W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jan 14, 2024 – 02:30 am GMT

PDB ID	:	6SZD
Title	:	Hydrogenase-2 variant R479K - hydrogen reduced form
Authors	:	Carr, S.B.; Beaton, S.E.; Evans, R.M.; Armstrong, F.A.
Deposited on	:	2019-10-02
Resolution	:	1.50 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
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)	:	2.36

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 1.50 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.
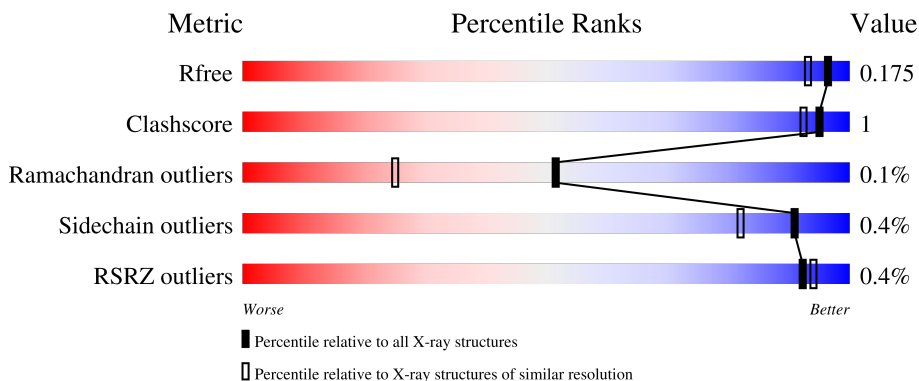
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)
RSRZ outliers	127900	2884 (1.50-1.50)

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	SSS	298	89% · 10%
1	TTT	298	.% 89% · 10%
2	LLL	567	92% 5% ·
2	MMM	567	.% 93% · ·

2 Entry composition (i)

There are 10 unique types of molecules in this entry. The entry contains 14042 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 Hydrogenase-2 small chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	SSS	268	Total 2050	C 1300	N 361	O 376	S 13	0	2	0
1	TTT	268	Total 2041	C 1293	N 360	O 375	S 13	0	1	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
SSS	291	HIS	-	expression tag	UNP P69741
SSS	292	HIS	-	expression tag	UNP P69741
SSS	293	HIS	-	expression tag	UNP P69741
SSS	294	HIS	-	expression tag	UNP P69741
SSS	295	HIS	-	expression tag	UNP P69741
SSS	296	HIS	-	expression tag	UNP P69741
TTT	291	HIS	-	expression tag	UNP P69741
TTT	292	HIS	-	expression tag	UNP P69741
TTT	293	HIS	-	expression tag	UNP P69741
TTT	294	HIS	-	expression tag	UNP P69741
TTT	295	HIS	-	expression tag	UNP P69741
TTT	296	HIS	-	expression tag	UNP P69741

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	LLL	551	Total 4319	C 2749	N 741	O 811	S 18	0	6	0
2	MMM	551	Total 4331	C 2756	N 742	O 815	S 18	0	8	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
LLL	479	LYS	ARG	engineered mutation	UNP V0V766
MMM	479	LYS	ARG	engineered mutation	UNP V0V766

