wwPDB X-ray Structure Validation Summary Report (i)

Jan 6, 2021 – 07:16 PM GMT

PDB ID	:	6YR4
Title	:	Dye-type peroxidase DtpB in the ferryl state: Spectroscopically Validated composite structure
Authors	:	Lucic, M.; Dworkowski, F.S.N.; Worrall, J.A.R.; Hough, M.A.
Deposited on	:	2020-04-19
Resolution	:	1.85 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

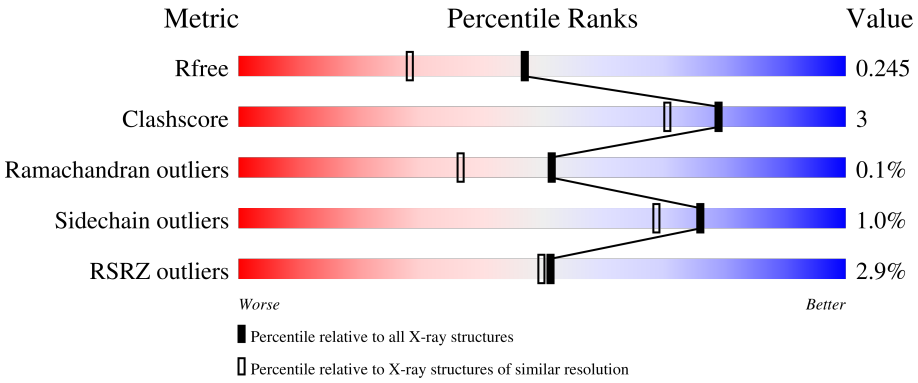
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.16
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.16

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.85 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.



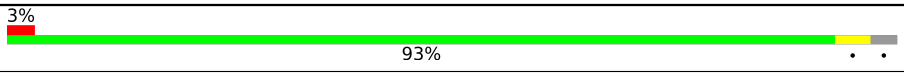
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	316	2% 89% 8%
1	B	316	5% 90% 6%
1	C	316	2% 92%
1	D	316	90% 6%
1	E	316	3% 91% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	316	

2 Entry composition (i)

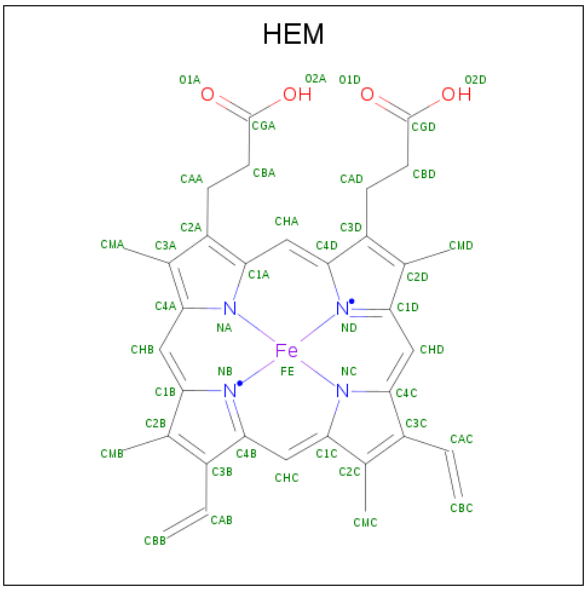
There are 6 unique types of molecules in this entry. The entry contains 15395 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Putative iron-dependent peroxidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	306	Total	C	N	O	S	0	3	0
			2333	1471	398	455	9			
1	B	305	Total	C	N	O	S	0	1	0
			2317	1460	403	445	9			
1	C	306	Total	C	N	O	S	0	2	0
			2341	1474	402	455	10			
1	D	303	Total	C	N	O	S	0	0	0
			2307	1451	399	448	9			
1	E	306	Total	C	N	O	S	0	0	0
			2309	1454	396	450	9			
1	F	306	Total	C	N	O	S	0	2	0
			2314	1458	394	453	9			

- Molecule 2 is PROTOPORPHYRIN IX CONTAINING FE (three-letter code: HEM) (formula: $C_{34}H_{32}FeN_4O_4$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		
2	B	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		
2	C	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		
2	D	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		
2	E	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		
2	F	1	Total	C	Fe	N	O	0	0
			43	34	1	4	4		

