Full wwPDB X-ray Structure Validation Report (i)

Mar 30, 2025 – 04:07 am BST

PDB ID	:	9ERP / pdb_00009erp
Title	:	Hydrogenase-2 Ni-SI state
Authors	:	Carr, S.B.; Li, W.; Wong, K.l.; Ash, P.A.; Vincent, K.A.
Deposited on	:	2024-03-25
Resolution	:	1.37 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:

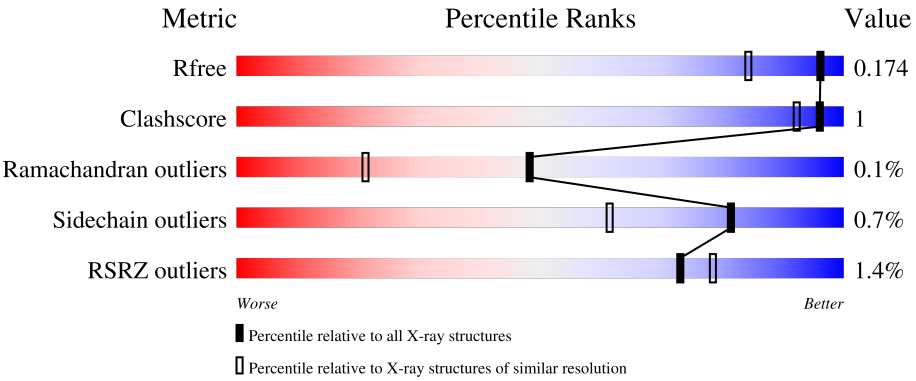
MolProbity	:	4.02b-467
Mogul	:	?.? (???), CSD ??CSD?? (????)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.42

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 1.37 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}	164625	3869 (1.40-1.36)
Clashscore	180529	4183 (1.40-1.36)
Ramachandran outliers	177936	4116 (1.40-1.36)
Sidechain outliers	177891	4115 (1.40-1.36)
RSRZ outliers	164620	3867 (1.40-1.36)

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	S	298	2% 89% · 10%
1	T	298	2% 87% · 10%
2	L	567	% 94% · ·
2	M	567	% 94% · ·

2 Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 26658 atoms, of which 12528 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 Hydrogenase-2 small chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	S	268	Total	C	H	N	O	S	0	5	0
			4045	1304	1987	362	379	13			
1	T	268	Total	C	H	N	O	S	0	6	0
			4072	1312	2005	362	380	13			

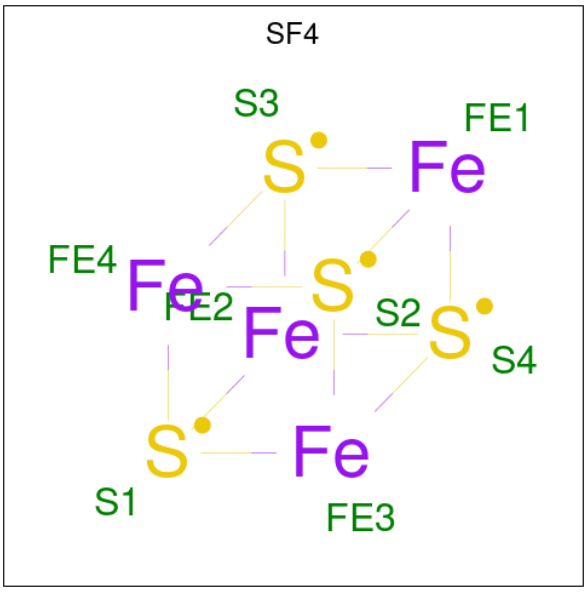
There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
S	294	HIS	-	expression tag	UNP P69741
S	295	HIS	-	expression tag	UNP P69741
S	296	HIS	-	expression tag	UNP P69741
S	297	HIS	-	expression tag	UNP P69741
S	298	HIS	-	expression tag	UNP P69741
S	299	HIS	-	expression tag	UNP P69741
T	294	HIS	-	expression tag	UNP P69741
T	295	HIS	-	expression tag	UNP P69741
T	296	HIS	-	expression tag	UNP P69741
T	297	HIS	-	expression tag	UNP P69741
T	298	HIS	-	expression tag	UNP P69741
T	299	HIS	-	expression tag	UNP P69741

- Molecule 2 is a protein called Hydrogenase-2 large chain.

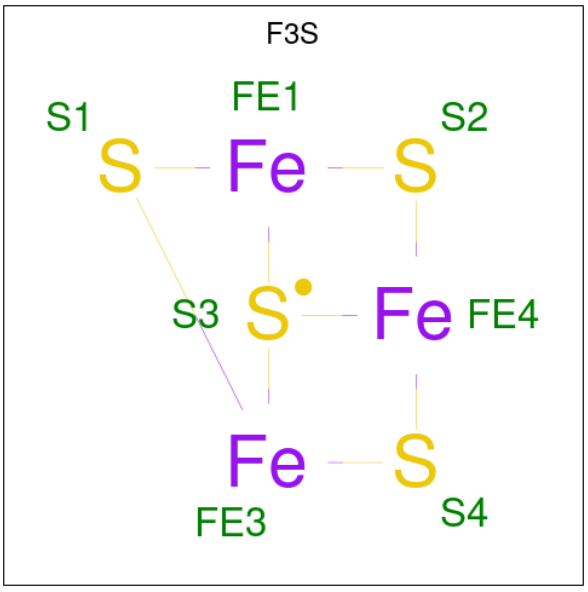
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	L	551	Total	C	H	N	O	S	0	4	0
			8550	2737	4247	740	808	18			
2	M	551	Total	C	H	N	O	S	0	7	0
			8581	2745	4265	742	811	18			

- Molecule 3 is IRON/SULFUR CLUSTER (CCD ID: SF4) (formula: Fe_4S_4).



