WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 28, 2014 – 11:06 PM GMT

PDB ID   :   1MIR
Title   :   RAT PROCATHEPSIN B
Authors   :   Cygler, M.; Sivaraman, J.; Grochulski, P.; Coulombe, R.; Storer, A.C.; Mort, J.S.
Deposited on :   1996-01-12
Resolution :   2.80 Å(reported)

---

1   Overall quality at a glance ⓘ

The reported resolution of this entry is 2.80 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		12
Ramachandran outliers		1.9%
Sidechain outliers		3.9%
RSRZ outliers		1.6%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
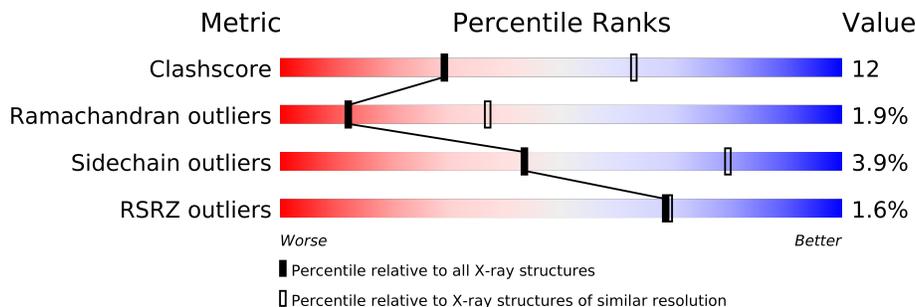
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	79885	2295 (2.80-2.80)
Ramachandran outliers	78287	2252 (2.80-2.80)
Sidechain outliers	78261	2254 (2.80-2.80)
RSRZ outliers	66119	1802 (2.80-2.80)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	322	
1	B	322	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6023 atoms, of which 1134 are hydrogens and 0 are deuterium.

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 PROCATHEPSIN B.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	313	Total	C	H	N	O	S	23	0	0
			2962	1524	534	417	468	19			
1	B	313	Total	C	H	N	O	S	23	0	0
			2962	1524	534	417	468	19			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	6P	PHE	SER	CONFLICT	UNP P00787
A	50P	LYS	ASN	CONFLICT	UNP P00787
A	29	SER	CYS	ENGINEERED	UNP P00787
A	115	ALA	SER	ENGINEERED	UNP P00787
A	223	ALA	VAL	VARIANT	UNP P00787
B	6P	PHE	SER	CONFLICT	UNP P00787
B	50P	LYS	ASN	CONFLICT	UNP P00787
B	29	SER	CYS	ENGINEERED	UNP P00787
B	115	ALA	SER	ENGINEERED	UNP P00787
B	223	ALA	VAL	VARIANT	UNP P00787

- Molecule 2 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	11	Total	H	O	0	0
			33	22	11		
2	B	22	Total	H	O	0	0
			66	44	22		

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.
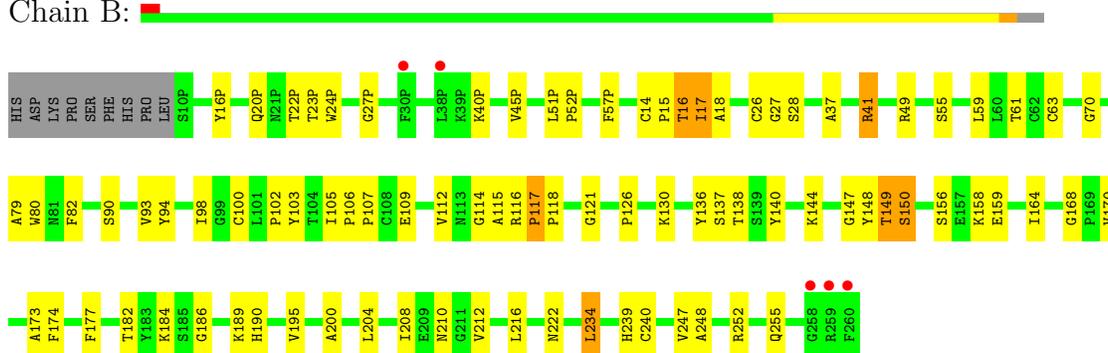
• Molecule 1: PROCATHEPSIN B