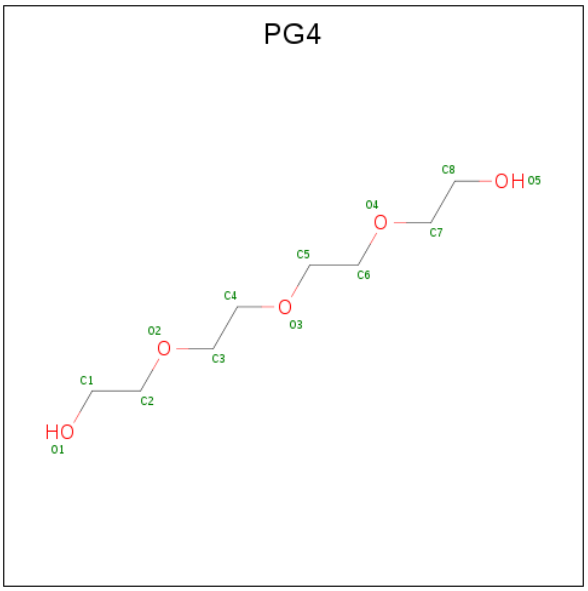
- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Mg	0	0
			1	1		
3	C	1	Total	Mg	0	0
			1	1		

- Molecule 4 is OXYGEN ATOM (three-letter code: O) (formula: O) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	D	1	Total	O	0	0
			1	1		
4	E	1	Total	O	0	0
			1	1		
4	B	1	Total	O	0	0
			1	1		
4	C	1	Total	O	0	0
			1	1		
4	A	1	Total	O	0	0
			1	1		
4	F	1	Total	O	0	0
			1	1		

- Molecule 5 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	D	1	Total C O 13 8 5	0	0

- Molecule 6 is water.

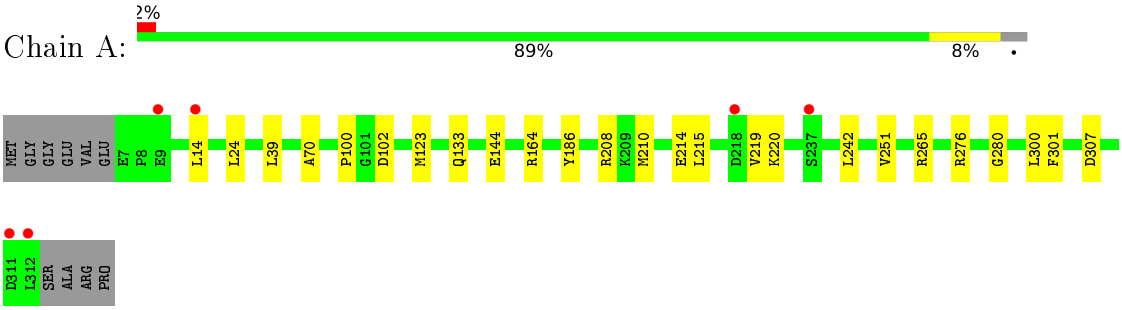
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	224	Total O 224 224	0	0
6	B	140	Total O 140 140	0	1
6	C	208	Total O 208 208	0	0
6	D	210	Total O 210 210	0	0
6	E	183	Total O 183 183	0	0
6	F	230	Total O 230 230	0	0

3 Residue-property plots

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 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.

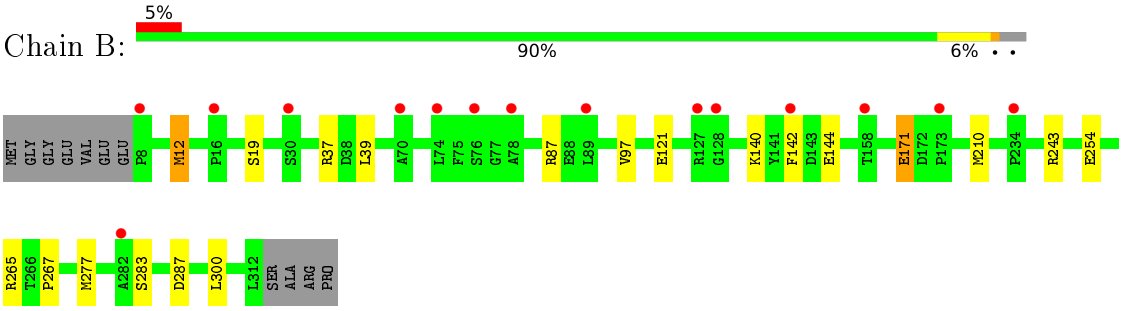
- Molecule 1: Putative iron-dependent peroxidase

Chain A:



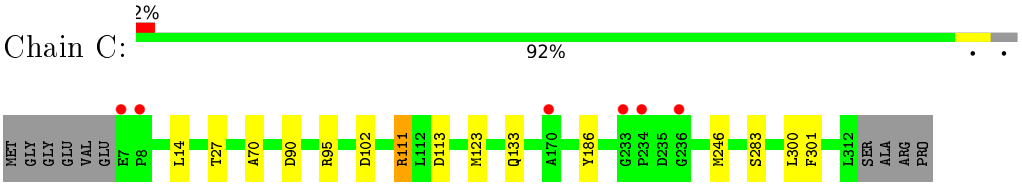
- Molecule 1: Putative iron-dependent peroxidase

