wwPDB X-ray Structure Validation Summary Report (i)

Oct 4, 2023 – 10:44 AM EDT

PDB ID	:	6O7N
Title	:	Nitrogenase MoFeP mutant F99Y/S188A from Azotobacter vinelandii in the indigo carmine oxidized state
Authors	:	Rutledge, H.L.; Tezcan, F.A.
Deposited on	:	2019-03-08
Resolution	:	1.75 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 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:

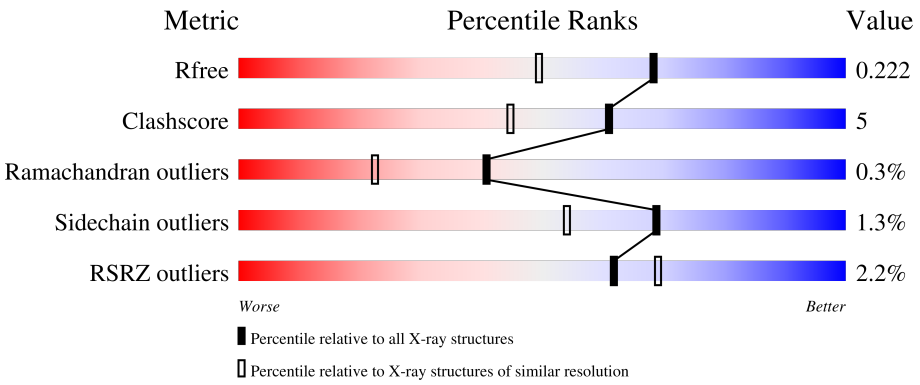
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 1.75 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	2340 (1.76-1.76)
Clashscore	141614	2466 (1.76-1.76)
Ramachandran outliers	138981	2437 (1.76-1.76)
Sidechain outliers	138945	2437 (1.76-1.76)
RSRZ outliers	127900	2298 (1.76-1.76)

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	492	3% 84% 12% • •
1	C	492	3% 88% 9% •
2	B	523	% 91% 9%
2	D	523	% 90% 9% •

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 33865 atoms, of which 15663 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 Nitrogenase molybdenum-iron protein alpha chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	477	7484	2409	3696	642	711	26	0	4	0
1	C	478	7497	2409	3712	644	706	26	0	4	1

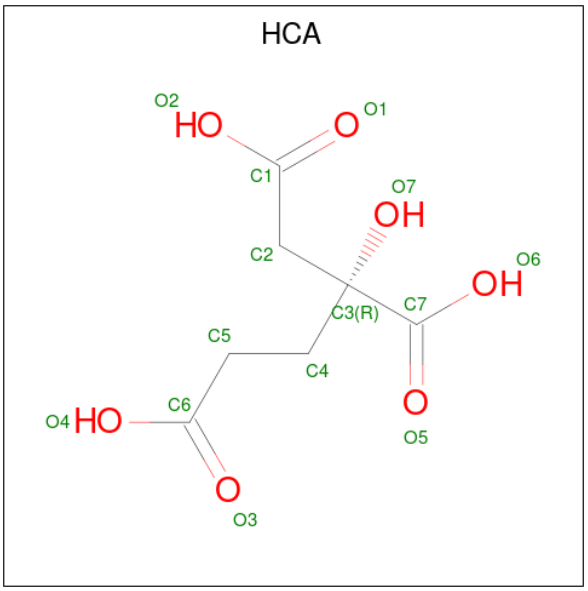
- Molecule 2 is a protein called Nitrogenase molybdenum-iron protein beta chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	B	522	8328	2687	4123	708	782	28	0	6	0
2	D	522	8321	2689	4120	705	778	29	0	8	0

There are 4 discrepancies between the modelled and reference sequences:

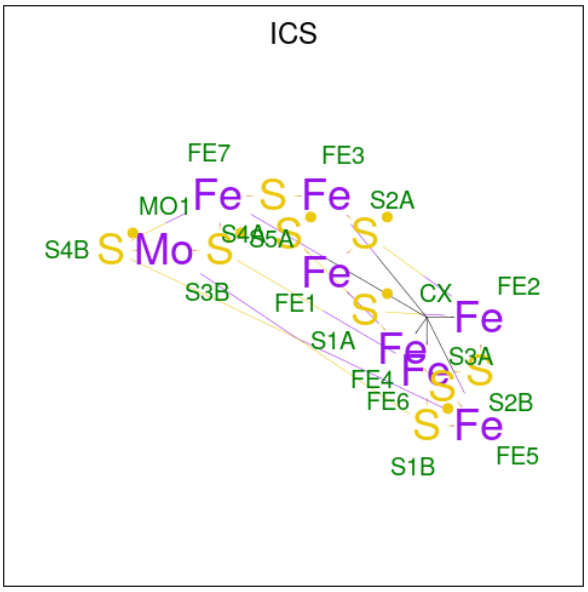
Chain	Residue	Modelled	Actual	Comment	Reference
B	99	TYR	PHE	engineered mutation	UNP P07329
B	188	ALA	SER	engineered mutation	UNP P07329
D	99	TYR	PHE	engineered mutation	UNP P07329
D	188	ALA	SER	engineered mutation	UNP P07329

- Molecule 3 is 3-HYDROXY-3-CARBOXY-ADIPIC ACID (three-letter code: HCA) (formula: $C_7H_{10}O_7$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
3	A	1	20	7	6	7	0	0
			Total	C	H	O		
3	C	1	20	7	6	7	0	0

- Molecule 4 is iron-sulfur-molybdenum cluster with interstitial carbon (three-letter code: ICS) (formula: CFe_7MoS_9).



