



wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 04:58 pm BST

PDB ID : 4QF0
Title : Structure of a 16 nm protein cage designed by fusing symmetric oligomeric domains, quadruple mutant, P21212 form
Authors : Lai, Y.-T.; Yeates, T.O.
Deposited on : 2014-05-19
Resolution : 6.49 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

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	:	2.11
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)	:	2.11

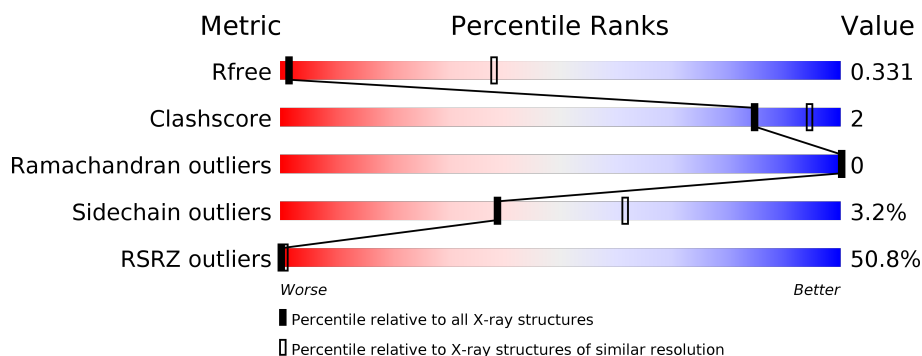
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 6.49 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	1000 (9.00-3.90)
Clashscore	141614	1064 (9.00-3.90)
Ramachandran outliers	138981	1012 (9.00-3.88)
Sidechain outliers	138945	1010 (9.00-3.84)
RSRZ outliers	127900	1002 (9.00-3.80)

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 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 $\leq 5\%$. 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	456	<div> <div>54%</div> <div> <div>88%</div> <div>9%</div> <div>.</div> </div> </div>
1	B	456	<div> <div>47%</div> <div> <div>88%</div> <div>9%</div> <div>.</div> </div> </div>
1	C	456	<div> <div>37%</div> <div> <div>88%</div> <div>9%</div> <div>.</div> </div> </div>
1	D	456	<div> <div>57%</div> <div> <div>88%</div> <div>9%</div> <div>.</div> </div> </div>
1	E	456	<div> <div>55%</div> <div> <div>88%</div> <div>9%</div> <div>.</div> </div> </div>
1	F	456	<div> <div>46%</div> <div> <div>88%</div> <div>9%</div> <div>.</div> </div> </div>

2 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 20256 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 Non-haem bromoperoxidase BPO-A2, Matrix protein 1 chimera.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	441	Total	C	N	O	S	0	0	0
			3376	2151	569	648	8			
1	B	441	Total	C	N	O	S	0	0	0
			3376	2151	569	648	8			
1	C	441	Total	C	N	O	S	0	0	0
			3376	2151	569	648	8			
1	D	441	Total	C	N	O	S	0	0	0
			3376	2151	569	648	8			
1	E	441	Total	C	N	O	S	0	0	0
			3376	2151	569	648	8			
1	F	441	Total	C	N	O	S	0	0	0
			3376	2151	569	648	8			

There are 114 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	24	THR	GLN	ENGINEERED MUTATION	UNP P29715
A	51	ALA	TYR	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

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	454	HIS	-	EXPRESSION TAG	UNP P03485
A	455	HIS	-	EXPRESSION TAG	UNP P03485
B	24	THR	GLN	ENGINEERED MUTATION	UNP P29715
B	51	ALA	TYR	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
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	51	ALA	TYR	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
D	24	THR	GLN	ENGINEERED MUTATION	UNP P29715
D	51	ALA	TYR	ENGINEERED MUTATION	UNP P29715

Continued on next page...

Continued from previous page...


