



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 6, 2016 – 05:35 PM EDT

PDB ID : 5KOJ
Title : Nitrogenase MoFeP protein in the IDS oxidized state
Authors : Owens, C.P.; Tezcan, F.A.
Deposited on : 2016-06-30
Resolution : 2.59 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

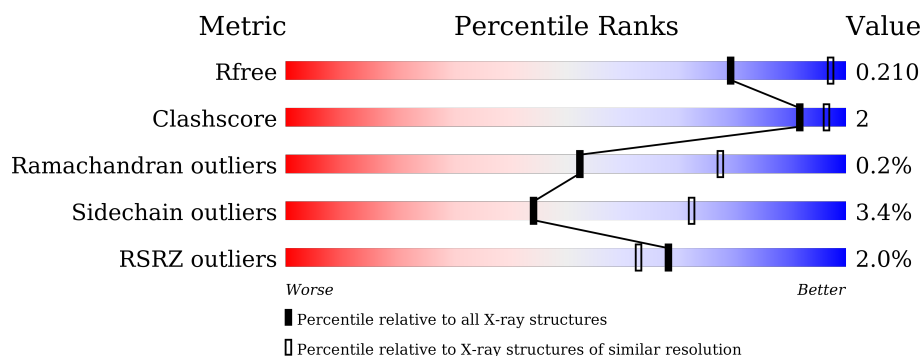
The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20027939
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20027939

i

X-RAY DIFFRACTION

A.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R _{free}	91344	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

Mol	Chain	Length	Quality of chain
1	A	499	<div> <div></div> <div>88%</div> <div>8%</div> <div>• •</div> </div>
1	C	499	<div> <div>3%</div> <div></div> <div>86%</div> <div>9%</div> <div>• •</div> </div>
2	B	511	<div> <div>2%</div> <div></div> <div>95%</div> <div>5%</div> <div>•</div> </div>
2	D	511	<div> <div></div> <div></div> <div>93%</div> <div>6%</div> <div>•</div> </div>

2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 16070 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.

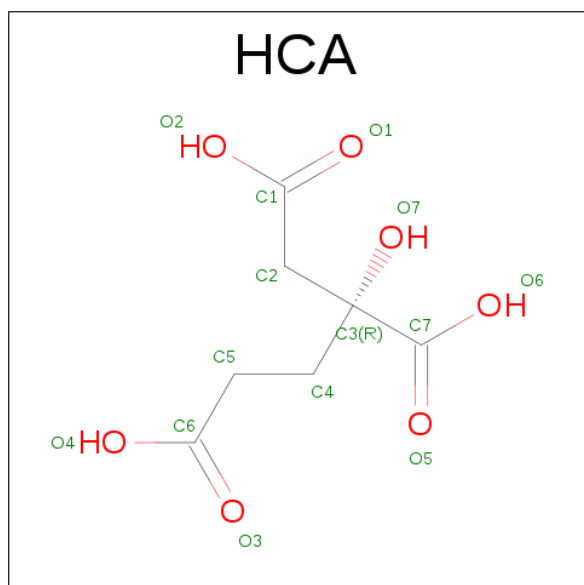
- Molecule 1 is a protein called Nitrogenase protein alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	481	Total	C	N	O	S	0	0	0
			3804	2420	663	701	20			
1	C	480	Total	C	N	O	S	0	0	0
			3800	2418	662	700	20			

- Molecule 2 is a protein called Nitrogenase FeMo beta subunit protein NifK.

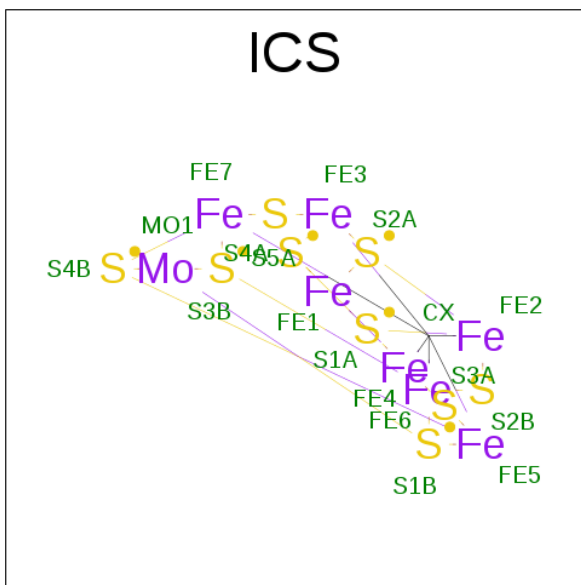
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	510	Total	C	N	O	S	0	0	0
			4006	2547	676	760	23			
2	D	510	Total	C	N	O	S	0	0	0
			4006	2547	676	760	23			