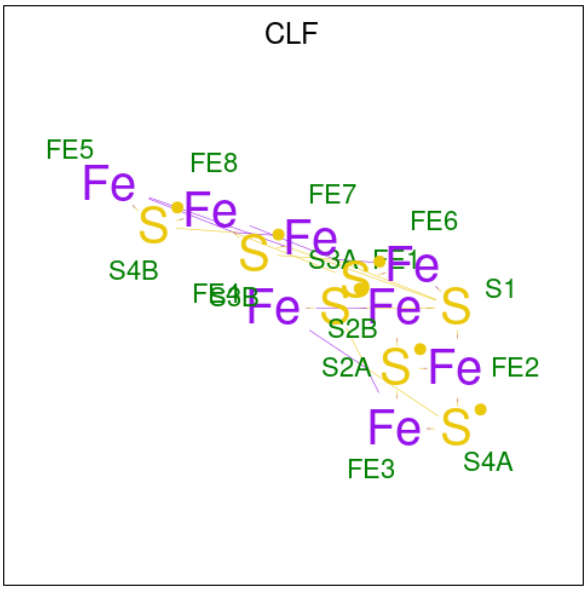
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	Fe	Mo	S		
4	A	1	18	1	7	1	9	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	Fe	Mo	S		
4	C	1	18	1	7	1	9	0	0

- Molecule 5 is FE(8)-S(7) CLUSTER (three-letter code: CLF) (formula: Fe_8S_7) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	Fe	S		
5	A	1	15	8	7	0	0
5	C	1	15	8	7	0	0

- Molecule 6 is FE (III) ION (three-letter code: FE) (formula: Fe).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Fe		
6	B	1	1	1	0	0
6	D	1	1	1	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	A	440	440	440	0	0

Continued on next page...

Continued from previous page...

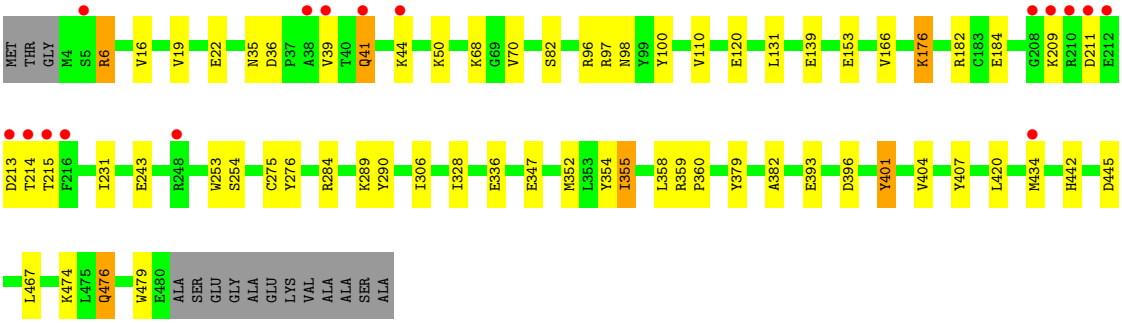
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	B	603	603	603	0	0
			Total	O		
7	C	457	457	457	0	0
			Total	O		
7	D	627	627	627	0	0

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.

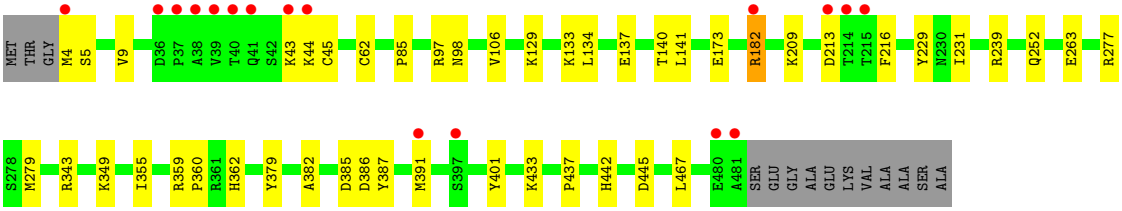
• Molecule 1: Nitrogenase molybdenum-iron protein alpha chain

Chain A:



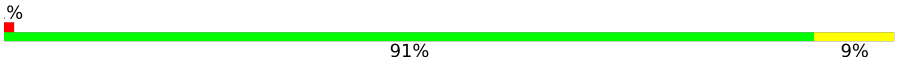
• Molecule 1: Nitrogenase molybdenum-iron protein alpha chain

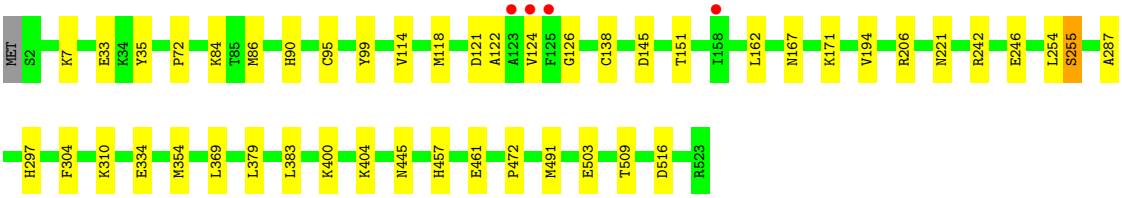
Chain C:



• Molecule 2: Nitrogenase molybdenum-iron protein beta chain

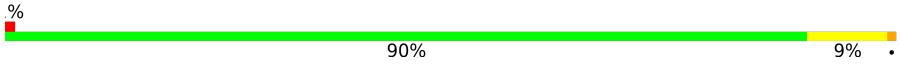
Chain B:

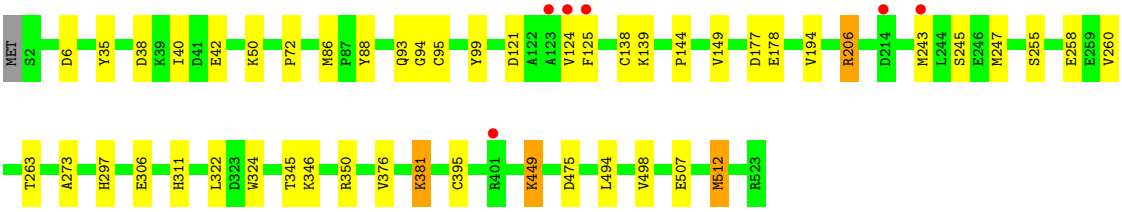




• Molecule 2: Nitrogenase molybdenum-iron protein beta chain

Chain D:




MET
S2
D6
Y35
D38
K39
I40
D41
E42
K50
P72
M86
P87
Y88
Q93
G94
C95
Y99
D121
A122
A123
V124
F125
C138
K139
P144
V149
D177
E178
V194
R206
D214
M243
L244
S245
E246
M247
S255
E258
E259
V260
T263
A273
H297
E306
H311
L322
D323
W324
T345
K346
R350
V376
K381
C395
H401
K449
D475
L494
V498
E507
M512
R523

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	76.34Å 127.96Å 107.51Å 90.00° 109.01° 90.00°	Depositor
Resolution (Å)	40.10 – 1.75 79.59 – 1.75	Depositor EDS
% Data completeness (in resolution range)	90.0 (40.10-1.75) 90.3 (79.59-1.75)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.31 (at 1.75Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.187 , 0.223 0.186 , 0.222	Depositor DCC
R_{free} test set	17685 reflections (9.99%)	wwPDB-VP
Wilson B-factor (Å^2)	17.3	Xtriage
Anisotropy	0.589	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 38.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.025 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	33865	wwPDB-VP
Average B, all atoms (Å^2)	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.98% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: HCA, ICS, CLF, FE

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	0.50	0/3888	0.55	0/5249
1	C	0.46	0/3885	0.55	0/5243
2	B	0.49	0/4329	0.55	0/5852
2	D	0.47	0/4331	0.55	0/5857
All	All	0.48	0/16433	0.55	0/22201

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	3788	3696	3698	48	0
1	C	3785	3712	3709	32	0
2	B	4205	4123	4119	30	0
2	D	4201	4120	4135	39	0
3	A	14	6	6	1	0
3	C	14	6	6	2	0
4	A	18	0	0	1	0
4	C	18	0	0	0	0
5	A	15	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	C	15	0	0	1	0
6	B	1	0	0	0	0
6	D	1	0	0	0	0
7	A	440	0	0	15	2
7	B	603	0	0	8	2
7	C	457	0	0	6	1
7	D	627	0	0	13	1
All	All	18202	15663	15673	144	3

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 5.

The worst 5 of 144 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:182:ARG:NH1	7:A:606:HOH:O	2.14	0.81
2:B:206:ARG:NH1	7:B:704:HOH:O	2.15	0.79
2:B:354:MET:SD	2:B:491:MET:HG2	2.28	0.74
2:B:503:GLU:OE2	7:B:702:HOH:O	2.07	0.73
2:D:42:GLU:OE1	7:D:702:HOH:O	2.08	0.72

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
7:A:986:HOH:O	7:B:1242:HOH:O[1_455]	2.09	0.11
7:A:943:HOH:O	7:B:1281:HOH:O[2_646]	2.14	0.06
7:C:1036:HOH:O	7:D:1177:HOH:O[2_655]	2.15	0.05

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	479/492 (97%)	459 (96%)	18 (4%)	2 (0%)	34	17
1	C	480/492 (98%)	457 (95%)	22 (5%)	1 (0%)	47	29
2	B	526/523 (101%)	516 (98%)	9 (2%)	1 (0%)	47	29
2	D	528/523 (101%)	513 (97%)	14 (3%)	1 (0%)	47	29
All	All	2013/2030 (99%)	1945 (97%)	63 (3%)	5 (0%)	41	29

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	255	SER
1	A	6	ARG
2	D	255	SER
1	A	355	ILE
1	C	355	ILE

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	404/415 (97%)	396 (98%)	8 (2%)	55	34
1	C	404/415 (97%)	398 (98%)	6 (2%)	65	49
2	B	459/454 (101%)	455 (99%)	4 (1%)	78	67
2	D	459/454 (101%)	454 (99%)	5 (1%)	73	60
All	All	1726/1738 (99%)	1703 (99%)	23 (1%)	69	54

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	182	ARG
1	C	445	ASP
1	C	401	TYR
2	D	177	ASP
1	A	445	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	D	359	HIS

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 8 ligands modelled in this entry, 2 are monoatomic - leaving 6 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	$\#\|Z\| > 2$
3	HCA	A	501	-	13,13,13	1.10	0	14,18,18	1.47	4 (28%)
4	ICS	A	502	1	18,30,30	2.90	13 (72%)	-		
3	HCA	C	501	-	13,13,13	1.13	0	14,18,18	1.53	2 (14%)
4	ICS	C	502	1	18,30,30	2.85	12 (66%)	-		
5	CLF	C	503	1,2	0,24,24	-	-	-		
5	CLF	A	503	1,2	0,24,24	-	-	-		