Chain	Residue	Modelled	Actual	Comment	Reference
D	118	ALA	LYS	ENGINEERED MUTATION	UNP P29715
D	278	ALA	-	LINKER	UNP P03485
D	279	GLN	-	LINKER	UNP P03485
D	280	GLU	-	LINKER	UNP P03485
D	281	ALA	-	LINKER	UNP P03485
D	282	GLN	-	LINKER	UNP P03485
D	283	LYS	-	LINKER	UNP P03485
D	284	GLN	-	LINKER	UNP P03485
D	285	LYS	-	LINKER	UNP P03485
D	448	LEU	-	EXPRESSION TAG	UNP P03485
D	449	GLU	-	EXPRESSION TAG	UNP P03485
D	450	HIS	-	EXPRESSION TAG	UNP P03485
D	451	HIS	-	EXPRESSION TAG	UNP P03485
D	452	HIS	-	EXPRESSION TAG	UNP P03485
D	453	HIS	-	EXPRESSION TAG	UNP P03485
D	454	HIS	-	EXPRESSION TAG	UNP P03485
D	455	HIS	-	EXPRESSION TAG	UNP P03485
E	24	THR	GLN	ENGINEERED MUTATION	UNP P29715
E	51	ALA	TYR	ENGINEERED MUTATION	UNP P29715
E	118	ALA	LYS	ENGINEERED MUTATION	UNP P29715
E	278	ALA	-	LINKER	UNP P03485
E	279	GLN	-	LINKER	UNP P03485
E	280	GLU	-	LINKER	UNP P03485
E	281	ALA	-	LINKER	UNP P03485
E	282	GLN	-	LINKER	UNP P03485
E	283	LYS	-	LINKER	UNP P03485
E	284	GLN	-	LINKER	UNP P03485
E	285	LYS	-	LINKER	UNP P03485
E	448	LEU	-	EXPRESSION TAG	UNP P03485
E	449	GLU	-	EXPRESSION TAG	UNP P03485
E	450	HIS	-	EXPRESSION TAG	UNP P03485
E	451	HIS	-	EXPRESSION TAG	UNP P03485
E	452	HIS	-	EXPRESSION TAG	UNP P03485
E	453	HIS	-	EXPRESSION TAG	UNP P03485
E	454	HIS	-	EXPRESSION TAG	UNP P03485
E	455	HIS	-	EXPRESSION TAG	UNP P03485
F	24	THR	GLN	ENGINEERED MUTATION	UNP P29715
F	51	ALA	TYR	ENGINEERED MUTATION	UNP P29715
F	118	ALA	LYS	ENGINEERED MUTATION	UNP P29715
F	278	ALA	-	LINKER	UNP P03485
F	279	GLN	-	LINKER	UNP P03485
F	280	GLU	-	LINKER	UNP P03485

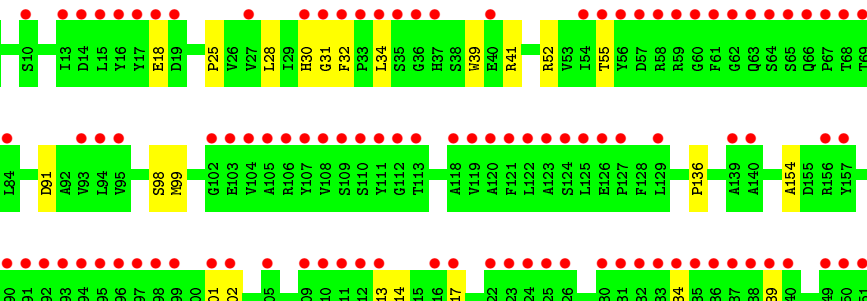
Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	281	ALA	-	LINKER	UNP P03485
F	282	GLN	-	LINKER	UNP P03485
F	283	LYS	-	LINKER	UNP P03485
F	284	GLN	-	LINKER	UNP P03485
F	285	LYS	-	LINKER	UNP P03485
F	448	LEU	-	EXPRESSION TAG	UNP P03485
F	449	GLU	-	EXPRESSION TAG	UNP P03485
F	450	HIS	-	EXPRESSION TAG	UNP P03485
F	451	HIS	-	EXPRESSION TAG	UNP P03485
F	452	HIS	-	EXPRESSION TAG	UNP P03485
F	453	HIS	-	EXPRESSION TAG	UNP P03485
F	454	HIS	-	EXPRESSION TAG	UNP P03485
F	455	HIS	-	EXPRESSION TAG	UNP P03485

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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 ($\text{RSRZ} > 2$). 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.