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



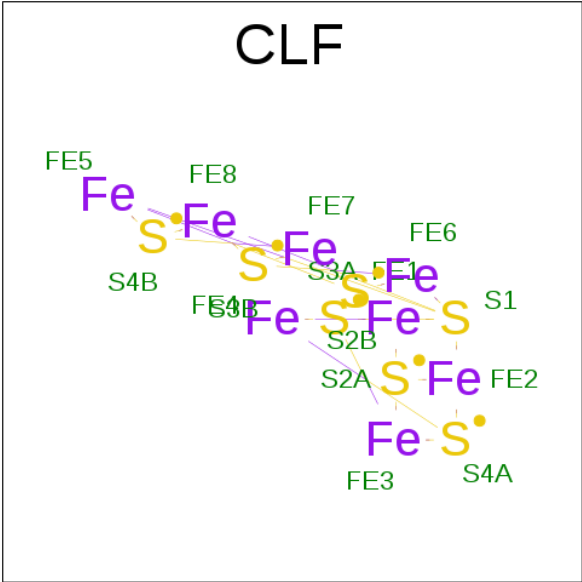
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			14	7	7		
3	C	1	Total	C	O	0	0
			14	7	7		

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



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

- Molecule 5 is FE(8)-S(7) CLUSTER (three-letter code: CLF) (formula: Fe_8S_7).



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

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

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

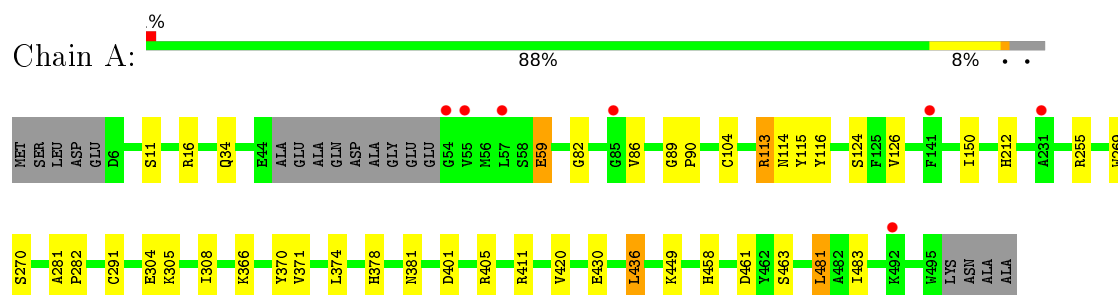
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	87	Total	O	0	0
			87	87		
7	B	93	Total	O	0	0
			93	93		
7	C	76	Total	O	0	0
			76	76		
7	D	102	Total	O	0	0
			102	102		

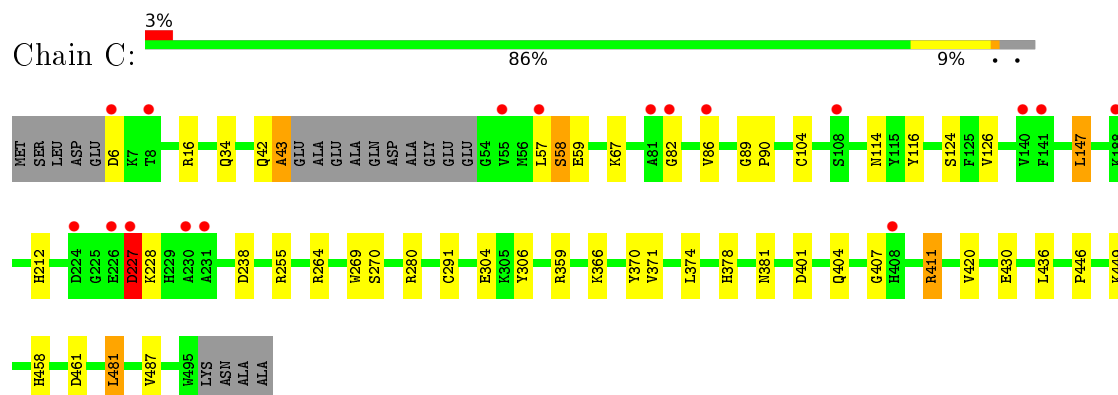
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