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
3	HCA	A	501	-	-	2/17/17/17	-
5	CLF	A	503	1,2	-	-	0/12/10/10
3	HCA	C	501	-	-	2/17/17/17	-
5	CLF	C	503	1,2	-	-	0/12/10/10

The worst 5 of 25 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	502	ICS	S4B-FE7	-5.28	2.19	2.32
4	A	502	ICS	S2A-FE2	-4.97	2.20	2.32
4	C	502	ICS	S4B-FE7	-4.66	2.20	2.32
4	C	502	ICS	S3B-FE7	-4.43	2.21	2.32
4	C	502	ICS	S2A-FE2	-4.32	2.21	2.32

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	501	HCA	O6-C7-C3	3.96	119.92	113.05
3	C	501	HCA	O5-C7-C3	-2.74	118.38	122.25
3	A	501	HCA	O6-C7-C3	2.60	117.56	113.05
3	A	501	HCA	O3-C6-C5	-2.31	115.66	123.08
3	A	501	HCA	O5-C7-C3	-2.08	119.30	122.25

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	501	HCA	C1-C2-C3-C4
3	C	501	HCA	C1-C2-C3-C7
3	A	501	HCA	C4-C5-C6-O3
3	A	501	HCA	C4-C5-C6-O4

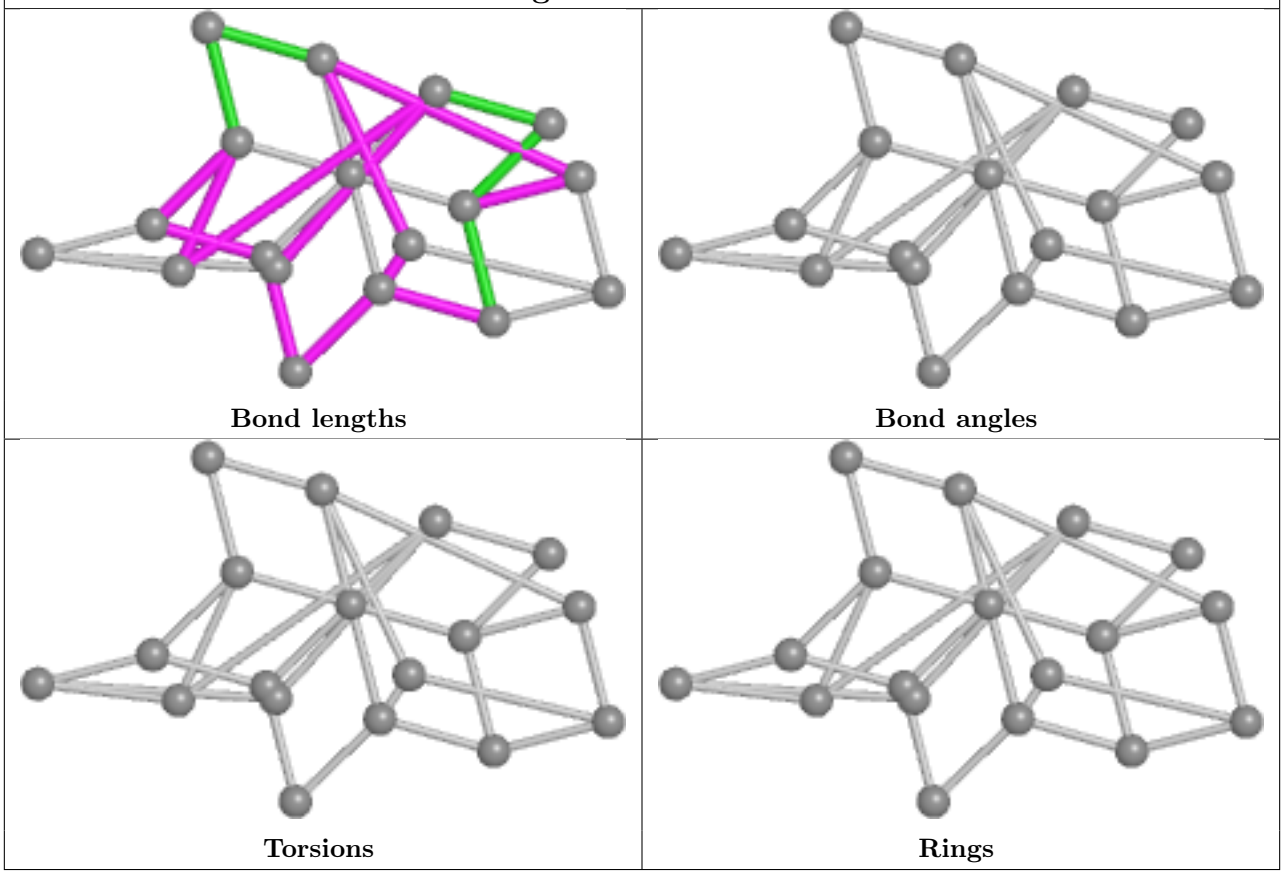
There are no ring outliers.

