Full wwPDB X-ray Structure Validation Report ⓘ

Apr 7, 2022 – 03:04 PM EDT

PDB ID	:	1SFX
Title	:	X-ray crystal structure of putative HTH transcription regulator from Archaeoglobus fulgidus
Authors	:	Osipiuk, J.; Skarina, T.; Savchenko, A.; Edwards, A.; Cymborowski, M.; Minor, W.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2004-02-20
Resolution	:	1.55 Å(reported)

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

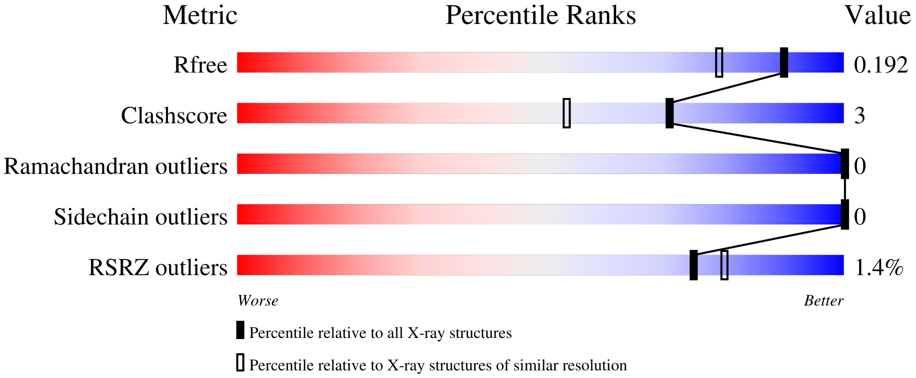
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.27
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

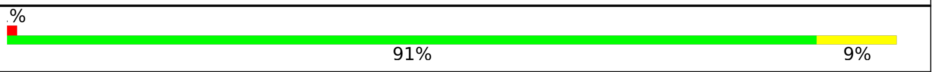
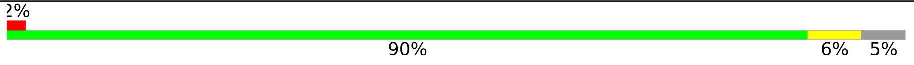
The reported resolution of this entry is 1.55 Å.

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	1483 (1.56-1.56)
Clashscore	141614	1529 (1.56-1.56)
Ramachandran outliers	138981	1498 (1.56-1.56)
Sidechain outliers	138945	1495 (1.56-1.56)
RSRZ outliers	127900	1465 (1.56-1.56)

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	109	
1	B	109	

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 2321 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 Conserved hypothetical protein AF2008.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	109	Total 979	C 631	N 172	O 173	Se 3	0	21	0
1	B	104	Total 925	C 597	N 163	O 163	Se 2	0	18	0

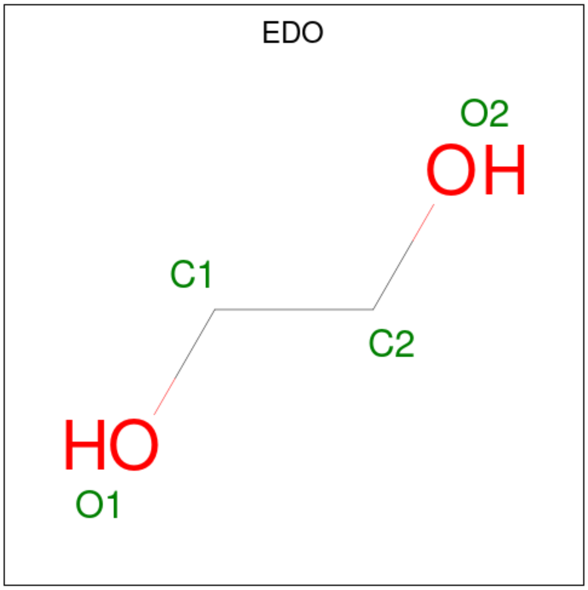
There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	HIS	-	cloning artifact	UNP O28271
A	1	MSE	MET	modified residue	UNP O28271
A	34	MSE	MET	modified residue	UNP O28271
A	103	MSE	MET	modified residue	UNP O28271
A	107	GLY	-	cloning artifact	UNP O28271
A	108	SER	-	cloning artifact	UNP O28271
B	0	HIS	-	cloning artifact	UNP O28271
B	1	MSE	MET	modified residue	UNP O28271
B	34	MSE	MET	modified residue	UNP O28271
B	103	MSE	MET	modified residue	UNP O28271
B	107	GLY	-	cloning artifact	UNP O28271
B	108	SER	-	cloning artifact	UNP O28271

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total 2	Cl 2	0	0
2	B	1	Total 1	Cl 1	0	0

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 8 4 4	0	1
3	B	1	Total C O 4 2 2	0	0

- Molecule 4 is water.

