WORLDWIDE PDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 16, 2023 – 03:06 PM EDT

PDB ID	:	4P6Y
Title	:	Crystal structure of the M42 aminopeptidase TmPep1050 from Thermotoga maritima
Authors	:	Dutoit, R.; Demarez, M.; Van Elder, D.; Bauvois, C.
Deposited on	:	2014-03-25
Resolution	:	2.20 Å(reported)

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 ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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.
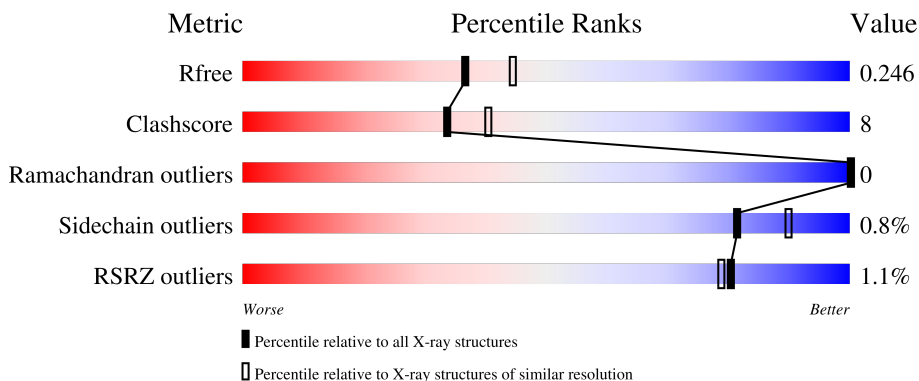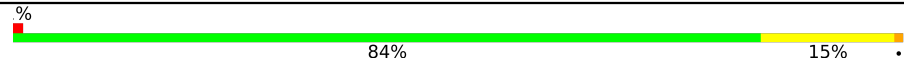

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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% 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	331	85% 14%
1	B	331	2% 82% 17%
1	C	331	86% 14%
1	D	331	2% 82% 17%
1	E	331	84% 15% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	331	84% 15%
1	G	331	84% 15%
1	H	331	83% 17%
1	I	331	85% 15%
1	J	331	79% 21%
1	K	331	82% 16%
1	L	331	85% 14%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 32192 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 Aminopeptidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	331	Total 2513	C 1588	N 425	O 489	S 11	0	0	0
1	B	330	Total 2491	C 1579	N 419	O 482	S 11	0	0	0
1	C	331	Total 2501	C 1584	N 420	O 486	S 11	0	0	0
1	D	331	Total 2496	C 1584	N 417	O 484	S 11	0	0	0
1	E	330	Total 2513	C 1589	N 426	O 487	S 11	0	0	0
1	F	331	Total 2522	C 1593	N 429	O 489	S 11	0	0	0
1	G	330	Total 2492	C 1579	N 423	O 479	S 11	0	0	0
1	H	330	Total 2480	C 1572	N 420	O 477	S 11	0	0	0
1	I	330	Total 2494	C 1582	N 420	O 481	S 11	0	0	0
1	J	330	Total 2493	C 1581	N 420	O 481	S 11	0	0	0
1	K	330	Total 2482	C 1573	N 421	O 477	S 11	0	0	0
1	L	330	Total 2492	C 1578	N 422	O 481	S 11	0	0	0

