wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 08:11 AM GMT

PDB ID	:	3DKT
Title	:	Crystal structure of Thermotoga maritima encapsulin
Authors	:	Sutter, M.; Boehringer, D.; Gutmann, S.; Weber-Ban, E.; Ban, N.
Deposited on	:	2008-06-26
Resolution	:	3.10 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 (i) symbol.

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

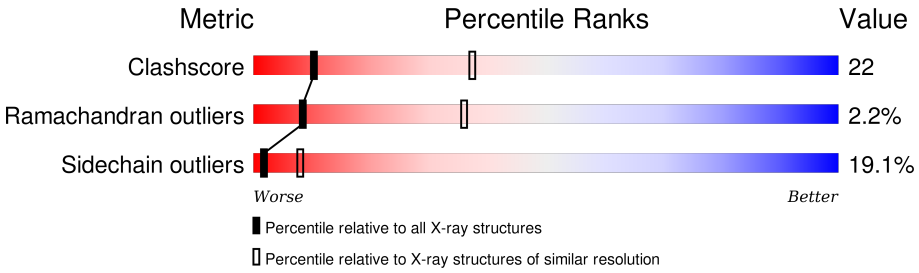
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	FAILED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 3.10 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))
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)

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 <=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.

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	265	55% 37% 7%
1	B	265	53% 38% 9%
1	C	265	52% 38% 9%
1	D	265	51% 40% 9%
1	E	265	54% 36% 9%
1	F	265	52% 38% 10%
1	G	265	54% 37% 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	H	265	53% 37% 9%
1	I	265	51% 38% 9%
1	J	265	52% 37% 9%
2	K	8	63% 38%
2	L	8	63% 25% 13%
2	M	8	63% 25% 13%
2	N	8	63% 38%
2	O	8	63% 25% 13%
2	P	8	63% 13% 25%
2	Q	8	63% 25% 13%
2	R	8	63% 13% 25%
2	S	8	63% 25% 13%
2	T	8	63% 25% 13%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 22070 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 Maritimacin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	264	Total	C	N	O	S	0	0	0
			2141	1372	358	407	4			
1	B	264	Total	C	N	O	S	0	0	0
			2141	1372	358	407	4			
1	C	264	Total	C	N	O	S	0	0	0
			2141	1372	358	407	4			
1	D	264	Total	C	N	O	S	0	0	0
			2141	1372	358	407	4			
1	E	264	Total	C	N	O	S	0	0	0
			2141	1372	358	407	4			
1	F	264	Total	C	N	O	S	0	0	0
			2141	1372	358	407	4			
1	G	264	Total	C	N	O	S	0	0	0
			2141	1372	358	407	4			
1	H	264	Total	C	N	O	S	0	0	0
			2141	1372	358	407	4			
1	I	264	Total	C	N	O	S	0	0	0
			2141	1372	358	407	4			
1	J	264	Total	C	N	O	S	0	0	0
			2141	1372	358	407	4			

- Molecule 2 is a protein called Putative uncharacterized protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	K	8	Total	C	N	O	0	0	0
			56	34	12	10			
2	L	8	Total	C	N	O	0	0	0
			56	34	12	10			
2	M	8	Total	C	N	O	0	0	0
			56	34	12	10			
2	N	8	Total	C	N	O	0	0	0
			56	34	12	10			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	O	8	Total 56	C 34	N 12	O 10	0	0	0
2	P	8	Total 56	C 34	N 12	O 10	0	0	0
2	Q	8	Total 56	C 34	N 12	O 10	0	0	0
2	R	8	Total 56	C 34	N 12	O 10	0	0	0
2	S	8	Total 56	C 34	N 12	O 10	0	0	0
2	T	8	Total 56	C 34	N 12	O 10	0	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	9	Total 9	O 9	0	0
3	B	11	Total 11	O 11	0	0
3	C	9	Total 9	O 9	0	0
3	D	12	Total 12	O 12	0	0
3	E	9	Total 9	O 9	0	0
3	F	11	Total 11	O 11	0	0
3	G	9	Total 9	O 9	0	0
3	H	11	Total 11	O 11	0	0
3	I	9	Total 9	O 9	0	0
3	J	10	Total 10	O 10	0	0

