



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 11, 2016 – 05:40 AM EDT

PDB ID : 5KOH
Title : Nitrogenase MoFeP from *Gluconacetobacter diazotrophicus* in dithionite reduced state
Authors : Owens, C.P.; Tezcan, F.A.
Deposited on : 2016-06-30
Resolution : 1.83 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

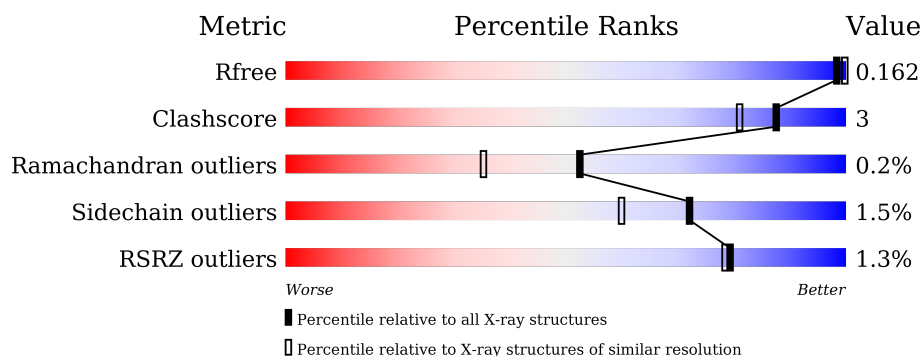
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.83 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}	91344	2634 (1.86-1.82)
Clashscore	102246	2862 (1.86-1.82)
Ramachandran outliers	100387	2831 (1.86-1.82)
Sidechain outliers	100360	2832 (1.86-1.82)
RSRZ outliers	91569	2639 (1.86-1.82)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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 $\leq 5\%$. 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	499	<div> <div>3%</div> <div> <div></div> <div>89%</div> <div>7%</div> <div>••</div> </div> </div>
1	C	499	<div> <div>%</div> <div> <div></div> <div>89%</div> <div>7%</div> <div>••</div> </div> </div>
2	B	511	<div> <div>%</div> <div> <div></div> <div>93%</div> <div>6%</div> <div>•</div> </div> </div>
2	D	511	<div> <div></div> <div> <div></div> <div>93%</div> <div>6%</div> <div>•</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	MRD	A	503	-	-	X	X
5	MRD	A	504	-	-	-	X
7	FE	B	602	-	-	-	X
8	MPD	C	503	-	-	-	X

2 Entry composition [i](#)

There are 9 unique types of molecules in this entry. The entry contains 17785 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.

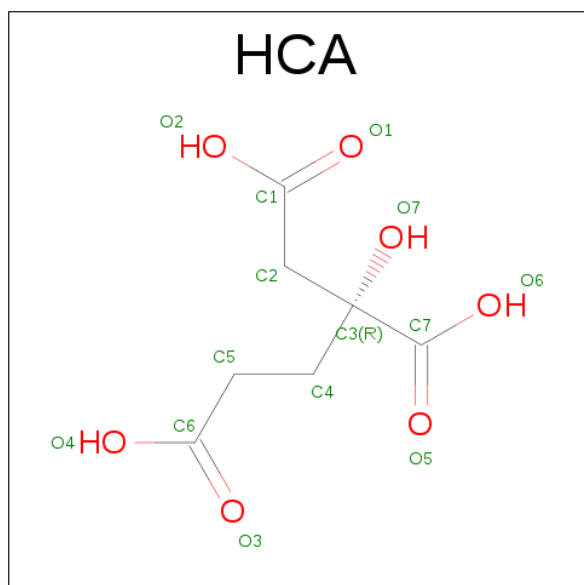
- Molecule 1 is a protein called Nitrogenase protein alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	483	Total	C	N	O	S	0	1	0
			3823	2433	669	701	20			
1	C	482	Total	C	N	O	S	0	1	0
			3816	2427	664	704	21			

