Full wwPDB X-ray Structure Validation Report (i)

Nov 20, 2023 – 11:33 PM JST

PDB ID	:	7E05
Title	:	Trans-3/4-proline-hydroxylase H11 apo structure
Authors	:	Gong, W.G.; Yang, L.Y.
Deposited on	:	2021-01-27
Resolution	:	1.88 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	:	2.36
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
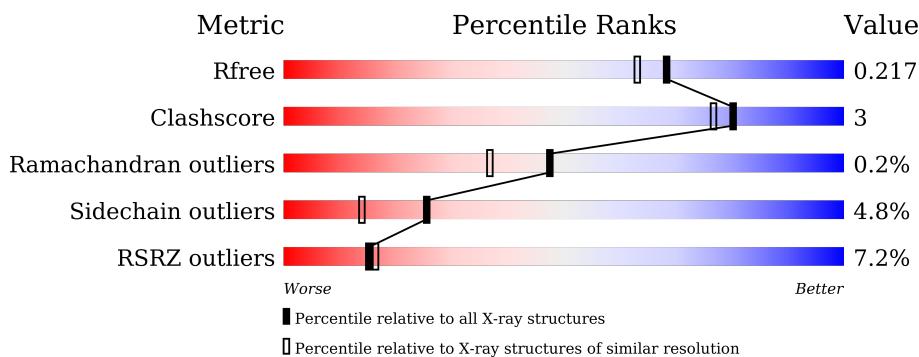
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
*X-RAY DIFFRACTION*
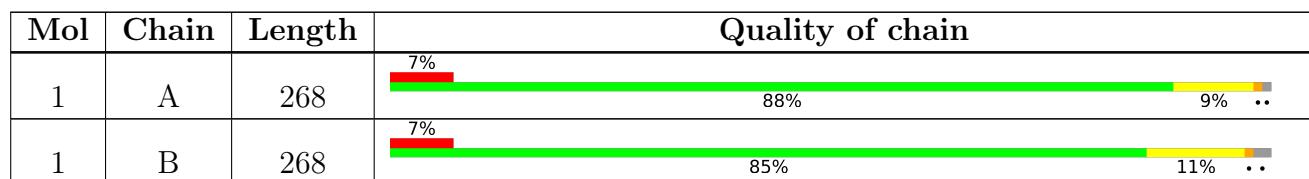
The reported resolution of this entry is 1.88 Å.

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		0.217
Clashscore		3
Ramachandran outliers		0.2%
Sidechain outliers		4.8%
RSRZ outliers		7.2%

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	9470 (1.90-1.86)
Clashscore	141614	10282 (1.90-1.86)
Ramachandran outliers	138981	10152 (1.90-1.86)
Sidechain outliers	138945	10152 (1.90-1.86)
RSRZ outliers	127900	9303 (1.90-1.86)

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	268	7% / 88% / 9% ••
1	B	268	7% / 85% / 11% ••

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 4756 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 Phytanoyl-CoA dioxygenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	264	Total 2101	C 1333	N 373	O 389	S 6	0	1	0
1	B	262	Total 2102	C 1333	N 375	O 388	S 6	0	2	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLN	-	expression tag	UNP S5TUM1
A	0	ILE	-	expression tag	UNP S5TUM1
B	-1	GLN	-	expression tag	UNP S5TUM1
B	0	ILE	-	expression tag	UNP S5TUM1
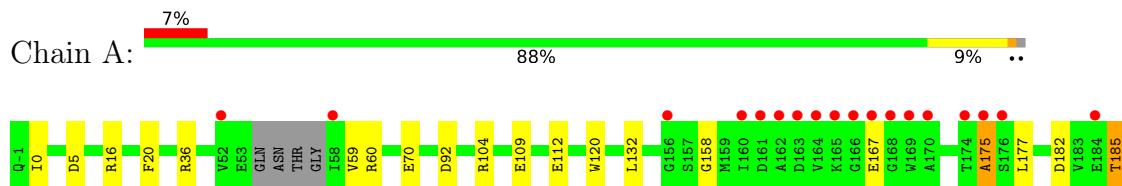
- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	284	Total 284	O 284	0	0
2	B	269	Total 269	O 269	0	0
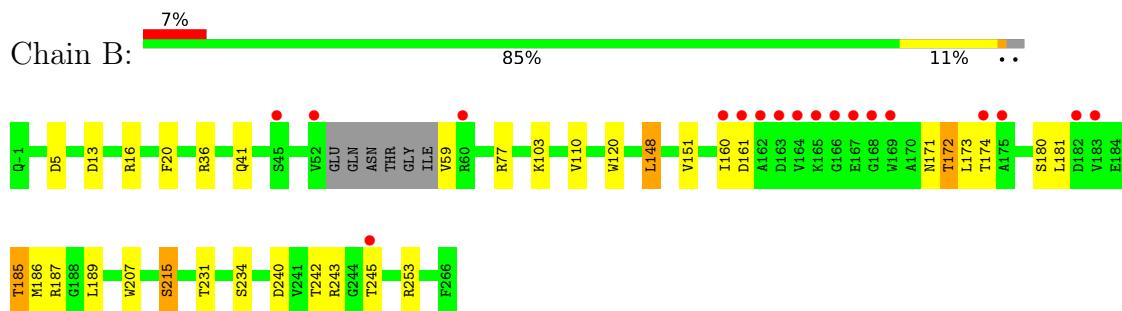
3 Residue-property plots ⓘ

