



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 03:11 am GMT

PDB ID : 3CDD
Title : Crystal structure of prophage MuSo2, 43 kDa tail protein from *Shewanella oneidensis*
Authors : Chang, C.; Evdokimova, E.; Kudritska, M.; Savchenko, A.; Edwards, A.M.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on : 2008-02-26
Resolution : 2.10 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

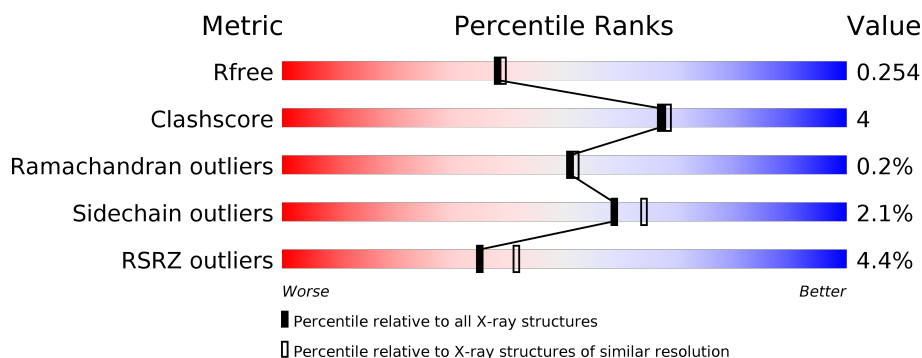
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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}	100719	4243 (2.10-2.10)
Clashscore	112137	4788 (2.10-2.10)
Ramachandran outliers	110173	4740 (2.10-2.10)
Sidechain outliers	110143	4741 (2.10-2.10)
RSRZ outliers	101464	4275 (2.10-2.10)

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 $\leq 5\%$. 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	361	<div> <div>6%</div> <div> <div></div> <div>79%</div> <div>10%</div> <div>9%</div> </div> </div>
1	B	361	<div> <div>2%</div> <div> <div></div> <div>83%</div> <div>7%</div> <div>10%</div> </div> </div>
1	C	361	<div> <div>%</div> <div> <div></div> <div>80%</div> <div>10%</div> <div>9%</div> </div> </div>
1	D	361	<div> <div>4%</div> <div> <div></div> <div>78%</div> <div>12%</div> <div>8%</div> </div> </div>
1	E	361	<div> <div>4%</div> <div> <div></div> <div>83%</div> <div>6%</div> <div>11%</div> </div> </div>
1	F	361	<div> <div>6%</div> <div> <div></div> <div>79%</div> <div>11%</div> <div>10%</div> </div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 16250 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 Prophage MuSo2, 43 kDa tail protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	328	Total	C	N	O	S	Se	1	8	0
			2539	1606	431	492	3	7			
1	B	326	Total	C	N	O	S	Se	0	5	0
			2517	1590	431	487	3	6			
1	C	328	Total	C	N	O	S	Se	2	6	0
			2546	1609	435	490	3	9			
1	D	331	Total	C	N	O	S	Se	1	4	0
			2540	1604	439	486	3	8			
1	E	322	Total	C	N	O	S	Se	1	8	0
			2508	1581	428	489	3	7			
1	F	325	Total	C	N	O	S	Se	0	4	0
			2510	1588	432	481	3	6			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLN	-	EXPRESSION TAG	UNP Q8EDP4
A	0	GLY	-	EXPRESSION TAG	UNP Q8EDP4
A	1	HIS	-	EXPRESSION TAG	UNP Q8EDP4
B	-1	GLN	-	EXPRESSION TAG	UNP Q8EDP4
B	0	GLY	-	EXPRESSION TAG	UNP Q8EDP4
B	1	HIS	-	EXPRESSION TAG	UNP Q8EDP4
C	-1	GLN	-	EXPRESSION TAG	UNP Q8EDP4
C	0	GLY	-	EXPRESSION TAG	UNP Q8EDP4
C	1	HIS	-	EXPRESSION TAG	UNP Q8EDP4
D	-1	GLN	-	EXPRESSION TAG	UNP Q8EDP4
D	0	GLY	-	EXPRESSION TAG	UNP Q8EDP4
D	1	HIS	-	EXPRESSION TAG	UNP Q8EDP4
E	-1	GLN	-	EXPRESSION TAG	UNP Q8EDP4
E	0	GLY	-	EXPRESSION TAG	UNP Q8EDP4
E	1	HIS	-	EXPRESSION TAG	UNP Q8EDP4
F	-1	GLN	-	EXPRESSION TAG	UNP Q8EDP4
F	0	GLY	-	EXPRESSION TAG	UNP Q8EDP4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	1	HIS	-	EXPRESSION TAG	UNP Q8EDP4