Chain B:



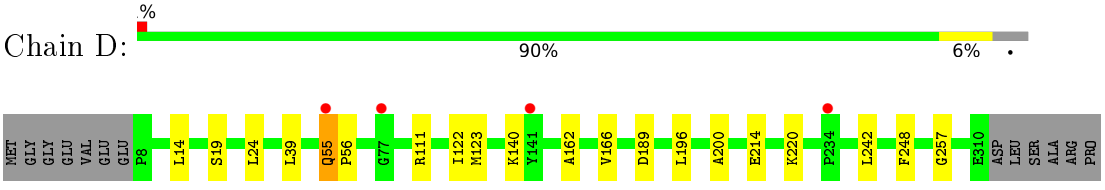
- Molecule 1: Putative iron-dependent peroxidase

Chain C:



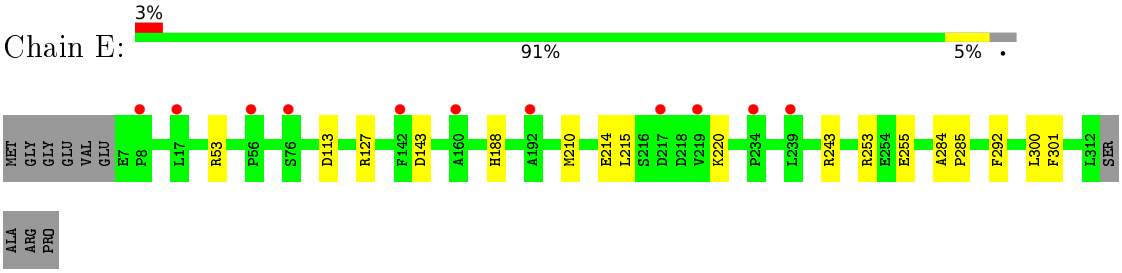
- Molecule 1: Putative iron-dependent peroxidase

Chain D:



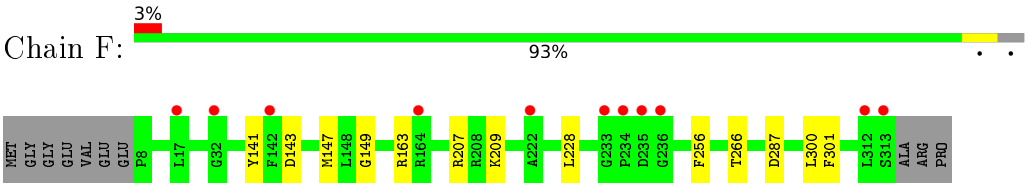
- Molecule 1: Putative iron-dependent peroxidase

Chain E:



• Molecule 1: Putative iron-dependent peroxidase

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	85.35Å 119.91Å 194.15Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.61 – 1.85 47.57 – 1.85	Depositor EDS
% Data completeness (in resolution range)	93.8 (47.61-1.85) 88.8 (47.57-1.85)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.15 (at 1.86Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.188 , 0.232 0.202 , 0.245	Depositor DCC
R_{free} test set	7982 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	22.3	Xtriage
Anisotropy	0.214	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 37.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	15395	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.35% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: HEM, MG, PG4, O

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	0.70	0/2389	0.85	0/3246
1	B	0.69	0/2367	0.83	1/3215 (0.0%)
1	C	0.71	0/2391	0.84	2/3248 (0.1%)
1	D	0.69	0/2357	0.85	0/3201
1	E	0.70	0/2358	0.83	0/3206
1	F	0.69	0/2363	0.84	1/3212 (0.0%)
All	All	0.70	0/14225	0.84	4/19328 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	111	ARG	NE-CZ-NH1	6.28	123.44	120.30
1	C	111	ARG	NE-CZ-NH2	-5.86	117.37	120.30
1	B	142	PHE	CB-CA-C	5.32	121.04	110.40
1	F	207	ARG	NE-CZ-NH2	5.23	122.92	120.30

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	2333	0	2251	15	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	2317	0	2247	20	0
1	C	2341	0	2264	11	0
1	D	2307	0	2237	13	0
1	E	2309	0	2225	10	0
1	F	2314	0	2212	9	0
2	A	43	0	30	3	0
2	B	43	0	30	2	0
2	C	43	0	30	2	0
2	D	43	0	30	1	0
2	E	43	0	30	0	0
2	F	43	0	30	1	0
3	A	1	0	0	0	0
3	C	1	0	0	0	0
4	A	1	0	0	0	0
4	B	1	0	0	1	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
4	E	1	0	0	1	0
4	F	1	0	0	0	0
5	D	13	0	18	0	0
6	A	224	0	0	1	0
6	B	140	0	0	2	0
6	C	208	0	0	4	0
6	D	210	0	0	2	0
6	E	183	0	0	1	0
6	F	230	0	0	0	0
All	All	15395	0	13634	78	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 3.