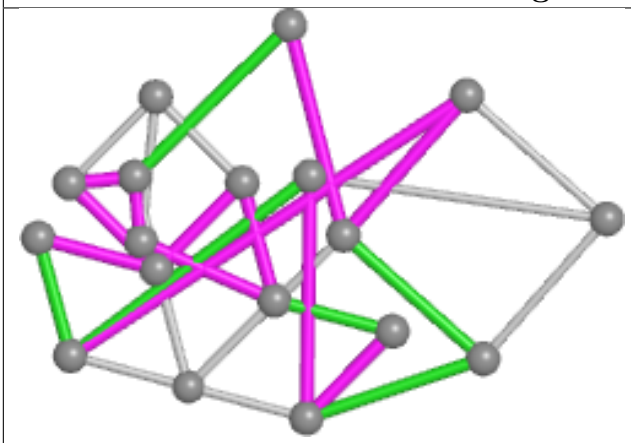
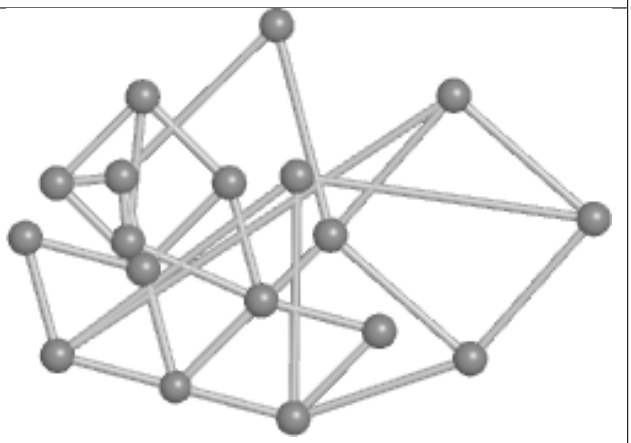
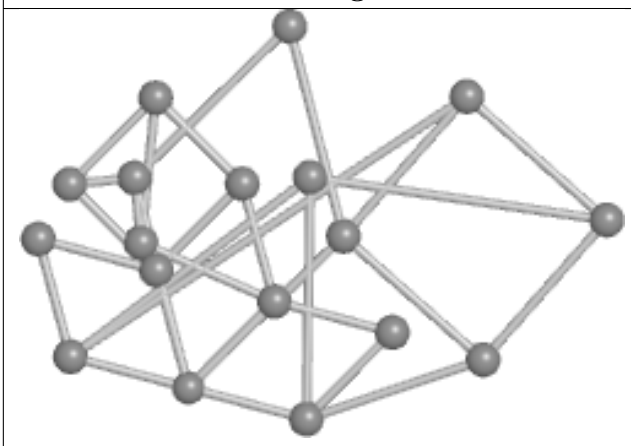
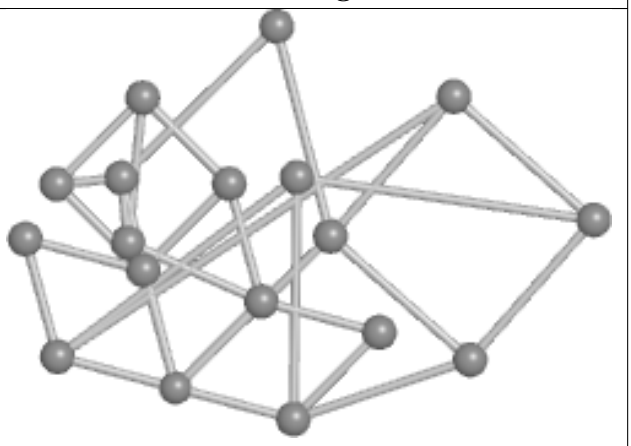
4 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	501	HCA	1	0
4	A	502	ICS	1	0
3	C	501	HCA	2	0
5	C	503	CLF	1	0

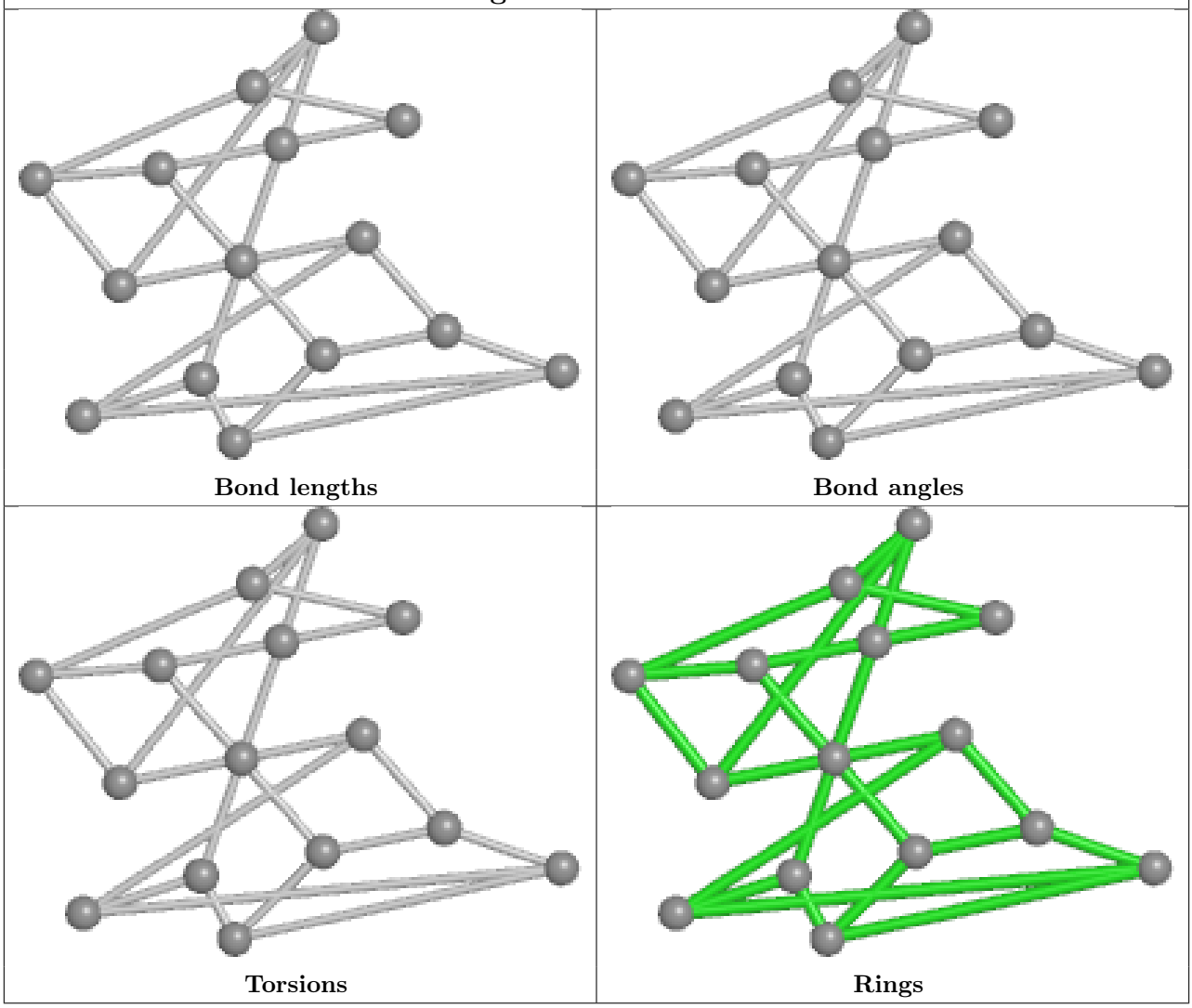
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. 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 ICS A 502

