Full wwPDB X-ray Structure Validation Report (i)

Mar 2, 2023 – 10:05 AM EST

PDB ID	:	8HFB
Title	:	Evolved variant of quercetin 2,4-dioxygenase from Bacillus subtilis
Authors	:	Eom, H.; Song, W.J.
Deposited on	:	2022-11-10
Resolution	:	2.24 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

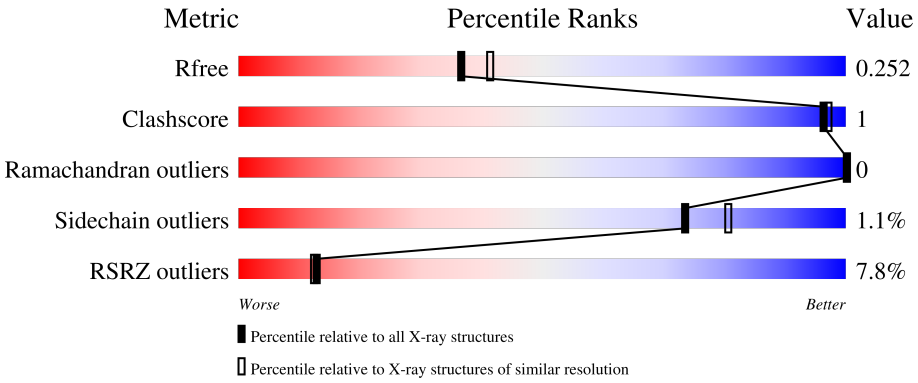
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.32.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.1

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 2.24 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	2391 (2.26-2.22)
Clashscore	141614	2539 (2.26-2.22)
Ramachandran outliers	138981	2489 (2.26-2.22)
Sidechain outliers	138945	2490 (2.26-2.22)
RSRZ outliers	127900	2353 (2.26-2.22)

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	337	7% 90% · 5%
1	B	337	7% 90% · 8%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 5143 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.

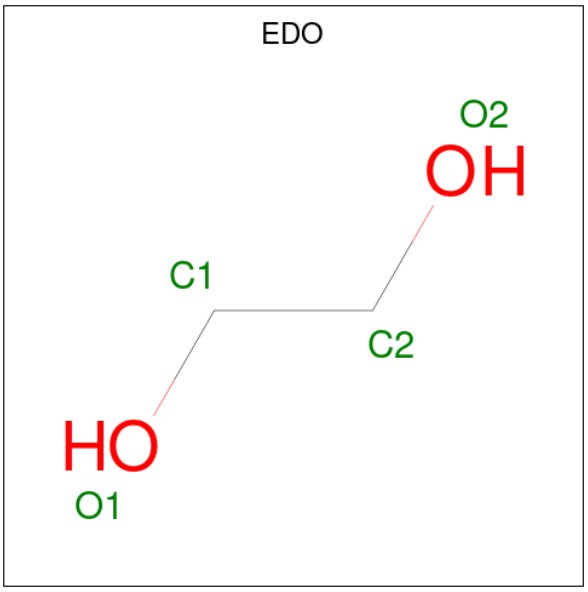
- Molecule 1 is a protein called Quercetin 2,3-dioxygenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	319	Total	C	N	O	S	0	0	0
			2490	1582	423	473	12			
1	B	311	Total	C	N	O	S	0	1	0
			2440	1548	417	464	11			

There are 16 discrepancies between the modelled and reference sequences:

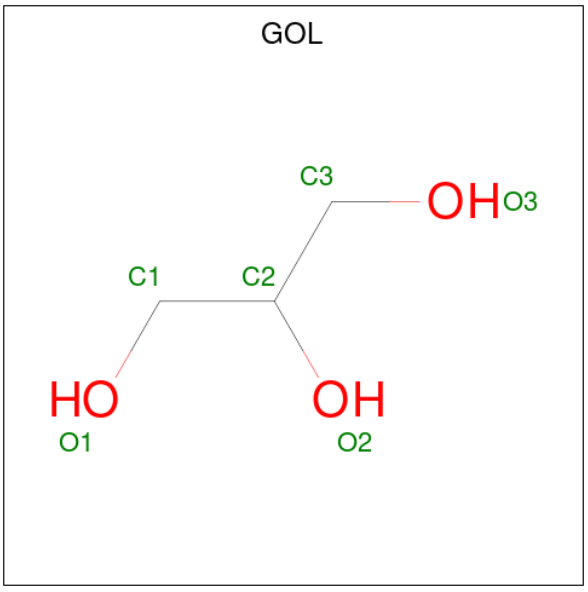
Chain	Residue	Modelled	Actual	Comment	Reference
A	160	THR	ASP	conflict	UNP P42106
A	221	LEU	VAL	engineered mutation	UNP P42106
A	231	VAL	ILE	engineered mutation	UNP P42106
A	239	CYS	HIS	engineered mutation	UNP P42106
A	269	ALA	VAL	engineered mutation	UNP P42106
A	286	PHE	MET	engineered mutation	UNP P42106
A	290	VAL	LEU	engineered mutation	UNP P42106
A	314	LYS	GLU	conflict	UNP P42106
B	160	THR	ASP	conflict	UNP P42106
B	221	LEU	VAL	engineered mutation	UNP P42106
B	231	VAL	ILE	engineered mutation	UNP P42106
B	239	CYS	HIS	engineered mutation	UNP P42106
B	269	ALA	VAL	engineered mutation	UNP P42106
B	286	PHE	MET	engineered mutation	UNP P42106
B	290	VAL	LEU	engineered mutation	UNP P42106
B	314	LYS	GLU	conflict	UNP P42106

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	B	1	Total C O 4 2 2	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 6 3 3	0	0

- Molecule 4 is NICKEL (II) ION (three-letter code: NI) (formula: Ni) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	2	Total Ni 2 2	0	0
4	B	2	Total Ni 2 2	0	0

- Molecule 5 is water.

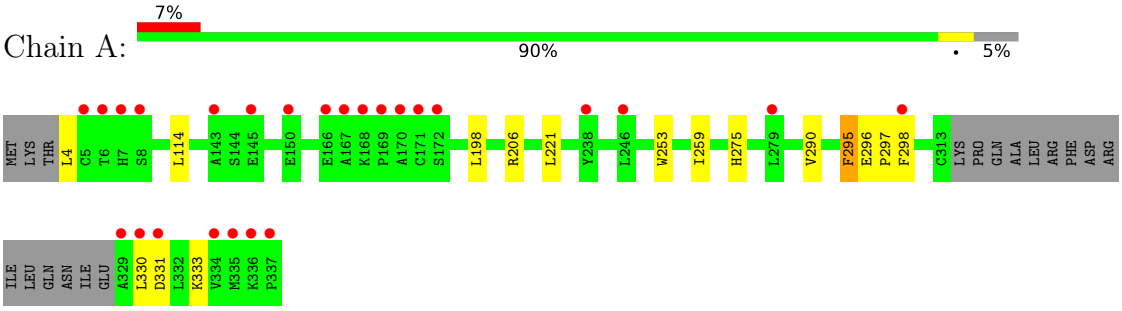
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	97	Total O 97 97	0	0
5	B	90	Total O 90 90	0	0

3 Residue-property plots (i)

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.

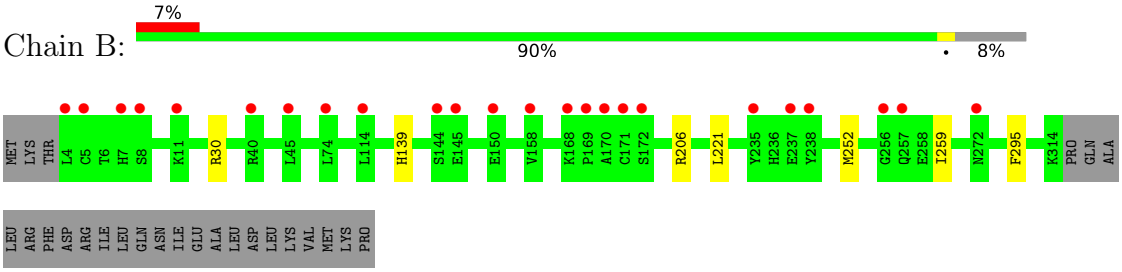
• Molecule 1: Quercetin 2,3-dioxygenase

Chain A:



• Molecule 1: Quercetin 2,3-dioxygenase

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	94.30Å 125.21Å 119.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.19 – 2.24 43.13 – 2.24	Depositor EDS
% Data completeness (in resolution range)	95.2 (47.19-2.24) 95.3 (43.13-2.24)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.47 (at 2.24Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.213 , 0.251 0.215 , 0.252	Depositor DCC
R_{free} test set	1626 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	48.6	Xtriage
Anisotropy	0.005	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 48.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5143	wwPDB-VP
Average B, all atoms (Å^2)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.37% 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: NI, GOL, EDO