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

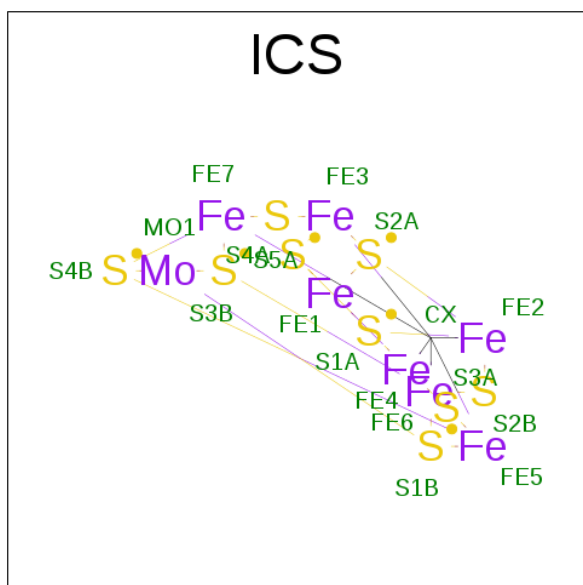
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	510	Total	C	N	O	S	0	5	0
			4032	2565	680	764	23			
2	D	510	Total	C	N	O	S	0	2	0
			4015	2553	676	763	23			

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



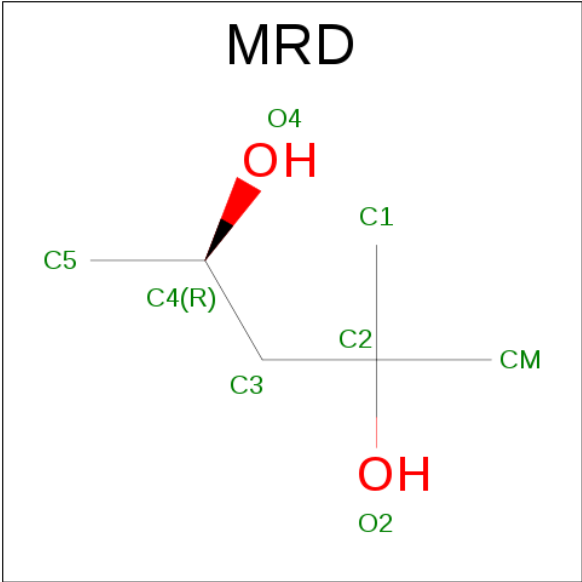
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			14	7	7		
3	C	1	Total	C	O	0	0
			14	7	7		

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



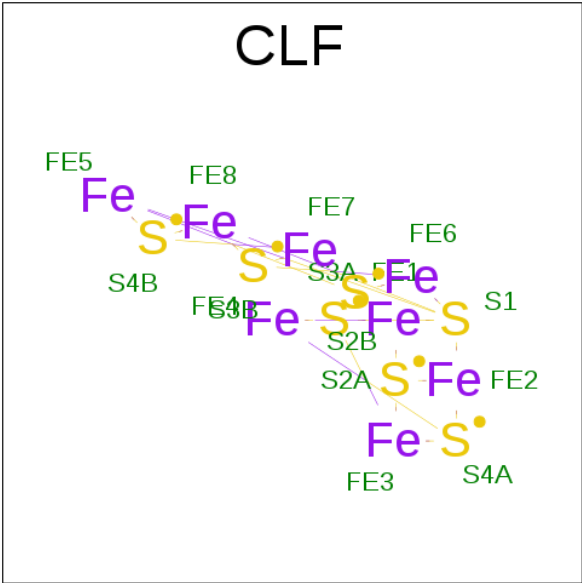
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	Fe	Mo	S	
			18	1	7	1	9	0
4	C	1	Total	C	Fe	Mo	S	
			18	1	7	1	9	0

- Molecule 5 is (4R)-2-METHYLPENTANE-2,4-DIOL (three-letter code: MRD) (formula: $\text{C}_6\text{H}_{14}\text{O}_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			8	6	2		
5	A	1	Total	C	O	0	0
			8	6	2		

- Molecule 6 is FE(8)-S(7) CLUSTER (three-letter code: CLF) (formula: Fe₈S₇).

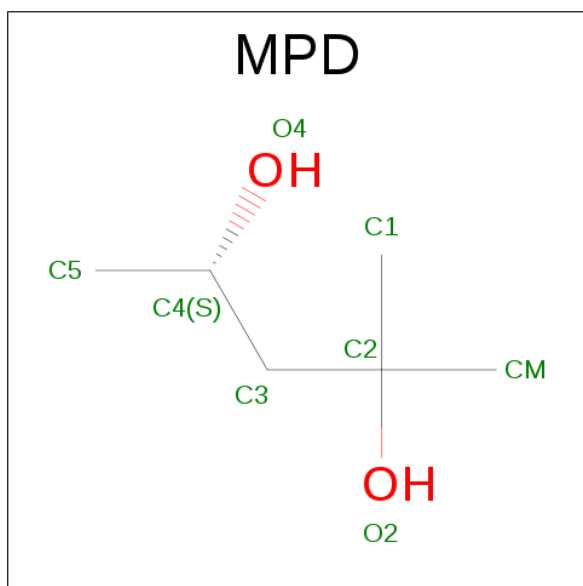


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	Fe	S	0	0
			15	8	7		
6	C	1	Total	Fe	S	0	0
			15	8	7		

