W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Jan 2, 2024 – 11:19 pm GMT

PDB ID	:	5A4M
Title	:	Mechanism of Hydrogen activation by NiFe-hydrogenases
Authors	:	Evans, R.M.; Brooke, E.J.; Wehlin, S.A.M.; Nomerotskaia, E.; Sergent, F.; Carr, S.B.; Philips, S.E.V.; Armstrong, F.A.
Deposited on	:	2015-06-10
Resolution	:	1.70 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
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 1.70 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.
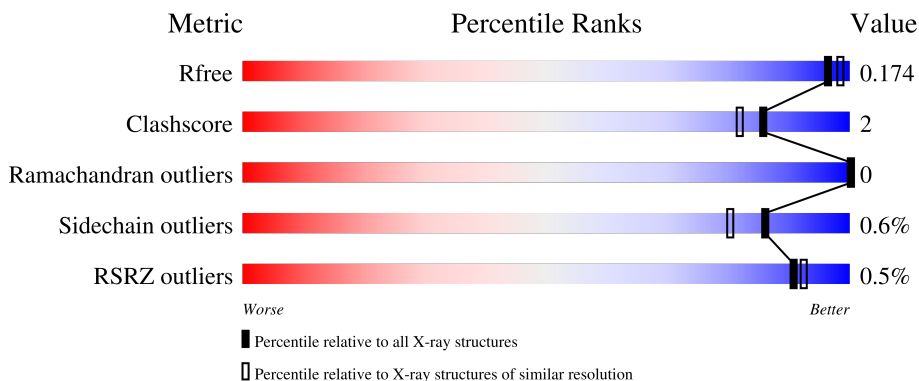
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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	L	582	95% ⋯ 5%
1	M	582	94% ⋯ 6%
2	S	278	88% ⋯ 6% · 5%
2	T	278	85% ⋯ 8% ·· 6%

2   Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 14827 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-1 LARGE CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	L	581	Total 4600	C 2919	N 809	O 844	S 28	0	8	0
1	M	581	Total 4578	C 2906	N 798	O 846	S 28	0	7	0

- Molecule 2 is a protein called HYDROGENASE-1 SMALL CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	S	263	Total 2045	C 1298	N 353	O 374	S 20	4	2	0
2	T	262	Total 2046	C 1301	N 346	O 378	S 21	0	6	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
S	267	GLY	-	expression tag	UNP P69739
S	268	SER	-	expression tag	UNP P69739
S	269	PHE	-	expression tag	UNP P69739
S	270	TYR	-	expression tag	UNP P69739
S	271	SER	-	expression tag	UNP P69739
S	272	ARG	-	expression tag	UNP P69739
S	273	HIS	-	expression tag	UNP P69739
S	274	HIS	-	expression tag	UNP P69739
S	275	HIS	-	expression tag	UNP P69739
S	276	HIS	-	expression tag	UNP P69739
S	277	HIS	-	expression tag	UNP P69739
S	278	HIS	-	expression tag	UNP P69739
T	267	GLY	-	expression tag	UNP P69739
T	268	SER	-	expression tag	UNP P69739
T	269	PHE	-	expression tag	UNP P69739
T	270	TYR	-	expression tag	UNP P69739

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
T	271	SER	-	expression tag	UNP P69739
T	272	ARG	-	expression tag	UNP P69739
T	273	HIS	-	expression tag	UNP P69739
T	274	HIS	-	expression tag	UNP P69739
T	275	HIS	-	expression tag	UNP P69739
T	276	HIS	-	expression tag	UNP P69739
T	277	HIS	-	expression tag	UNP P69739
T	278	HIS	-	expression tag	UNP P69739