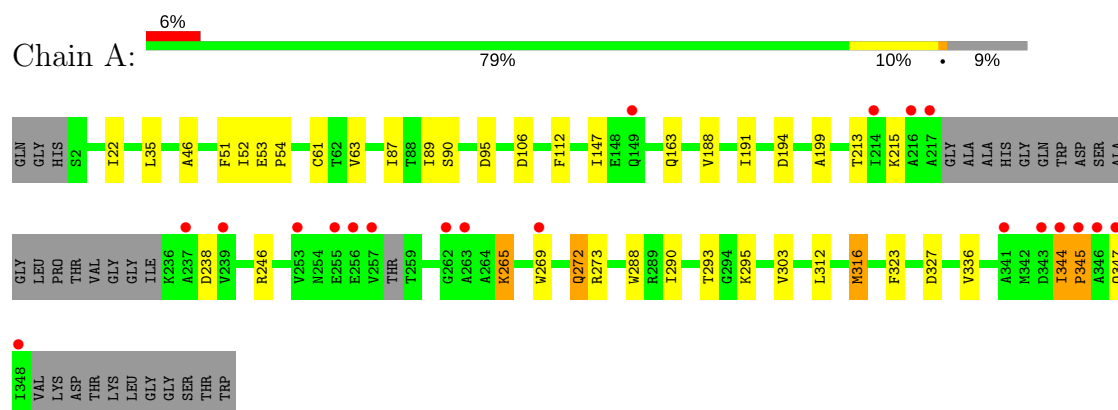
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	225	Total 225	O 225	0	0
2	B	166	Total 166	O 166	0	0
2	C	174	Total 174	O 174	0	0
2	D	183	Total 183	O 183	0	0
2	E	199	Total 199	O 199	0	0
2	F	143	Total 143	O 143	0	0

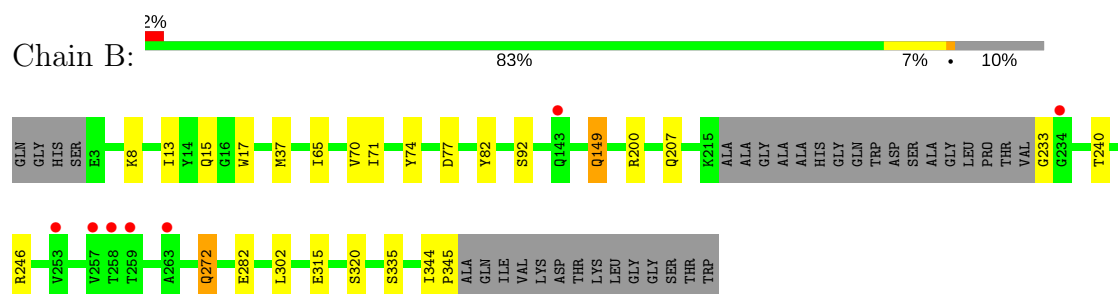
3 Residue-property plots [i](#)

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.

