W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Feb 27, 2014 – 04:45 PM GMT

PDB ID	:	4IVJ
Title	:	Structure of a 16 nm protein cage designed by fusing symmetric oligomeric domains, triple mutant, I222 form
Authors	:	Lai, Y.-T.; Sawaya, M.R.; Yeates, T.O.
Deposited on	:	2013-01-23
Resolution	:	7.35 Å(reported)

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

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

1   Overall quality at a glance ⓘ

The reported resolution of this entry is 7.35 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
Rfree		0.282
Clashscore		21
Ramachandran outliers		0
Sidechain outliers		5.1%
RSRZ outliers		10.6%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
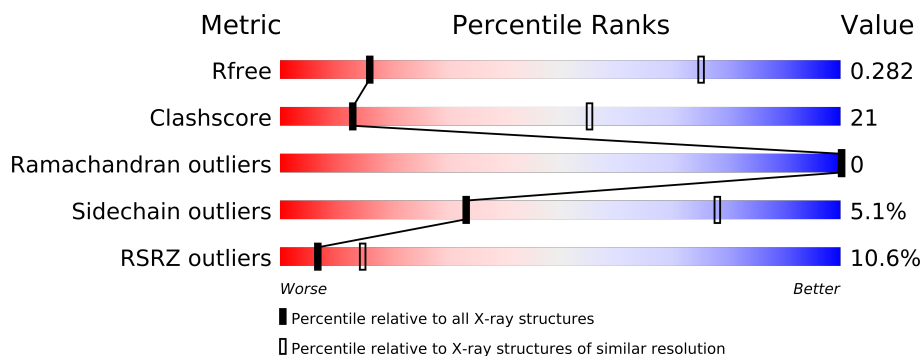
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	66092	1098 (10.00-3.50)
Clashscore	79885	1039 (10.00-3.52)
Ramachandran outliers	78287	1291 (9.50-3.50)
Sidechain outliers	78261	1265 (9.50-3.50)
RSRZ outliers	66119	1097 (10.00-3.50)

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	456	
1	B	456	
1	C	456	

2   Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 10194 atoms, of which 0 are hydrogen 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 Non-haem bromoperoxidase BPO-A2, Matrix protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	440	Total 3398	C 2166	N 572	O 652	S 8	0	0	0
1	B	440	Total 3398	C 2166	N 572	O 652	S 8	0	0	0
1	C	440	Total 3398	C 2166	N 572	O 652	S 8	0	0	0

There are 54 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	24	THR	GLN	ENGINEERED MUTATION	UNP P29715
A	118	ALA	LYS	ENGINEERED MUTATION	UNP P29715
A	278	ALA	-	LINKER	UNP P03485
A	279	GLN	-	LINKER	UNP P03485
A	280	GLU	-	LINKER	UNP P03485
A	281	ALA	-	LINKER	UNP P03485
A	282	GLN	-	LINKER	UNP P03485
A	283	LYS	-	LINKER	UNP P03485
A	284	GLN	-	LINKER	UNP P03485
A	285	LYS	-	LINKER	UNP P03485
A	448	LEU	-	EXPRESSION TAG	UNP P03485
A	449	GLU	-	EXPRESSION TAG	UNP P03485
A	450	HIS	-	EXPRESSION TAG	UNP P03485
A	451	HIS	-	EXPRESSION TAG	UNP P03485
A	452	HIS	-	EXPRESSION TAG	UNP P03485
A	453	HIS	-	EXPRESSION TAG	UNP P03485
A	454	HIS	-	EXPRESSION TAG	UNP P03485
A	455	HIS	-	EXPRESSION TAG	UNP P03485
B	24	THR	GLN	ENGINEERED MUTATION	UNP P29715
B	118	ALA	LYS	ENGINEERED MUTATION	UNP P29715
B	278	ALA	-	LINKER	UNP P03485
B	279	GLN	-	LINKER	UNP P03485
B	280	GLU	-	LINKER	UNP P03485

