WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Nov 2, 2021 – 01:37 AM EDT

PDB ID : 2AZJ

Title : Crystal structure for the mutant D81C of Sulfolobus solfataricus hexaprenyl pyrophosphate synthase

Authors : Sun, H.Y.; Ko, T.P.; Kuo, C.J.; Guo, R.T.; Chou, C.C.; Liang, P.H.; Wang, A.H.J.

Deposited on : 2005-09-11

Resolution : 2.40 Å(reported)

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

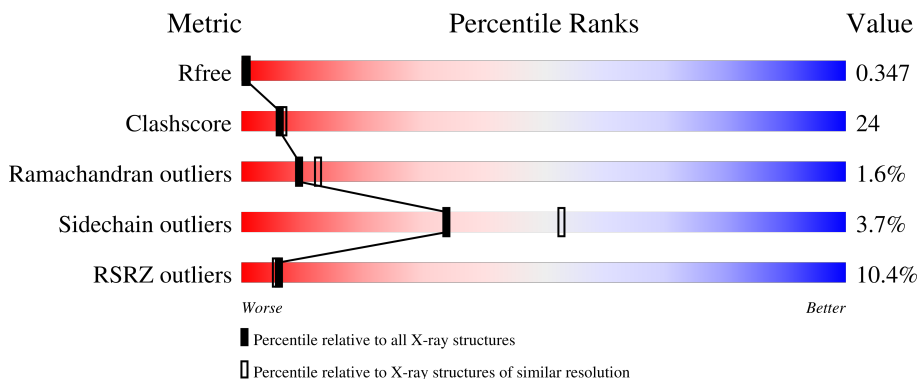|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.40 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	Percentile Ranks	Value
Rfree		0.347
Clashscore		24
Ramachandran outliers		1.6%
Sidechain outliers		3.7%
RSRZ outliers		10.4%

Worse   *Better*

▐ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	289	4% 70% 23% • •
1	B	289	16% 45% 45% • 5%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 4853 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 Geranylgeranyl pyrophosphate synthetase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	276	Total 2237	C 1443	N 363	O 426	S 5	0	0	0
1	B	274	Total 2223	C 1435	N 361	O 423	S 4	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-7	MET	-	expression tag	GB 15899101
A	-6	ALA	-	expression tag	GB 15899101
A	-5	HIS	-	expression tag	GB 15899101
A	-4	HIS	-	expression tag	GB 15899101
A	-3	HIS	-	expression tag	GB 15899101
A	-2	HIS	-	expression tag	GB 15899101
A	-1	HIS	-	expression tag	GB 15899101
A	0	HIS	-	expression tag	GB 15899101
A	81	CYS	ASP	engineered mutation	GB 15899101
B	-7	MET	-	expression tag	GB 15899101
B	-6	ALA	-	expression tag	GB 15899101
B	-5	HIS	-	expression tag	GB 15899101
B	-4	HIS	-	expression tag	GB 15899101
B	-3	HIS	-	expression tag	GB 15899101
B	-2	HIS	-	expression tag	GB 15899101
B	-1	HIS	-	expression tag	GB 15899101
B	0	HIS	-	expression tag	GB 15899101
B	81	CYS	ASP	engineered mutation	GB 15899101

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	216	Total 216	O 216	0	0

Continued on next page...

Continued from previous page...

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	177	Total 177	O 177	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: Geranylgeranyl pyrophosphate synthetase

Chain A:

• Molecule 1: Geranylgeranyl pyrophosphate synthetase

Chain B:
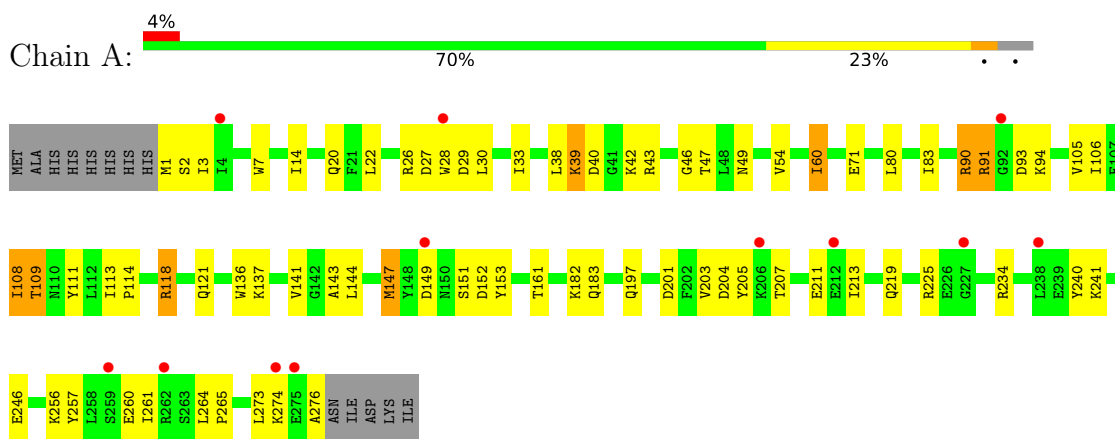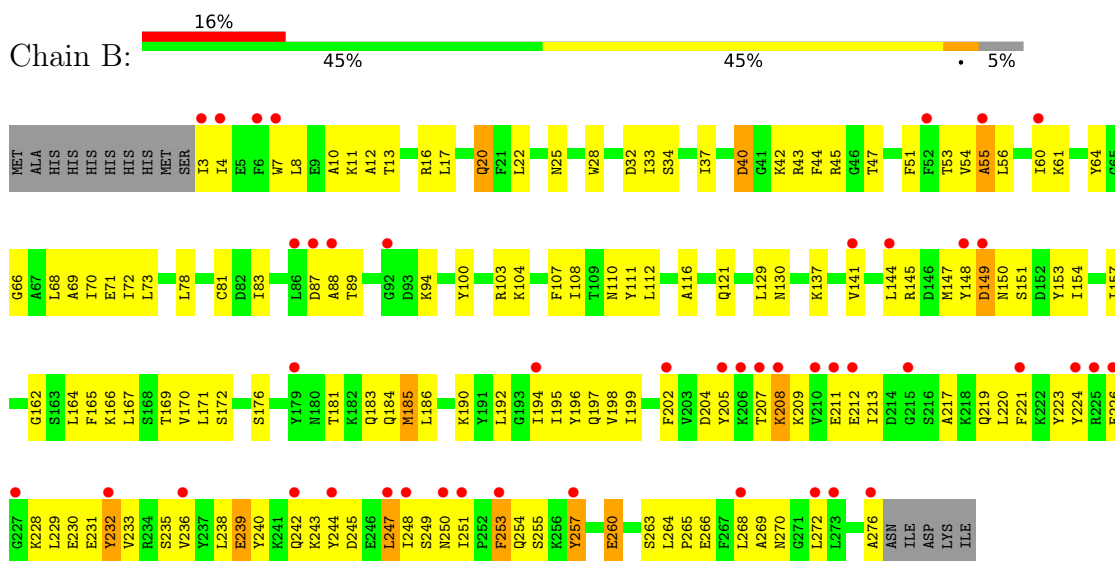
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	91.93Å 91.93Å 121.81Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.37 – 2.40 48.37 – 2.38	Depositor EDS
% Data completeness (in resolution range)	97.3 (48.37-2.40) 97.2 (48.37-2.38)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
$< I/\sigma(I) >$ [1]	5.59 (at 2.39Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.317 , 0.347 0.316 , 0.347	Depositor DCC
R_{free} test set	1120 reflections (4.75%)	wwPDB-VP
Wilson B-factor (Å^2)	41.2	Xtriage
Anisotropy	0.054	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 59.0	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.032 for -h,-k,l	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	4853	wwPDB-VP
Average B, all atoms (Å^2)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.91% 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 ⓘ

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	0.40	0/2279	0.57	0/3077
1	B	0.32	0/2265	0.52	0/3059
All	All	0.36	0/4544	0.55	0/6136

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

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	240	TYR	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	2237	0	2249	73	0
1	B	2223	0	2232	159	2
2	A	216	0	0	7	6

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	177	0	0	10	0
All	All	4853	0	4481	216	7

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 24.

The worst 5 of 216 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:118:ARG:NH2	1:B:130:ASN:HD22	1.50	1.10
1:B:148:TYR:HA	2:B:445:HOH:O	1.59	1.01
1:A:60:ILE:H	1:A:60:ILE:HD12	1.25	0.99
1:B:70:ILE:HD11	1:B:171:LEU:HD12	1.45	0.96
1:A:1:MET:HG3	1:A:2:SER:H	1.32	0.94

The worst 5 of 7 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)
2:A:284:HOH:O	2:A:284:HOH:O[5_555]	0.50	1.70
2:A:283:HOH:O	2:A:283:HOH:O[4_556]	0.75	1.45
2:A:282:HOH:O	2:A:282:HOH:O[4_556]	0.85	1.35
1:B:249:SER:CB	1:B:249:SER:CB[6_555]	2.06	0.14
2:A:355:HOH:O	2:A:480:HOH:O[4_556]	2.07	0.13

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	A	274/289 (95%)	254 (93%)	16 (6%)	4 (2%)	10	14
1	B	272/289 (94%)	218 (80%)	49 (18%)	5 (2%)	8	10

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	546/578 (94%)	472 (86%)	65 (12%)	9 (2%)	9	13

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	27	ASP
1	A	39	LYS
1	B	149	ASP
1	B	213	ILE
1	B	253	PHE

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	242/254 (95%)	233 (96%)	9 (4%)	34	53
1	B	240/254 (94%)	231 (96%)	9 (4%)	33	51
All	All	482/508 (95%)	464 (96%)	18 (4%)	34	53

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	239	GLU
1	B	260	GLU
1	B	257	TYR
1	A	274	LYS
1	B	232	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	242	GLN
1	B	270	ASN
1	B	20	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	25	ASN
1	B	110	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 ⓘ

There are no ligands in this entry.

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	A	276/289 (95%)	0.61	12 (4%)	35	33	20, 37, 68, 88	0
1	B	274/289 (94%)	0.85	45 (16%)	1	1	28, 66, 109, 115	0
All	All	550/578 (95%)	0.73	57 (10%)	6	6	20, 49, 103, 115	0

The worst 5 of 57 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	210	VAL	6.1
1	B	3	ILE	5.1
1	B	215	GLY	4.8
1	A	28	TRP	4.7
1	B	272	LEU	4.2

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 monosaccharides in this entry.

6.4   Ligands ⓘ

There are no ligands in this entry.

6.5   Other polymers ⓘ

There are no such residues in this entry.

WORLDWIDE
wwPDB
PROTEIN DATA BANK