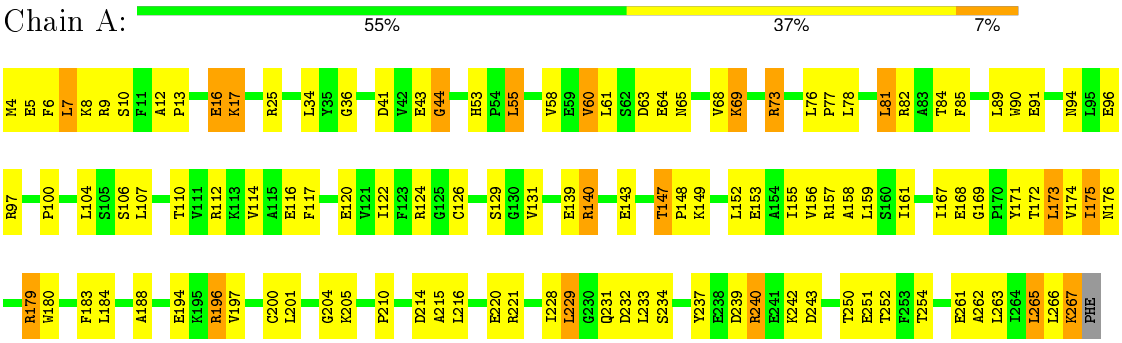
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

Note EDS failed to run properly.

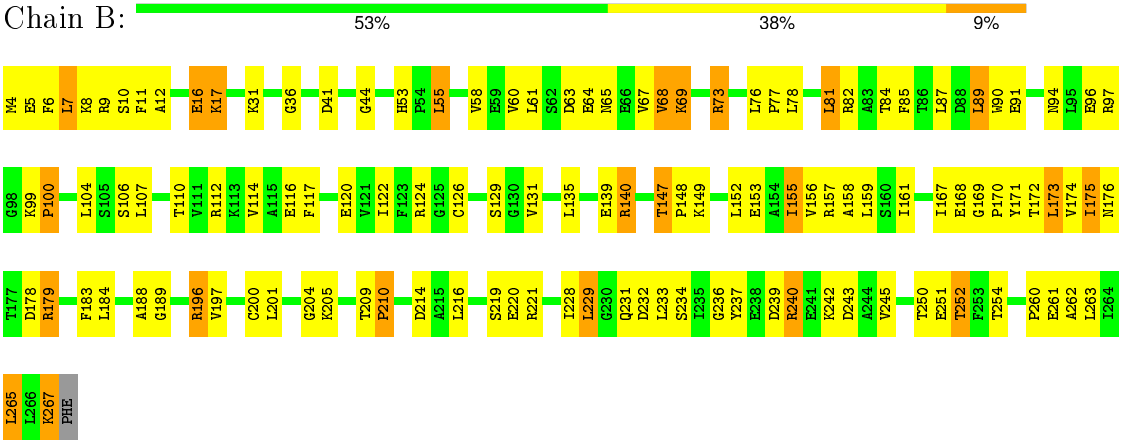
- Molecule 1: Maritimacin

Chain A:



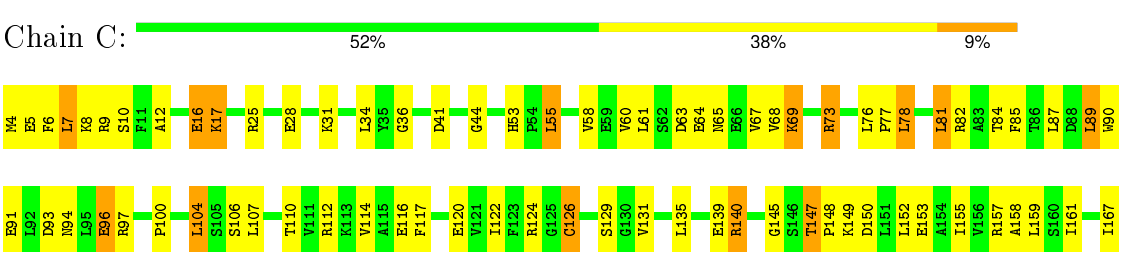
- Molecule 1: Maritimacin

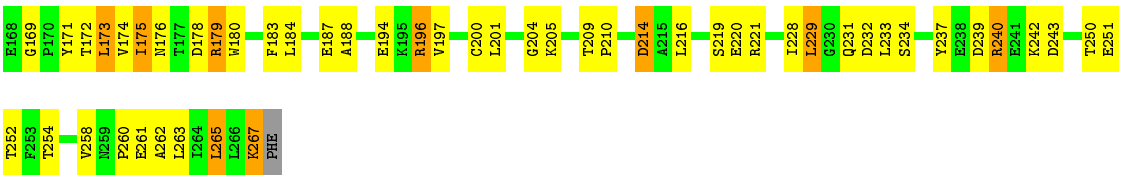
Chain B:



- Molecule 1: Maritimacin

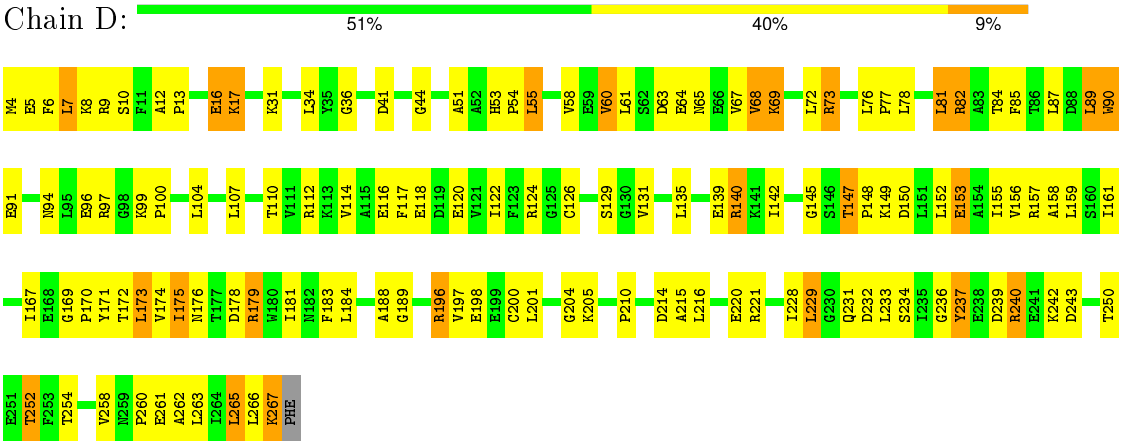
Chain C:





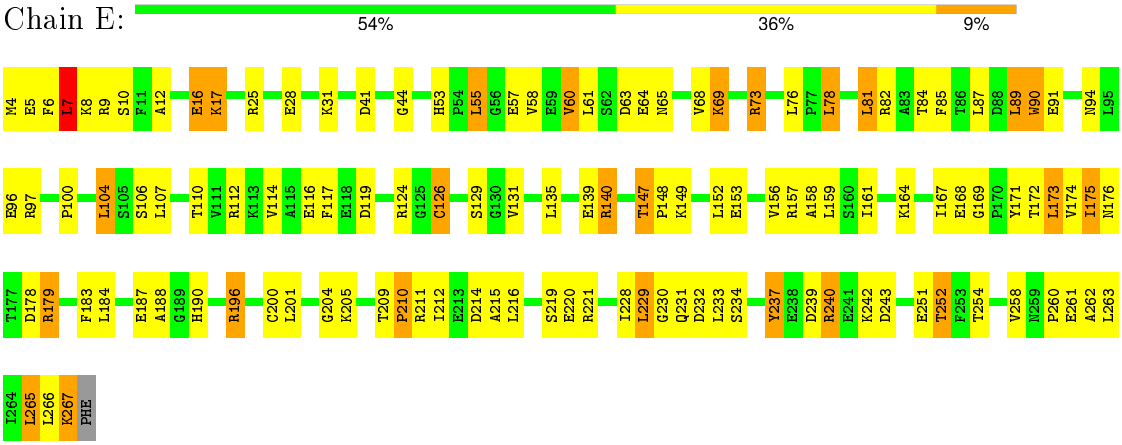
- Molecule 1: Maritimacin

Chain D:



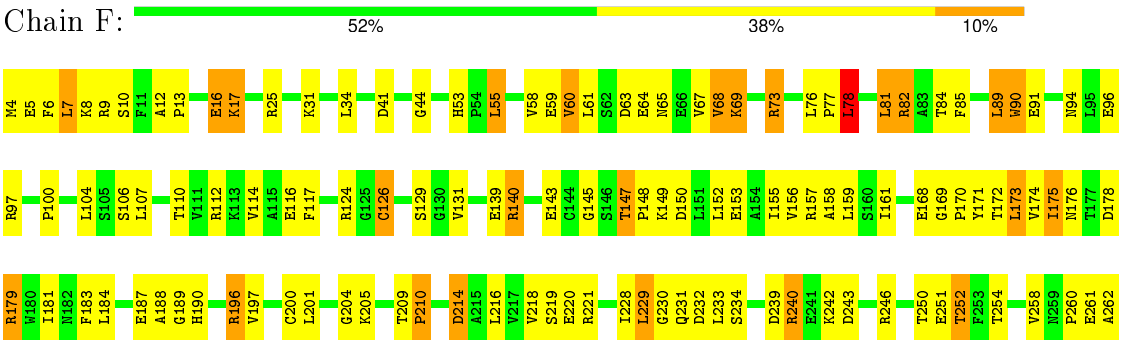
- Molecule 1: Maritimacin

Chain E:



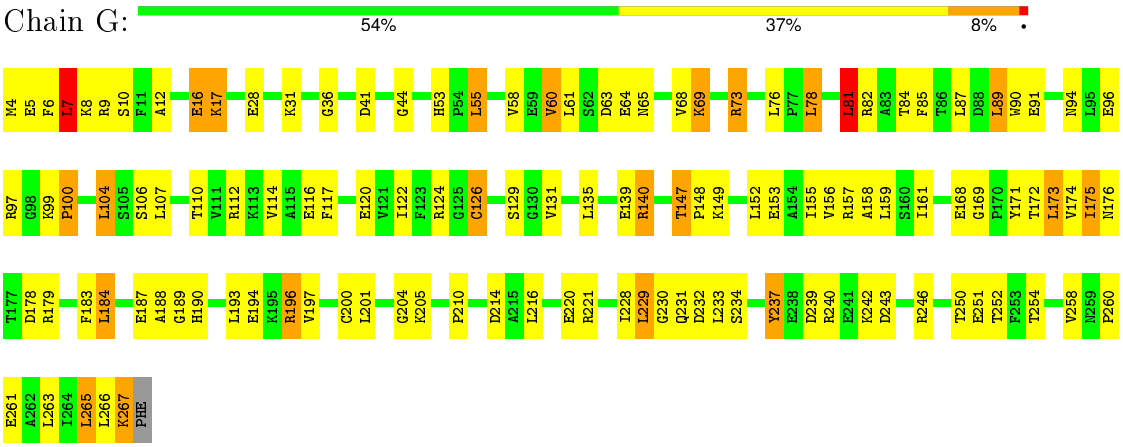
- Molecule 1: Maritimacin

Chain F:



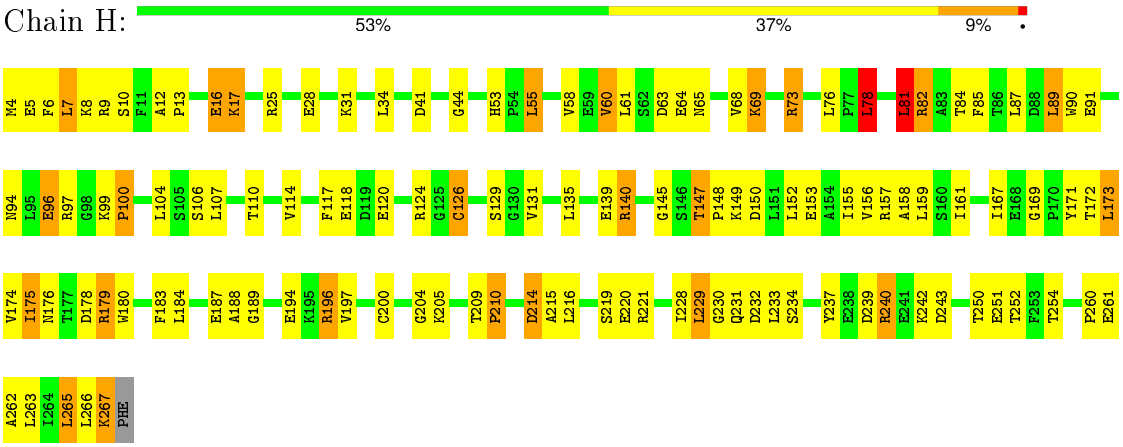
- Molecule 1: Maritimacin

Chain G:



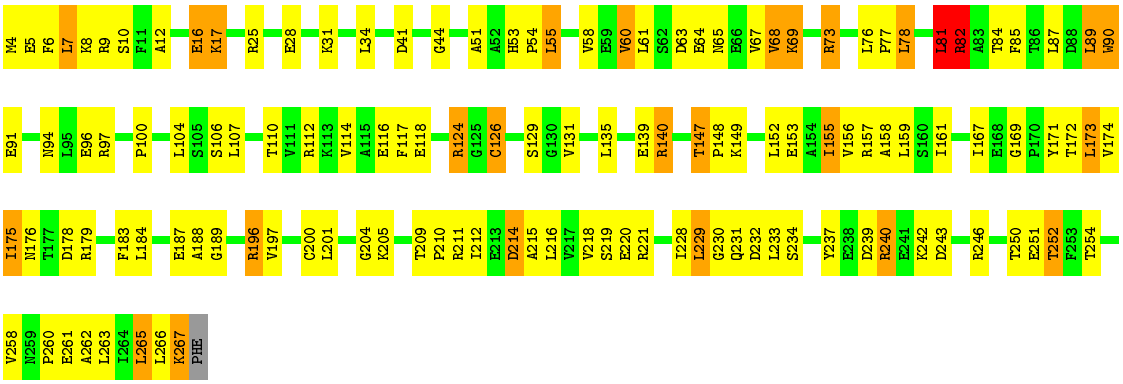
- Molecule 1: Maritimacin

Chain H:



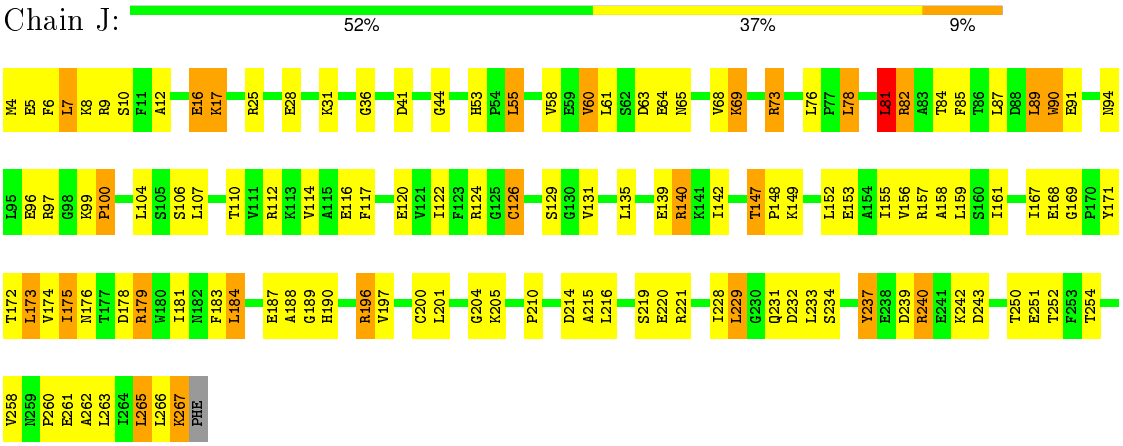
- Molecule 1: Maritimacin

Chain I:



- Molecule 1: Maritimacin

Chain J:



- Molecule 2: Putative uncharacterized protein

Chain K:

- Molecule 2: Putative uncharacterized protein

Chain L:

- Molecule 2: Putative uncharacterized protein

Chain M:

- Molecule 2: Putative uncharacterized protein

Chain N:

- Molecule 2: Putative uncharacterized protein

Chain O:

- Molecule 2: Putative uncharacterized protein

Chain P:

- Molecule 2: Putative uncharacterized protein

Chain Q:

- Molecule 2: Putative uncharacterized protein

Chain R:

- Molecule 2: Putative uncharacterized protein

Chain S:

- Molecule 2: Putative uncharacterized protein

Chain T:

4 Data and refinement statistics (i)

EDS failed to run properly - this section will therefore be incomplete.