The worst 5 of 78 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:B:37[B]:ARG:HB2	1:B:37[B]:ARG:HH11	1.36	0.90
1:A:214:GLU:OE2	1:A:220:LYS:NZ	2.12	0.77
1:B:37[B]:ARG:HH11	1:B:37[B]:ARG:CB	1.99	0.75
1:A:14:LEU:HD23	1:A:242:LEU:HD22	1.72	0.72
1:A:144:GLU:OE1	6:A:501:HOH:O	2.10	0.68

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	307/316 (97%)	301 (98%)	6 (2%)	0	100 100
1	B	304/316 (96%)	295 (97%)	9 (3%)	0	100 100
1	C	306/316 (97%)	300 (98%)	6 (2%)	0	100 100
1	D	301/316 (95%)	292 (97%)	8 (3%)	1 (0%)	41 26
1	E	304/316 (96%)	297 (98%)	7 (2%)	0	100 100
1	F	306/316 (97%)	298 (97%)	8 (3%)	0	100 100
All	All	1828/1896 (96%)	1783 (98%)	44 (2%)	1 (0%)	51 36

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	55	GLN

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	242/252 (96%)	240 (99%)	2 (1%)	81 76
1	B	240/252 (95%)	235 (98%)	5 (2%)	53 38
1	C	244/252 (97%)	242 (99%)	2 (1%)	81 76
1	D	242/252 (96%)	241 (100%)	1 (0%)	91 89
1	E	239/252 (95%)	236 (99%)	3 (1%)	69 58
1	F	238/252 (94%)	236 (99%)	2 (1%)	81 76

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1445/1512 (96%)	1430 (99%)	15 (1%)	76 69

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	283	SER
1	C	14	LEU
1	E	215	LEU
1	B	171	GLU
1	E	210	MET

Sometimes 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 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

Of 15 ligands modelled in this entry, 8 are monoatomic - leaving 7 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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

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, 95^{th} 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	306/316 (96%)	-0.02	6 (1%) 65 64	17, 23, 39, 56	0
1	B	305/316 (96%)	0.36	15 (4%) 29 28	18, 27, 47, 70	0
1	C	306/316 (96%)	0.16	6 (1%) 65 64	17, 24, 41, 67	0
1	D	303/316 (95%)	0.07	4 (1%) 77 78	16, 24, 42, 58	0
1	E	306/316 (96%)	0.32	11 (3%) 42 40	19, 26, 45, 67	0
1	F	306/316 (96%)	0.13	11 (3%) 42 40	17, 24, 40, 68	0
All	All	1832/1896 (96%)	0.17	53 (2%) 51 50	16, 25, 43, 70	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	234	PRO	8.2
1	F	313	SER	5.6
1	A	218	ASP	5.4
1	D	77	GLY	5.3
1	F	234	PRO	5.1

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 (i)

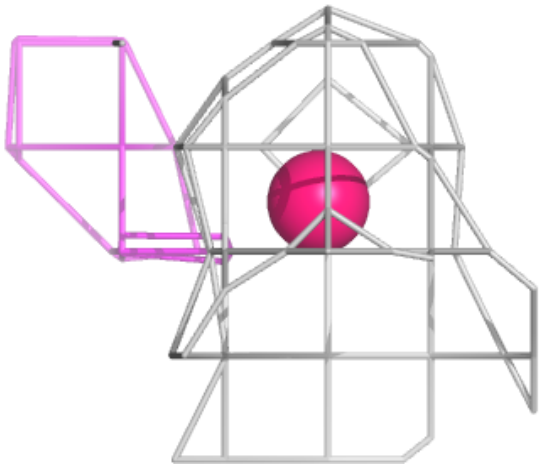
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

