W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 07:29 PM GMT

PDB ID	:	1FT8
Title	:	CRYSTAL STRUCTURE OF THE RNA-BINDING DOMAIN OF THE MRNA EXPORT FACTOR TAP
Authors	:	Liker, E.; Fernandez, E.; Izaurralde, E.; Conti, E.
Deposited on	:	2000-09-12
Resolution	:	3.15 Å(reported)

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

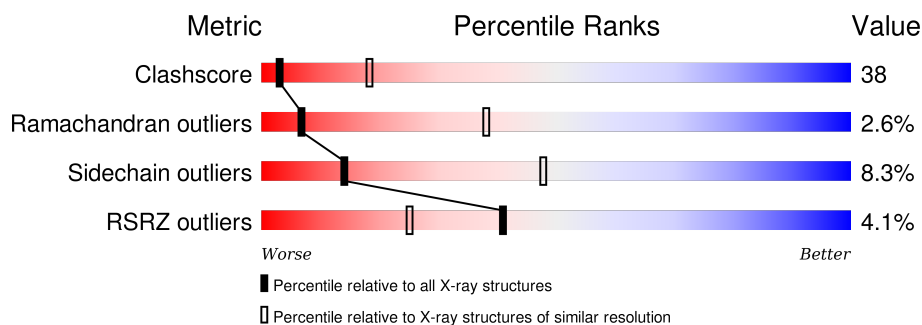|		
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 3.15 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	Percentile Ranks	Value
Clashscore		38
Ramachandran outliers		2.6%
Sidechain outliers		8.3%
RSRZ outliers		4.1%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	102246	1249 (3.20-3.12)
Ramachandran outliers	100387	1222 (3.20-3.12)
Sidechain outliers	100360	1221 (3.20-3.12)
RSRZ outliers	91569	1117 (3.20-3.12)

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.

Mol	Chain	Length	Quality of chain
1	A	271	3% — 42% 38% 8% · 11%
1	B	271	2% — 32% 25% · 39%
1	C	271	2% — 49% 34% 6% · 10%
1	D	271	3% — 28% 27% 5% · 40%
1	E	271	3% — 6% 8% · 84%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 6913 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 TIP ASSOCIATING PROTEIN.

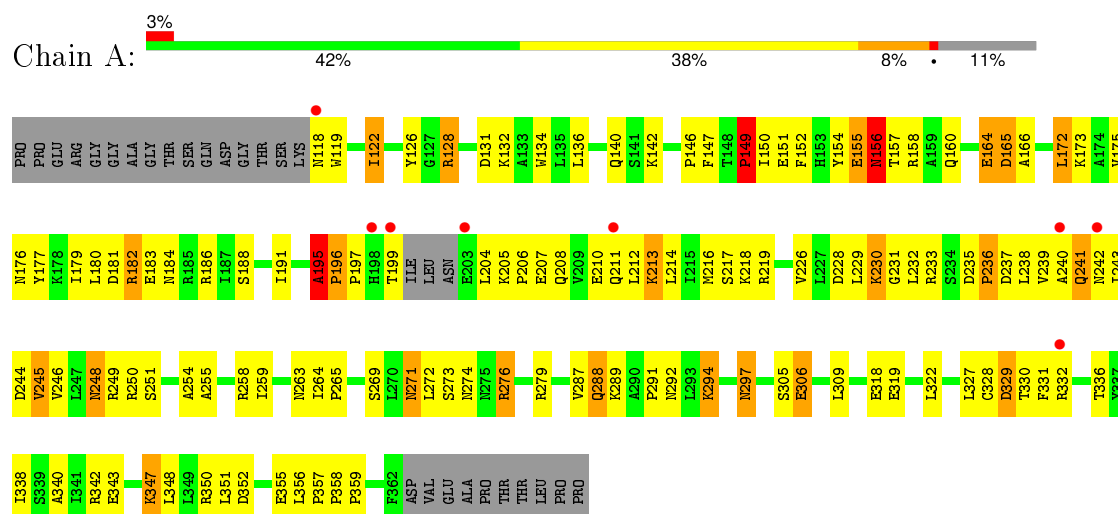| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	242	Total	C	N	O	S	0	0	0
			1961	1245	343	366	7			
1	B	165	Total	C	N	O	S	0	0	0
			1323	831	232	255	5			
1	C	244	Total	C	N	O	S	0	0	0
			1973	1253	346	367	7			
1	D	162	Total	C	N	O	S	0	0	0
			1302	818	229	250	5			
1	E	44	Total	C	N	O	S	0	0	0
			354	236	54	62	2			