Property	Value	Source
Space group	F 41 3 2	Depositor
Cell constants a, b, c, α, β, γ	669.04Å 669.04Å 669.04Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.87 – 3.10	Depositor
% Data completeness (in resolution range)	96.5 (49.87-3.10)	Depositor
R_{merge}	0.27	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.32 (at 3.12Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.219 , 0.239	Depositor
Wilson B-factor (Å^2)	94.4	Xtriage
Anisotropy	0.000	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.24$	Xtriage
Outliers	1 of 219258 reflections (0.000%)	Xtriage
Total number of atoms	22070	wwPDB-VP
Average B, all atoms (Å^2)	128.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.40% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.44	0/2182	0.66	0/2945
1	B	0.45	0/2182	0.66	0/2945
1	C	0.45	0/2182	0.67	0/2945
1	D	0.45	0/2182	0.66	0/2945
1	E	0.44	0/2182	0.66	1/2945 (0.0%)
1	F	0.51	0/2182	0.72	2/2945 (0.1%)
1	G	0.47	0/2182	0.68	1/2945 (0.0%)
1	H	0.46	0/2182	0.67	3/2945 (0.1%)
1	I	0.47	0/2182	0.69	3/2945 (0.1%)
1	J	0.45	0/2182	0.66	3/2945 (0.1%)
2	K	0.67	0/55	0.91	0/70
2	L	0.73	0/55	1.03	0/70
2	M	0.68	0/55	0.89	0/70
2	N	0.66	0/55	0.89	0/70
2	O	0.66	0/55	1.00	0/70
2	P	0.65	0/55	0.94	0/70
2	Q	0.70	0/55	1.02	0/70
2	R	0.69	0/55	1.04	0/70
2	S	0.78	0/55	1.00	0/70
2	T	0.67	0/55	0.95	0/70
All	All	0.47	0/22370	0.68	13/30150 (0.0%)

There are no bond length outliers.

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	F	82	ARG	NE-CZ-NH1	-6.71	116.94	120.30
1	F	78	LEU	CA-CB-CG	6.32	129.83	115.30
1	I	82	ARG	NE-CZ-NH1	-5.63	117.48	120.30
1	J	82	ARG	NE-CZ-NH1	-5.59	117.51	120.30
1	E	78	LEU	CA-CB-CG	5.47	127.87	115.30

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	2141	0	2148	87	0
1	B	2141	0	2148	97	0
1	C	2141	0	2148	92	0
1	D	2141	0	2148	111	0
1	E	2141	0	2148	95	0
1	F	2141	0	2148	106	0
1	G	2141	0	2148	102	0
1	H	2141	0	2148	94	0
1	I	2141	0	2148	104	0
1	J	2141	0	2148	97	0
2	K	56	0	60	6	0
2	L	56	0	60	5	0
2	M	56	0	60	5	0
2	N	56	0	60	6	0
2	O	56	0	60	5	0
2	P	56	0	60	6	0
2	Q	56	0	60	8	0
2	R	56	0	60	6	0
2	S	56	0	60	7	0
2	T	56	0	60	6	0
3	A	9	0	0	2	0
3	B	11	0	0	4	0
3	C	9	0	0	2	0
3	D	12	0	0	4	0
3	E	9	0	0	2	0
3	F	11	0	0	4	0
3	G	9	0	0	2	0
3	H	11	0	0	2	0
3	I	9	0	0	3	0
3	J	10	0	0	2	0
All	All	22070	0	22080	953	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 22.