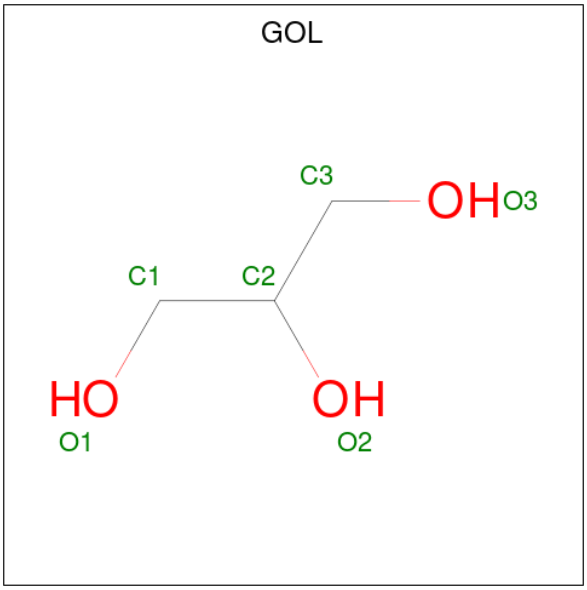
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	S	1	Total Fe S 8 4 4	0	0
3	S	1	Total Fe S 8 4 4	0	0
3	T	1	Total Fe S 8 4 4	0	0
3	T	1	Total Fe S 8 4 4	0	0

- Molecule 4 is FE3-S4 CLUSTER (CCD ID: F3S) (formula: Fe_3S_4).



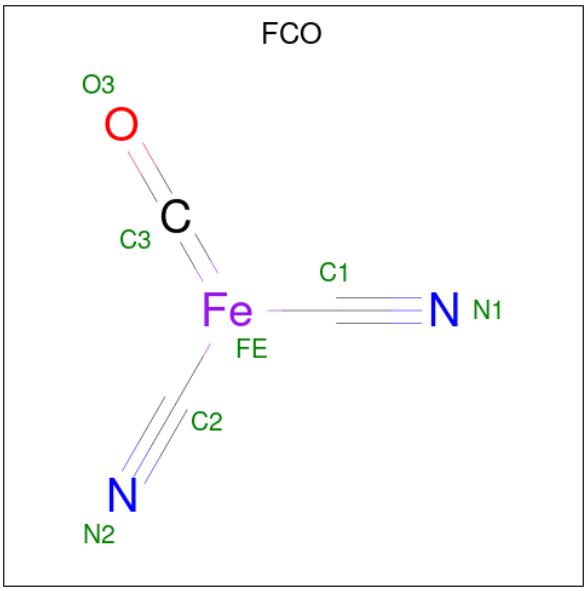
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	S	1	Total Fe S 7 3 4	0	0
4	T	1	Total Fe S 7 3 4	0	0

- Molecule 5 is GLYCEROL (CCD ID: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	S	1	Total C H O 14 3 8 3	0	0
5	T	1	Total C H O 14 3 8 3	0	0
5	M	1	Total C H O 14 3 8 3	0	0

- Molecule 6 is CARBONMONOXIDE-(DICYANO) IRON (CCD ID: FCO) (formula: C_3FeN_2O) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	L	1	Total	C	Fe	N	O	0	0
			7	3	1	2	1		
6	M	1	Total	C	Fe	N	O	0	0
			7	3	1	2	1		

- Molecule 7 is NICKEL (II) ION (CCD ID: NI) (formula: Ni) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	L	1	Total	Ni	0	0
			1	1		
7	M	1	Total	Ni	0	0
			1	1		

- Molecule 8 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	L	2	Total	Mg	0	0
			2	2		
8	M	2	Total	Mg	0	0
			2	2		

- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	S	210	Total	O	0	0
			210	210		

Continued on next page...

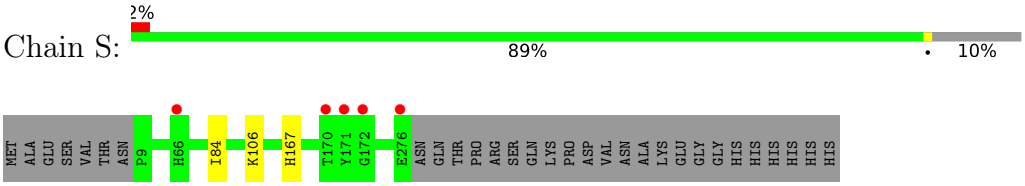
Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
9	L	471	Total O 471 471	0	0
9	T	197	Total O 197 197	0	0
9	M	424	Total O 424 424	0	0

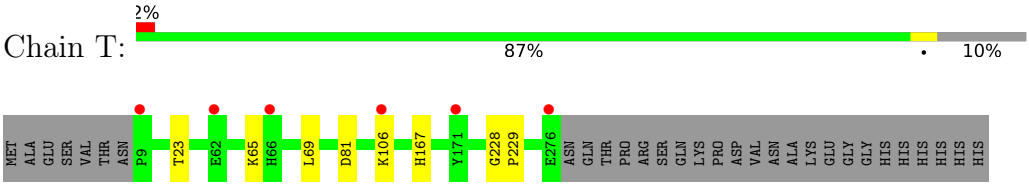
3 Residue-property plots (i)

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.