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	0.65	0/2554	0.71	0/3468
1	B	0.65	0/2505	0.70	0/3403
All	All	0.65	0/5059	0.71	0/6871

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

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	2490	0	2407	7	0
1	B	2440	0	2345	3	0
2	A	12	0	18	0	0
2	B	4	0	6	0	0
3	A	6	0	8	0	0
4	A	2	0	0	0	0
4	B	2	0	0	0	0
5	A	97	0	0	0	0
5	B	90	0	0	0	0
All	All	5143	0	4784	10	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 1.

All (10) 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:298:PHE:HB3	1:A:331:ASP:HB3	1.90	0.53
1:B:206:ARG:O	1:B:221:LEU:HA	2.12	0.49
1:A:290:VAL:HG11	1:A:295:PHE:CE2	2.50	0.47
1:B:252:MET:HB2	1:B:259:ILE:HD11	1.97	0.46
1:A:295:PHE:CD1	1:A:295:PHE:C	2.89	0.46
1:A:206:ARG:O	1:A:221:LEU:HA	2.16	0.46
1:A:296:GLU:HB3	1:A:297:PRO:HD3	2.00	0.43
1:A:253:TRP:O	1:A:275:HIS:HA	2.20	0.42
1:B:30:ARG:O	1:B:139:HIS:HB2	2.20	0.42
1:A:4:LEU:HD22	1:A:259:ILE:HD11	2.02	0.41

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	315/337 (94%)	304 (96%)	11 (4%)	0	100	100
1	B	310/337 (92%)	302 (97%)	8 (3%)	0	100	100
All	All	625/674 (93%)	606 (97%)	19 (3%)	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	267/284 (94%)	262 (98%)	5 (2%)	57	64
1	B	261/284 (92%)	260 (100%)	1 (0%)	91	93
All	All	528/568 (93%)	522 (99%)	6 (1%)	73	80

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	114	LEU
1	A	198	LEU
1	A	295	PHE
1	A	330	LEU
1	A	333	LYS
1	B	295	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	211	GLN
1	B	260	GLN
1	B	262	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 9 ligands modelled in this entry, 4 are monoatomic - leaving 5 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	EDO	A	1001	-	3,3,3	0.06	0	2,2,2	0.20	0
2	EDO	A	1002	-	3,3,3	0.06	0	2,2,2	0.25	0
3	GOL	A	1004	-	5,5,5	0.11	0	5,5,5	0.34	0
2	EDO	B	501	-	3,3,3	0.06	0	2,2,2	0.16	0
2	EDO	A	1003	-	3,3,3	0.06	0	2,2,2	0.19	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	EDO	A	1001	-	-	0/1/1/1	-
2	EDO	A	1002	-	-	1/1/1/1	-
3	GOL	A	1004	-	-	2/4/4/4	-
2	EDO	B	501	-	-	0/1/1/1	-
2	EDO	A	1003	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1004	GOL	O1-C1-C2-C3
2	A	1002	EDO	O1-C1-C2-O2
2	A	1003	EDO	O1-C1-C2-O2
3	A	1004	GOL	O1-C1-C2-O2

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	319/337 (94%)	0.43	25 (7%)	13	12	46, 55, 92, 105	0
1	B	311/337 (92%)	0.42	24 (7%)	13	12	47, 58, 78, 94	0
All	All	630/674 (93%)	0.43	49 (7%)	13	12	46, 56, 85, 105	0