- Molecule 3 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe_4S_4).
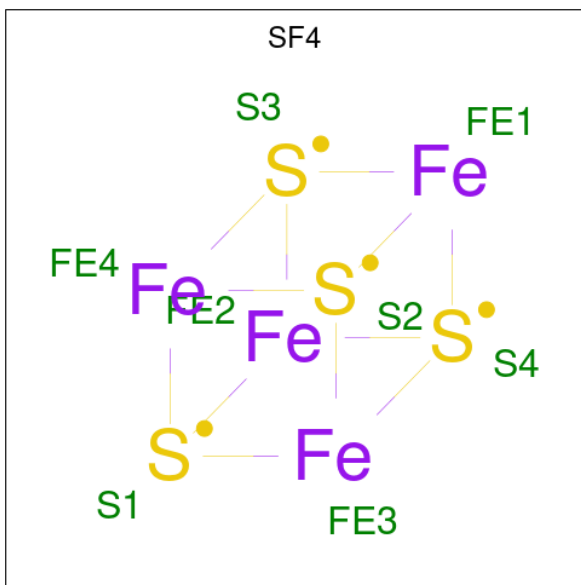
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	SSS	1	Total 8	Fe 4	S 4	0	0
3	SSS	1	Total 8	Fe 4	S 4	0	0
3	TTT	1	Total 8	Fe 4	S 4	0	0
3	TTT	1	Total 8	Fe 4	S 4	0	0

- Molecule 4 is FE3-S4 CLUSTER (three-letter code: F3S) (formula: Fe_3S_4).
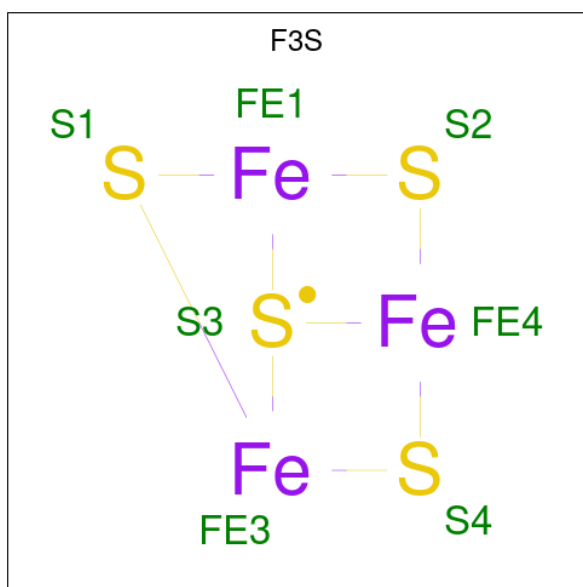
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	SSS	1	Total Fe S 7 3 4		0	0
4	TTT	1	Total Fe S 7 3 4		0	0

- Molecule 5 is CARBONMONOXIDE-(DICYANO) IRON (three-letter code: FCO) (formula: C_3FeN_2O).
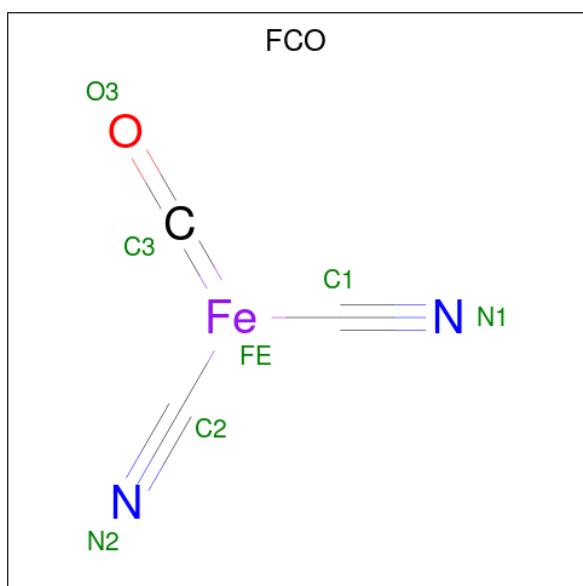
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	LLL	1	Total C Fe N O 7 3 1 2 1					0	0

Continued on next page...

Continued from previous page...