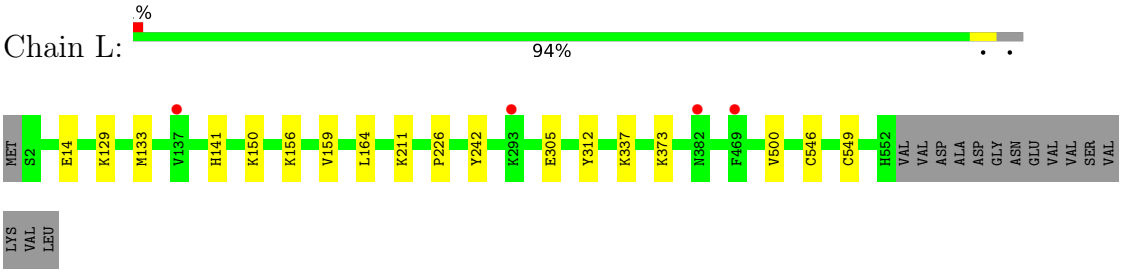
• Molecule 1: Hydrogenase-2 small chain



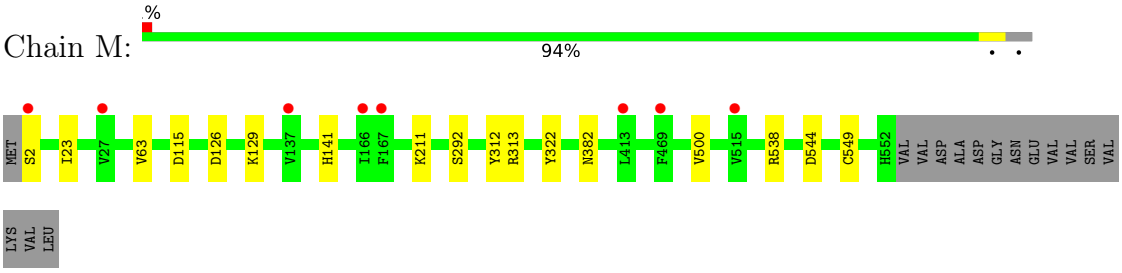
• Molecule 1: Hydrogenase-2 small chain



• Molecule 2: Hydrogenase-2 large chain



• Molecule 2: Hydrogenase-2 large chain



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	99.18Å 100.00Å 167.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	64.28 – 1.37 64.28 – 1.37	Depositor EDS
% Data completeness (in resolution range)	99.9 (64.28-1.37) 100.0 (64.28-1.37)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.02 (at 1.37Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487, PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.157 , 0.174 0.157 , 0.174	Depositor DCC
R_{free} test set	17452 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	16.3	Xtriage
Anisotropy	0.121	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 34.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.005 for k,h,-l	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	26658	wwPDB-VP
Average B, all atoms (Å^2)	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.55% 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: SF4, MG, NI, FCO, GOL, F3S

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	S	0.45	0/2130	0.65	0/2900
1	T	0.42	0/2141	0.64	1/2913 (0.0%)
2	L	0.46	0/4423	0.68	0/6029
2	M	0.44	0/4444	0.67	1/6058 (0.0%)
All	All	0.44	0/13138	0.66	2/17900 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	M	544	ASP	CB-CG-OD1	5.25	123.02	118.30
1	T	81	ASP	CB-CG-OD1	5.07	122.86	118.30

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	S	2058	1987	1975	2	0
1	T	2067	2005	2000	8	0
2	L	4303	4247	4234	9	0
2	M	4316	4265	4244	7	0
3	S	16	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	T	16	0	0	0	0
4	S	7	0	0	0	0
4	T	7	0	0	0	0
5	M	6	8	7	0	0
5	S	6	8	8	0	0
5	T	6	8	8	0	0
6	L	7	0	0	0	0
6	M	7	0	0	0	0
7	L	1	0	0	0	0
7	M	1	0	0	0	0
8	L	2	0	0	0	0
8	M	2	0	0	0	0
9	L	471	0	0	3	0
9	M	424	0	0	2	0
9	S	210	0	0	1	0
9	T	197	0	0	3	0
All	All	14130	12528	12476	25	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 (25) 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:T:167:HIS:NE2	9:T:401:HOH:O	2.15	0.80
1:S:167:HIS:NE2	9:S:401:HOH:O	2.07	0.76
2:L:150:LYS:HE3	9:M:844:HOH:O	1.88	0.73
2:M:126:ASP:OD2	2:M:129:LYS:HE3	2.04	0.57
2:L:305:GLU:OE2	2:L:373:LYS:NZ	2.34	0.56
1:T:106:LYS:CE	9:T:478:HOH:O	2.56	0.54
2:M:313:ARG:HD3	2:M:322:TYR:CE2	2.43	0.54
1:T:69:LEU:CD1	1:T:106:LYS:HE3	2.37	0.53
2:M:292[B]:SER:OG	9:M:701:HOH:O	2.20	0.49
2:L:129:LYS:HE2	9:L:1046:HOH:O	2.13	0.48
2:M:500:VAL:HG11	2:M:549:CYS:HB3	1.95	0.48
1:T:106:LYS:HE2	1:T:106:LYS:HB3	1.74	0.47
2:L:129:LYS:O	2:L:133:MET:HG3	2.14	0.47
1:T:69:LEU:HD11	1:T:106:LYS:HE3	1.96	0.47
2:M:2:SER:HA	2:M:23:ILE:O	2.15	0.47
2:L:500:VAL:HG11	2:L:549:CYS:HB3	2.00	0.43
2:L:159:VAL:HG22	2:L:164:LEU:HD11	2.02	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:T:106:LYS:HE2	9:T:478:HOH:O	2.19	0.42
2:L:156:LYS:HE2	9:L:705:HOH:O	2.19	0.42
1:S:84:ILE:HD12	1:S:84:ILE:N	2.34	0.42
1:T:23:THR:HB	2:M:63:VAL:HG23	2.03	0.41
1:T:228:GLY:N	1:T:229:PRO:CD	2.84	0.40
2:M:115:ASP:HB3	2:M:538:ARG:HG2	2.03	0.40
2:L:14:GLU:HB3	2:L:546:CYS:HA	2.03	0.40
2:L:337:LYS:NZ	9:L:707:HOH:O	2.49	0.40