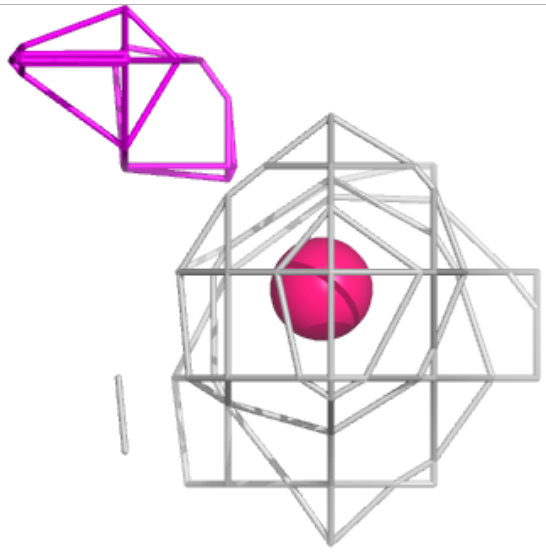
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	PG4	D	402	13/13	0.79	0.19	41,49,52,53	0
2	HEM	F	401	43/43	0.95	0.12	17,19,21,22	0
2	HEM	D	401	43/43	0.96	0.11	15,17,20,21	0
2	HEM	A	401	43/43	0.97	0.10	17,19,22,27	0
2	HEM	C	401	43/43	0.97	0.10	17,20,22,28	0
2	HEM	E	401	43/43	0.98	0.09	18,21,23,24	0
3	MG	C	402	1/1	0.98	0.03	24,24,24,24	0
2	HEM	B	401	43/43	0.98	0.10	18,19,22,25	0
4	O	D	403	1/1	0.99	0.15	18,18,18,18	0
3	MG	A	402	1/1	0.99	0.05	21,21,21,21	0
4	O	C	403	1/1	0.99	0.06	20,20,20,20	0
4	O	E	402	1/1	0.99	0.14	21,21,21,21	0
4	O	A	403	1/1	0.99	0.07	17,17,17,17	0
4	O	B	402	1/1	0.99	0.13	24,24,24,24	0
4	O	F	402	1/1	0.99	0.13	20,20,20,20	0

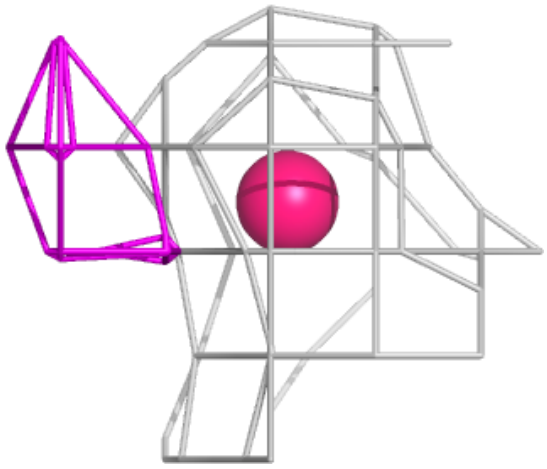
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around O D 403:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

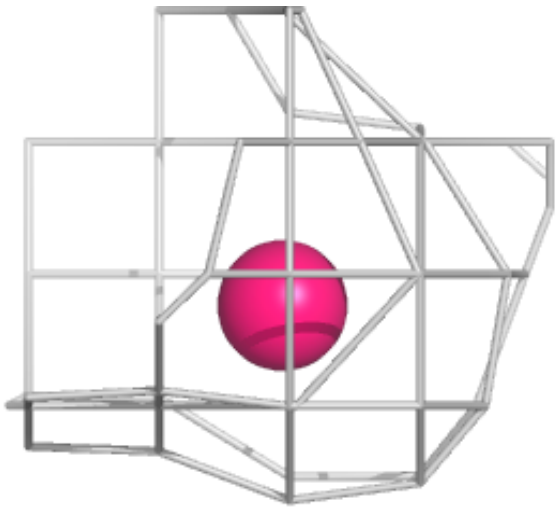


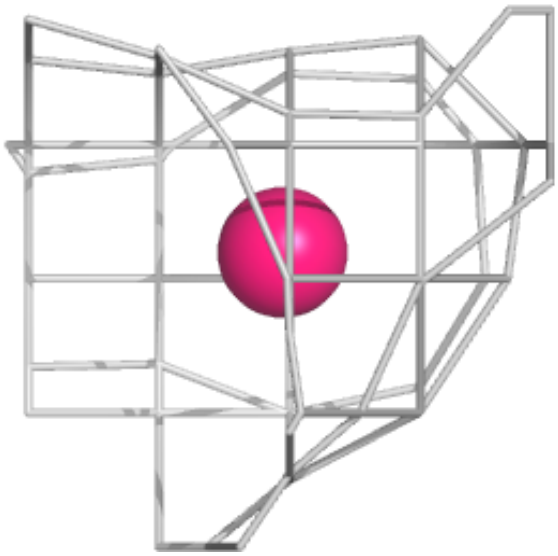


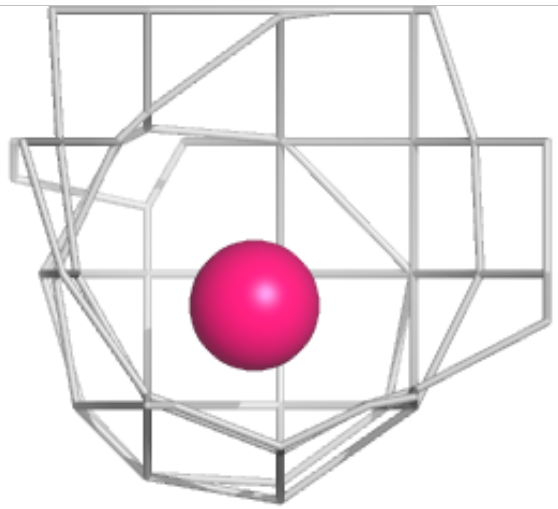


Electron density around O C 403:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