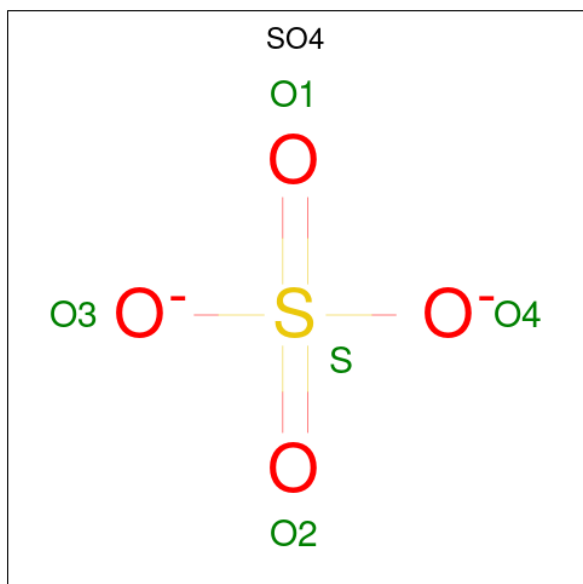| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	MMM	1	Total 7	C 3	Fe 1	N 2	O 1	0	0

- Molecule 6 is NICKEL (II) ION (three-letter code: NI) (formula: Ni) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	LLL	1	Total 1	Ni 1	0	0
6	MMM	1	Total 1	Ni 1	0	0

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	LLL	1	Total 1	Mg 1	0	0
7	MMM	1	Total 1	Mg 1	0	0

- Molecule 8 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	TTT	1	Total 5	O 4	S 1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	MMM	1	Total 5	O 4	S 1	0	0

- Molecule 9 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	MMM	1	Total 1	Cl 1	0	0

- Molecule 10 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	SSS	205	Total 205	O 205	0	0
10	LLL	454	Total 454	O 454	0	0
10	TTT	178	Total 178	O 178	0	0
10	MMM	389	Total 389	O 389	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.
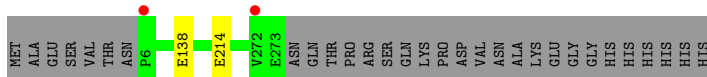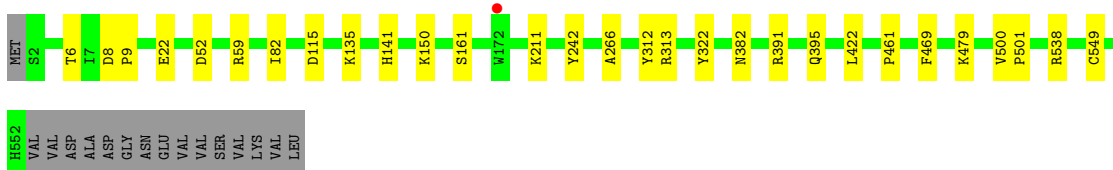
• Molecule 1: Hydrogenase-2 small chain

Chain SSS:
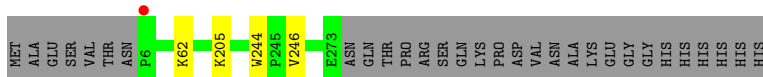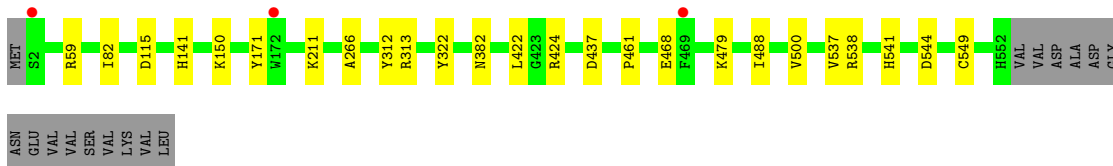

• Molecule 1: Hydrogenase-2 small chain

Chain TTT:

• Molecule 2: Hydrogenase-2 large chain

Chain LLL:

• Molecule 2: Hydrogenase-2 large chain

Chain MMM:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	99.42Å　　100.56Å　　168.66Å 90.00°　　90.00°　　90.00°	Depositor
Resolution (Å)	86.38　–　1.50 86.38　–　1.50	Depositor EDS
% Data completeness (in resolution range)	97.6 (86.38-1.50) 97.6 (86.38-1.50)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.52 (at 1.50Å)	Xtriage
Refinement program	REFMAC 5.8.0253	Depositor
R, R_{free}	0.152　,　0.173 0.153　,　0.175	Depositor DCC
R_{free} test set	13281 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	15.3	Xtriage
Anisotropy	0.034	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 47.1	EDS
L-test for twinning[2]	$< |L| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.004 for k,h,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	14042	wwPDB-VP
Average B, all atoms (Å^2)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.16% 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: MG, FCO, NI, F3S, CL, SF4, SO4

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	SSS	0.76	0/2114	0.84	0/2877
1	TTT	0.77	2/2102 (0.1%)	0.84	0/2862
2	LLL	0.73	0/4441	0.85	4/6052 (0.1%)
2	MMM	0.73	0/4456	0.85	5/6073 (0.1%)
All	All	0.74	2/13113 (0.0%)	0.85	9/17864 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	TTT	214	GLU	CD-OE1	6.84	1.33	1.25
1	TTT	138	GLU	CD-OE1	5.08	1.31	1.25

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	LLL	59	ARG	NE-CZ-NH2	-5.83	117.38	120.30
2	MMM	424	ARG	NE-CZ-NH1	5.57	123.08	120.30
2	MMM	382	ASN	N-CA-CB	-5.39	100.90	110.60
2	LLL	382	ASN	N-CA-CB	-5.33	101.01	110.60
2	LLL	469	PHE	CB-CA-C	5.32	121.04	110.40

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	SSS	2050	0	1981	5	0
1	TTT	2041	0	1967	0	0
2	LLL	4319	0	4268	16	0
2	MMM	4331	0	4278	9	0
3	SSS	16	0	0	0	0
3	TTT	16	0	0	0	0
4	SSS	7	0	0	0	0
4	TTT	7	0	0	0	0
5	LLL	7	0	0	0	0
5	MMM	7	0	0	0	0
6	LLL	1	0	0	0	0
6	MMM	1	0	0	0	0
7	LLL	1	0	0	0	0
7	MMM	1	0	0	0	0
8	MMM	5	0	0	0	0
8	TTT	5	0	0	0	0
9	MMM	1	0	0	0	0
10	LLL	454	0	0	7	0
10	MMM	389	0	0	2	0
10	SSS	205	0	0	5	0
10	TTT	178	0	0	0	0
All	All	14042	0	12494	30	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.

The worst 5 of 30 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)
2:LLL:150:LYS:HD3	10:LLL:857:HOH:O	1.61	1.00
2:LLL:22[A]:GLU:OE1	10:LLL:701:HOH:O	1.87	0.92
2:LLL:6[B]:THR:HG21	10:LLL:913:HOH:O	1.89	0.72
1:SSS:205:LYS:CE	10:SSS:504:HOH:O	2.42	0.66
2:MMM:437[B]:ASP:OD2	10:MMM:701:HOH:O	2.14	0.65

There are no symmetry-related clashes.

5.3   Torsion angles ⓘ

5.3.1   Protein backbone ⓘ

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	SSS	268/298 (90%)	259 (97%)	9 (3%)	0	100	100
1	TTT	267/298 (90%)	257 (96%)	10 (4%)	0	100	100
2	LLL	555/567 (98%)	531 (96%)	23 (4%)	1 (0%)	47	23
2	MMM	557/567 (98%)	536 (96%)	20 (4%)	1 (0%)	47	23
All	All	1647/1730 (95%)	1583 (96%)	62 (4%)	2 (0%)	51	25

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	MMM	211	LYS
2	LLL	211	LYS

5.3.2   Protein sidechains ⓘ

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	SSS	215/239 (90%)	215 (100%)	0	100	100
1	TTT	214/239 (90%)	214 (100%)	0	100	100
2	LLL	471/479 (98%)	468 (99%)	3 (1%)	86	74
2	MMM	473/479 (99%)	470 (99%)	3 (1%)	86	74
All	All	1373/1436 (96%)	1367 (100%)	6 (0%)	91	82