The worst 5 of 953 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:H:196:ARG:HG3	1:H:196:ARG:HH11	1.11	1.16
1:I:196:ARG:HH11	1:I:196:ARG:HG3	1.12	1.15
1:B:196:ARG:HG3	1:B:196:ARG:HH11	1.12	1.14
1:J:196:ARG:HG3	1:J:196:ARG:HH11	1.12	1.13
2:R:109:LEU:HD12	2:R:111:ILE:HD11	1.27	1.13

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	262/265 (99%)	232 (88%)	26 (10%)	4 (2%)	13	46
1	B	262/265 (99%)	233 (89%)	24 (9%)	5 (2%)	10	40
1	C	262/265 (99%)	232 (88%)	27 (10%)	3 (1%)	17	55
1	D	262/265 (99%)	231 (88%)	26 (10%)	5 (2%)	10	40
1	E	262/265 (99%)	232 (88%)	24 (9%)	6 (2%)	8	35
1	F	262/265 (99%)	235 (90%)	22 (8%)	5 (2%)	10	40
1	G	262/265 (99%)	231 (88%)	26 (10%)	5 (2%)	10	40
1	H	262/265 (99%)	232 (88%)	26 (10%)	4 (2%)	13	46
1	I	262/265 (99%)	234 (89%)	23 (9%)	5 (2%)	10	40
1	J	262/265 (99%)	233 (89%)	24 (9%)	5 (2%)	10	40
2	K	6/8 (75%)	3 (50%)	2 (33%)	1 (17%)	0	0
2	L	6/8 (75%)	3 (50%)	2 (33%)	1 (17%)	0	0
2	M	6/8 (75%)	3 (50%)	2 (33%)	1 (17%)	0	0
2	N	6/8 (75%)	3 (50%)	2 (33%)	1 (17%)	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	O	6/8 (75%)	3 (50%)	2 (33%)	1 (17%)	0 0
2	P	6/8 (75%)	3 (50%)	1 (17%)	2 (33%)	0 0
2	Q	6/8 (75%)	3 (50%)	2 (33%)	1 (17%)	0 0
2	R	6/8 (75%)	3 (50%)	2 (33%)	1 (17%)	0 0
2	S	6/8 (75%)	3 (50%)	2 (33%)	1 (17%)	0 0
2	T	6/8 (75%)	3 (50%)	2 (33%)	1 (17%)	0 0
All	All	2680/2730 (98%)	2355 (88%)	267 (10%)	58 (2%)	8 36

5 of 58 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	44	GLY
1	B	44	GLY
1	C	44	GLY
1	D	44	GLY
1	F	44	GLY

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	233/234 (100%)	192 (82%)	41 (18%)	2 10
1	B	233/234 (100%)	192 (82%)	41 (18%)	2 10
1	C	233/234 (100%)	189 (81%)	44 (19%)	2 8
1	D	233/234 (100%)	192 (82%)	41 (18%)	2 10
1	E	233/234 (100%)	189 (81%)	44 (19%)	2 8
1	F	233/234 (100%)	187 (80%)	46 (20%)	1 7
1	G	233/234 (100%)	191 (82%)	42 (18%)	2 10
1	H	233/234 (100%)	189 (81%)	44 (19%)	2 8
1	I	233/234 (100%)	190 (82%)	43 (18%)	2 9
1	J	233/234 (100%)	190 (82%)	43 (18%)	2 9

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	K	5/5 (100%)	3 (60%)	2 (40%)	0	0
2	L	5/5 (100%)	2 (40%)	3 (60%)	0	0
2	M	5/5 (100%)	2 (40%)	3 (60%)	0	0
2	N	5/5 (100%)	3 (60%)	2 (40%)	0	0
2	O	5/5 (100%)	2 (40%)	3 (60%)	0	0
2	P	5/5 (100%)	2 (40%)	3 (60%)	0	0
2	Q	5/5 (100%)	2 (40%)	3 (60%)	0	0
2	R	5/5 (100%)	4 (80%)	1 (20%)	1	7
2	S	5/5 (100%)	2 (40%)	3 (60%)	0	0
2	T	5/5 (100%)	2 (40%)	3 (60%)	0	0
All	All	2380/2390 (100%)	1925 (81%)	455 (19%)	2	8

5 of 455 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	240	ARG
1	F	261	GLU
1	J	219	SER
1	F	4	MET
1	F	106	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	176	ASN
1	F	176	ASN
1	H	176	ASN
1	D	176	ASN
1	G	176	ASN

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

EDS failed to run properly - this section will therefore be empty.

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

EDS failed to run properly - this section will therefore be empty.

6.3 Carbohydrates

EDS failed to run properly - this section will therefore be empty.

6.4 Ligands

EDS failed to run properly - this section will therefore be empty.

6.5 Other polymers

EDS failed to run properly - this section will therefore be empty.