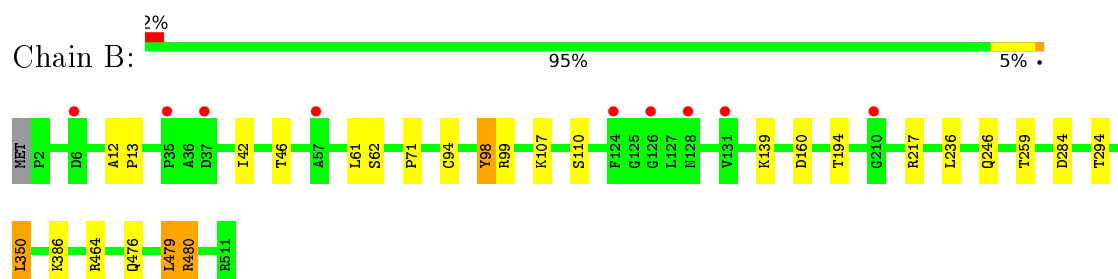
- Molecule 1: Nitrogenase protein alpha chain



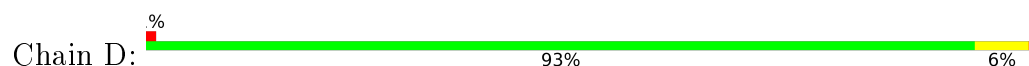
- Molecule 1: Nitrogenase protein alpha chain

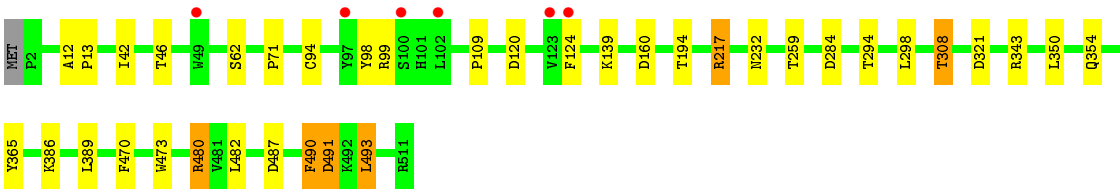


- Molecule 2: Nitrogenase FeMo beta subunit protein NifK



- Molecule 2: Nitrogenase FeMo beta subunit protein NifK





4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	201.57Å 201.57Å 132.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.59 – 2.59 39.59 – 2.59	Depositor EDS
% Data completeness (in resolution range)	99.5 (39.59-2.59) 99.6 (39.59-2.59)	Depositor EDS
R_{merge}	0.41	Depositor
R_{sym}	0.42	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.02 (at 2.58Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.168 , 0.208 0.174 , 0.210	Depositor DCC
R_{free} test set	4244 reflections (5.31%)	DCC
Wilson B-factor (Å ²)	39.1	Xtriage
Anisotropy	0.335	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 28.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	16070	wwPDB-VP
Average B, all atoms (Å ²)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.14% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: ICS, CLF, HCA, 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	0.70	0/3896	0.86	5/5263 (0.1%)
1	C	0.70	1/3892 (0.0%)	0.88	8/5257 (0.2%)
2	B	0.72	1/4109 (0.0%)	0.81	2/5566 (0.0%)
2	D	0.71	0/4109	0.83	9/5566 (0.2%)
All	All	0.71	2/16006 (0.0%)	0.84	24/21652 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	C	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	98	TYR	CZ-OH	5.35	1.47	1.37
1	C	43	ALA	N-CA	5.08	1.56	1.46

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	99	ARG	NE-CZ-NH1	7.28	123.94	120.30
1	A	255	ARG	NE-CZ-NH2	-7.04	116.78	120.30
1	A	411	ARG	NE-CZ-NH1	6.54	123.57	120.30
1	C	280	ARG	NE-CZ-NH2	-6.38	117.11	120.30
1	C	147	LEU	CA-CB-CG	6.37	129.95	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	227	ASP	Peptide

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	3804	0	3710	19	0
1	C	3800	0	3710	19	0
2	B	4006	0	3869	15	0
2	D	4006	0	3869	20	0
3	A	14	0	6	1	0
3	C	14	0	6	1	0
4	A	18	0	0	0	0
4	C	18	0	0	0	0
5	A	15	0	0	0	0
5	C	15	0	0	0	0
6	B	1	0	0	0	0
6	D	1	0	0	0	0
7	A	87	0	0	2	0
7	B	93	0	0	1	0
7	C	76	0	0	2	0
7	D	102	0	0	2	0
All	All	16070	0	15170	64	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 2.