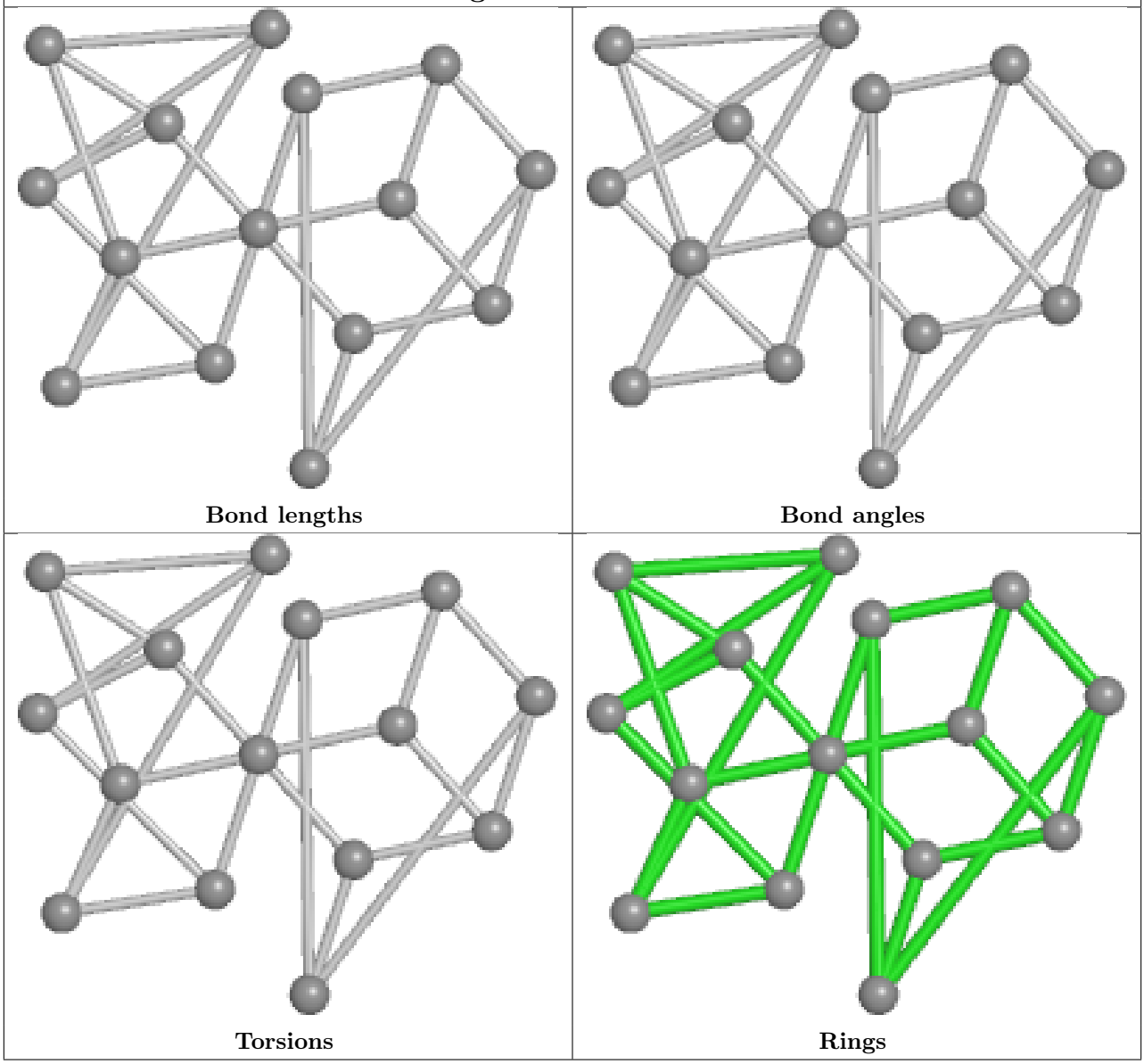


Ligand ICS C 502	
 Bond lengths	 Bond angles
 Torsions	 Rings

Ligand CLF C 503



Ligand CLF A 503



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	477/492 (96%)	0.32	16 (3%)	45	51	15, 23, 40, 60	0
1	C	478/492 (97%)	0.24	17 (3%)	42	49	15, 23, 39, 56	0
2	B	522/523 (99%)	0.09	4 (0%)	86	90	13, 20, 30, 44	0
2	D	522/523 (99%)	0.08	6 (1%)	80	86	14, 21, 32, 41	0
All	All	1999/2030 (98%)	0.18	43 (2%)	62	69	13, 21, 35, 60	0