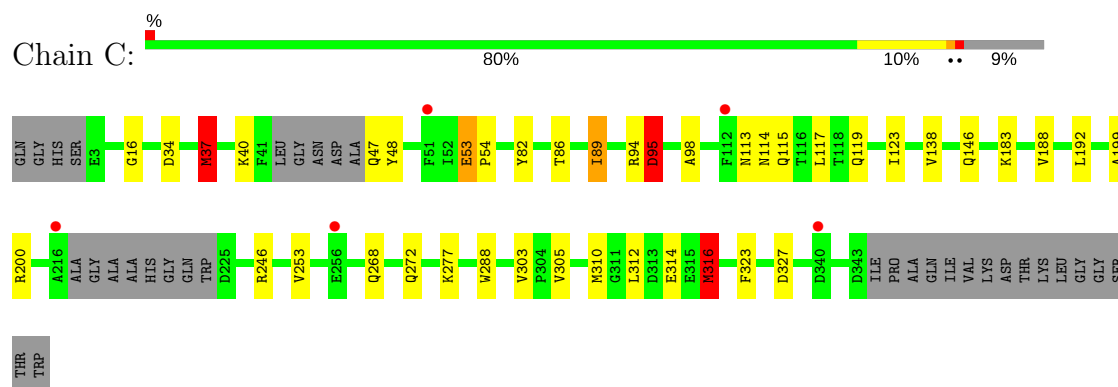
- Molecule 1: Prophage MuSo2, 43 kDa tail protein



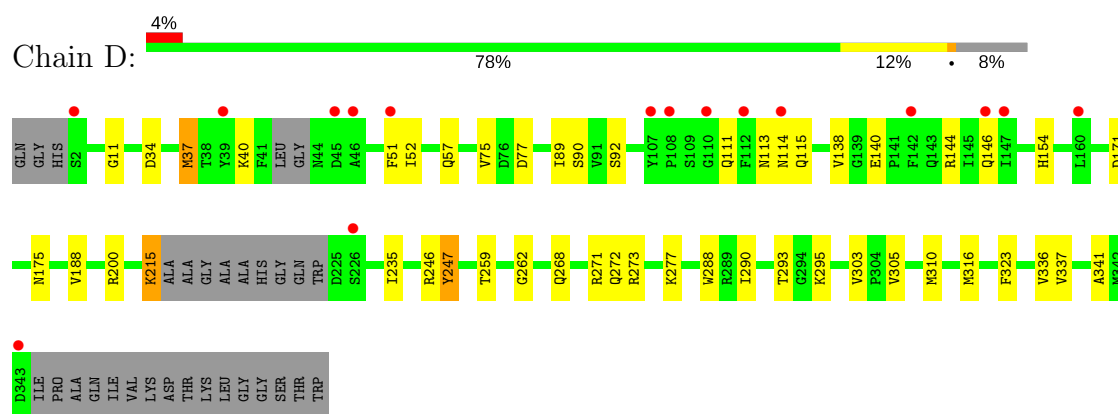
- Molecule 1: Prophage MuSo2, 43 kDa tail protein



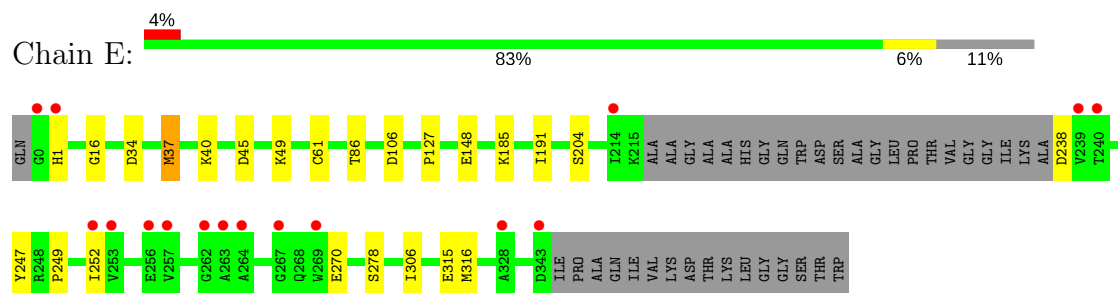
- Molecule 1: Prophage MuSo2, 43 kDa tail protein