- Molecule 3 is NI-FE OXIDIZED ACTIVE CENTER (three-letter code: NFV) (formula: $C_3FeN_2NiO_2$).
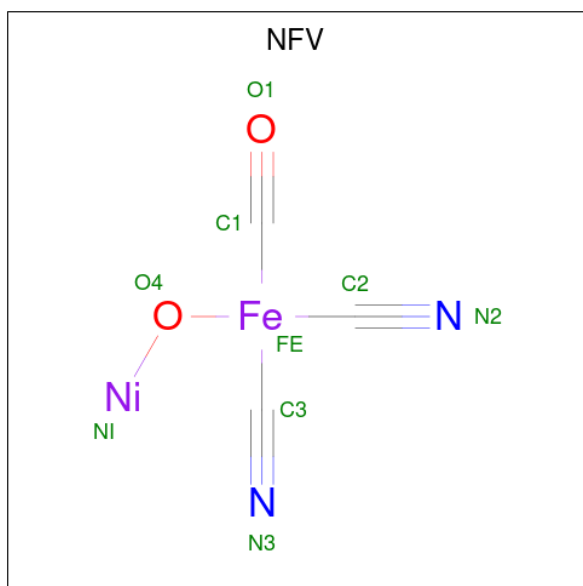
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	L	1	Total	C	Fe	N	Ni	O	0	0
			9	3	1	2	1	2		
3	M	1	Total	C	Fe	N	Ni	O	0	0
			9	3	1	2	1	2		

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	L	1	Total	Mg	0	0
			1	1		
4	M	1	Total	Mg	0	0
			1	1		

- Molecule 5 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe_4S_4).
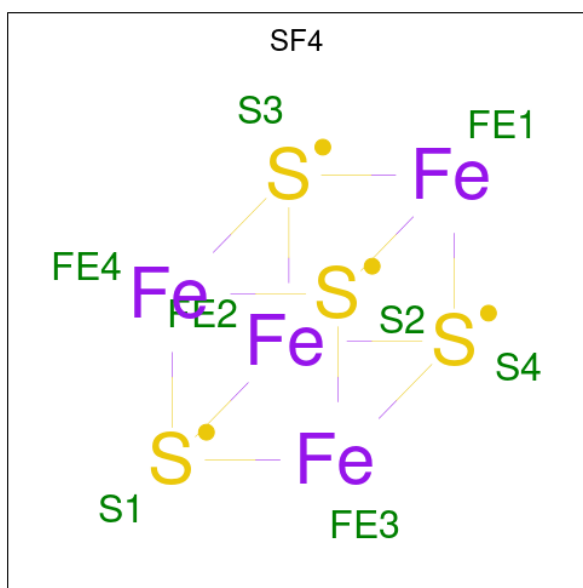
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	S	1	Total	Fe	S	0	0
			8	4	4		
5	T	1	Total	Fe	S	0	0
			8	4	4		

- Molecule 6 is FE3-S4 CLUSTER (three-letter code: F3S) (formula: Fe_3S_4).
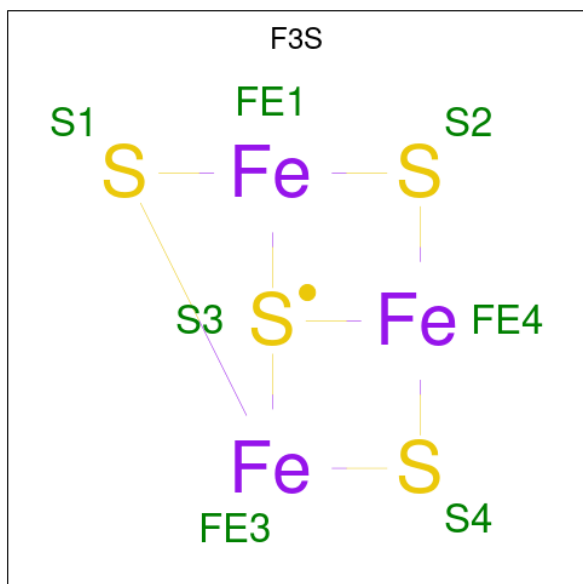
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	S	1	Total	Fe	S	0	0
			7	3	4		

Continued on next page...

Continued from previous page...