- Molecule 7 is FE (III) ION (three-letter code: FE) (formula: Fe).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	B	2	Total	Fe	0	0
			2	2		

- Molecule 8 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: C₆H₁₄O₂).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	C	1	Total	C	O	0	0
			8	6	2		

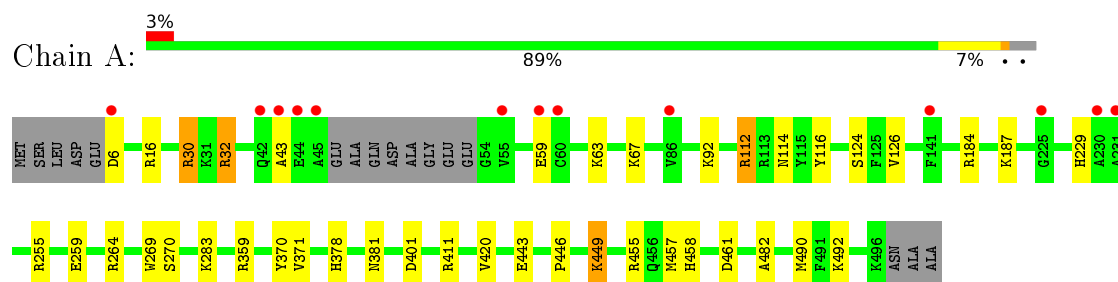
- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	443	Total	O	0	0
			443	443		
9	B	545	Total	O	0	0
			545	545		
9	C	440	Total	O	0	0
			440	440		
9	D	551	Total	O	0	0
			551	551		

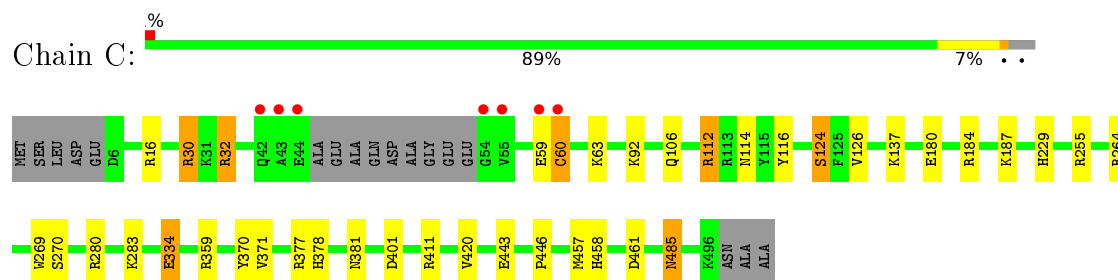
3 Residue-property plots [i](#)

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.

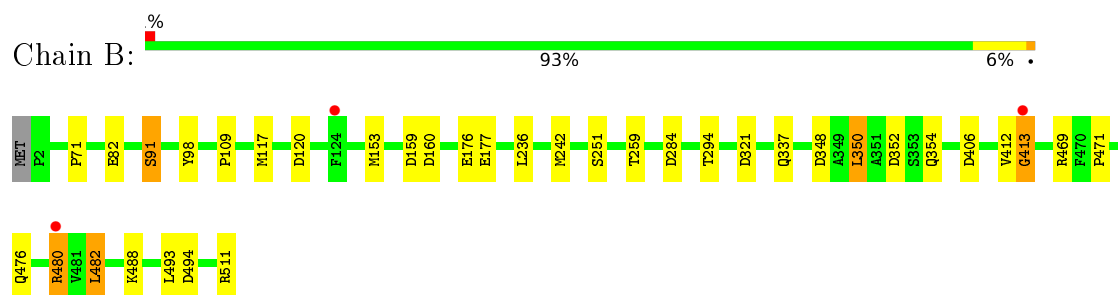
- Molecule 1: Nitrogenase protein alpha chain



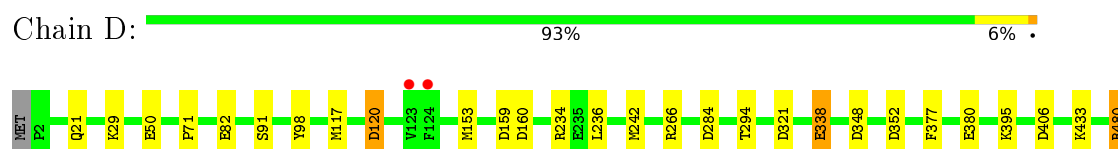
- Molecule 1: Nitrogenase protein alpha chain