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	221	Total 221	O 221	0	0
2	B	168	Total 168	O 168	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	192	Total 192	O 192	0	0
2	D	206	Total 206	O 206	0	0
2	E	204	Total 204	O 204	0	0
2	F	176	Total 176	O 176	0	0
2	G	195	Total 195	O 195	0	0
2	H	165	Total 165	O 165	0	0
2	I	199	Total 199	O 199	0	0
2	J	147	Total 147	O 147	0	0
2	K	176	Total 176	O 176	0	0
2	L	174	Total 174	O 174	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 seco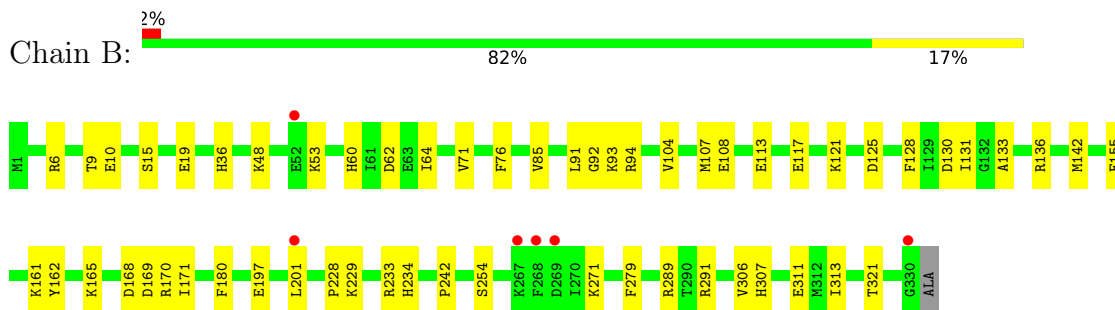nd graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above 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.
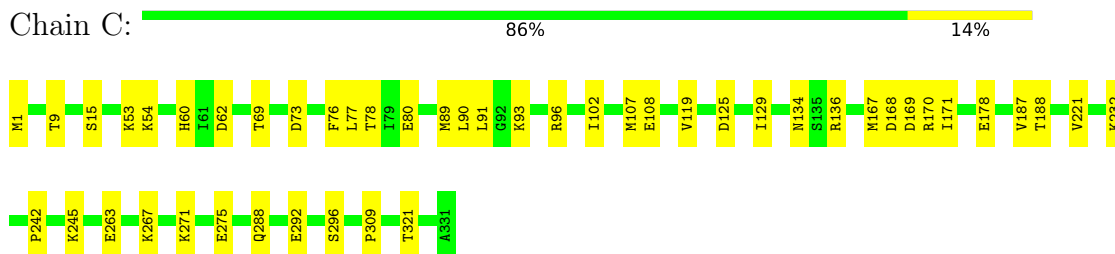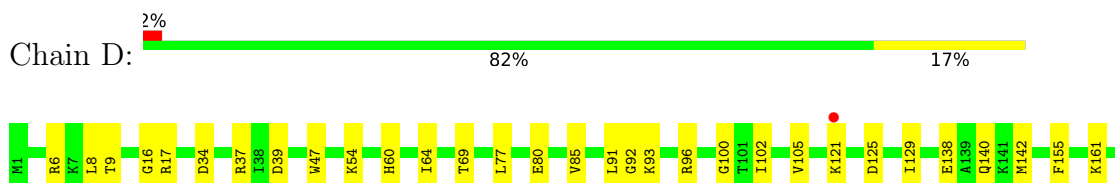
- Molecule 1: Aminopeptidase

Chain A:

- Molecule 1: Aminopeptidase

Chain B:

- Molecule 1: Aminopeptidase

Chain C:

- Molecule 1: Aminopeptidase

Chain D:

K165 A166 M167 D168 D169 R170 I171 T188 L189 E197 L201 V202 V206 D220 P228 K229 H234 A235 M236 P242 E263 K267 F268 D269 I270 K271 F279 Q288 S296 P309 S310 E311 V322 I326 A331

• Molecule 1: Aminopeptidase

Chain E:

84%　　　　　15%

M1 K2 E3 T9 S23 L26 E27 E30 R37 K53 H60 T69 D72 F76 E80 M89 L90 L91 G92 K93 E110 Q116 V119 R120 D125 I131 R136 Q140 K141 M142 C143 K161 M167 D168 D169 R170 I171 K184

V187 V202 G203 A204 S205 D220 V221 P228 H234 A235 M236 E263 F279 N283 A284 M285 S300 R304 Y305 V306 H307 S308 P309 R327 G330 ALA

• Molecule 1: Aminopeptidase

Chain F:

84%　　　　　15%

M1 R6 V21 I25 L26 E27 E30 I33 K53 K54 I64 F76 V85 M89 L90 L91 G92 K93 R96 I102 M107 E108 E113 Q116 E117 N118 V119 R120 N134 E138 M142 F155 V156 E157 M167 D168 I171 V187

T188 V202 S205 P228 K229 R233 H234 P242 A243 R249 K267 F268 K271 V276 F279 R291 S300 E311 P315 V318 R327 A331

• Molecule 1: Aminopeptidase

Chain G:

84%　　　　　15%

M1 K2 E3 R6 T9 L26 L29 E30 H60 I64 F76 V85 M89 L90 L91 G92 K93 R96 F97 E98 N99 I102 E113 Q116 E117 R120 D130 N134 E138 E157 M167 D168 D169 R170 I171 E178 G208 T222

P228 P242 A243 K247 D248 R249 T278 M285 Q288 R291 S296 S300 P309 S310 E311 M312 I313 D316 D317 T321 V322 D323 R327 G330 ALA

• Molecule 1: Aminopeptidase

Chain H:

83%　　　　　17%

M1 T9 E18 E19 R22 G31 L41 G42 E52 K53 H60 I61 D62 T69 F76 L77 T78 M89 L90 L91 G92 K93 R96 F97 E98 E108 E113 E117 N118 V119 N134 M167 D168 D169 R170 I171 I183 F193 E197 L201

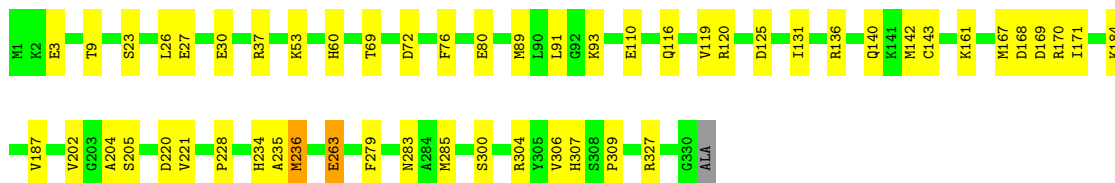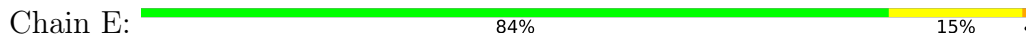V206 Y209 A213 D214 E215 A216 I217 P228 P242 I257 L258 E259 E263 K267 K271 V276 Q288 R291 S296 P309 I313 Y328 L329 G330 ALA

• Molecule 1: Aminopeptidase

Chain I:

85%　　　　　15%

M1 L8 R22 G31 R37 L44 K53 H60 I61 I64 D73 K74 G75 F76 P81 V85 L91 G92 K93 E108 T112 Q116 F124 D125 K126 L127 F128 R136 Q140 K161 A166 M167 D168 D169 R170 I171 K184 P185 A186
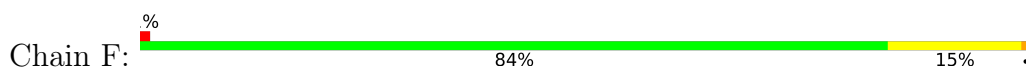
V187 E197 L201 D223 P228 P242 A243 R249 A250 S251 I264 K271 V276 M285 S300 P309 I313 G330 ALA

• Molecule 1: Aminopeptidase

Chain J:

M1 K7 L8 T9 I24 I25 L26 H32 R37 K48 K53 K54 H60 I61 D62 E63 I64 F76 V85 M89 L90 L91 G92 K93 R94 I95 R96 N99 G100 T101 I102 E113 E117 D125 E137 E138 A139 M142 C143 S153 K161

K165 A166 M167 D168 D169 R170 I171 I183 V187 T188 F193 E197 L201 S205 E215 V221 T222 P228 K229 A243 L244 K245 I257 L258 E266 K267 F268 D269 T282 N283 E292 S300 R304 Y305 V306 H307 E311 M312 D316 V322 D323

I326 G330 ALA

• Molecule 1: Aminopeptidase

Chain K:

M1 K2 E3 L4 I5 R6 K7 L8 L26 E30 G31 H60 I64 D73 V85 M89 G92 K93 E98 E108 T112 Q116 V119 D125 R136 E137 E138 A139 Q140 K141 M142 Y151 F155 V156 E157 M167 D168 D169 R170 I171 I183

Q196 E197 V202 S205 E215 T222 P228 K229 R237 R249 E259 N260 E263 F268 K271 V276 L277 T278 F279 N283 Q288 S300 P309 S310 E311 V322 I326 R327 G330 ALA

• Molecule 1: Aminopeptidase

Chain L:

M1 K7 L8 T9 E10 L26 G31 H32 I33 D34 G35 I45 S50 K53 H60 I64 T69 N70 F76 V85 L91 G92 K93 T112 E113 Q116 N134 E138 M142 E157 V158 Y162 K165 A166 M167 D168 D169 R170 K181