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

- Molecule 7 is FE4-S3 CLUSTER (three-letter code: SF3) (formula: Fe_4S_3).
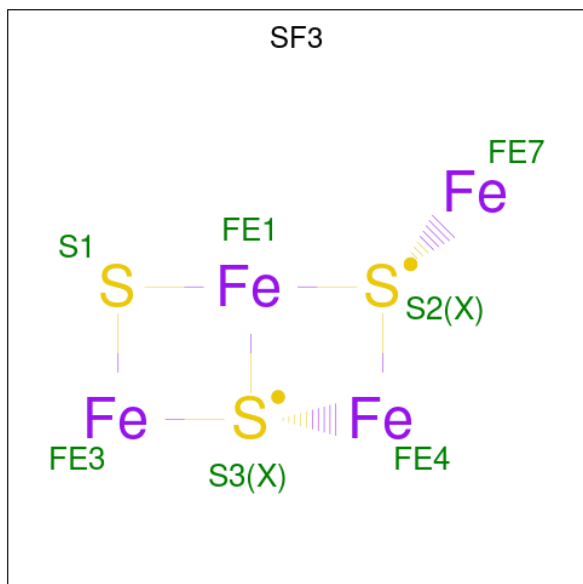
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	S	1	Total 7	Fe 4	S 3	0	0
7	T	1	Total 7	Fe 4	S 3	0	0

- Molecule 8 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	S	1	Total 1	Cl 1	0	0
8	T	1	Total 1	Cl 1	0	0

- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	L	570	Total 570	O 570	0	0
9	M	560	Total 560	O 560	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	S	189	Total 189	O 189	0	0
9	T	173	Total 173	O 173	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.
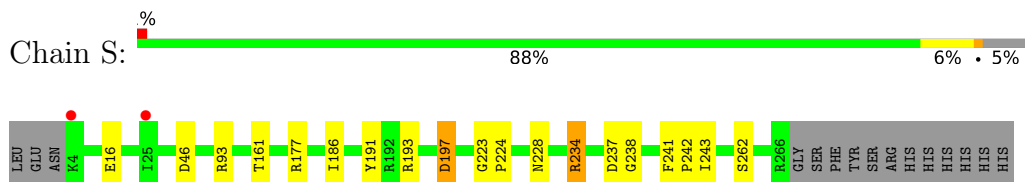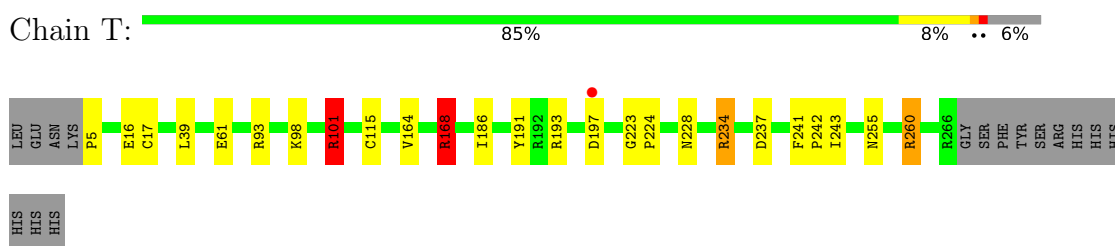
- Molecule 1: HYDROGENASE-1 LARGE CHAIN
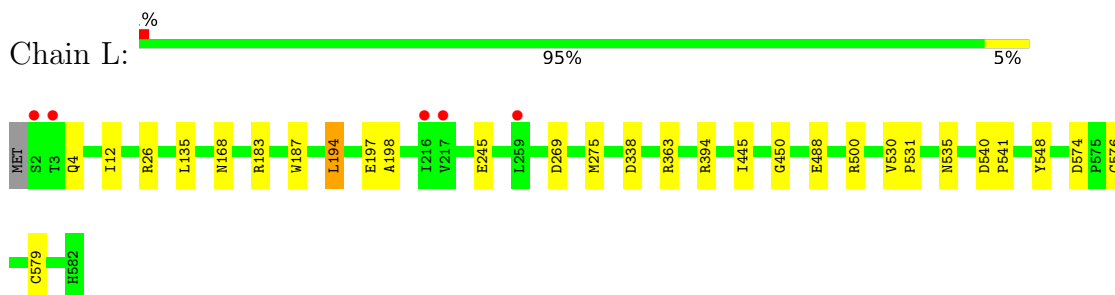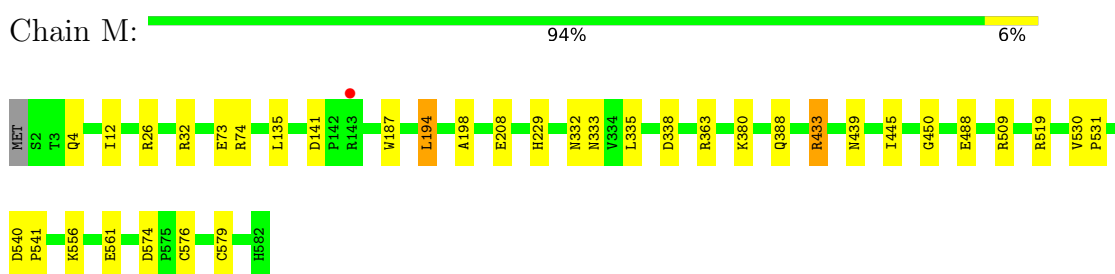
Chain L:

- Molecule 1: HYDROGENASE-1 LARGE CHAIN

Chain M:

- Molecule 2: HYDROGENASE-1 SMALL CHAIN

Chain S:

- Molecule 2: HYDROGENASE-1 SMALL CHAIN

Chain T:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	93.67Å 97.16Å 182.91Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	91.62 – 1.70 46.95 – 1.70	Depositor EDS
% Data completeness (in resolution range)	98.9 (91.62-1.70) 98.9 (46.95-1.70)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.85 (at 1.70Å)	Xtriage
Refinement program	REFMAC 5.8.0123	Depositor
R, R_{free}	0.134 , 0.162 0.149 , 0.174	Depositor DCC
R_{free} test set	8979 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	17.9	Xtriage
Anisotropy	0.076	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 47.1	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.018 for k,h,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	14827	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 5.16% 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: SF4, SF3, MG, F3S, CL, NFV

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	L	0.90	3/4740 (0.1%)	0.94	11/6448 (0.2%)
1	M	0.90	3/4718 (0.1%)	0.93	13/6421 (0.2%)
2	S	0.95	2/2109 (0.1%)	1.00	12/2864 (0.4%)
2	T	0.94	0/2119	1.01	11/2878 (0.4%)
All	All	0.91	8/13686 (0.1%)	0.96	47/18611 (0.3%)

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	L	245	GLU	CD-OE2	7.11	1.33	1.25
1	M	73	GLU	CD-OE1	6.45	1.32	1.25
2	S	262	SER	CB-OG	5.70	1.49	1.42
2	S	197	ASP	CB-CG	5.32	1.62	1.51
1	L	197	GLU	CD-OE1	5.29	1.31	1.25

The worst 5 of 47 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	T	260	ARG	NE-CZ-NH1	13.41	127.00	120.30
2	T	260	ARG	NE-CZ-NH2	-11.41	114.59	120.30
2	S	234	ARG	NE-CZ-NH1	9.95	125.28	120.30
2	T	168	ARG	NE-CZ-NH1	8.86	124.73	120.30
2	T	193	ARG	NE-CZ-NH2	-8.12	116.24	120.30

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	L	4600	0	4497	17	0
1	M	4578	0	4465	18	0
2	S	2045	0	1988	10	0
2	T	2046	0	1986	13	0
3	L	9	0	0	0	0
3	M	9	0	0	1	0
4	L	1	0	0	0	0
4	M	1	0	0	0	0
5	S	8	0	0	0	0
5	T	8	0	0	0	0
6	S	7	0	0	0	0
6	T	7	0	0	0	0
7	S	7	0	0	0	0
7	T	7	0	0	0	0
8	S	1	0	0	0	0
8	T	1	0	0	0	0
9	L	570	0	0	9	0
9	M	560	0	0	6	0
9	S	189	0	0	6	0
9	T	173	0	0	4	0
All	All	14827	0	12936	57	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 57 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:T:5:PRO:N	9:T:2002:HOH:O	2.09	0.86
1:M:208[A]:GLU:OE2	9:M:2192:HOH:O	2.00	0.80
9:L:2511:HOH:O	2:T:168:ARG:NH2	1.95	0.76
2:S:234:ARG:NH2	9:S:2177:HOH:O	2.18	0.76
1:L:579:CYS:CB	9:L:2106:HOH:O	2.37	0.72