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: Phytanoyl-CoA dioxygenase

Chain A:

- Molecule 1: Phytanoyl-CoA dioxygenase

Chain B:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	106.50Å 106.50Å 143.70Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.78 – 1.88 42.78 – 1.88	Depositor EDS
% Data completeness (in resolution range)	99.7 (42.78-1.88) 99.8 (42.78-1.88)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.66 (at 1.88Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.171 , 0.211 0.183 , 0.217	Depositor DCC
R_{free} test set	3381 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	22.4	Xtriage
Anisotropy	0.045	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 46.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4756	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 20.90 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 7.8559e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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	1.12	0/2156	1.06	9/2929 (0.3%)
1	B	1.11	0/2157	1.12	11/2929 (0.4%)
All	All	1.11	0/4313	1.09	20/5858 (0.3%)

There are no bond length outliers.

All (20) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	36[A]	ARG	NE-CZ-NH2	-9.93	115.33	120.30
1	B	36[B]	ARG	NE-CZ-NH2	-9.93	115.33	120.30
1	B	148	LEU	CB-CG-CD1	8.21	124.96	111.00
1	A	253	ARG	NE-CZ-NH1	7.20	123.90	120.30
1	A	243	ARG	NE-CZ-NH2	6.92	123.76	120.30
1	A	36	ARG	NE-CZ-NH1	6.79	123.70	120.30
1	B	16	ARG	NE-CZ-NH2	-6.74	116.93	120.30
1	A	92	ASP	CB-CG-OD2	-6.27	112.66	118.30
1	B	77	ARG	NE-CZ-NH2	-6.12	117.24	120.30
1	B	16	ARG	NE-CZ-NH1	6.08	123.34	120.30
1	B	13	ASP	CB-CG-OD1	6.04	123.73	118.30
1	A	224	ASP	CB-CG-OD2	-5.59	113.27	118.30
1	A	104	ARG	NE-CZ-NH1	5.57	123.09	120.30
1	A	16	ARG	NE-CZ-NH1	5.57	123.08	120.30
1	B	187	ARG	NE-CZ-NH2	-5.39	117.61	120.30
1	A	224	ASP	CB-CG-OD1	5.32	123.08	118.30
1	A	70	GLU	OE1-CD-OE2	5.30	129.66	123.30
1	B	77	ARG	NE-CZ-NH1	5.30	122.95	120.30
1	B	215	SER	CA-CB-OG	-5.12	97.37	111.20
1	B	243	ARG	NE-CZ-NH2	5.09	122.84	120.30

There are no chirality outliers.

There are no planarity outliers.

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	2101	0	2028	9	0
1	B	2102	0	2036	13	0
2	A	284	0	0	2	0
2	B	269	0	0	2	0
All	All	4756	0	4064	22	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 3.