A186 V187 V192 D220 V221 A225 P228 K229 A243 K247 R256 I262 E263 I264 A265 E266 K267 F268 D269 S300 V306 H307 E311 M312 I313 D317 G330 ALA

4 Data and refinement statistics (i)

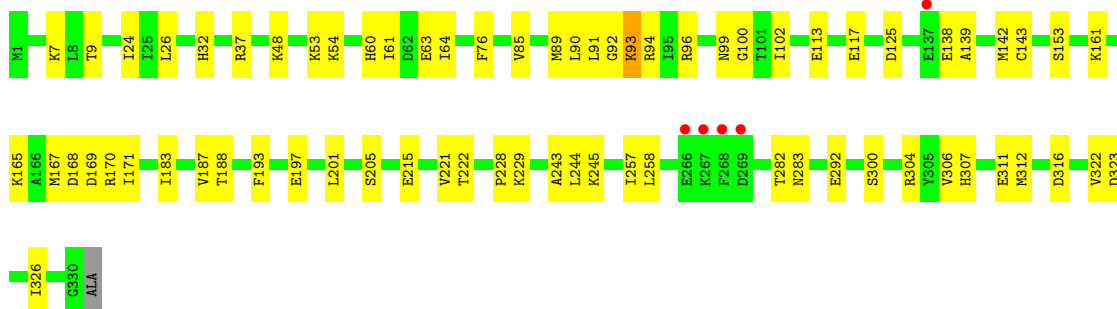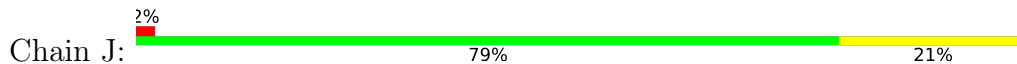| Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	114.26Å 114.57Å 114.04Å 114.46° 91.71° 105.69°	Depositor
Resolution (Å)	44.05 – 2.20 47.53 – 2.20	Depositor EDS
% Data completeness (in resolution range)	93.5 (44.05-2.20) 93.5 (47.53-2.20)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.60 (at 2.20Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1539)	Depositor
R, R_{free}	0.212 , 0.247 0.212 , 0.246	Depositor DCC
R_{free} test set	11858 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	32.5	Xtriage
Anisotropy	0.164	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 45.3	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	32192	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.89% 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 (i)

5.1 Standard geometry (i)

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	0/2551	0.62	0/3442
1	B	0.42	0/2529	0.60	0/3415
1	C	0.44	0/2539	0.61	0/3427
1	D	0.45	0/2534	0.63	0/3420
1	E	0.46	0/2551	0.63	0/3442
1	F	0.47	0/2560	0.64	0/3453
1	G	0.43	0/2530	0.64	1/3416 (0.0%)
1	H	0.42	0/2518	0.61	0/3403
1	I	0.43	0/2532	0.62	0/3418
1	J	0.41	0/2531	0.61	0/3417
1	K	0.42	0/2520	0.61	0/3404
1	L	0.42	0/2530	0.60	0/3417
All	All	0.44	0/30425	0.62	1/41074 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	G	327	ARG	NE-CZ-NH2	-7.26	116.67	120.30

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	2513	0	2526	38	0
1	B	2491	0	2500	49	0
1	C	2501	0	2512	47	0
1	D	2496	0	2507	46	0
1	E	2513	0	2535	49	0
1	F	2522	0	2542	39	0
1	G	2492	0	2510	46	0
1	H	2480	0	2481	53	0
1	I	2494	0	2511	50	0
1	J	2493	0	2509	62	0
1	K	2482	0	2488	47	0
1	L	2492	0	2503	51	0
2	A	221	0	0	16	1
2	B	168	0	0	11	0
2	C	192	0	0	8	0
2	D	206	0	0	9	0
2	E	204	0	0	11	0
2	F	176	0	0	5	0
2	G	195	0	0	12	0
2	H	165	0	0	11	0
2	I	199	0	0	17	1
2	J	147	0	0	13	0
2	K	176	0	0	12	0
2	L	174	0	0	17	0
All	All	32192	0	30124	509	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 8.