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	281	ALA	-	LINKER	UNP P03485
B	282	GLN	-	LINKER	UNP P03485
B	283	LYS	-	LINKER	UNP P03485
B	284	GLN	-	LINKER	UNP P03485
B	285	LYS	-	LINKER	UNP P03485
B	448	LEU	-	EXPRESSION TAG	UNP P03485
B	449	GLU	-	EXPRESSION TAG	UNP P03485
B	450	HIS	-	EXPRESSION TAG	UNP P03485
B	451	HIS	-	EXPRESSION TAG	UNP P03485
B	452	HIS	-	EXPRESSION TAG	UNP P03485
B	453	HIS	-	EXPRESSION TAG	UNP P03485
B	454	HIS	-	EXPRESSION TAG	UNP P03485
B	455	HIS	-	EXPRESSION TAG	UNP P03485
C	24	THR	GLN	ENGINEERED MUTATION	UNP P29715
C	118	ALA	LYS	ENGINEERED MUTATION	UNP P29715
C	278	ALA	-	LINKER	UNP P03485
C	279	GLN	-	LINKER	UNP P03485
C	280	GLU	-	LINKER	UNP P03485
C	281	ALA	-	LINKER	UNP P03485
C	282	GLN	-	LINKER	UNP P03485
C	283	LYS	-	LINKER	UNP P03485
C	284	GLN	-	LINKER	UNP P03485
C	285	LYS	-	LINKER	UNP P03485
C	448	LEU	-	EXPRESSION TAG	UNP P03485
C	449	GLU	-	EXPRESSION TAG	UNP P03485
C	450	HIS	-	EXPRESSION TAG	UNP P03485
C	451	HIS	-	EXPRESSION TAG	UNP P03485
C	452	HIS	-	EXPRESSION TAG	UNP P03485
C	453	HIS	-	EXPRESSION TAG	UNP P03485
C	454	HIS	-	EXPRESSION TAG	UNP P03485
C	455	HIS	-	EXPRESSION TAG	UNP P03485