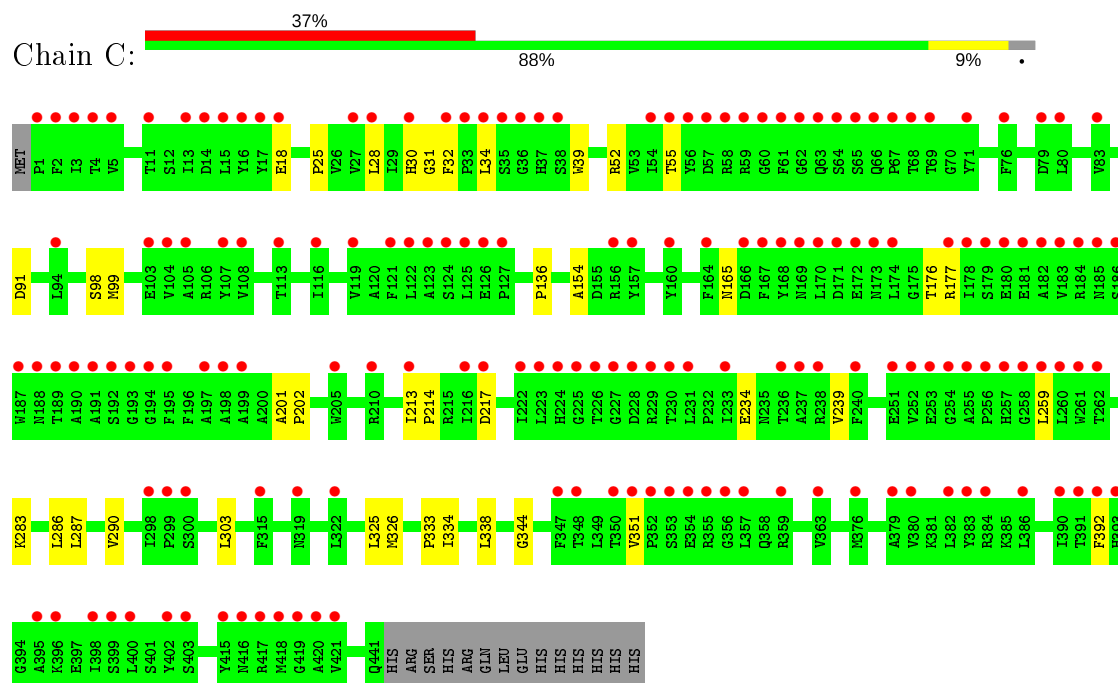
- Chain A: 