There are 10 discrepancies between the modelled and reference sequences:

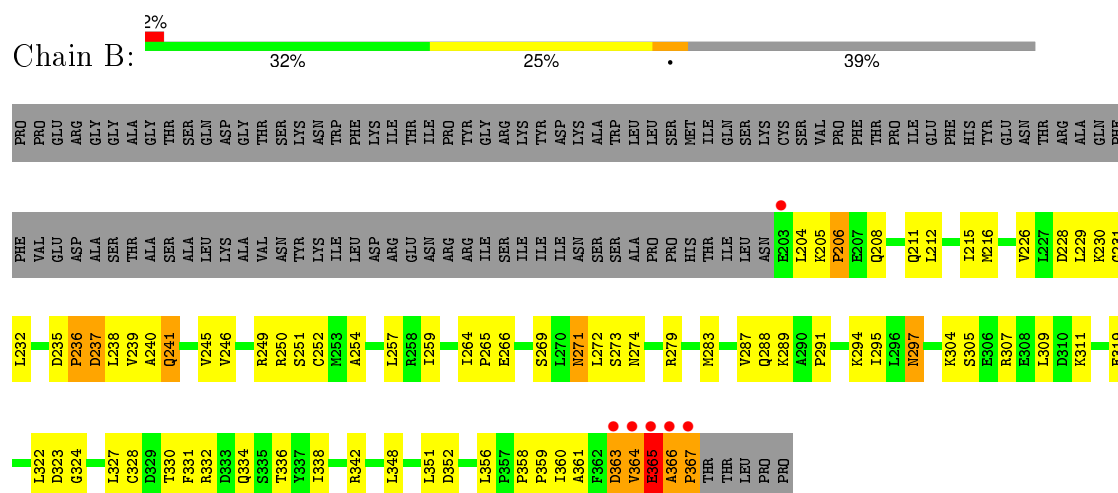| Chain	Residue	Modelled	Actual	Comment	Reference
A	119	TRP	CYS	ENGINEERED	UNP Q9UBU9
A	226	VAL	ALA	ENGINEERED	UNP Q9UBU9
B	119	TRP	CYS	ENGINEERED	UNP Q9UBU9
B	226	VAL	ALA	ENGINEERED	UNP Q9UBU9
C	119	TRP	CYS	ENGINEERED	UNP Q9UBU9
C	226	VAL	ALA	ENGINEERED	UNP Q9UBU9
D	119	TRP	CYS	ENGINEERED	UNP Q9UBU9
D	226	VAL	ALA	ENGINEERED	UNP Q9UBU9
E	119	TRP	CYS	ENGINEERED	UNP Q9UBU9
E	226	VAL	ALA	ENGINEERED	UNP Q9UBU9