- Molecule 2: Nitrogenase FeMo beta subunit protein NifK



- Molecule 2: Nitrogenase FeMo beta subunit protein NifK



L481	L482	L493	T497	D504	R511
------	------	------	------	------	------

4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	202.60 Å 202.60 Å 132.57 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	143.26 – 1.83 39.73 – 1.83	Depositor EDS
% Data completeness (in resolution range)	100.0 (143.26-1.83) 100.0 (39.73-1.83)	Depositor EDS
R_{merge}	0.42	Depositor
R_{sym}	0.43	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.21 (at 1.83 Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.129 , 0.150 0.142 , 0.162	Depositor DCC
R_{free} test set	11943 reflections (5.23%)	DCC
Wilson B-factor (Å ²)	14.5	Xtriage
Anisotropy	0.271	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 54.4	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	0.97	EDS
Total number of atoms	17785	wwPDB-VP
Average B, all atoms (Å ²)	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.69% 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

5.1 Standard geometry

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

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	0.92	3/3918 (0.1%)	1.05	19/5291 (0.4%)
1	C	1.00	11/3911 (0.3%)	1.13	19/5283 (0.4%)
2	B	1.03	10/4150 (0.2%)	1.00	18/5621 (0.3%)
2	D	0.96	9/4124 (0.2%)	0.95	14/5586 (0.3%)
All	All	0.98	33/16103 (0.2%)	1.03	70/21781 (0.3%)

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	A	0	1
1	C	0	1
2	B	0	1
All	All	0	3

The worst 5 of 33 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	480[A]	ARG	NE-CZ	-18.38	1.09	1.33
2	B	480[B]	ARG	NE-CZ	-18.38	1.09	1.33
2	B	511	ARG	C-O	11.47	1.45	1.23
1	A	449	LYS	CD-CE	11.23	1.79	1.51
2	D	511	ARG	C-O	9.83	1.42	1.23

The worst 5 of 70 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	255	ARG	NE-CZ-NH2	-26.46	107.07	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	457	MET	CG-SD-CE	-25.45	59.47	100.20
1	C	457	MET	CG-SD-CE	-25.22	59.85	100.20
1	C	255	ARG	NE-CZ-NH1	21.84	131.22	120.30
1	A	449	LYS	CD-CE-NZ	-20.22	65.19	111.70

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	112	ARG	Sidechain
2	B	413	GLY	Peptide
1	C	112	ARG	Sidechain

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	3823	0	3739	30	0
1	C	3816	0	3721	14	0
2	B	4032	0	3909	27	0
2	D	4015	0	3881	16	0
3	A	14	0	6	2	0
3	C	14	0	6	1	0
4	A	18	0	0	0	0
4	C	18	0	0	0	0
5	A	16	0	28	6	0
6	A	15	0	0	0	0
6	C	15	0	0	0	0
7	B	2	0	0	0	0
8	C	8	0	14	4	0
9	A	443	0	0	5	1
9	B	545	0	0	2	0
9	C	440	0	0	3	0
9	D	551	0	0	2	0
All	All	17785	0	15304	80	1

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 3.

The worst 5 of 80 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:449:LYS:CD	1:A:449:LYS:CE	1.79	1.58
1:A:449:LYS:CD	1:A:449:LYS:NZ	1.75	1.46
1:A:449:LYS:NZ	1:A:449:LYS:HD2	1.58	1.16
2:B:480[A]:ARG:NH2	2:D:480:ARG:HH22	1.51	1.07
2:B:480[A]:ARG:HH21	2:D:480:ARG:NH2	1.57	1.01

All (1) 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)
9:A:1033:HOH:O	9:A:1033:HOH:O[7_556]	2.09	0.11

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	480/499 (96%)	460 (96%)	20 (4%)	0	100	100
1	C	479/499 (96%)	461 (96%)	18 (4%)	0	100	100
2	B	513/511 (100%)	505 (98%)	6 (1%)	2 (0%)	39	22
2	D	510/511 (100%)	502 (98%)	7 (1%)	1 (0%)	52	35
All	All	1982/2020 (98%)	1928 (97%)	51 (3%)	3 (0%)	52	35

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	294	THR
2	D	294	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	251	SER