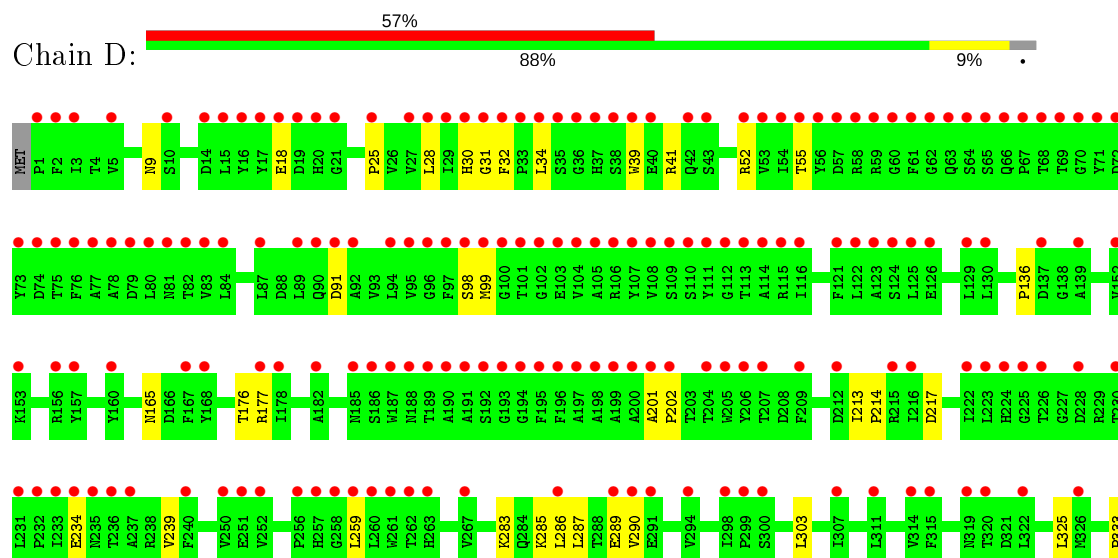
Category	Value
L413	54%
I414	88%
Y415	88%
M416	88%
R417	88%
A418	88%
G419	88%
C420	88%
V421	88%
F422	88%
D423	88%
B424	88%
N425	88%
H426	88%
P427	88%
J428	88%
K429	88%
X430	88%
Q431	88%
W432	88%
E433	88%
T434	88%
A435	88%
C436	88%
I437	88%
L438	88%
M439	88%
N440	88%
O441	88%
P442	88%
Q443	88%
R444	88%
S445	88%
T446	88%
U447	88%
V448	88%
W449	88%
X450	88%
Y451	88%
Z452	88%
A453	88%
B454	88%
C455	88%
D456	88%
E457	88%
F458	88%
G459	88%
H460	88%
I461	88%
J462	88%
K463	88%
L464	88%
M465	88%
N466	88%
O467	88%
P468	88%
Q469	88%
R470	88%
S471	88%
T472	88%
U473	88%
V474	88%
W475	88%
X476	88%
Y477	88%
Z478	88%
A479	88%
B480	88%
C481	88%
D482	88%
E483	88%
F484	88%
G485	88%
H486	88%
I487	88%
J488	88%
K489	88%
L490	88%
M491	88%
N492	88%
O493	88%
P494	88%
Q495	88%
R496	88%
S497	88%
T498	88%
U499	88%
V500	88%

- Chain B:
-
- 47%
- 88%
- 9%
-

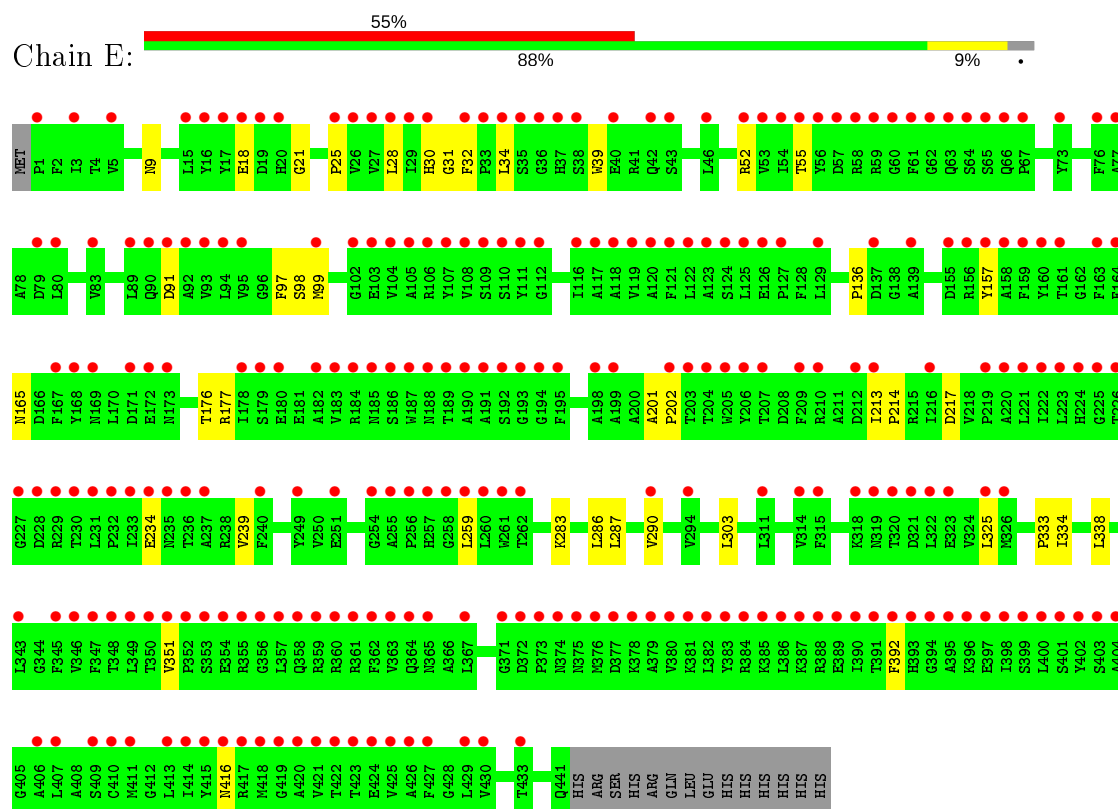
- Molecule 1: Non-haem bromoperoxidase BPO-A2, Matrix protein 1 chimera