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	S	271/298 (91%)	262 (97%)	9 (3%)	0	100	100
1	T	272/298 (91%)	263 (97%)	9 (3%)	0	100	100
2	L	553/567 (98%)	526 (95%)	26 (5%)	1 (0%)	44	20
2	M	556/567 (98%)	532 (96%)	23 (4%)	1 (0%)	44	20
All	All	1652/1730 (96%)	1583 (96%)	67 (4%)	2 (0%)	48	21

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	L	211	LYS
2	M	211	LYS

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	S	217/239 (91%)	216 (100%)	1 (0%)	86	71
1	T	218/239 (91%)	217 (100%)	1 (0%)	86	71
2	L	469/479 (98%)	465 (99%)	4 (1%)	75	52
2	M	471/479 (98%)	468 (99%)	3 (1%)	84	66
All	All	1375/1436 (96%)	1366 (99%)	9 (1%)	81	61

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	S	106	LYS
2	L	141	HIS
2	L	226	PRO
2	L	242	TYR
2	L	312	TYR
1	T	65	LYS
2	M	141	HIS
2	M	312	TYR
2	M	382	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	T	66	HIS
2	M	47	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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 17 ligands modelled in this entry, 6 are monoatomic - leaving 11 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

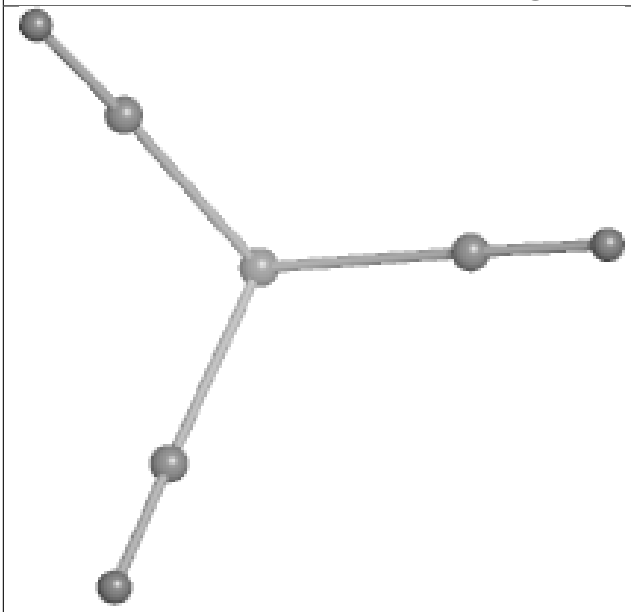
There are no torsion outliers.

There are no ring outliers.

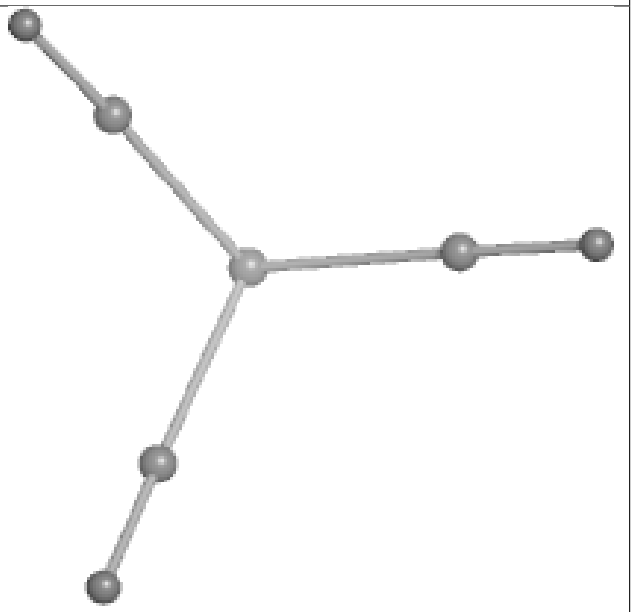
No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