All (49) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	331	ASP	11.6
1	A	336	LYS	7.2
1	A	5	CYS	6.7
1	A	170	ALA	6.4
1	B	168	LYS	6.1
1	A	143	ALA	5.1
1	A	335	MET	5.1
1	B	170	ALA	4.9
1	A	168	LYS	4.8
1	B	7	HIS	4.8
1	B	169	PRO	4.6
1	A	169	PRO	4.5
1	A	334	VAL	4.4
1	B	238	TYR	4.2
1	A	167	ALA	4.0
1	A	7	HIS	3.9
1	A	330	LEU	3.6
1	B	171	CYS	3.5
1	A	6	THR	3.4
1	A	171	CYS	3.4
1	B	150	GLU	3.3
1	A	329	ALA	3.2
1	A	337	PRO	3.1
1	B	256	GLY	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	5	CYS	3.0
1	B	172	SER	3.0
1	B	235	TYR	2.9
1	B	11	LYS	2.9
1	B	4	LEU	2.9
1	A	8	SER	2.8
1	B	45	LEU	2.7
1	A	172	SER	2.6
1	B	145	GLU	2.6
1	B	40	ARG	2.6
1	B	144	SER	2.5
1	B	158	VAL	2.5
1	B	74	LEU	2.3
1	A	238	TYR	2.3
1	B	272	ASN	2.3
1	B	257	GLN	2.3
1	A	246	LEU	2.3
1	B	8	SER	2.3
1	B	114	LEU	2.2
1	A	298	PHE	2.1
1	A	145	GLU	2.1
1	A	150	GLU	2.1
1	A	279	LEU	2.1
1	A	166	GLU	2.1
1	B	237	GLU	2.0

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 monosaccharides in this entry.

6.4 Ligands (i)

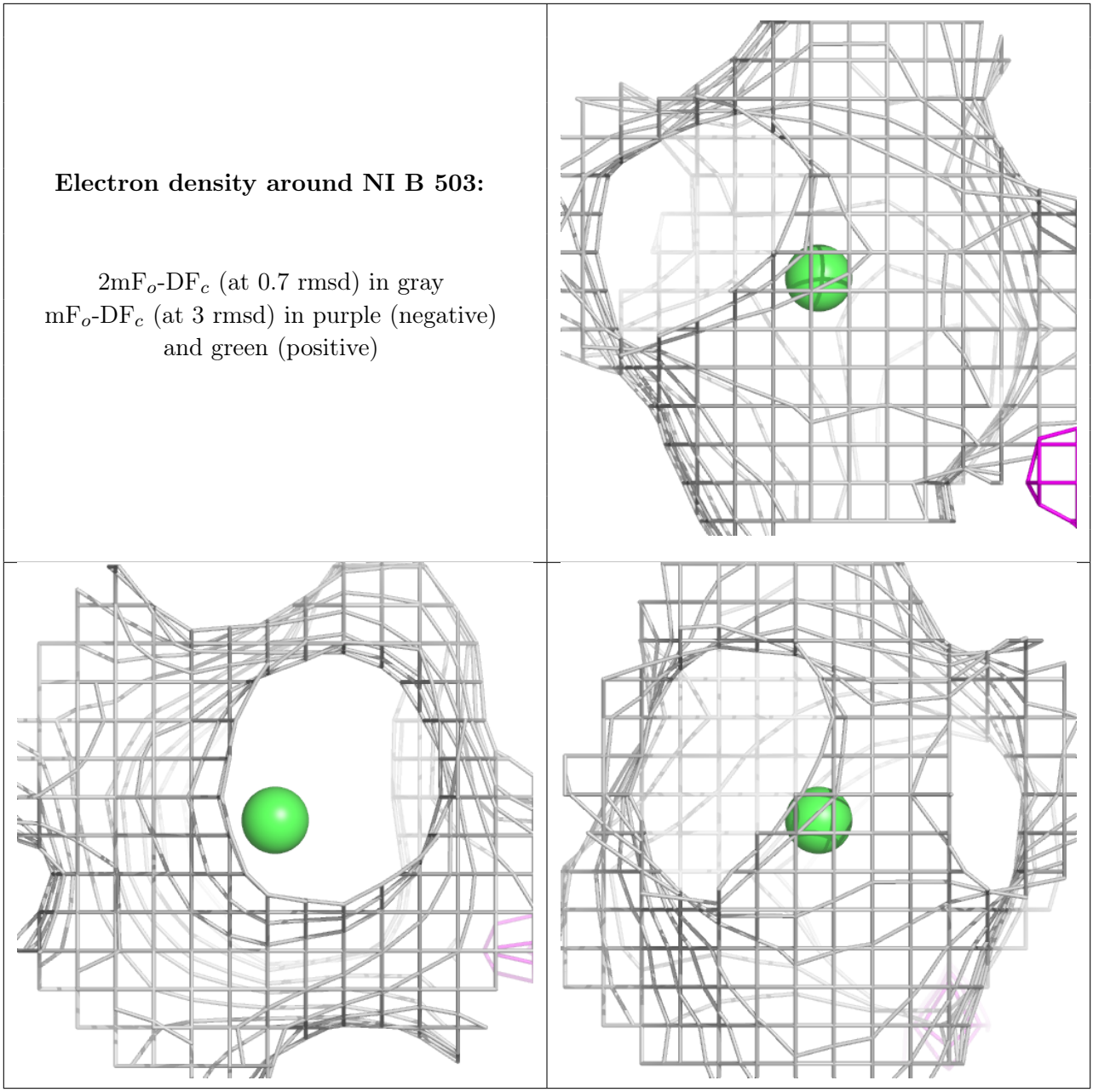
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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
2	EDO	A	1001	4/4	0.75	0.17	74,75,75,75	0
3	GOL	A	1004	6/6	0.80	0.25	77,78,78,79	0
2	EDO	A	1002	4/4	0.90	0.15	64,64,64,65	0
2	EDO	A	1003	4/4	0.91	0.43	66,67,67,67	0
2	EDO	B	501	4/4	0.93	0.58	69,69,69,70	0
4	NI	B	503	1/1	0.98	0.09	58,58,58,58	0
4	NI	A	1006	1/1	0.99	0.12	50,50,50,50	0
4	NI	B	502	1/1	0.99	0.14	51,51,51,51	0
4	NI	A	1005	1/1	0.99	0.10	52,52,52,52	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.