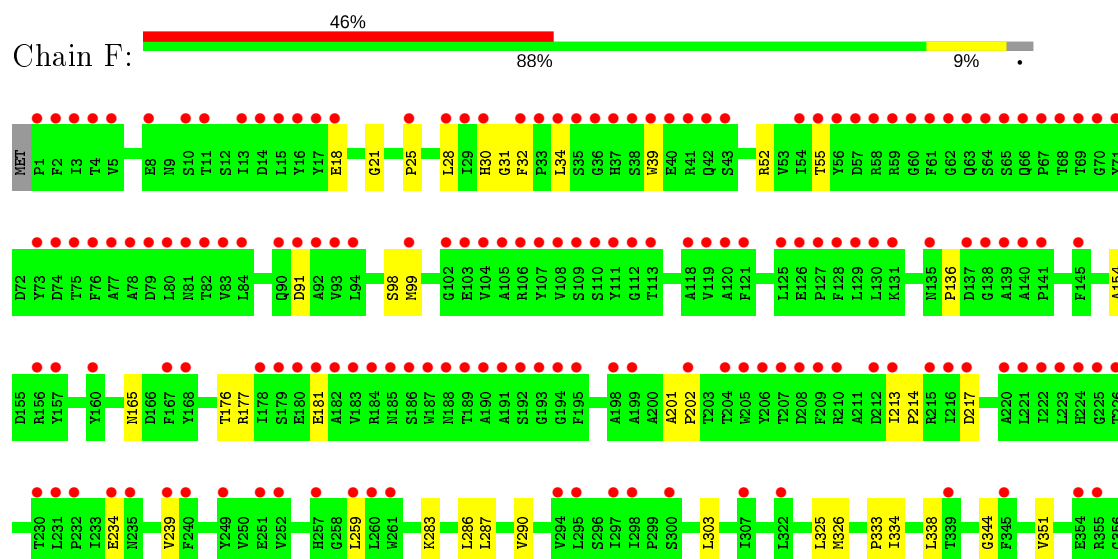
- Molecule 1: Non-haem bromoperoxidase BPO-A2, Matrix protein 1 chimera