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	MMM	141	HIS
2	MMM	312	TYR
2	MMM	479	LYS
2	LLL	312	TYR
2	LLL	141	HIS

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, 5 are monoatomic - leaving 10 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	$\#|Z| > 2$
8	SO4	TTT	404	-	4,4,4	0.20	0	6,6,6	0.20	0
8	SO4	MMM	605	-	4,4,4	0.21	0	6,6,6	0.09	0
3	SF4	SSS	401	1	0,12,12	-	-	-		
3	SF4	SSS	403	1	0,12,12	-	-	-		
5	FCO	MMM	601	2,6	0,6,6	-	-	-		
4	F3S	TTT	402	1	0,9,9	-	-	-		
3	SF4	TTT	403	1	0,12,12	-	-	-		

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\lvert Z \rvert > 2$	Counts	RMSZ	$\#\lvert Z \rvert > 2$
5	FCO	LLL	601	2,6	0,6,6	-	-	-		
4	F3S	SSS	402	1	0,9,9	-	-	-		
3	SF4	TTT	401	1	0,12,12	-	-	-		

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	SF4	SSS	401	1	-	-	0/6/5/5
3	SF4	SSS	403	1	-	-	0/6/5/5
4	F3S	TTT	402	1	-	-	0/3/3/3
3	SF4	TTT	403	1	-	-	0/6/5/5
4	F3S	SSS	402	1	-	-	0/3/3/3
3	SF4	TTT	401	1	-	-	0/6/5/5

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 ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 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	SSS	268/298 (89%)	-0.39	1 (0%)	92	94	11, 16, 30, 54	0
1	TTT	268/298 (89%)	-0.29	2 (0%)	87	90	12, 18, 34, 65	0
2	LLL	551/567 (97%)	-0.40	1 (0%)	95	95	11, 16, 30, 41	0
2	MMM	551/567 (97%)	-0.32	3 (0%)	91	93	11, 18, 36, 52	0
All	All	1638/1730 (94%)	-0.35	7 (0%)	92	94	11, 17, 33, 65	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	MMM	172	TRP	3.3
1	TTT	6	PRO	2.9
1	SSS	6	PRO	2.8
2	LLL	172	TRP	2.6
2	MMM	2	SER	2.6