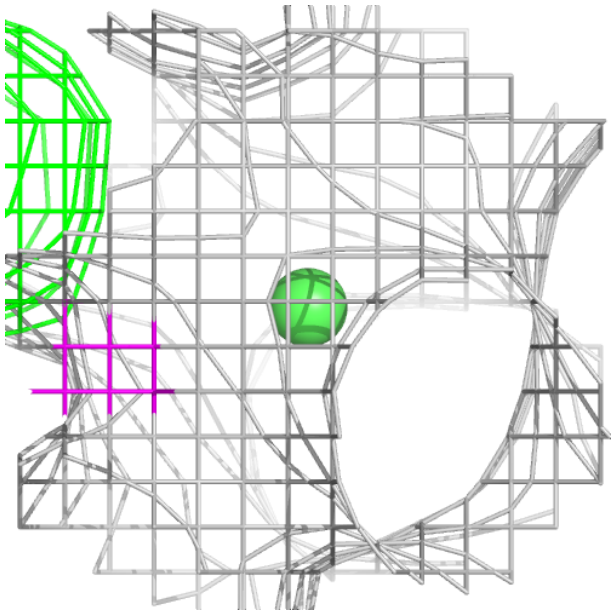
Electron density around NI B 503:

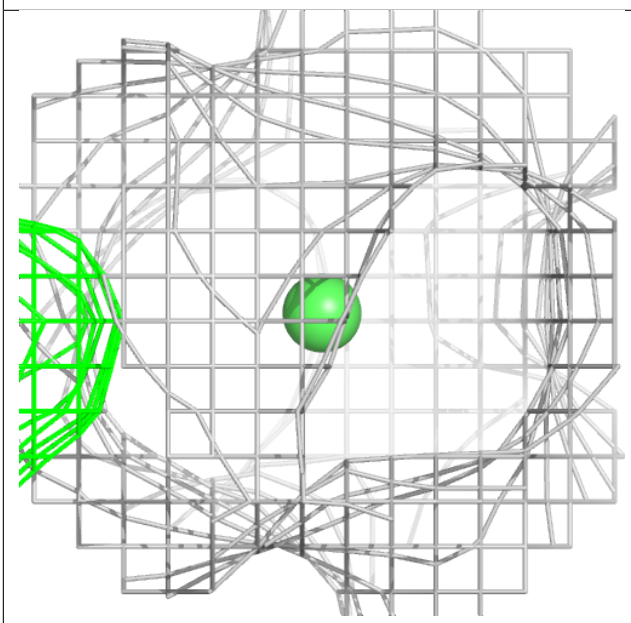
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