3 Residue-property plots ⓘ

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.
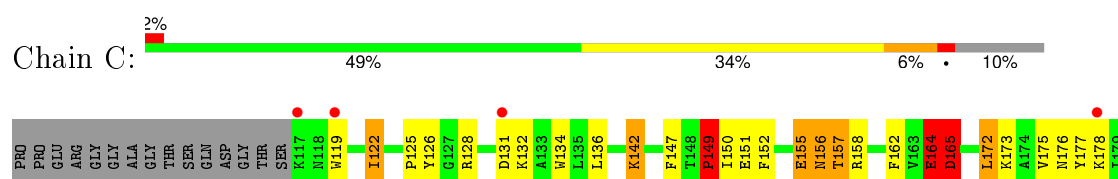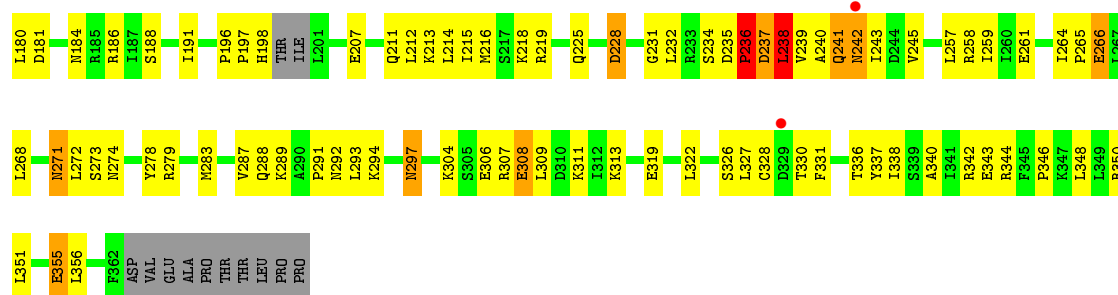
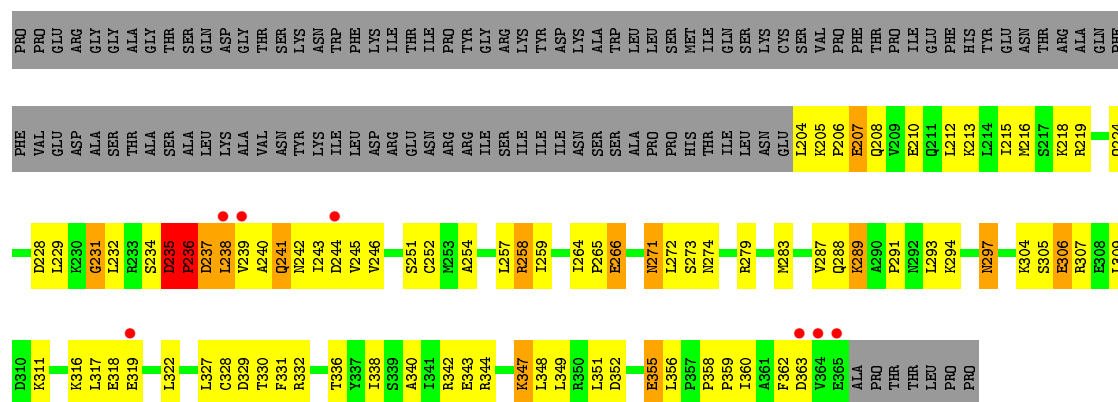- Molecule 1: TIP ASSOCIATING PROTEIN
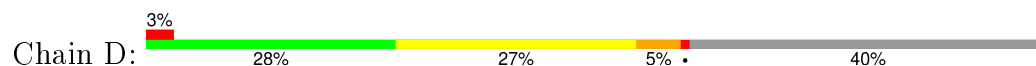
Chain A:

- Molecule 1: TIP ASSOCIATING PROTEIN

Chain B:

- Molecule 1: TIP ASSOCIATING PROTEIN

Chain C:

L180 D181 N184 R185 R186 I187 S188 I191 P196 P197 H198 THR ILE L201 E207 Q211 L212 K213 L214 I215 M216 S217 K218 R219 Q225 D228 G231 L232 R233 D235 F236 D237 L238 V239 A240 Q241 N242 I243 D244 V245 L257 R258 I259 I260 E261 I264 P265 L266 L267

L268 N271 L272 S273 N274 Y278 R279 M283 V287 Q288 K289 A290 P291 N292 L293 K294 N297 K304 S305 E306 R307 E308 L309 D310 K311 I312 K313 E319 L322 S326 L327 C328 D329 T330 F331 T336 Y337 I338 S339 A340 I341 R342 E343 R344 F345 P346 K347 L348 L349 R350

L351 E355 L356 F362 ASP VAL GLU ALA PRO THR THR LEU PRO PRO

● Molecule 1: TIP ASSOCIATING PROTEIN

Chain D:

PRO PRO GLU ARG GLY GLY ALA GLY THR SER GLN ASP GLY THR SER LYS ASN TRP PHE LYS ILE THR ILE PRO TYR GLY ARG LYS TYR ASP LYS ALA TRP LEU LEU SER MET ILE GLN SER LYS CYS SER VAL PRO PHE THR PRO ILE GLU PHE HIS TYR GLU ASN THR ARG ALA GLN PHE

