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

wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 06:47 pm BST

PDB ID	:	2CXI
Title	:	Crystal Structure Of An N-terminal Fragment Of The Phenylalanyl-tRNA Synthetase Beta-Subunit From Pyrococcus Horikoshii
Authors	:	Sasaki, H.; Sekine, S.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2005-06-29
Resolution	:	1.94 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	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.94 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		0.231
Clashscore		7
Ramachandran outliers		0.2%
Sidechain outliers		4.5%
RSRZ outliers		0.9%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
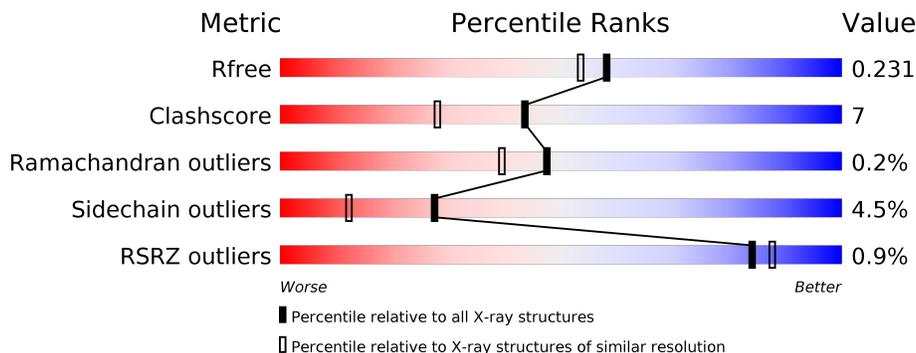
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4310 (1.96-1.92)
Clashscore	141614	1023 (1.94-1.94)
Ramachandran outliers	138981	1007 (1.94-1.94)
Sidechain outliers	138945	1007 (1.94-1.94)
RSRZ outliers	127900	4250 (1.96-1.92)

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 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	348	81% 17% •
1	B	348	83% 14% •
1	C	348	81% 16% •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 9228 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 Phenylalanyl-tRNA synthetase beta chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	347	Total	C	N	O	S	Se	0	2	0
			2848	1833	466	538	1	10			
1	B	347	Total	C	N	O	S	Se	0	2	0
			2848	1833	466	538	1	10			
1	C	347	Total	C	N	O	S	Se	0	2	0
			2848	1833	466	538	1	10			

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP O73984
A	122	MSE	MET	MODIFIED RESIDUE	UNP O73984
A	174	MSE	MET	MODIFIED RESIDUE	UNP O73984
A	212	MSE	MET	MODIFIED RESIDUE	UNP O73984
A	244	MSE	MET	MODIFIED RESIDUE	UNP O73984
A	250	MSE	MET	MODIFIED RESIDUE	UNP O73984
A	310	MSE	MET	MODIFIED RESIDUE	UNP O73984
A	311	MSE	MET	MODIFIED RESIDUE	UNP O73984
A	333	MSE	MET	MODIFIED RESIDUE	UNP O73984
B	1	MSE	MET	MODIFIED RESIDUE	UNP O73984
B	122	MSE	MET	MODIFIED RESIDUE	UNP O73984
B	174	MSE	MET	MODIFIED RESIDUE	UNP O73984
B	212	MSE	MET	MODIFIED RESIDUE	UNP O73984
B	244	MSE	MET	MODIFIED RESIDUE	UNP O73984
B	250	MSE	MET	MODIFIED RESIDUE	UNP O73984
B	310	MSE	MET	MODIFIED RESIDUE	UNP O73984
B	311	MSE	MET	MODIFIED RESIDUE	UNP O73984
B	333	MSE	MET	MODIFIED RESIDUE	UNP O73984
C	1	MSE	MET	MODIFIED RESIDUE	UNP O73984
C	122	MSE	MET	MODIFIED RESIDUE	UNP O73984
C	174	MSE	MET	MODIFIED RESIDUE	UNP O73984
C	212	MSE	MET	MODIFIED RESIDUE	UNP O73984
C	244	MSE	MET	MODIFIED RESIDUE	UNP O73984