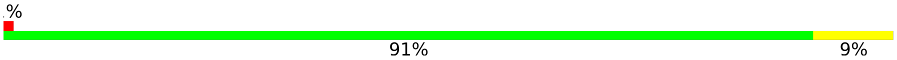
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	180	Total O 180 180	0	0
4	B	217	Total O 218 218	0	1

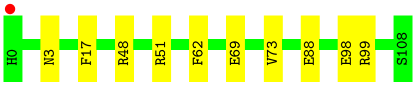
3 Residue-property plots ⓘ

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: Conserved hypothetical protein AF2008

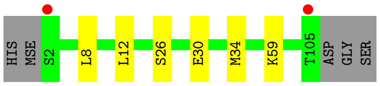
Chain A:





- Molecule 1: Conserved hypothetical protein AF2008

Chain B:



4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	64.27Å 59.63Å 63.50Å 90.00° 90.11° 90.00°	Depositor
Resolution (Å)	40.00 – 1.55 21.73 – 1.55	Depositor EDS
% Data completeness (in resolution range)	97.9 (40.00-1.55) 97.8 (21.73-1.55)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.46 (at 1.55Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.148 , 0.183 0.163 , 0.192	Depositor DCC
R_{free} test set	1710 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	12.0	Xtriage
Anisotropy	0.478	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 49.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.020 for -h,-k,l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	2321	wwPDB-VP
Average B, all atoms (Å^2)	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.62% 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: CL, EDO

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	0.42	0/1072	0.68	0/1415
1	B	0.45	1/1008 (0.1%)	0.63	0/1334
All	All	0.44	1/2080 (0.0%)	0.66	0/2749

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	34	MSE	SE-CE	-7.71	1.50	1.95

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	979	0	1049	9	0
1	B	925	0	977	4	0
2	A	2	0	0	0	0
2	B	1	0	0	0	0
3	A	12	0	18	1	0
3	B	4	0	6	0	0
4	A	180	0	0	4	1
4	B	218	0	0	3	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	2321	0	2050	14	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.

All (14) 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)
3:A:402:EDO:H21	4:A:640:HOH:O	1.95	0.65
1:B:30[A]:GLU:HG2	4:B:602:HOH:O	1.99	0.62
1:A:3:ASN:ND2	1:B:26[B]:SER:OG	2.37	0.58
1:A:48[A]:ARG:CZ	1:A:51[A]:ARG:HE	2.23	0.51
1:A:98[A]:GLU:HG3	4:A:612:HOH:O	2.12	0.49
1:A:69:GLU:HA	1:A:73[A]:VAL:HG12	1.96	0.48
1:A:99[B]:ARG:NH1	4:A:611:HOH:O	2.48	0.46
1:A:17:PHE:HE1	1:A:62:PHE:CE2	2.33	0.46
1:B:8:LEU:O	1:B:12[B]:LEU:HD13	2.17	0.45
1:A:99[B]:ARG:NH2	4:A:670:HOH:O	2.50	0.44
1:B:59[A]:LYS:HD2	4:B:619:HOH:O	2.20	0.40

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)
4:A:565:HOH:O	4:A:635:HOH:O[2_654]	2.19	0.01

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	128/109 (117%)	126 (98%)	2 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	120/109 (110%)	116 (97%)	4 (3%)	0	100	100
All	All	248/218 (114%)	242 (98%)	6 (2%)	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	A	117/94 (124%)	117 (100%)	0	100	100
1	B	110/94 (117%)	110 (100%)	0	100	100
All	All	227/188 (121%)	227 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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 7 ligands modelled in this entry, 3 are monoatomic - leaving 4 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	EDO	A	403[A]	-	3,3,3	0.31	0	2,2,2	0.17	0
3	EDO	A	403[B]	-	3,3,3	0.26	0	2,2,2	0.43	0
3	EDO	B	401	-	3,3,3	0.32	0	2,2,2	0.20	0
3	EDO	A	402	-	3,3,3	0.35	0	2,2,2	0.34	0

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	EDO	A	403[A]	-	-	0/1/1/1	-
3	EDO	A	403[B]	-	-	0/1/1/1	-
3	EDO	B	401	-	-	1/1/1/1	-
3	EDO	A	402	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	401	EDO	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	402	EDO	1	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 ⓘ

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, 95^{th} 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	106/109 (97%)	-0.19	1 (0%)	84	87	7, 11, 21, 25	0
1	B	102/109 (93%)	-0.07	2 (1%)	65	71	7, 11, 25, 28	1 (0%)
All	All	208/218 (95%)	-0.13	3 (1%)	75	80	7, 11, 23, 28	1 (0%)

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	105	THR	3.0
1	B	2	SER	2.4
1	A	0	HIS	2.1

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, 95^{th} 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	EDO	A	403[A]	4/4	0.86	0.22	16,19,20,20	4
3	EDO	A	403[B]	4/4	0.86	0.22	22,23,24,25	4

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	EDO	A	402	4/4	0.94	0.09	13,17,20,21	0
3	EDO	B	401	4/4	0.94	0.08	13,15,18,19	0
2	CL	A	501	1/1	0.99	0.08	20,20,20,20	0
2	CL	A	502	1/1	0.99	0.04	17,17,17,17	1
2	CL	B	503	1/1	1.00	0.06	12,12,12,12	1

6.5 Other polymers ⓘ

There are no such residues in this entry.