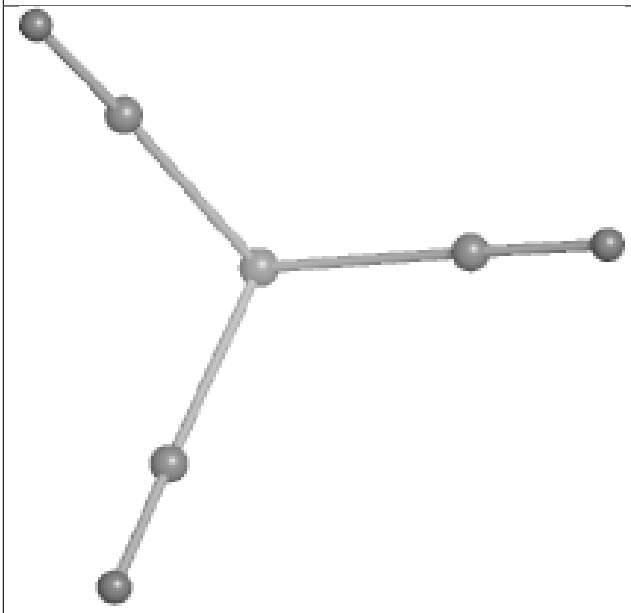
Ligand FCO L 601



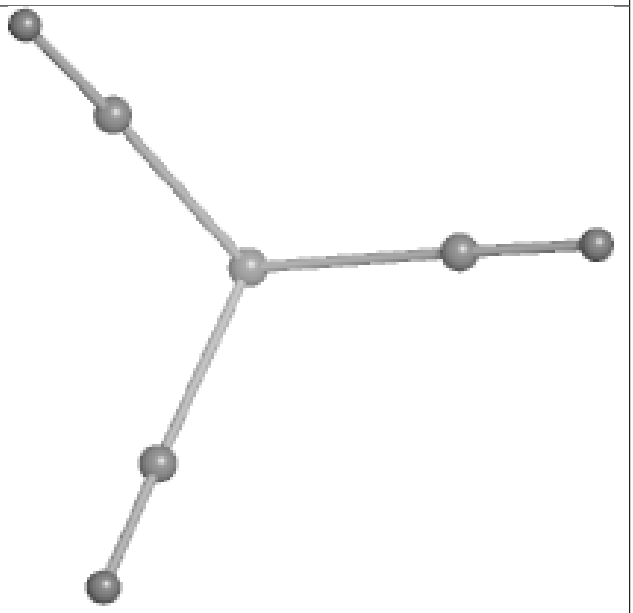
Bond lengths



Bond angles

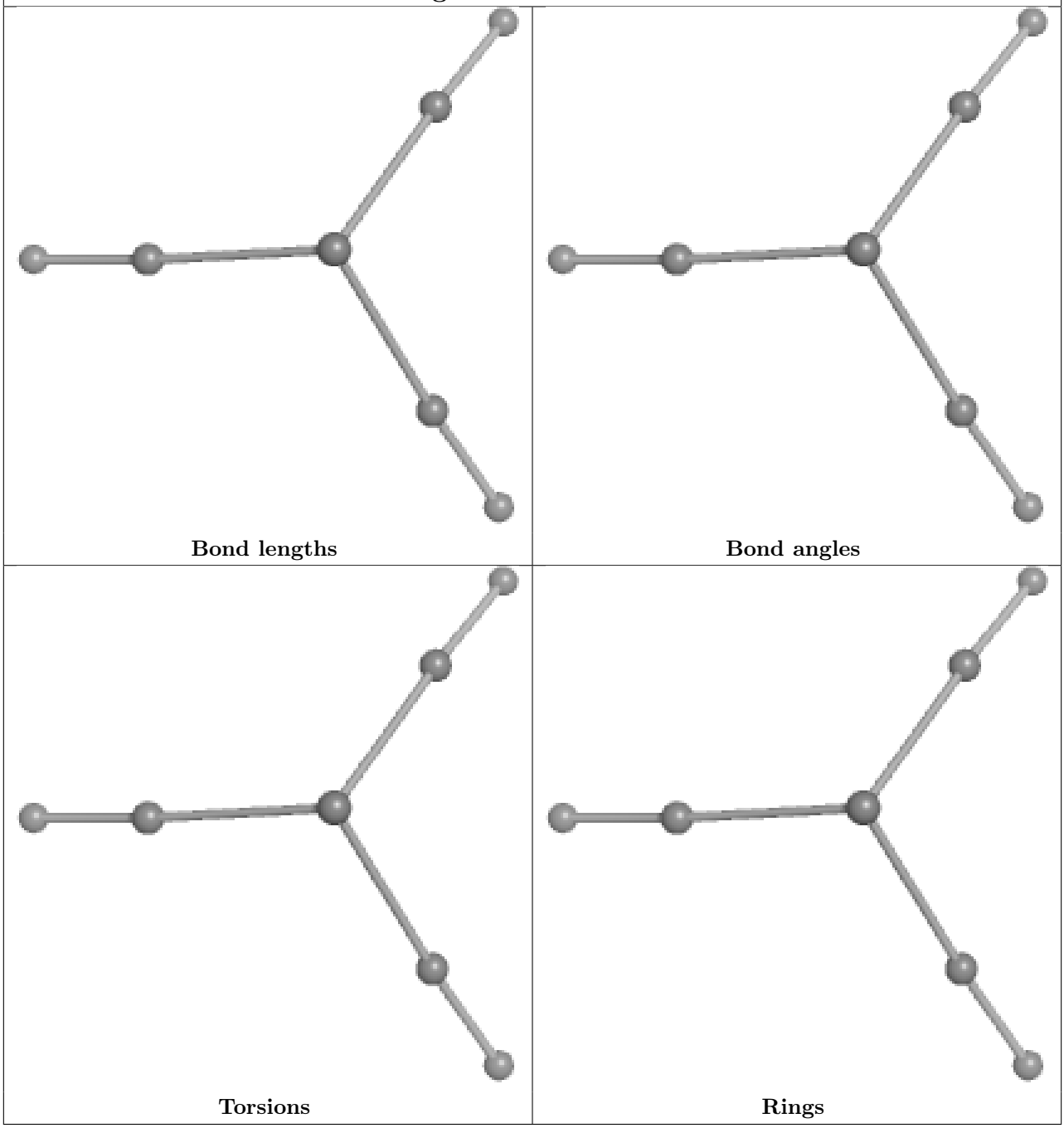


Torsions



Rings

Ligand FCO M 601



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	S	268/298 (89%)	-0.25	5 (1%)	66	71	8, 19, 35, 51	3 (1%)
1	T	268/298 (89%)	-0.17	6 (2%)	62	67	11, 20, 37, 55	5 (1%)
2	L	551/567 (97%)	-0.45	4 (0%)	84	88	10, 18, 30, 42	3 (0%)
2	M	551/567 (97%)	-0.25	8 (1%)	71	77	9, 20, 35, 51	5 (0%)
All	All	1638/1730 (94%)	-0.30	23 (1%)	73	78	8, 19, 34, 55	16 (0%)

All (23) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	S	171	TYR	4.5
1	T	9	PRO	4.2
1	S	172	GLY	3.6
2	M	469	PHE	3.4
2	L	469	PHE	3.1
2	M	2	SER	3.1
2	L	137	VAL	3.0
2	M	137	VAL	2.9
2	M	27	VAL	2.7
2	M	413	LEU	2.5
2	L	382	ASN	2.4
1	S	170	THR	2.4
2	M	515	VAL	2.4
1	T	106	LYS	2.3
1	T	66	HIS	2.3
1	S	276	GLU	2.2
2	M	166	ILE	2.2
1	S	66	HIS	2.1
2	M	167	PHE	2.1
1	T	171	TYR	2.1
1	T	62	GLU	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	T	276	GLU	2.1
2	L	293	LYS	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)