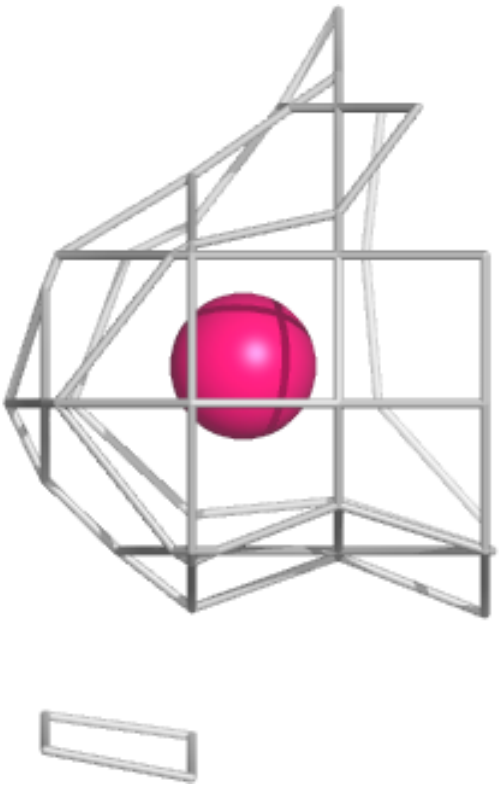


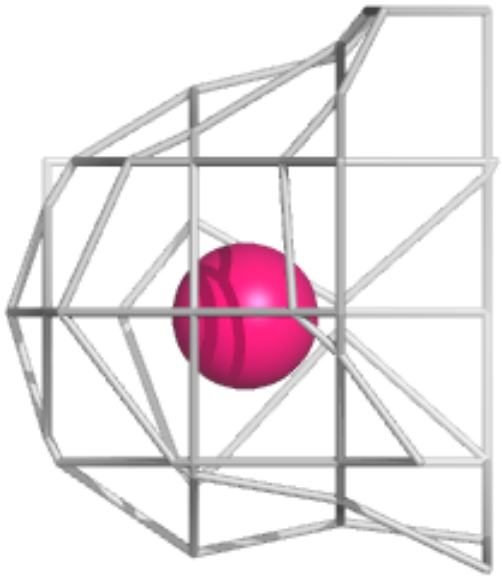


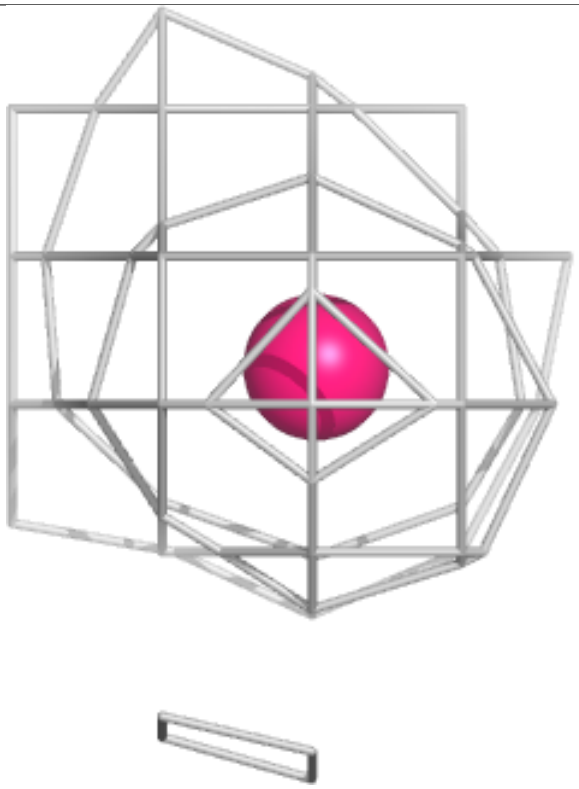


Electron density around O E 402:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