The worst 5 of 43 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	215	THR	11.3
1	A	214	THR	11.2
1	A	212	GLU	8.4
1	A	5	SER	7.5
1	C	38	ALA	7.1

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)

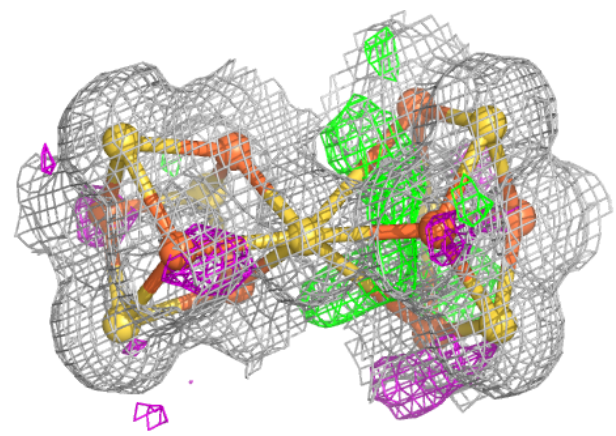
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

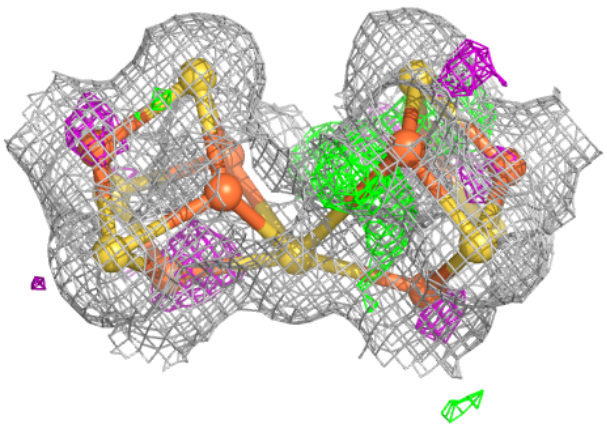
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	HCA	A	501	14/14	0.87	0.14	15,19,23,24	0
3	HCA	C	501	14/14	0.88	0.12	17,20,24,26	0
5	CLF	C	503	15/15	0.97	0.06	14,17,20,23	3
5	CLF	A	503	15/15	0.98	0.08	14,17,20,23	3
4	ICS	A	502	18/18	0.99	0.07	13,17,18,19	0
4	ICS	C	502	18/18	0.99	0.06	15,18,20,20	0
6	FE	B	601	1/1	0.99	0.08	22,22,22,22	1
6	FE	D	601	1/1	0.99	0.05	23,23,23,23	1

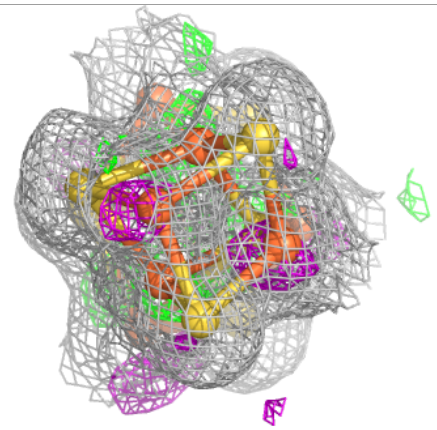
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around CLF C 503:

2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative) and green (positive)