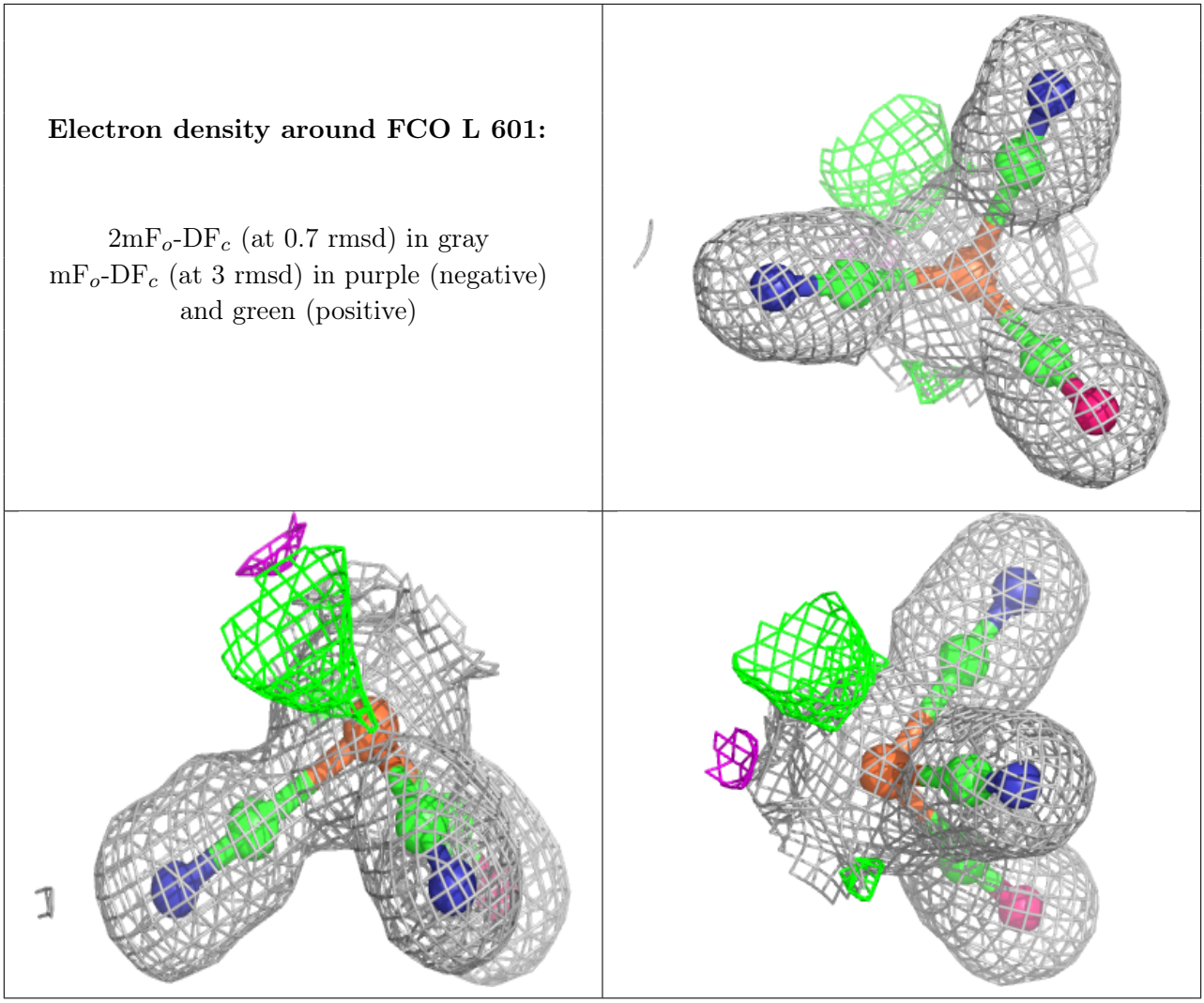
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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
8	MG	L	604	1/1	0.89	0.10	40,40,40,40	0
5	GOL	M	605	6/6	0.90	0.12	20,32,45,48	0
5	GOL	T	304	6/6	0.92	0.10	22,30,36,36	0
5	GOL	S	304	6/6	0.93	0.09	22,29,33,34	0
8	MG	M	604	1/1	0.96	0.21	35,35,35,35	0
8	MG	M	603	1/1	0.99	0.10	11,11,11,11	0
3	SF4	S	301	8/8	1.00	0.02	14,14,15,15	0
3	SF4	S	303	8/8	1.00	0.02	12,13,14,14	0
3	SF4	T	301	8/8	1.00	0.02	14,14,14,15	0
6	FCO	L	601	7/7	1.00	0.02	11,12,13,13	0
6	FCO	M	601	7/7	1.00	0.04	13,13,14,15	0
7	NI	L	602	1/1	1.00	0.02	15,15,15,15	0
7	NI	M	602	1/1	1.00	0.03	16,16,16,16	0
8	MG	L	603	1/1	1.00	0.07	9,9,9,9	0
3	SF4	T	303	8/8	1.00	0.02	13,14,15,15	0
4	F3S	S	302	7/7	1.00	0.02	12,13,13,13	0
4	F3S	T	302	7/7	1.00	0.02	13,13,14,15	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.