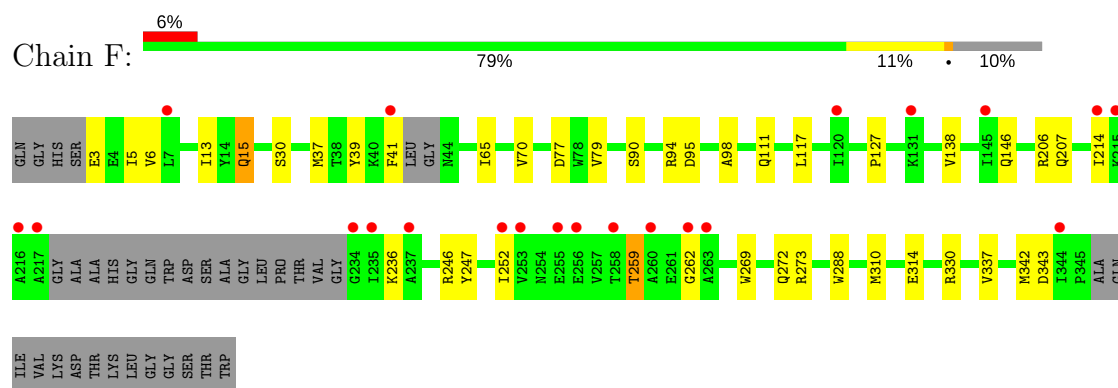
- Molecule 1: Prophage MuSo2, 43 kDa tail protein



- Molecule 1: Prophage MuSo2, 43 kDa tail protein



- Molecule 1: Prophage MuSo2, 43 kDa tail protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	96.34Å 82.44Å 151.83Å 90.00° 100.20° 90.00°	Depositor
Resolution (Å)	47.69 – 2.10 47.69 – 2.10	Depositor EDS
% Data completeness (in resolution range)	96.6 (47.69-2.10) 96.6 (47.69-2.10)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.38 (at 2.10Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.201 , 0.257 0.200 , 0.254	Depositor DCC
R_{free} test set	6655 reflections (5.29%)	DCC
Wilson B-factor (Å ²)	36.5	Xtriage
Anisotropy	0.562	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 36.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	16250	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 25.62 % 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 3.0621e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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

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	0.83	6/2590 (0.2%)	0.75	6/3497 (0.2%)
1	B	0.58	1/2564 (0.0%)	0.63	1/3464 (0.0%)
1	C	1.42	9/2585 (0.3%)	1.06	9/3487 (0.3%)
1	D	1.96	8/2576 (0.3%)	0.84	5/3477 (0.1%)
1	E	0.56	1/2555 (0.0%)	0.69	3/3450 (0.1%)
1	F	0.51	0/2553	0.62	1/3444 (0.0%)
All	All	1.12	25/15423 (0.2%)	0.78	25/20819 (0.1%)

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	C	1	1

The worst 5 of 25 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	37[A]	MSE	SE-CE	53.91	5.13	1.95
1	D	37[B]	MSE	SE-CE	53.91	5.13	1.95
1	D	37[A]	MSE	CG-SE	-38.91	0.63	1.95
1	D	37[B]	MSE	CG-SE	-38.91	0.63	1.95
1	C	37[A]	MSE	CG-SE	-36.08	0.72	1.95