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	250	MSE	MET	MODIFIED RESIDUE	UNP O73984
C	310	MSE	MET	MODIFIED RESIDUE	UNP O73984
C	311	MSE	MET	MODIFIED RESIDUE	UNP O73984
C	333	MSE	MET	MODIFIED RESIDUE	UNP O73984

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	249	Total O 249 249	0	0
2	B	208	Total O 208 208	0	0
2	C	227	Total O 227 227	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA 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: Phenylalanyl-tRNA synthetase beta chain
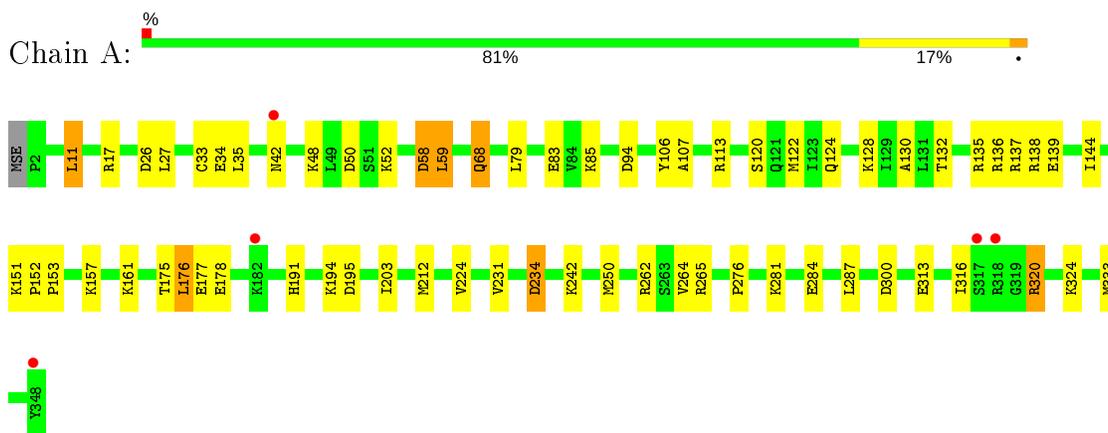
- Molecule 1: Phenylalanyl-tRNA synthetase beta chain
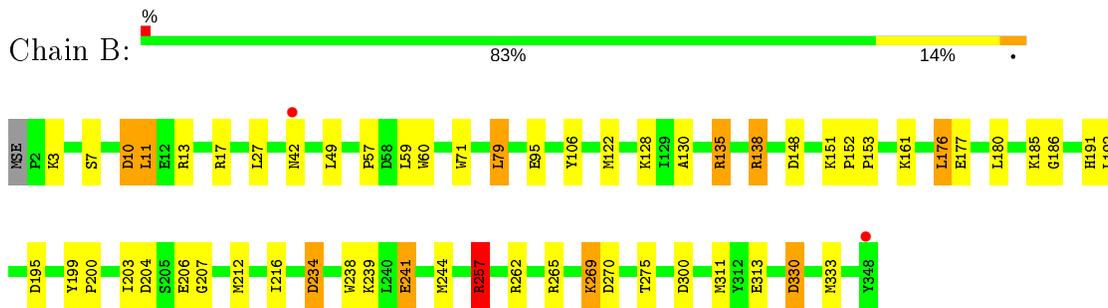
- Molecule 1: Phenylalanyl-tRNA synthetase beta chain
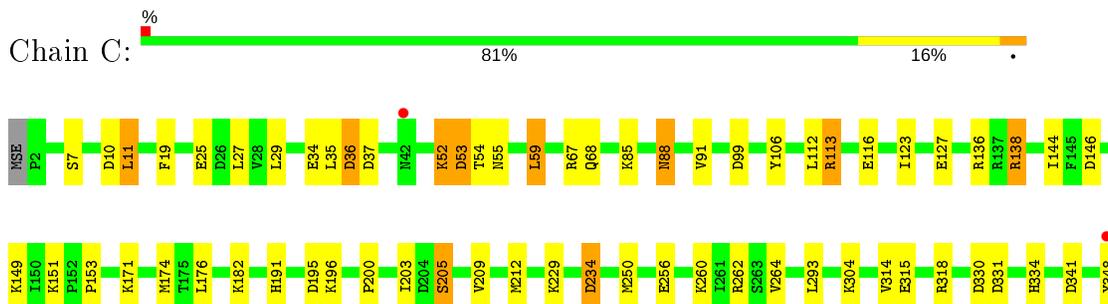
4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	142.78Å 81.97Å 139.10Å 90.00° 116.79° 90.00°	Depositor
Resolution (Å)	41.52 – 1.94 41.39 – 1.94	Depositor EDS
% Data completeness (in resolution range)	99.2 (41.52-1.94) 99.2 (41.39-1.94)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.59 (at 1.94Å)	Xtriage
Refinement program	REFMAC 5.1.9999	Depositor
R, R_{free}	0.188 , 0.231 0.188 , 0.231	Depositor DCC
R_{free} test set	5267 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	25.3	Xtriage
Anisotropy	0.059	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 47.5	EDS
L-test for twinning[2]	$<|L|> = 0.49$, $<L^2> = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	9228	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.55% 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)