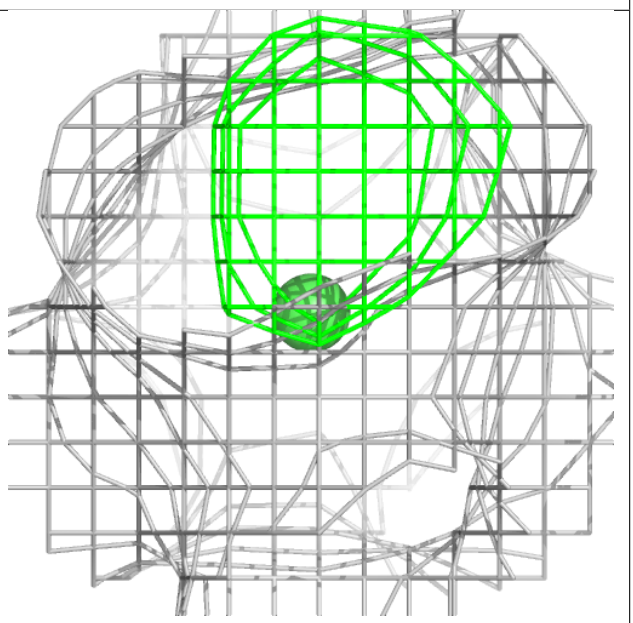


Electron density around NI A 1006:

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

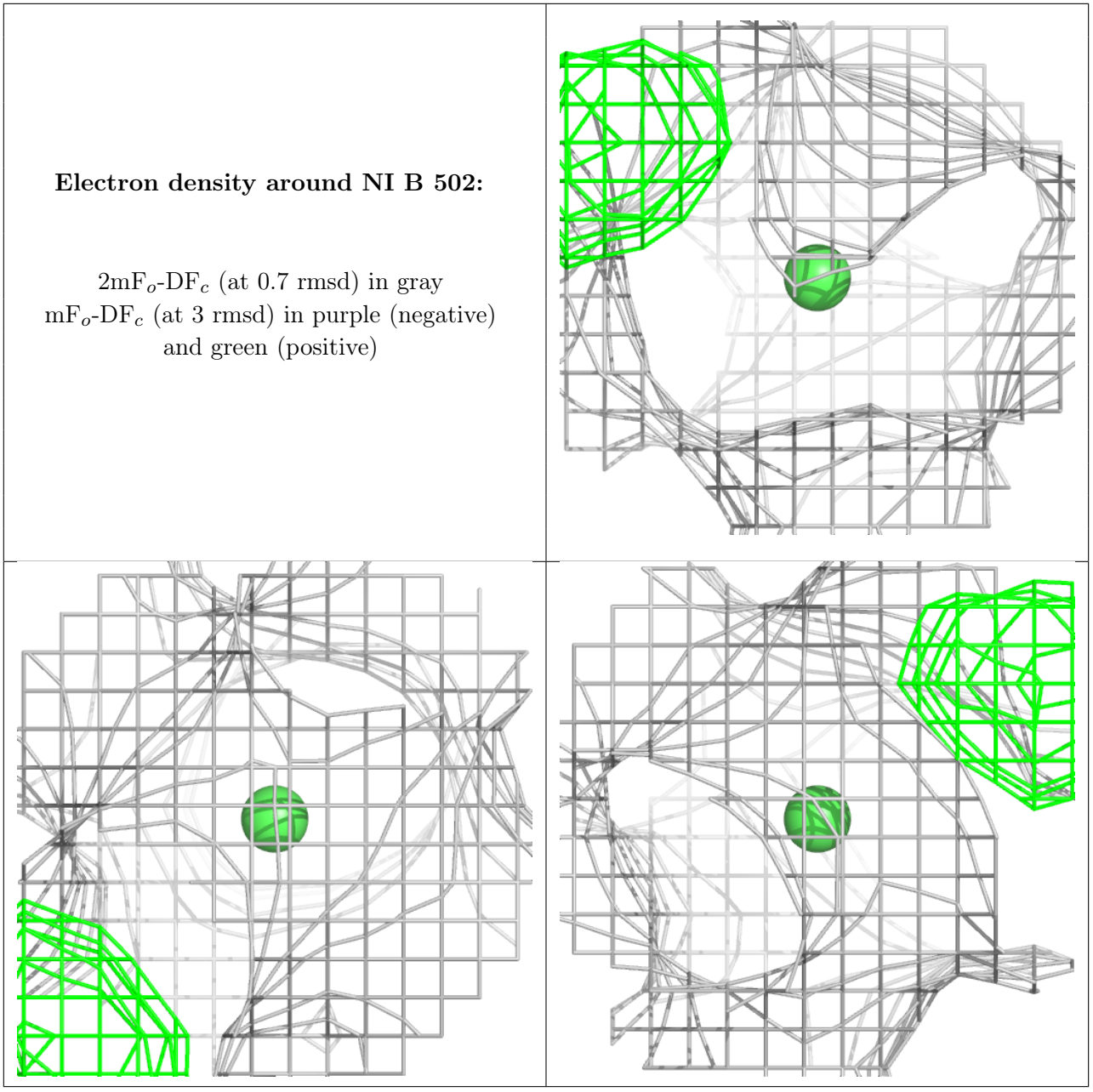






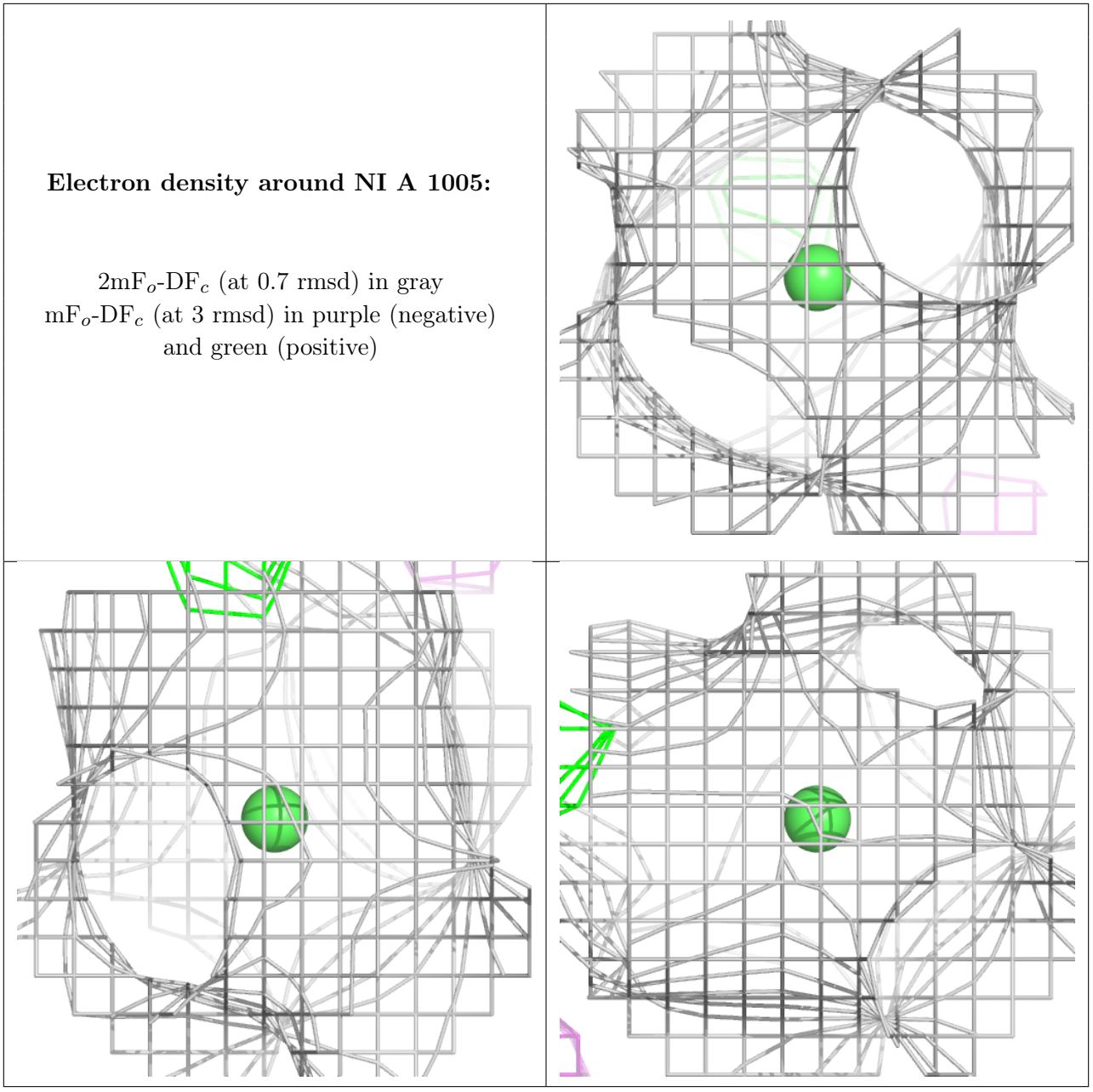
Electron density around NI B 502:

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



Electron density around NI A 1005:

$2mF_o$-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 (i)

There are no such residues in this entry.