PHE VAL GLU ASP ALA SER THR ALA SER ALA LEU LYS ALA VAL ASN TYR LYS ILE LEU ASP PRO ARG GLU ASN ARG ARG ILE SER ILE ILE ASN SER SER ALA PRO PRO HIS THR ILE LEU ASN GLU L204 K205 P206 E207 Q208 V209 E210 Q211 L212 K213 L214 I215 M216 S217 K218 R219 Q224

D228 L229 K230 G231 L232 R233 S234 D235 P236 D237 L238 V239 A240 Q241 N242 I243 D244 V245 V246 S251 C252 M253 A254 L257 R258 I259 I264 P265 E266 N271 L272 S273 N274 R279 M283 V287 Q288 K289 A290 P291 N292 L293 K294 N297 K304 S305 E306 R307 E308 L309

D310 K311 K316 L317 E318 E319 L322 L327 C328 D329 T330 F331 R332 T336 Y337 I338 S339 A340 I341 R342 E343 R344 K347 L348 L349 R350 L351 D352 E355 L356 P357 P358 P359 I360 A361 F362 D363 V364 E365 ALA PRO THR THR LEU PRO PRO

● Molecule 1: TIP ASSOCIATING PROTEIN

Chain E:

PRO PRO GLU ARG GLY GLY ALA GLY THR SER GLN ASP GLY THR SER LYS ASN W119 F120 K121 I122 T123 ILE PRO TYR GLY ARG LYS TYR D131 K132 A133 W134 L135 L136 S137 M138 I139 Q140 S141 K142 C143 S144 V145 P146 P149 I150 E151 F152 H153 TYR GLU ASN THR ARG ALA Q160 F161 F162

V163 E164 D165 A166 S170 A171 L172 K173 A174 V175 ASN TYR LYS ILE LEU ASP PRO ARG GLU ASN GLU ILE ASP VAL VAL LEU ASN SER ARG SER MET ALA ALA THR PRO PRO HIS THR ILE LEU ASN GLU GLU GLU GLN LEU TRP LEU ASP GLY ASN SER LEU ASP CYS ASP THR PHE ARG ASP GLN SER THR TYR ILE SER ALA ILE ARG GLU ARG

GLN VAL LEU ASP LEU LYS GLY ARG LEU ARG ASP PRO ASP LEU VAL ALA GLN ASN ILE LEU ASP VAL VAL LEU ASN ARG ARG SER MET GLU ALA ALA THR PRO PRO HIS THR ILE LEU LEU GLU GLU GLU GLN LEU TRP LEU ASP GLY ASN SER LEU ASP CYS ASP THR PHE ARG ASP GLN SER THR TYR ILE SER ALA ILE ARG GLU ARG

SER ILE VAL GLN LYS ALA ASN LEU LYS ILE LEU LEU ASN ALA LEU GLY GLU LEU LYS SER GLU ARG GLU ASP LEU ASP LYS LYS GLY ILE LEU LYS LYS GLU GLU GLU TRP LEU ASP ASP GLY ASN SER ARG SER ASP LEU ASP THR PHE ARG ASP ASP GLN SER THR TYR ILE SER ALA ILE ARG GLU ARG

PHE PRO LYS LEU LEU ARG ASP GLY ASP GLY HIS GLU LEU LEU PRO PRO ILE ALA PHE ASP VAL GLU ALA PRO THR THR LEU PRO PRO

4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	139.92Å 139.92Å 206.70Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.15 29.94 – 3.16	Depositor EDS
% Data completeness (in resolution range)	(Not available) (30.00-3.15) 99.6 (29.94-3.16)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	6.10 (at 3.18Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.303 , 0.303 0.301 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	80.0	Xtriage
Anisotropy	0.046	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 35.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.33$	Xtriage
Outliers	0 of 35832 reflections	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	6913	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.48% 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	0.73	4/1996 (0.2%)	0.91	10/2696 (0.4%)
1	B	0.60	1/1340 (0.1%)	0.79	2/1808 (0.1%)
1	C	0.75	7/2008 (0.3%)	0.98	8/2711 (0.3%)
1	D	0.61	1/1318 (0.1%)	0.92	8/1777 (0.5%)
1	E	0.72	0/363	0.95	2/489 (0.4%)
All	All	0.69	13/7025 (0.2%)	0.92	30/9481 (0.3%)