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	A	399/413 (97%)	392 (98%)	7 (2%)	66	52
1	C	400/413 (97%)	390 (98%)	10 (2%)	55	37
2	B	432/428 (101%)	428 (99%)	4 (1%)	84	78
2	D	429/428 (100%)	424 (99%)	5 (1%)	78	69
All	All	1660/1682 (99%)	1634 (98%)	26 (2%)	72	57

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	60[A]	CYS
1	C	106	GLN
2	D	482	LEU
1	C	60[B]	CYS
1	C	63	LYS

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

Mol	Chain	Res	Type
1	A	106	GLN
2	D	302	GLN
2	D	476	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 11 ligands modelled in this entry, 2 are monoatomic - leaving 9 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	1.41	1 (25%)	3,18,18	1.87	2 (66%)
4	ICS	A	502	1	6,30,30	1.48	1 (16%)	0,78,78	0.00	-
5	MRD	A	503	-	6,7,7	0.70	0	6,10,10	1.00	0
5	MRD	A	504	-	6,7,7	0.37	0	6,10,10	0.37	0
6	CLF	A	505	1,2	0,24,24	0.00	-	0,57,57	0.00	-
3	HCA	C	501	-	4,13,13	1.74	1 (25%)	3,18,18	1.70	1 (33%)
4	ICS	C	502	1	6,30,30	2.15	4 (66%)	0,78,78	0.00	-
8	MPD	C	503	-	6,7,7	0.92	0	6,10,10	1.22	1 (16%)
6	CLF	C	504	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	MRD	A	503	-	-	0/5/5/5	0/0/0/0
5	MRD	A	504	-	-	0/5/5/5	0/0/0/0
6	CLF	A	505	1,2	-	0/0/132/132	0/12/10/10

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
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
8	MPD	C	503	-	-	0/5/5/5	0/0/0/0
6	CLF	C	504	1,2	-	0/0/132/132	0/12/10/10

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	502	ICS	S2B-FE2	-3.00	2.17	2.24
4	C	502	ICS	S5A-FE7	-2.70	2.18	2.24
4	A	502	ICS	S2B-FE6	-2.57	2.18	2.24
4	C	502	ICS	S5A-FE3	-2.37	2.19	2.24
4	C	502	ICS	S2B-FE6	-2.35	2.19	2.24

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	C	503	MPD	CM-C2-C1	-2.64	104.11	110.41
3	C	501	HCA	C3-C2-C1	-2.46	111.11	114.95
3	A	501	HCA	C3-C2-C1	-2.34	111.30	114.95
3	A	501	HCA	C4-C5-C6	2.16	121.17	112.33

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 13 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	501	HCA	2	0
5	A	503	MRD	6	0
3	C	501	HCA	1	0
8	C	503	MPD	4	0

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 [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	483/499 (96%)	-0.28	13 (2%) 58 54	10, 15, 36, 81	0
1	C	482/499 (96%)	-0.36	7 (1%) 76 74	9, 15, 31, 59	0
2	B	510/511 (99%)	-0.44	3 (0%) 90 89	9, 14, 29, 50	0
2	D	510/511 (99%)	-0.45	2 (0%) 93 92	9, 14, 28, 44	0
All	All	1985/2020 (98%)	-0.38	25 (1%) 79 78	9, 15, 31, 81	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	45	ALA	8.6
1	C	43	ALA	4.3
1	C	44	GLU	3.7
1	C	55	VAL	3.7
1	A	231	ALA	3.6

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 carbohydrates 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. 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
5	MRD	A	504	8/8	0.90	0.12	6.33	27,33,35,43	0
5	MRD	A	503	8/8	0.89	0.18	3.98	19,25,30,31	0
8	MPD	C	503	8/8	0.87	0.18	3.95	18,25,30,32	0
7	FE	B	602	1/1	0.99	0.14	2.45	31,31,31,31	0
7	FE	B	601	1/1	0.98	0.10	0.40	31,31,31,31	0
3	HCA	C	501	14/14	0.98	0.11	-0.41	8,11,14,14	0
3	HCA	A	501	14/14	0.98	0.11	-0.57	9,11,14,16	0
6	CLF	A	505	15/15	0.99	0.06	-1.54	9,11,13,15	2
4	ICS	A	502	18/18	1.00	0.06	-2.08	9,11,12,12	0
4	ICS	C	502	18/18	0.99	0.07	-2.12	9,10,11,12	0
6	CLF	C	504	15/15	1.00	0.05	-2.30	10,10,12,14	0

6.5 Other polymers [i](#)

There are no such residues in this entry.