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	316[A]	MSE	CG-SE-CE	27.48	159.37	98.90
1	C	316[B]	MSE	CG-SE-CE	27.48	159.37	98.90
1	D	37[A]	MSE	CB-CG-SE	18.60	168.51	112.70
1	D	37[B]	MSE	CB-CG-SE	18.60	168.51	112.70
1	C	37[A]	MSE	CB-CG-SE	17.70	165.79	112.70

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	C	95	ASP	CA

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	94	ARG	Peptide

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	2539	0	2527	33	1
1	B	2517	0	2499	18	0
1	C	2546	0	2544	28	0
1	D	2540	0	2526	27	1
1	E	2508	0	2466	14	0
1	F	2510	0	2506	25	0
2	A	225	0	0	2	0
2	B	166	0	0	2	0
2	C	174	0	0	4	0
2	D	183	0	0	5	0
2	E	199	0	0	1	0
2	F	143	0	0	3	0
All	All	16250	0	15068	136	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 4.

The worst 5 of 136 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:A:265:LYS:CD	1:A:265:LYS:CE	1.90	1.48
1:A:265:LYS:NZ	1:A:265:LYS:CE	2.02	1.22
1:D:316[B]:MSE:HE2	1:D:336:VAL:HG23	1.43	1.00
1:A:213[A]:THR:HG22	1:A:238:ASP:OD1	1.62	0.98

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:293[A]:THR:HG23	1:A:295:LYS:H	1.38	0.87

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:A:327:ASP:OD2	1:D:11:GLY:N[2_746]	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	330/361 (91%)	314 (95%)	15 (4%)	1 (0%)	44	44
1	B	327/361 (91%)	314 (96%)	13 (4%)	0	100	100
1	C	328/361 (91%)	311 (95%)	16 (5%)	1 (0%)	44	44
1	D	329/361 (91%)	314 (95%)	13 (4%)	2 (1%)	28	24
1	E	326/361 (90%)	308 (94%)	18 (6%)	0	100	100
1	F	323/361 (90%)	313 (97%)	10 (3%)	0	100	100
All	All	1963/2166 (91%)	1874 (96%)	85 (4%)	4 (0%)	51	52

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	345	PRO
1	C	95	ASP
1	D	247	TYR
1	D	341	ALA

5.3.2 Protein sidechains ⓘ

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	270/290 (93%)	266 (98%)	4 (2%)	70	76
1	B	267/290 (92%)	263 (98%)	4 (2%)	70	76
1	C	271/290 (93%)	261 (96%)	10 (4%)	39	39
1	D	266/290 (92%)	256 (96%)	10 (4%)	38	38
1	E	266/290 (92%)	262 (98%)	4 (2%)	70	76
1	F	265/290 (91%)	259 (98%)	6 (2%)	56	60
All	All	1605/1740 (92%)	1567 (98%)	38 (2%)	59	59

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	327	ASP
1	D	77	ASP
1	F	259	THR
1	D	37[B]	MSE
1	D	90	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	272	GLN
1	B	47	GLN
1	B	149	GLN
1	D	154	HIS
1	F	207	GLN

5.3.3 RNA ⓘ

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 [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	322/361 (89%)	0.51	20 (6%) 21 27	25, 32, 43, 70	0
1	B	320/361 (88%)	0.32	7 (2%) 62 67	25, 34, 42, 55	0
1	C	322/361 (89%)	0.28	5 (1%) 72 76	16, 31, 42, 58	0
1	D	325/361 (90%)	0.50	16 (4%) 30 37	15, 32, 49, 57	0
1	E	316/361 (87%)	0.52	16 (5%) 29 35	26, 32, 51, 61	0
1	F	319/361 (88%)	0.45	21 (6%) 19 24	26, 34, 44, 58	0
All	All	1924/2166 (88%)	0.43	85 (4%) 35 41	15, 33, 46, 70	0

The worst 5 of 85 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	217	ALA	12.9
1	A	348	ILE	6.5
1	A	344	ILE	6.3
1	E	262	GLY	5.7
1	B	258	THR	5.3

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 carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.