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	L	587/582 (101%)	570 (97%)	17 (3%)	0	100	100
1	M	586/582 (101%)	571 (97%)	15 (3%)	0	100	100
2	S	264/278 (95%)	253 (96%)	11 (4%)	0	100	100
2	T	266/278 (96%)	254 (96%)	12 (4%)	0	100	100
All	All	1703/1720 (99%)	1648 (97%)	55 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	L	488/481 (102%)	488 (100%)	0	100	100
1	M	487/481 (101%)	487 (100%)	0	100	100
2	S	218/230 (95%)	215 (99%)	3 (1%)	67	53
2	T	220/230 (96%)	214 (97%)	6 (3%)	44	26
All	All	1413/1422 (99%)	1404 (99%)	9 (1%)	86	80

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	T	191	TYR
2	T	242	PRO
2	T	16	GLU

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Res	Type
2	T	17	CYS
2	T	101	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	L	332	ASN
1	L	392	GLN
1	M	333	ASN
1	M	387	GLN
1	M	439	ASN

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 12 ligands modelled in this entry, 4 are monoatomic - leaving 8 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	NFV	L	601	1	3,8,8	0.91	0	-		
6	F3S	S	402	2	0,9,9	-	-	-		

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
5	SF4	S	401	2	0,12,12	-	-	-		
7	SF3	S	404	9,2	0,8,8	-	-	-		
3	NFV	M	603	1	3,8,8	1.50	1 (33%)	-		
5	SF4	T	401	2	0,12,12	-	-	-		
7	SF3	T	404	9,2	0,8,8	-	-	-		
6	F3S	T	402	2	0,9,9	-	-	-		

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
6	F3S	S	402	2	-	-	0/3/3/3
5	SF4	S	401	2	-	-	0/6/5/5
7	SF3	S	404	9,2	-	-	0/2/2/2
5	SF4	T	401	2	-	-	0/6/5/5
7	SF3	T	404	9,2	-	-	0/2/2/2
6	F3S	T	402	2	-	-	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	M	603	NFV	C3-N3	2.18	1.18	1.13

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	M	603	NFV	1	0

5.7 Other polymers ⓘ

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	L	581/582 (99%)	-0.46	5 (0%)	84	87	13, 19, 32, 51	3 (0%)
1	M	581/582 (99%)	-0.54	1 (0%)	95	95	13, 19, 30, 53	1 (0%)
2	S	263/278 (94%)	-0.40	2 (0%)	86	88	13, 18, 28, 66	1 (0%)
2	T	262/278 (94%)	-0.34	1 (0%)	92	93	14, 19, 33, 44	1 (0%)
All	All	1687/1720 (98%)	-0.46	9 (0%)	91	92	13, 19, 31, 66	6 (0%)

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	L	2	SER	3.1
2	S	4	LYS	2.9
2	T	197	ASP	2.4
1	L	217	VAL	2.4
1	L	216	ILE	2.3

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(Å^2)	Q<0.9
5	SF4	S	401	8/8	0.98	0.05	15,17,18,20	0
5	SF4	T	401	8/8	0.98	0.05	15,16,17,20	0
7	SF3	S	404	7/7	0.98	0.04	19,23,32,34	0
7	SF3	T	404	7/7	0.98	0.04	20,24,30,36	0
8	CL	T	409	1/1	0.98	0.12	32,32,32,32	0
6	F3S	T	402	7/7	0.99	0.03	16,16,17,18	0
4	MG	L	602	1/1	0.99	0.06	13,13,13,13	0
3	NFV	L	601	9/9	0.99	0.06	14,15,19,26	0
8	CL	S	408	1/1	0.99	0.06	29,29,29,29	0
3	NFV	M	603	9/9	0.99	0.06	14,17,19,26	0
6	F3S	S	402	7/7	1.00	0.03	15,15,16,17	0
4	MG	M	604	1/1	1.00	0.06	13,13,13,13	0

6.5 Other polymers ⓘ

There are no such residues in this entry.