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 ⓘ

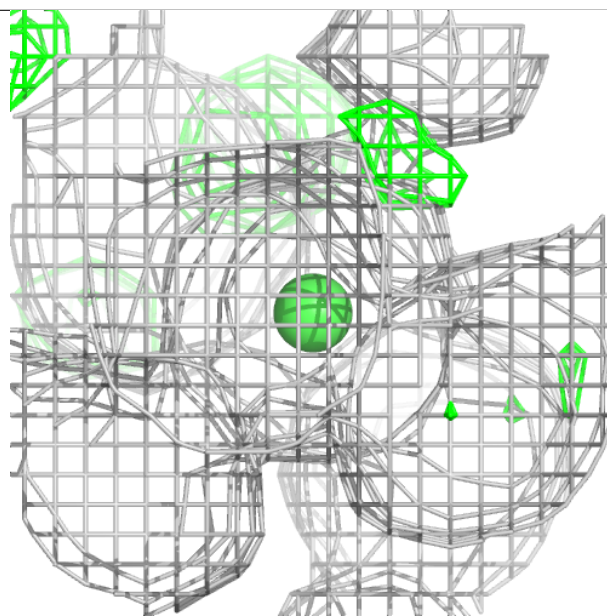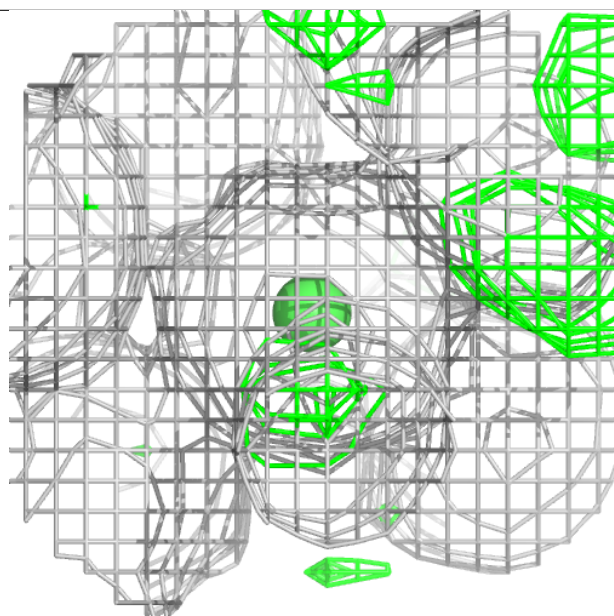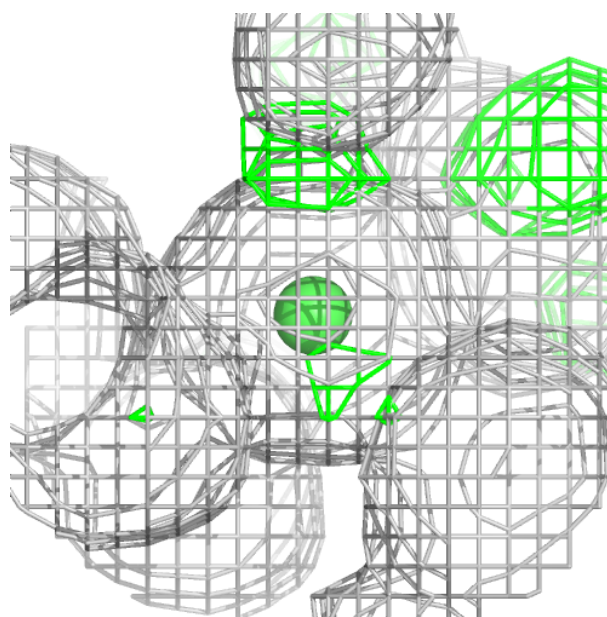There are no monosaccharides in this entry.