All (22) 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:182:ASP:HB3	1:A:185:THR:OG1	1.95	0.66
1:A:158:GLY:C	1:A:211:ASN:HD21	2.02	0.63
1:B:171:ASN:OD1	1:B:173:LEU:N	2.28	0.63
1:B:253:ARG:NH1	2:B:303:HOH:O	2.32	0.63
1:A:112:GLU:OE1	1:A:175:ALA:O	2.19	0.60
1:B:59:VAL:O	1:B:59:VAL:HG23	2.02	0.59
1:B:172:THR:HG21	2:B:389:HOH:O	2.04	0.57
1:B:240:ASP:OD1	1:B:242:THR:HB	2.06	0.56
1:B:181:LEU:HD13	1:B:186:MET:CE	2.36	0.56
1:B:160:ILE:HG21	1:B:185:THR:HG21	1.87	0.56
1:A:59:VAL:O	1:A:59:VAL:HG23	2.05	0.55
1:B:160:ILE:CG2	1:B:185:THR:HG21	2.39	0.53
1:A:132:LEU:HD12	1:A:132:LEU:C	2.35	0.47
1:B:181:LEU:HD13	1:B:186:MET:HE3	1.95	0.47
1:A:0:ILE:HG23	2:A:831:HOH:O	2.16	0.46
1:B:181:LEU:HD13	1:B:186:MET:HE1	1.99	0.43
1:B:120:TRP:CH2	1:B:231:THR:HG21	2.53	0.43
1:B:20:PHE:HA	1:B:207:TRP:O	2.19	0.43
1:A:20:PHE:HA	1:A:207:TRP:O	2.19	0.43
1:B:151:VAL:HG11	1:B:189:LEU:HD22	2.02	0.41
1:A:253:ARG:NH2	2:A:816:HOH:O	2.51	0.40
1:A:120:TRP:CH2	1:A:231:THR:HG21	2.57	0.40

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	A	261/268 (97%)	255 (98%)	5 (2%)	1 (0%)	34	22
1	B	260/268 (97%)	256 (98%)	4 (2%)	0	100	100
All	All	521/536 (97%)	511 (98%)	9 (2%)	1 (0%)	47	37

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	175	ALA

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	221/225 (98%)	213 (96%)	8 (4%)	35	23
1	B	222/225 (99%)	209 (94%)	13 (6%)	19	9
All	All	443/450 (98%)	422 (95%)	21 (5%)	25	14

All (21) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	5	ASP
1	A	60	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	109	GLU
1	A	167	GLU
1	A	177	LEU
1	A	185	THR
1	A	189	LEU
1	A	245	THR
1	B	5	ASP
1	B	41	GLN
1	B	103	LYS
1	B	110	VAL
1	B	148	LEU
1	B	161	ASP
1	B	172	THR
1	B	174	THR
1	B	180	SER
1	B	185	THR
1	B	215	SER
1	B	234	SER
1	B	245	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	97	GLN
1	A	121	HIS
1	A	211	ASN
1	B	97	GLN
1	B	121	HIS
1	B	193	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	264/268 (98%)	0.25	20 (7%)	13	15	14, 24, 73, 124	0
1	B	262/268 (97%)	0.14	18 (6%)	16	18	14, 25, 68, 105	0
All	All	526/536 (98%)	0.19	38 (7%)	15	16	14, 25, 72, 124	0

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	166	GLY	7.7
1	A	164	VAL	7.3
1	A	169	TRP	7.2
1	A	167	GLU	6.7
1	B	175	ALA	6.0
1	B	167	GLU	5.6
1	A	165	LYS	5.3
1	A	168	GLY	5.1
1	B	164	VAL	4.7
1	A	170	ALA	4.5
1	A	160	ILE	4.5
1	A	161	ASP	4.4
1	B	163	ASP	4.4
1	B	165	LYS	4.4
1	A	162	ALA	4.2
1	B	174	THR	3.9
1	A	175	ALA	3.9
1	A	174	THR	3.9
1	A	176	SER	3.9
1	B	169	TRP	3.9
1	A	163	ASP	3.8
1	B	183	VAL	3.8
1	B	52	VAL	3.8
1	B	168	GLY	3.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	245	THR	3.2
1	A	156	GLY	3.1
1	B	166	GLY	3.0
1	B	182	ASP	2.9
1	B	161	ASP	2.6
1	A	52	VAL	2.5
1	A	58	ILE	2.4
1	B	60	ARG	2.3
1	B	160	ILE	2.3
1	B	45	SER	2.3
1	A	243	ARG	2.2
1	A	184	GLU	2.2
1	A	245	THR	2.2
1	B	162	ALA	2.2

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.