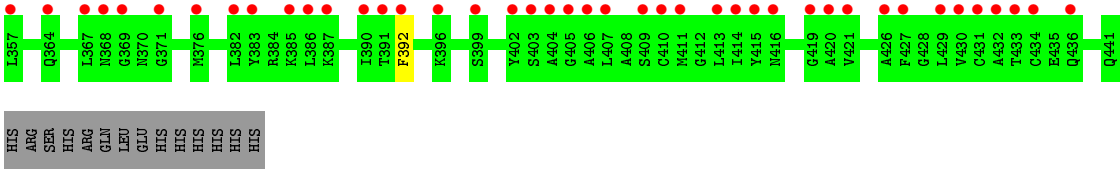


- Molecule 1: Non-haem bromoperoxidase BPO-A2, Matrix protein 1 chimera



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





4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	175.50Å 147.73Å 167.63Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	93.71 – 6.49 93.71 – 6.49	Depositor EDS
% Data completeness (in resolution range)	99.6 (93.71-6.49) 99.6 (93.71-6.49)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.07 (at 6.72Å)	Xtriage
Refinement program	REFMAC 5.8.0071	Depositor
R, R_{free}	0.271 , 0.324 0.273 , 0.331	Depositor DCC
R_{free} test set	451 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	398.2	Xtriage
Anisotropy	0.081	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 410.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.003 for l,-k,h	Xtriage
F_o, F_c correlation	0.86	EDS
Total number of atoms	20256	wwPDB-VP
Average B, all atoms (Å ²)	313.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.70% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	0.41	0/3452	0.55	0/4702
1	B	0.40	0/3452	0.56	0/4702
1	C	0.41	0/3452	0.56	0/4702
1	D	0.40	0/3452	0.56	0/4702
1	E	0.40	0/3452	0.56	0/4702
1	F	0.41	0/3452	0.56	0/4702
All	All	0.40	0/20712	0.56	0/28212

There are no bond length outliers.

There are no bond angle outliers.

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	3376	0	3286	16	0
1	B	3376	0	3286	16	0
1	C	3376	0	3286	15	0
1	D	3376	0	3286	16	0
1	E	3376	0	3286	18	0
1	F	3376	0	3286	18	0
All	All	20256	0	19716	92	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 2.

The worst 5 of 92 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:F:30:HIS:CE1	1:F:34:LEU:O	2.28	0.85
1:B:30:HIS:CE1	1:B:34:LEU:O	2.32	0.82
1:A:30:HIS:CE1	1:A:34:LEU:O	2.33	0.81
1:C:30:HIS:CE1	1:C:34:LEU:O	2.33	0.81
1:E:30:HIS:CE1	1:E:34:LEU:O	2.38	0.77

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	439/456 (96%)	413 (94%)	26 (6%)	0	100	100
1	B	439/456 (96%)	413 (94%)	26 (6%)	0	100	100
1	C	439/456 (96%)	414 (94%)	25 (6%)	0	100	100
1	D	439/456 (96%)	413 (94%)	26 (6%)	0	100	100
1	E	439/456 (96%)	414 (94%)	25 (6%)	0	100	100
1	F	439/456 (96%)	414 (94%)	25 (6%)	0	100	100
All	All	2634/2736 (96%)	2481 (94%)	153 (6%)	0	100	100

There are no Ramachandran outliers to report.

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	348/369 (94%)	336 (97%)	12 (3%)	37	60
1	B	348/369 (94%)	337 (97%)	11 (3%)	39	61
1	C	348/369 (94%)	337 (97%)	11 (3%)	39	61
1	D	348/369 (94%)	337 (97%)	11 (3%)	39	61
1	E	348/369 (94%)	337 (97%)	11 (3%)	39	61
1	F	348/369 (94%)	337 (97%)	11 (3%)	39	61
All	All	2088/2214 (94%)	2021 (97%)	67 (3%)	39	61

5 of 67 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	303	LEU
1	D	217	ASP
1	F	239	VAL
1	C	325	LEU
1	D	91	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	188	ASN
1	F	188	ASN
1	D	188	ASN
1	B	188	ASN
1	E	188	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 [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(Å ²)	Q<0.9
1	A	441/456 (96%)	2.63	244 (55%) 0 0	121, 325, 340, 340	0
1	B	441/456 (96%)	2.21	215 (48%) 0 1	122, 314, 340, 340	0
1	C	441/456 (96%)	1.87	169 (38%) 0 1	122, 313, 340, 340	0
1	D	441/456 (96%)	2.96	258 (58%) 0 0	122, 321, 340, 340	0
1	E	441/456 (96%)	2.71	251 (56%) 0 0	122, 320, 340, 340	0
1	F	441/456 (96%)	2.18	208 (47%) 0 1	121, 317, 340, 340	0
All	All	2646/2736 (96%)	2.43	1345 (50%) 0 1	121, 318, 340, 340	0

The worst 5 of 1345 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	358	GLN	14.3
1	A	359	ARG	14.1
1	D	33	PRO	14.0
1	A	66	GLN	13.5
1	B	359	ARG	13.2

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.