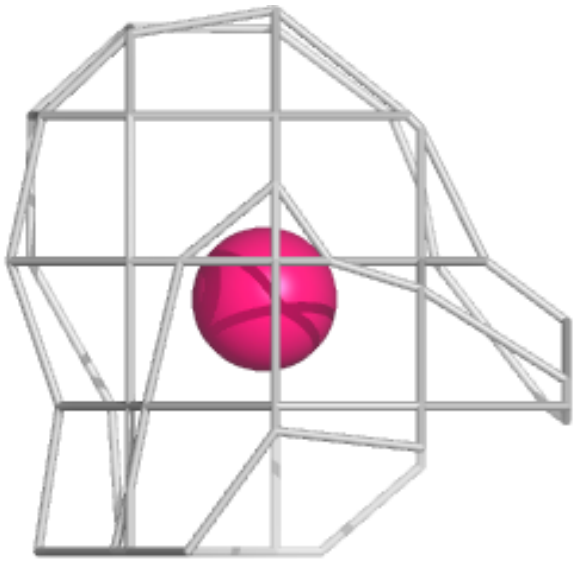


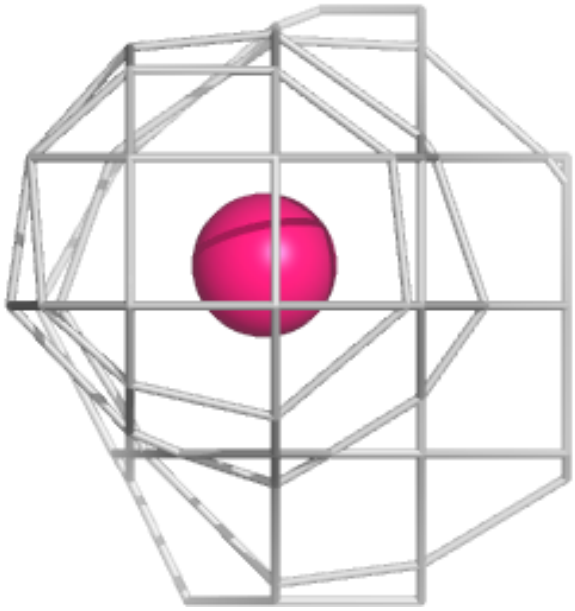


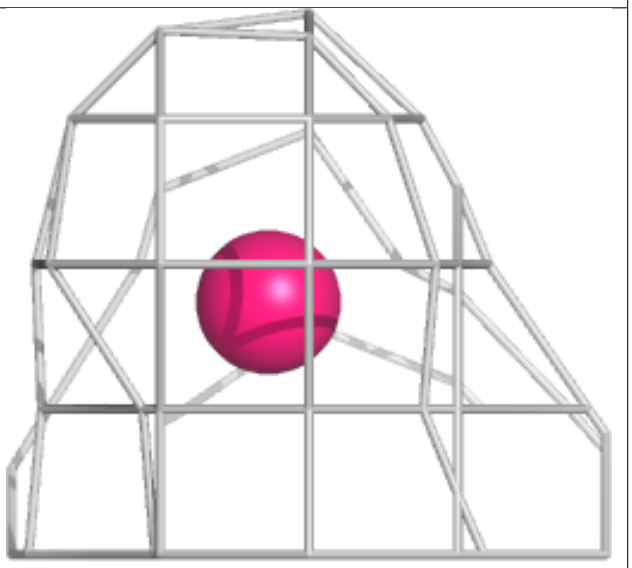


Electron density around O A 403:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