The worst 5 of 64 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:D:490:PHE:O	2:D:491:ASP:CB	2.11	0.98
2:D:490:PHE:O	2:D:491:ASP:HB3	1.79	0.81
1:C:212:HIS:NE2	7:C:601:HOH:O	1.93	0.76
2:D:490:PHE:O	2:D:491:ASP:HB2	1.90	0.72
2:B:99:ARG:HD2	2:B:110:SER:O	1.92	0.69

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	477/499 (96%)	455 (95%)	22 (5%)	0	100	100
1	C	476/499 (95%)	453 (95%)	22 (5%)	1 (0%)	52	77
2	B	508/511 (99%)	496 (98%)	11 (2%)	1 (0%)	52	77
2	D	508/511 (99%)	497 (98%)	9 (2%)	2 (0%)	39	65
All	All	1969/2020 (98%)	1901 (96%)	64 (3%)	4 (0%)	52	77

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	227	ASP
2	D	491	ASP
2	B	294	THR
2	D	294	THR

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	398/413 (96%)	381 (96%)	17 (4%)	35	64
1	C	398/413 (96%)	378 (95%)	20 (5%)	30	56
2	B	427/428 (100%)	418 (98%)	9 (2%)	61	85

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	D	427/428 (100%)	417 (98%)	10 (2%)	58	83
All	All	1650/1682 (98%)	1594 (97%)	56 (3%)	44	72

5 of 56 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	480	ARG
1	C	104	CYS
2	D	350	LEU
1	C	6	ASP
1	C	58	SER

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

Mol	Chain	Res	Type
2	B	21	GLN
1	C	448	GLN

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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

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	-	4,13,13	0.52	0	3,18,18	0.80	0
4	ICS	A	502	1	6,30,30	1.29	0	0,78,78	0.00	-
5	CLF	A	503	1,2	0,24,24	0.00	-	0,57,57	0.00	-
3	HCA	C	501	-	4,13,13	0.49	0	3,18,18	0.55	0
4	ICS	C	502	1	6,30,30	1.93	3 (50%)	0,78,78	0.00	-
5	CLF	C	503	1,2	0,24,24	0.00	-	0,57,57	0.00	-

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	-	-	0/7/17/17	0/0/0/0
4	ICS	A	502	1	-	0/0/204/204	0/0/13/13
5	CLF	A	503	1,2	-	0/0/132/132	0/12/10/10
3	HCA	C	501	-	-	0/7/17/17	0/0/0/0
4	ICS	C	502	1	-	0/0/204/204	0/0/13/13
5	CLF	C	503	1,2	-	0/0/132/132	0/12/10/10

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	502	ICS	S5A-FE3	-2.14	2.19	2.24
4	C	502	ICS	S2B-FE6	-2.11	2.19	2.24
4	C	502	ICS	S3A-FE5	3.07	2.32	2.24

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	501	HCA	1	0
3	C	501	HCA	1	0

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, 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	481/499 (96%)	0.30	7 (1%) 76 71	28, 38, 61, 79	0
1	C	480/499 (96%)	0.44	17 (3%) 48 40	28, 39, 64, 100	0
2	B	510/511 (99%)	0.22	9 (1%) 71 66	25, 36, 56, 76	0
2	D	510/511 (99%)	0.23	6 (1%) 81 77	25, 37, 57, 86	0
All	All	1981/2020 (98%)	0.30	39 (1%) 68 63	25, 38, 59, 100	0

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	227	ASP	4.2
2	B	124	PHE	3.8
2	D	124	PHE	3.4
1	C	6	ASP	3.3
1	C	226	GLU	3.2

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
3	HCA	C	501	14/14	0.97	0.22	-0.16	31,37,39,43	0
3	HCA	A	501	14/14	0.97	0.19	-1.23	28,33,38,43	0
5	CLF	C	503	15/15	0.94	0.10	-1.92	34,38,49,54	0
5	CLF	A	503	15/15	0.95	0.10	-1.99	28,34,38,44	2
4	ICS	C	502	18/18	0.99	0.12	-2.09	31,34,38,40	0
4	ICS	A	502	18/18	0.99	0.11	-2.26	27,32,34,37	0
6	FE	D	601	1/1	0.92	0.09	-3.13	62,62,62,62	0
6	FE	B	601	1/1	0.91	0.06	-3.94	81,81,81,81	0

6.5 Other polymers [i](#)

There are no such residues in this entry.