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.
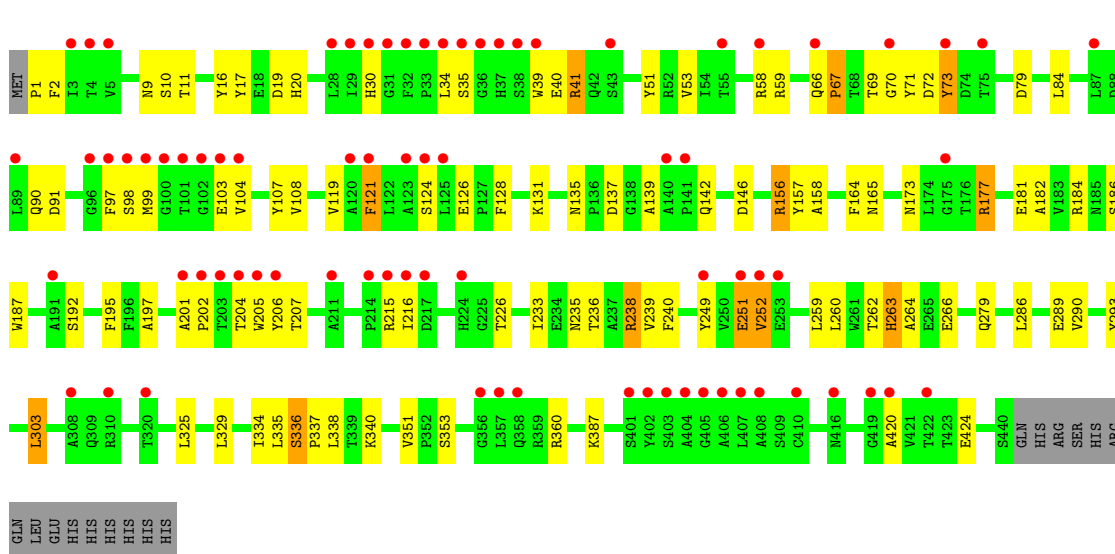
• Molecule 1: Non-haem bromoperoxidase BPO-A2, Matrix protein 1
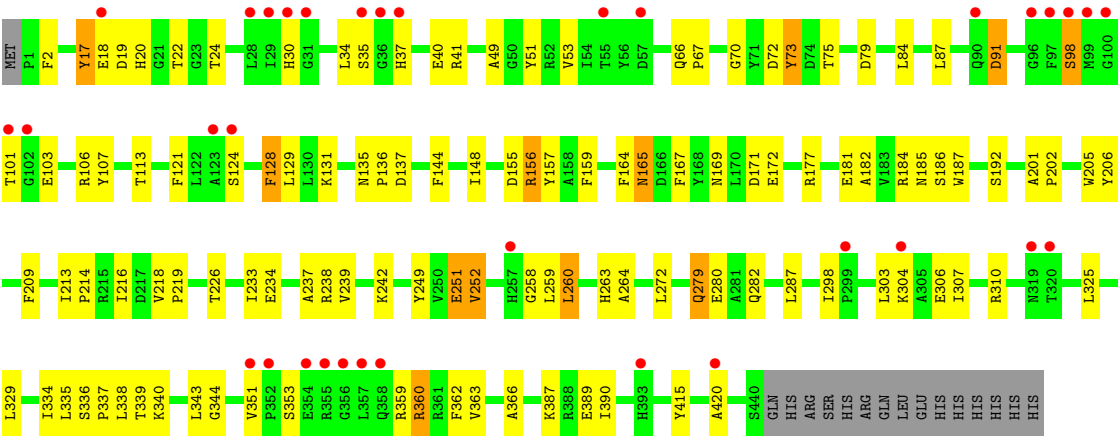
Chain A:
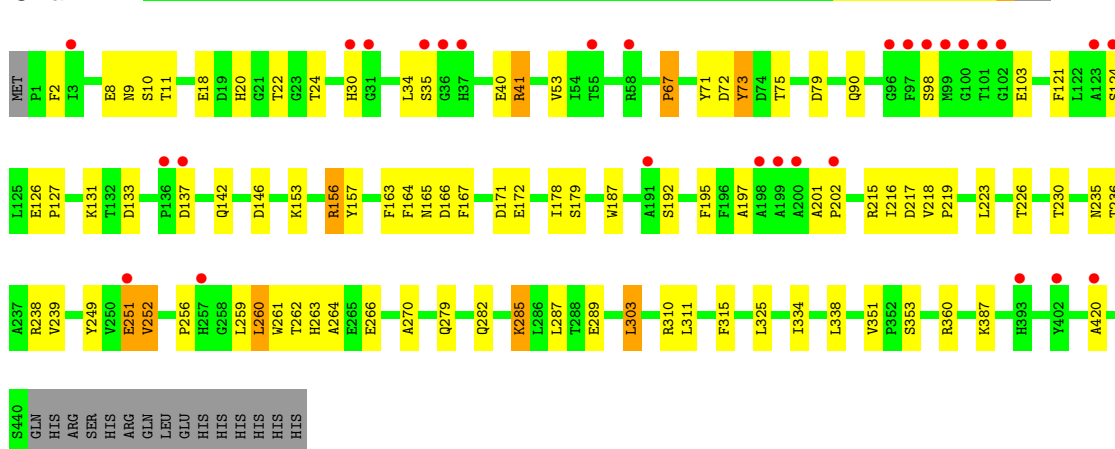
• Molecule 1: Non-haem bromoperoxidase BPO-A2, Matrix protein 1

Chain B:

• Molecule 1: Non-haem bromoperoxidase BPO-A2, Matrix protein 1

Chain C:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	129.67Å 160.98Å 167.63Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	28.42 – 7.35 29.28 – 7.35	Depositor EDS
% Data completeness (in resolution range)	97.0 (28.42-7.35) 97.2 (29.28-7.35)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.21 (at 7.22Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, R_{free}	0.276 , 0.288 0.274 , 0.282	Depositor DCC
R_{free} test set	121 reflections (4.98%)	DCC
Wilson B-factor (Å^2)	400.2	Xtriage
Anisotropy	0.357	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.24 , 210.1	EDS
Estimated twinning fraction	0.038 for -h,-l,-k	Xtriage
L-test for twinning	$< |L| > = 0.45, < L^2 > = 0.28$	Xtriage
Outliers	0 of 2439 reflections	Xtriage
F_o,F_c correlation	0.79	EDS
Total number of atoms	10194	wwPDB-VP
Average B, all atoms (Å^2)	318.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.27% 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	0.79	2/3475 (0.1%)	1.23	22/4729 (0.5%)
1	B	0.75	2/3475 (0.1%)	1.17	20/4729 (0.4%)
1	C	0.77	3/3475 (0.1%)	1.16	21/4729 (0.4%)
All	All	0.77	7/10425 (0.1%)	1.19	63/14187 (0.4%)

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	192	SER	CB-OG	5.35	1.49	1.42
1	C	136	PRO	N-CD	5.28	1.55	1.47
1	A	192	SER	CB-OG	5.16	1.49	1.42
1	C	192	SER	CB-OG	5.15	1.49	1.42
1	B	337	PRO	N-CD	5.07	1.54	1.47

The worst 5 of 63 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	17	TYR	CZ-CE2-CD2	-5.99	114.41	119.80
1	C	53	VAL	CA-CB-CG2	5.98	119.87	110.90
1	C	98	SER	CB-CA-C	5.98	121.46	110.10
1	A	98	SER	CB-CA-C	5.97	121.45	110.10
1	A	53	VAL	CA-CB-CG2	5.96	119.84	110.90

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	3398	0	0	62	0
1	B	3398	0	0	90	0
1	C	3398	0	0	95	0
All	All	10194	0	0	211	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 21.

The worst 5 of 211 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:99:MET:SD	1:B:205:TRP:NE1	2.29	1.06
1:B:303:LEU:CD1	1:B:334:ILE:CB	2.35	1.04
1:B:10:SER:CB	1:C:87:LEU:CD2	2.37	1.01
1:C:24:THR:CG2	1:C:51:TYR:CD1	2.46	0.97
1:A:11:THR:CG2	1:B:17:TYR:CE2	2.51	0.94

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	438/456 (96%)	427 (98%)	11 (2%)	0	100	100
1	B	438/456 (96%)	428 (98%)	10 (2%)	0	100	100
1	C	438/456 (96%)	427 (98%)	11 (2%)	0	100	100
All	All	1314/1368 (96%)	1282 (98%)	32 (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	354/370 (96%)	337 (95%)	17 (5%)	35	79
1	B	354/370 (96%)	334 (94%)	20 (6%)	30	75
1	C	354/370 (96%)	337 (95%)	17 (5%)	35	79
All	All	1062/1110 (96%)	1008 (95%)	54 (5%)	33	78

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	165	ASN
1	B	290	VAL
1	C	338	LEU
1	B	177	ARG
1	B	238	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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	440/456 (96%)	0.61	29 (6%)	18	25	184, 286, 365, 403	0
1	B	440/456 (96%)	1.20	77 (17%)	2	7	268, 397, 443, 467	0
1	C	440/456 (96%)	0.62	34 (7%)	13	21	165, 260, 336, 373	0
All	All	1320/1368 (96%)	0.81	140 (10%)	7	15	165, 306, 426, 467	0

The worst 5 of 140 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	30	HIS	9.6
1	A	30	HIS	9.5
1	B	100	GLY	9.5
1	B	101	THR	9.1
1	B	96	GLY	8.5

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 (i)

There are no such residues in this entry.