (15%) 2 2	18, 35, 73, 91	0
All	All	1376/1452 (94%)	0.86	186 (13%) 3 5	10, 35, 75, 112	0

The worst 5 of 186 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	998	GLY	10.8
1	D	1183	ALA	9.8
1	C	1201	THR	7.8
1	A	1013	GLN	7.8
1	F	1204	GLU	7.6

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
2	XE	D	14	1/1	0.75	0.12	122,122,122,122	0
2	XE	D	11	1/1	0.77	0.12	121,121,121,121	0
2	XE	A	7	1/1	0.92	0.06	96,96,96,96	0
2	XE	F	10	1/1	0.94	0.04	65,65,65,65	0
4	CL	D	4	1/1	0.97	0.11	33,33,33,33	0
3	CA	A	1	1/1	0.98	0.07	21,21,21,21	0
2	XE	D	13	1/1	0.99	0.05	35,35,35,35	0
2	XE	C	4	1/1	0.99	0.06	73,73,73,73	0
2	XE	D	8	1/1	0.99	0.02	33,33,33,33	0
2	XE	A	1234	1/1	0.99	0.05	59,59,59,59	0
4	CL	A	2	1/1	0.99	0.13	28,28,28,28	0
2	XE	D	12	1/1	0.99	0.02	35,35,35,35	0
2	XE	A	5	1/1	1.00	0.05	33,33,33,33	0
2	XE	D	9	1/1	1.00	0.04	32,32,32,32	0
2	XE	A	6	1/1	1.00	0.04	31,31,31,31	0
3	CA	D	3	1/1	1.00	0.10	23,23,23,23	0
2	XE	A	1235	1/1	1.00	0.05	34,34,34,34	0
2	XE	A	3	1/1	1.00	0.03	39,39,39,39	0

6.5 Other polymers [i](#)

There are no such residues in this entry.