Chain A:
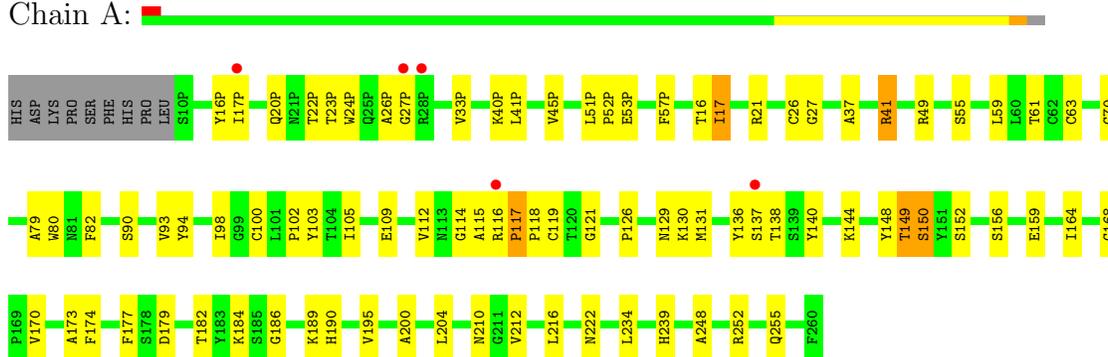
• Molecule 1: PROCATHEPSIN B

Chain B:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	99.60Å 99.60Å 141.40Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	8.00 – 2.80 73.64 – 2.80	Depositor EDS
% Data completeness (in resolution range)	92.0 (8.00-2.80) 92.5 (73.64-2.80)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.62 (at 2.82Å)	Xtriage
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.220 , 0.270 0.213 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	41.8	Xtriage
Anisotropy	0.228	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 41.8	EDS
Estimated twinning fraction	0.059 for -h,-k,l	Xtriage
L-test for twinning	$<|L|> = 0.45, <L^2> = 0.27$	Xtriage
Outliers	0 of 19015 reflections	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	6023	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.40% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

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	0.40	0/2496	0.63	0/3386
1	B	0.41	0/2496	0.63	0/3386
All	All	0.41	0/4992	0.63	0/6772

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2428	534	1718	57	0
1	B	2428	534	1718	57	0
2	A	11	22	0	5	0
2	B	22	44	0	3	0
All	All	4889	1134	3436	114	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 12.

The worst 5 of 114 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:93:VAL:HG23	2:A:409:HOH:H1	1.48	0.77

Continued on next page...

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

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:239:HIS:CE1	1:B:240:CYS:HG	2.10	0.69
1:A:23(P):THR:CG2	1:A:212:VAL:HG21	2.24	0.68
1:B:23(P):THR:CG2	1:B:212:VAL:HG21	2.26	0.65
1:A:21:ARG:HH11	2:A:426:HOH:H1	1.45	0.63

There are no symmetry-related clashes.

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	311/322 (97%)	260 (84%)	45 (14%)	6 (2%)	12	37
1	B	311/322 (97%)	261 (84%)	44 (14%)	6 (2%)	12	37
All	All	622/644 (97%)	521 (84%)	89 (14%)	12 (2%)	12	37

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	117	PRO
1	A	117	PRO
1	A	130	LYS
1	B	130	LYS
1	B	121	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	257/266 (97%)	247 (96%)	10 (4%)	43	80

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	B	257/266 (97%)	247 (96%)	10 (4%)	43 80
All	All	514/532 (97%)	494 (96%)	20 (4%)	43 80

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	255	GLN
1	B	45(P)	VAL
1	B	210	ASN
1	A	222	ASN
1	A	234	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	239	HIS
1	B	222	ASN
1	B	129	ASN
1	A	210	ASN
1	B	210	ASN

5.3.3 RNA ⓘ

There are no RNA chains 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 (i)

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(Å^2)	Q<0.9
1	A	313/322 (97%)	-0.06	5 (1%)	68	69	6, 29, 62, 81	5 (1%)
1	B	313/322 (97%)	-0.04	5 (1%)	68	69	9, 29, 63, 82	5 (1%)
All	All	626/644 (97%)	-0.05	10 (1%)	68	69	6, 29, 62, 82	10 (1%)

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	260	PHE	4.9
1	B	259	ARG	4.0
1	B	258	GLY	3.8
1	A	17(P)	ILE	3.6
1	A	116	ARG	2.9

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 carbohydrates 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.