The worst 5 of 509 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:E:120:ARG:NH1	2:E:502:HOH:O	1.85	1.09
1:J:316:ASP:HB2	2:J:545:HOH:O	1.53	1.09
1:J:89:MET:O	1:J:93:LYS:HE2	1.54	1.07
1:I:184:LYS:HG2	2:I:598:HOH:O	1.56	1.05
1:L:225:ALA:HA	2:L:570:HOH:O	1.60	0.99

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)
2:A:425:HOH:O	2:I:404:HOH:O[1_655]	2.18	0.02

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	329/331 (99%)	320 (97%)	9 (3%)	0	100	100
1	B	328/331 (99%)	320 (98%)	8 (2%)	0	100	100
1	C	329/331 (99%)	322 (98%)	7 (2%)	0	100	100
1	D	329/331 (99%)	320 (97%)	9 (3%)	0	100	100
1	E	328/331 (99%)	320 (98%)	8 (2%)	0	100	100
1	F	329/331 (99%)	320 (97%)	9 (3%)	0	100	100
1	G	328/331 (99%)	319 (97%)	9 (3%)	0	100	100
1	H	328/331 (99%)	320 (98%)	8 (2%)	0	100	100
1	I	328/331 (99%)	319 (97%)	9 (3%)	0	100	100
1	J	328/331 (99%)	321 (98%)	7 (2%)	0	100	100
1	K	328/331 (99%)	320 (98%)	8 (2%)	0	100	100
1	L	328/331 (99%)	319 (97%)	9 (3%)	0	100	100
All	All	3940/3972 (99%)	3840 (98%)	100 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	267/273 (98%)	265 (99%)	2 (1%)	84	91
1	B	263/273 (96%)	263 (100%)	0	100	100
1	C	265/273 (97%)	264 (100%)	1 (0%)	91	96
1	D	263/273 (96%)	261 (99%)	2 (1%)	81	90
1	E	269/273 (98%)	265 (98%)	4 (2%)	65	78
1	F	269/273 (98%)	266 (99%)	3 (1%)	73	85
1	G	264/273 (97%)	264 (100%)	0	100	100
1	H	260/273 (95%)	258 (99%)	2 (1%)	81	90
1	I	264/273 (97%)	263 (100%)	1 (0%)	91	96
1	J	264/273 (97%)	262 (99%)	2 (1%)	81	90
1	K	261/273 (96%)	255 (98%)	6 (2%)	50	63
1	L	264/273 (97%)	263 (100%)	1 (0%)	91	96
All	All	3173/3276 (97%)	3149 (99%)	24 (1%)	81	90

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	140	GLN
1	K	116	GLN
1	J	143	CYS
1	K	119	VAL
1	E	184	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	L	116	GLN

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 monosaccharides 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 ⓘ

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(Å²)	Q<0.9
1	A	331/331 (100%)	-0.17	1 (0%)	94	93	20, 32, 46, 60	0
1	B	330/331 (99%)	0.18	6 (1%)	68	66	24, 35, 49, 62	0
1	C	331/331 (100%)	-0.21	0	100	100	21, 32, 45, 53	0
1	D	331/331 (100%)	0.06	7 (2%)	63	61	21, 30, 44, 55	0
1	E	330/331 (99%)	-0.09	0	100	100	23, 32, 45, 56	0
1	F	331/331 (100%)	-0.13	2 (0%)	89	88	22, 32, 45, 56	0
1	G	330/331 (99%)	-0.13	2 (0%)	89	88	25, 33, 45, 59	0
1	H	330/331 (99%)	-0.10	3 (0%)	84	83	22, 34, 46, 58	0
1	I	330/331 (99%)	-0.06	2 (0%)	89	88	24, 32, 45, 54	0
1	J	330/331 (99%)	-0.19	5 (1%)	73	72	25, 34, 46, 62	0
1	K	330/331 (99%)	-0.08	3 (0%)	84	83	25, 36, 48, 57	0
1	L	330/331 (99%)	0.05	14 (4%)	36	34	28, 37, 52, 63	0
All	All	3964/3972 (99%)	-0.07	45 (1%)	80	79	20, 33, 48, 63	0

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	331	ALA	4.7
1	D	331	ALA	4.6
1	B	268	PHE	3.5
1	D	268	PHE	3.4
1	L	265	ALA	3.2

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.