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	0.81	0/2905	0.90	7/3895 (0.2%)
1	B	0.79	0/2905	0.92	13/3895 (0.3%)
1	C	0.76	0/2905	0.89	11/3895 (0.3%)
All	All	0.79	0/8715	0.90	31/11685 (0.3%)

There are no bond length outliers.

The worst 5 of 31 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	257	ARG	NE-CZ-NH2	-8.82	115.89	120.30
1	B	257	ARG	NE-CZ-NH1	7.77	124.18	120.30
1	C	341	ASP	CB-CG-OD2	7.42	124.98	118.30
1	B	195	ASP	CB-CG-OD2	7.33	124.89	118.30
1	B	265	ARG	NE-CZ-NH2	-7.07	116.76	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	A	2848	0	2887	48	1
1	B	2848	0	2887	33	1
1	C	2848	0	2887	39	0
2	A	249	0	0	8	0
2	B	208	0	0	5	2

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	227	0	0	11	2
All	All	9228	0	8661	115	3

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 7.

The worst 5 of 115 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:250:MSE:HE3	2:A:490:HOH:O	1.56	1.02
1:B:135:ARG:HH21	1:C:36:ASP:HA	1.41	0.84
1:A:34:GLU:HG3	1:A:52:LYS:NZ	1.97	0.78
1:A:59:LEU:HD11	1:A:68:GLN:NE2	1.98	0.78
1:A:59:LEU:CD1	1:A:68:GLN:HE22	1.98	0.76

All (3) 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:B:552:HOH:O	2:C:575:HOH:O[2_655]	1.59	0.61
2:B:552:HOH:O	2:C:480:HOH:O[2_655]	2.01	0.19
1:A:113:ARG:NH2	1:B:148:ASP:O[3_455]	2.16	0.04

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	347/348 (100%)	337 (97%)	10 (3%)	0	100	100
1	B	347/348 (100%)	336 (97%)	10 (3%)	1 (0%)	41	32
1	C	347/348 (100%)	339 (98%)	7 (2%)	1 (0%)	41	32

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1041/1044 (100%)	1012 (97%)	27 (3%)	2 (0%)	47 39

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	42	ASN
1	C	53	ASP

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	311/301 (103%)	297 (96%)	14 (4%)	27 12
1	B	311/301 (103%)	298 (96%)	13 (4%)	30 14
1	C	311/301 (103%)	295 (95%)	16 (5%)	24 9
All	All	933/903 (103%)	890 (95%)	43 (5%)	27 12

5 of 43 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	79	LEU
1	B	241	GLU
1	C	260	LYS
1	B	122	MSE
1	B	135	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	191	HIS
1	B	208	ASN
1	C	126	GLN
1	B	88	ASN
1	C	88	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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 (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	339/348 (97%)	-0.30	5 (1%)	73	79	13, 23, 34, 50	0
1	B	339/348 (97%)	-0.34	2 (0%)	89	92	13, 25, 36, 41	0
1	C	339/348 (97%)	-0.27	2 (0%)	89	92	15, 26, 39, 49	0
All	All	1017/1044 (97%)	-0.30	9 (0%)	84	87	13, 25, 37, 50	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	318	ARG	3.9
1	C	348	TYR	3.4
1	C	42	ASN	2.8
1	B	42	ASN	2.5
1	A	317	SER	2.4

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.