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	165	ASP	CB-CG	-11.57	1.27	1.51
1	C	228	ASP	CB-CG	-10.18	1.30	1.51
1	A	164	GLU	CB-CG	-10.11	1.32	1.52
1	C	165	ASP	CB-CG	-9.04	1.32	1.51
1	C	164	GLU	CB-CG	-7.82	1.37	1.52

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	228	ASP	CB-CG-OD1	-19.55	100.71	118.30
1	A	165	ASP	CB-CG-OD2	-18.87	101.32	118.30
1	C	165	ASP	CB-CG-OD2	-18.30	101.83	118.30
1	C	155	GLU	OE1-CD-OE2	-8.83	112.70	123.30
1	C	165	ASP	CB-CG-OD1	8.22	125.70	118.30

There are no chirality outliers.

There are no planarity outliers.

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	1961	0	2003	171	0
1	B	1323	0	1363	107	0
1	C	1973	0	2013	135	0
1	D	1302	0	1345	113	0
1	E	354	0	349	50	1
All	All	6913	0	7073	536	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 38.

The worst 5 of 536 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:230:LYS:HD2	1:A:273:SER:OG	1.29	1.32
1:B:323:ASP:HB2	1:C:177:TYR:CE2	1.78	1.19
1:C:119:TRP:HZ3	1:C:150:ILE:CD1	1.56	1.19
1:E:144:SER:O	1:E:145:VAL:HG23	1.39	1.19
1:E:145:VAL:HG22	1:E:146:PRO:CD	1.73	1.17

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:E:142:LYS:NZ	1:E:142:LYS:NZ[8_665]	1.44	0.76

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.

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	238/271 (88%)	213 (90%)	20 (8%)	5 (2%)	9	44
1	B	163/271 (60%)	149 (91%)	10 (6%)	4 (2%)	7	39
1	C	240/271 (89%)	216 (90%)	17 (7%)	7 (3%)	6	35
1	D	160/271 (59%)	139 (87%)	16 (10%)	5 (3%)	5	32
1	E	38/271 (14%)	35 (92%)	2 (5%)	1 (3%)	7	38
All	All	839/1355 (62%)	752 (90%)	65 (8%)	22 (3%)	7	38

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	236	PRO
1	C	157	THR
1	C	236	PRO
1	C	238	LEU
1	C	308	GLU

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	222/245 (91%)	201 (90%)	21 (10%)	11	38
1	B	152/245 (62%)	143 (94%)	9 (6%)	24	63
1	C	222/245 (91%)	205 (92%)	17 (8%)	16	52
1	D	150/245 (61%)	137 (91%)	13 (9%)	13	44
1	E	39/245 (16%)	34 (87%)	5 (13%)	5	24
All	All	785/1225 (64%)	720 (92%)	65 (8%)	14	48

5 of 65 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	367	PRO
1	C	172	LEU
1	E	121	LYS
1	C	122	ILE
1	C	156	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	263	ASN
1	B	297	ASN
1	D	297	ASN
1	B	271	ASN
1	C	156	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 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(Å²)	Q<0.9
1	A	242/271 (89%)	0.02	8 (3%)	50	33	55, 66, 89, 109	0
1	B	165/271 (60%)	0.03	6 (3%)	46	29	49, 63, 82, 97	0
1	C	244/271 (90%)	0.18	6 (2%)	61	45	55, 74, 106, 106	0
1	D	162/271 (59%)	0.14	7 (4%)	39	23	52, 65, 85, 100	0
1	E	44/271 (16%)	1.01	8 (18%)	2	1	89, 94, 94, 94	0
All	All	857/1355 (63%)	0.14	35 (4%)	41	25	49, 68, 106, 109	0

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	239	VAL	4.3
1	E	164	GLU	4.1
1	B	367	PRO	3.4
1	A	242	ASN	3.2
1	D	365	GLU	3.2

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.