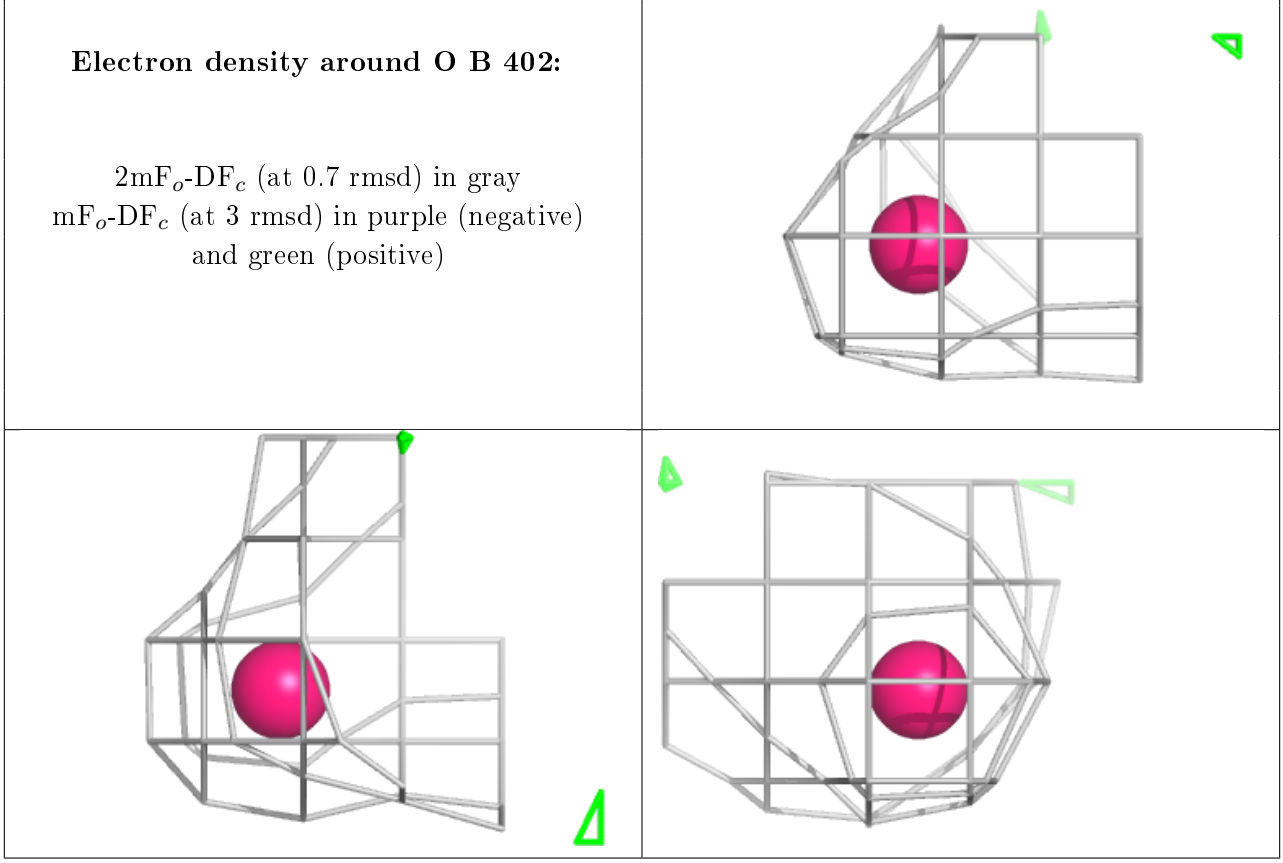






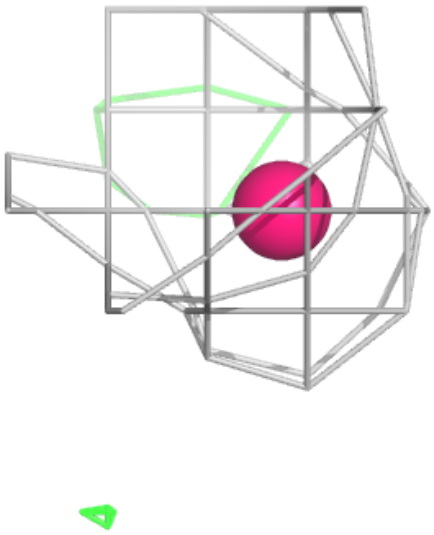
Electron density around O B 402:

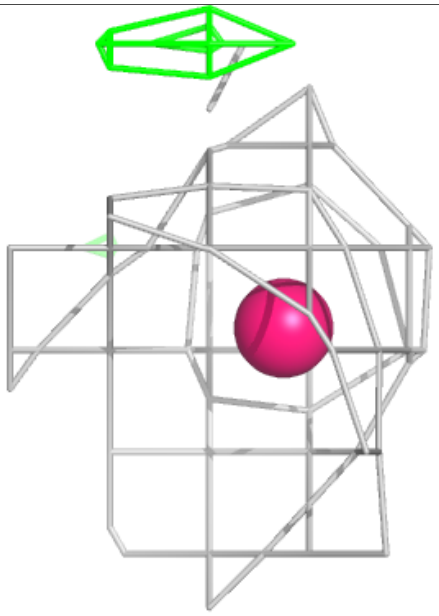
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

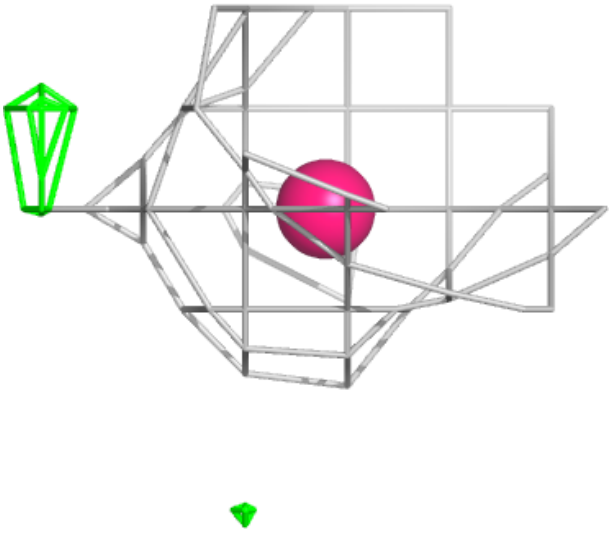


Electron density around O F 402:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)







6.5 Other polymers

There are no such residues in this entry.