6.4   Ligands ⓘ

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.
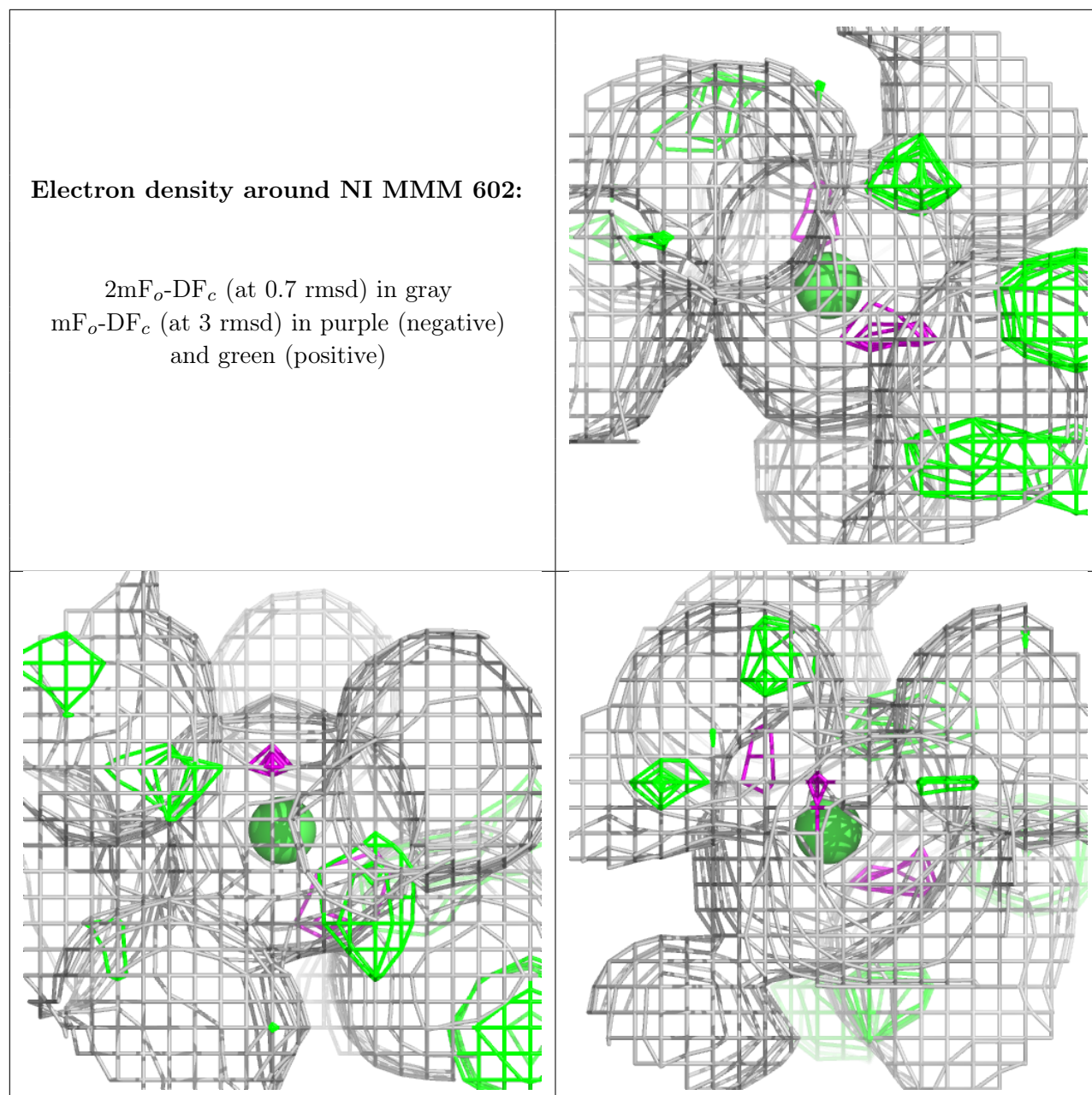
WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å²)	Q<0.9
8	SO4	TTT	404	5/5	0.87	0.24	34,43,46,55	0
8	SO4	MMM	605	5/5	0.90	0.21	36,39,49,50	0
7	MG	MMM	603	1/1	0.99	0.16	8,8,8,8	0
3	SF4	SSS	401	8/8	0.99	0.07	13,13,14,14	0
3	SF4	TTT	403	8/8	0.99	0.07	12,13,14,14	0
9	CL	MMM	604	1/1	0.99	0.06	22,22,22,22	0
5	FCO	LLL	601	7/7	1.00	0.07	10,11,12,12	0
5	FCO	MMM	601	7/7	1.00	0.07	12,12,14,14	0
6	NI	LLL	602	1/1	1.00	0.07	12,12,12,12	0
6	NI	MMM	602	1/1	1.00	0.05	14,14,14,14	0
7	MG	LLL	603	1/1	1.00	0.17	6,6,6,6	0
3	SF4	TTT	401	8/8	1.00	0.06	13,14,14,15	0
3	SF4	SSS	403	8/8	1.00	0.07	11,11,11,12	0
4	F3S	SSS	402	7/7	1.00	0.07	11,12,12,13	0
4	F3S	TTT	402	7/7	1.00	0.06	13,13,13,13	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 LLL 602:

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

Electron density around NI MMM 602:

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

6.5 Other polymers ⓘ

There are no such residues in this entry.