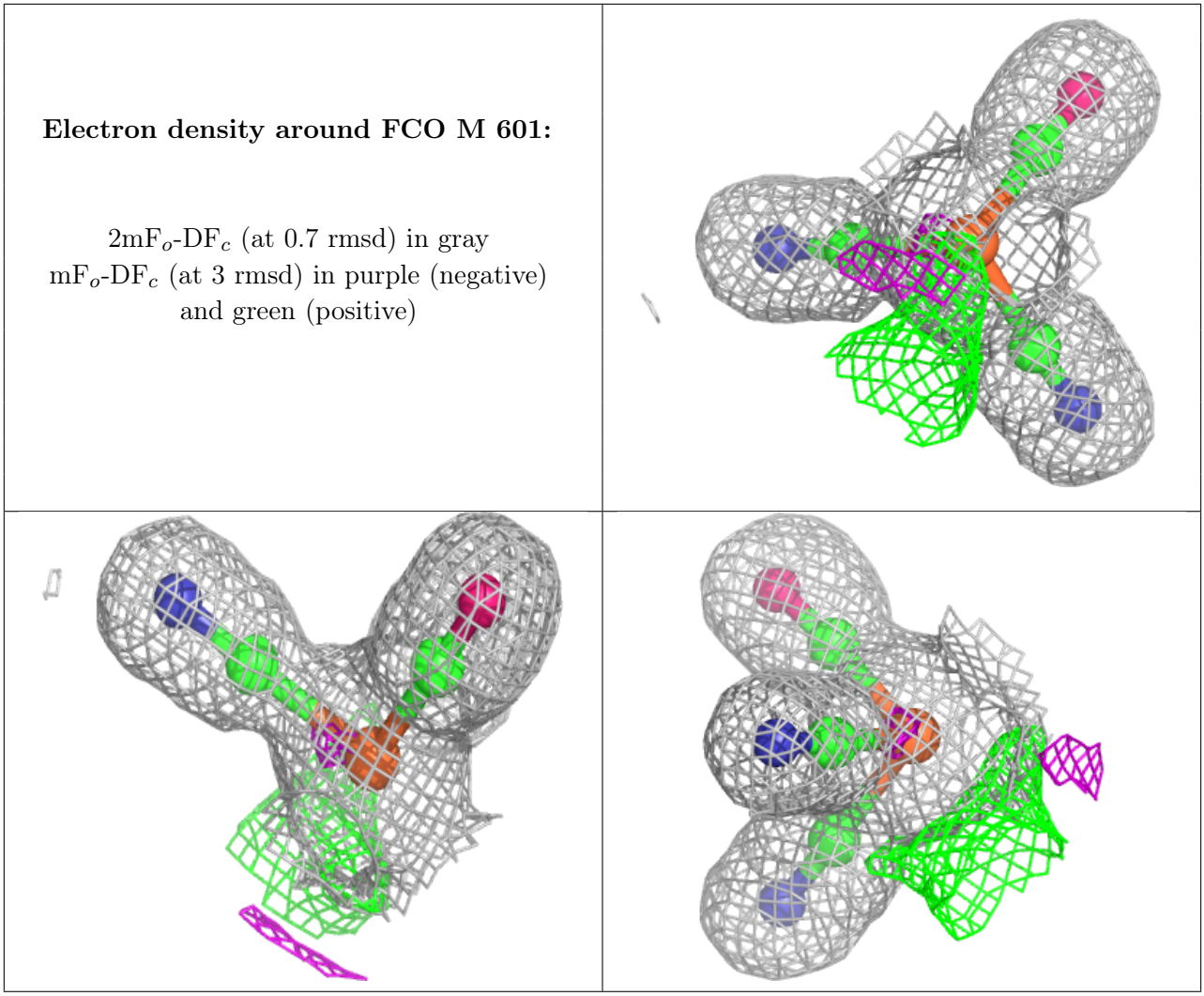
Electron density around FCO L 601:

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



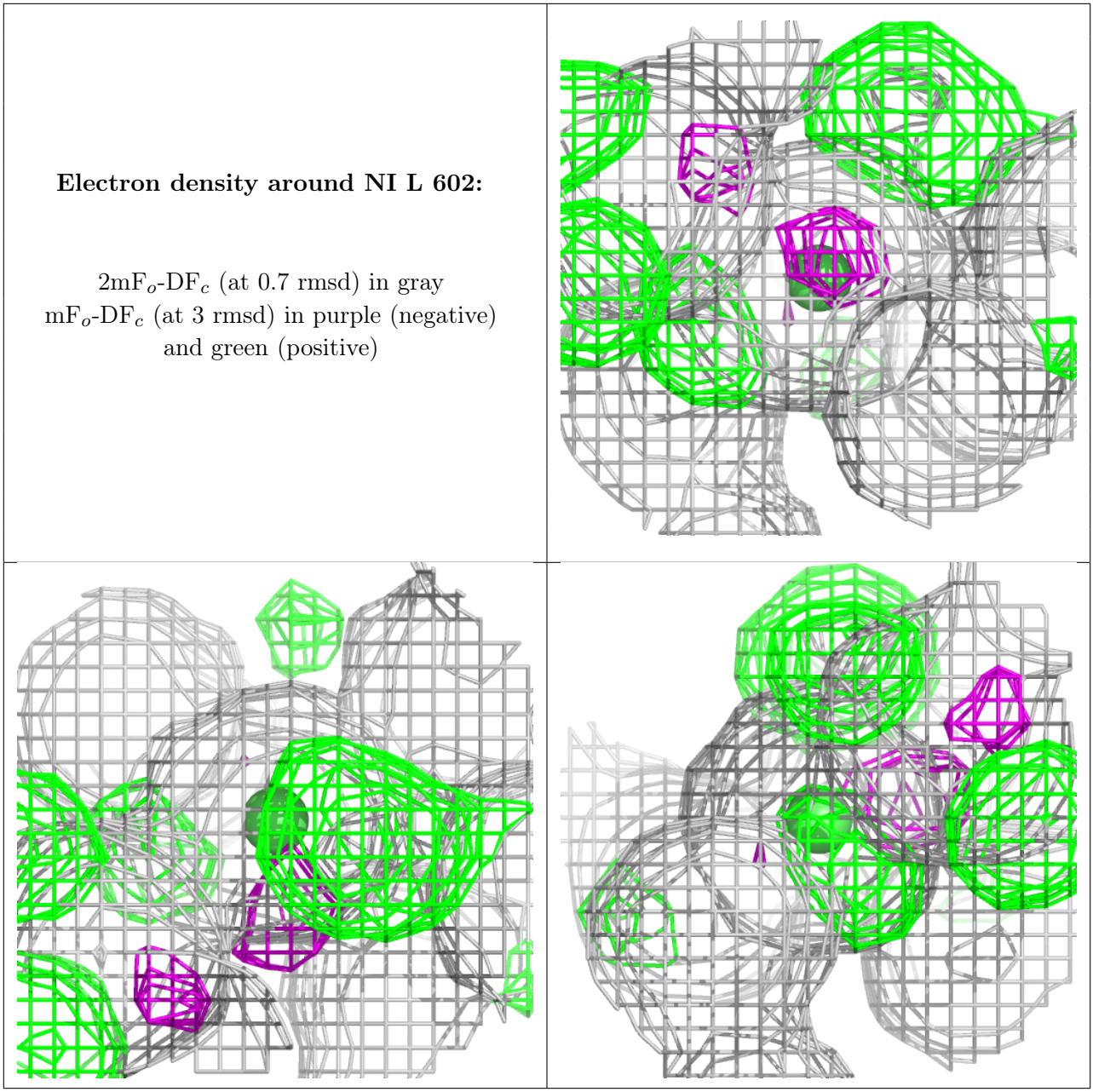
Electron density around FCO M 601:

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



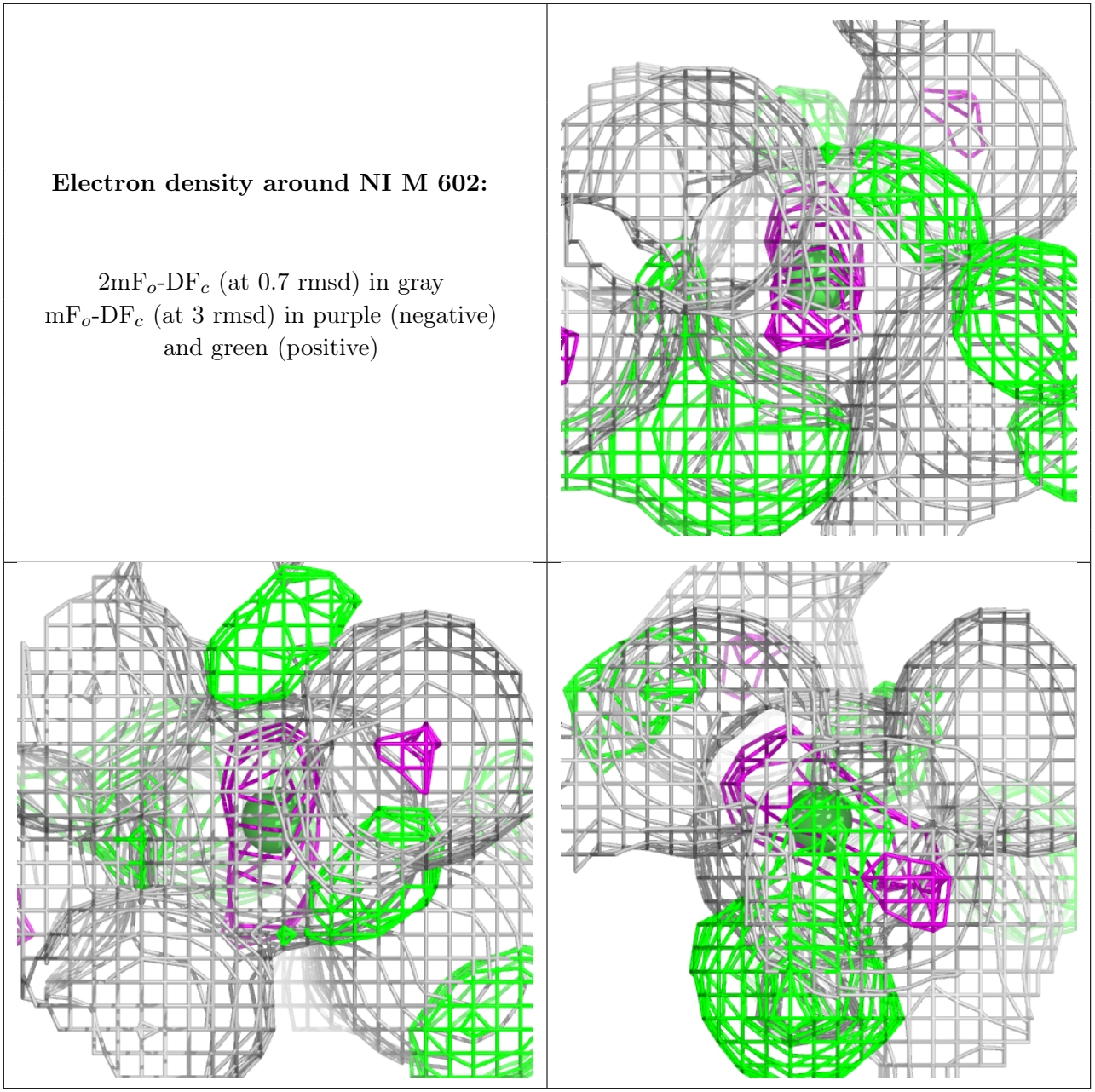
Electron density around NI L 602:

$2mF_o$-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 M 602:

$2mF_o$-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.