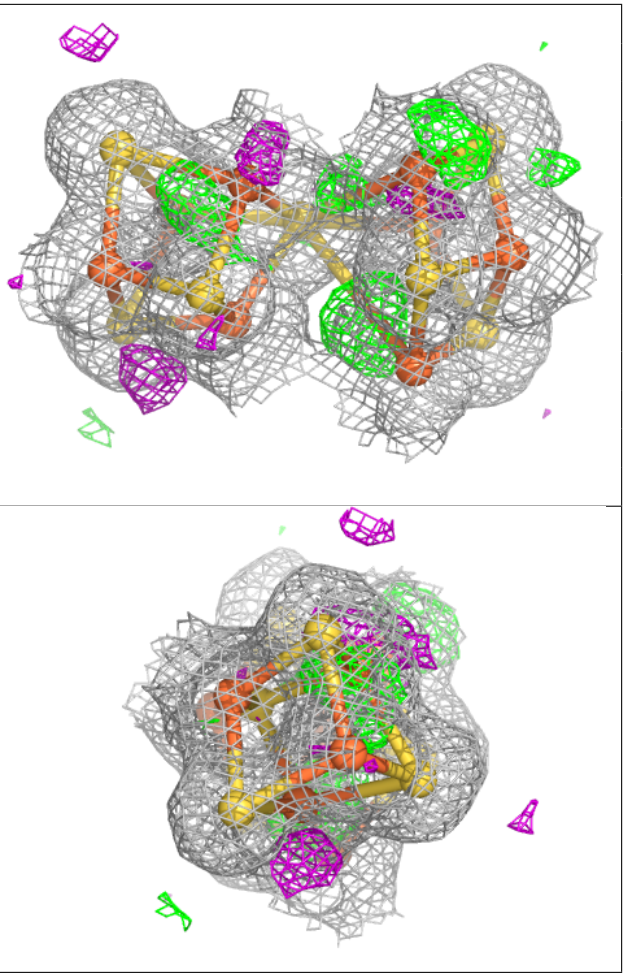






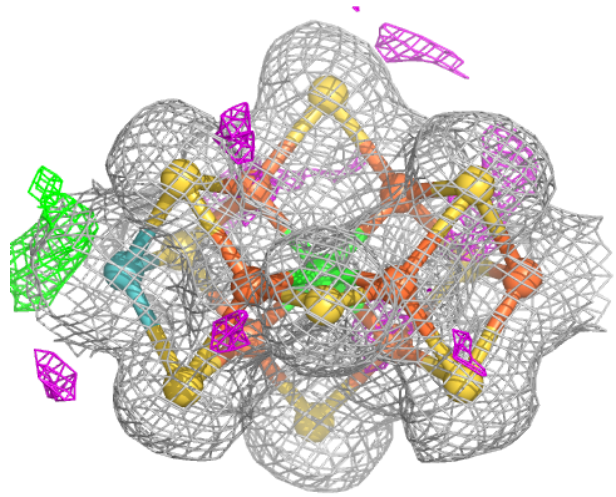
Electron density around CLF A 503:

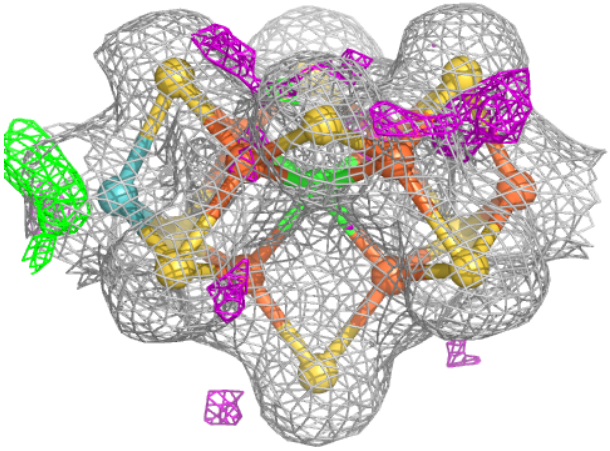
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)

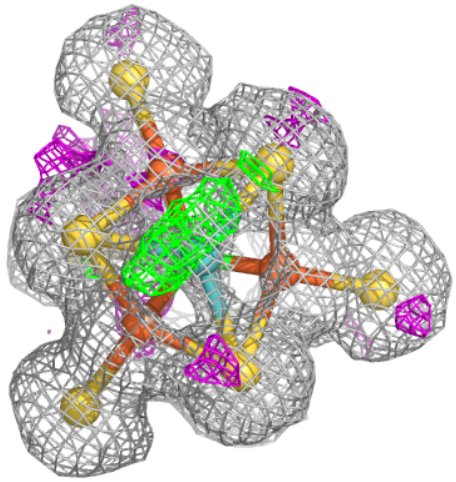


Electron density around ICS A 502:

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

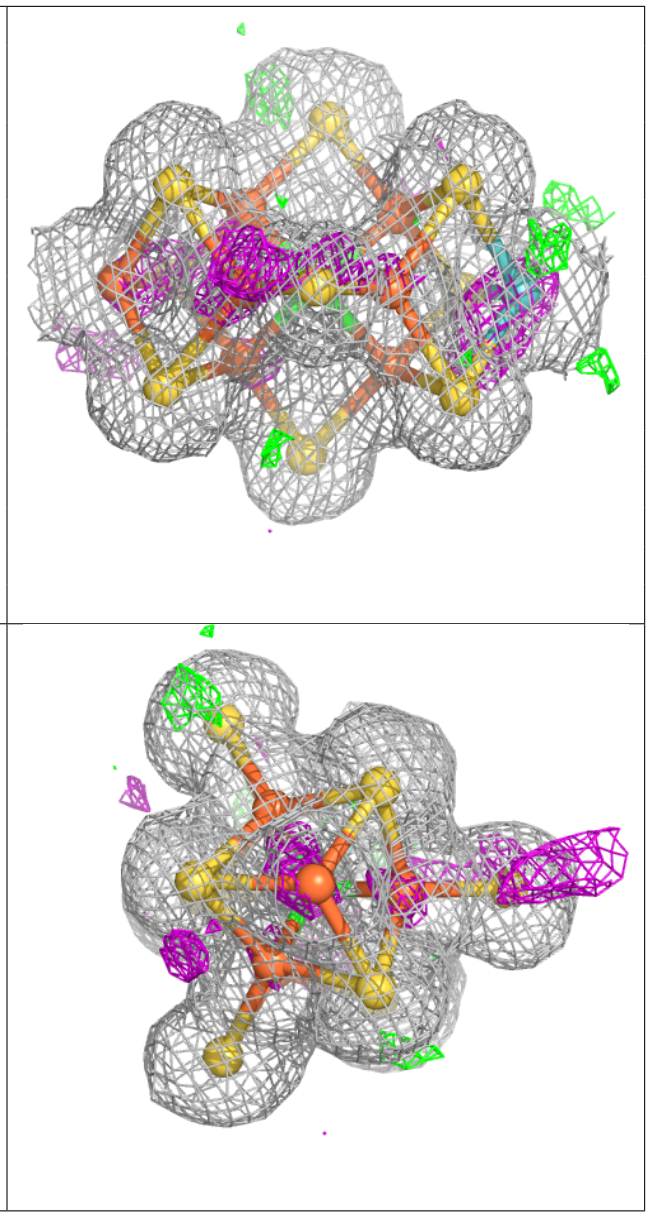






Electron density around ICS C 502:

$2mF_o$-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.