



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 4, 2026 – 09:59 PM UTC

PDB ID : 1PQY / pdb_00001pqy
Title : Crystal structure of formyl-coA transferase yfdW from E. coli
Authors : Gogos, A.; Gorman, J.; Shapiro, L.; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on : 2003-06-19
Resolution : 2.35 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

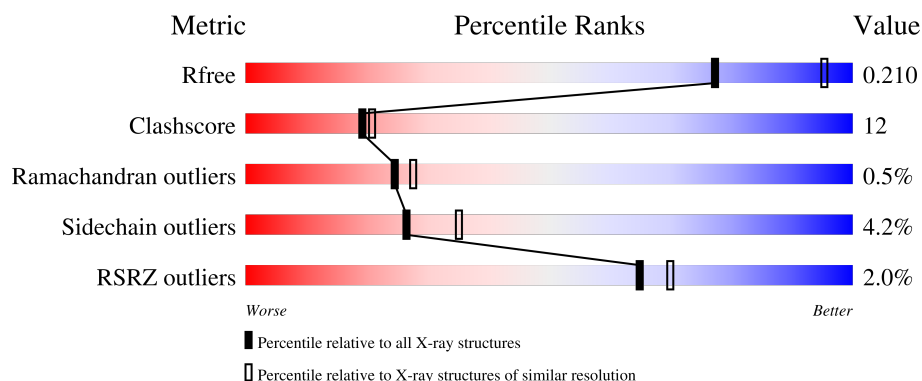
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

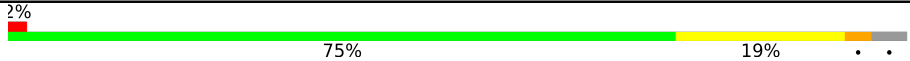
The reported resolution of this entry is 2.35 Å.

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}	180053	1596 (2.36-2.36)
Clashscore	190562	1663 (2.36-2.36)
Ramachandran outliers	187476	1646 (2.36-2.36)
Sidechain outliers	187428	1646 (2.36-2.36)
RSRZ outliers	180081	1598 (2.36-2.36)

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 $\leq 5\%$. 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	428	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3423 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 Hypothetical protein yfdW.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	412	Total	C	N	O	S	Se	0	0	0
			3163	2009	536	602	7	9			

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	cloning artifact	UNP P69902
A	2	SER	-	cloning artifact	UNP P69902
A	3	LEU	-	cloning artifact	UNP P69902
A	27	MSE	MET	modified residue	UNP P69902
A	64	MSE	MET	modified residue	UNP P69902
A	85	MSE	MET	modified residue	UNP P69902
A	107	MSE	MET	modified residue	UNP P69902
A	176	MSE	MET	modified residue	UNP P69902
A	200	MSE	MET	modified residue	UNP P69902
A	202	MSE	MET	modified residue	UNP P69902
A	343	MSE	MET	modified residue	UNP P69902
A	376	MSE	MET	modified residue	UNP P69902
A	412	MSE	MET	modified residue	UNP P69902
A	419	GLU	-	expression tag	UNP P69902
A	420	GLY	-	expression tag	UNP P69902
A	421	GLY	-	expression tag	UNP P69902
A	422	SER	-	expression tag	UNP P69902
A	423	HIS	-	expression tag	UNP P69902
A	424	HIS	-	expression tag	UNP P69902
A	425	HIS	-	expression tag	UNP P69902
A	426	HIS	-	expression tag	UNP P69902
A	427	HIS	-	expression tag	UNP P69902
A	428	HIS	-	expression tag	UNP P69902

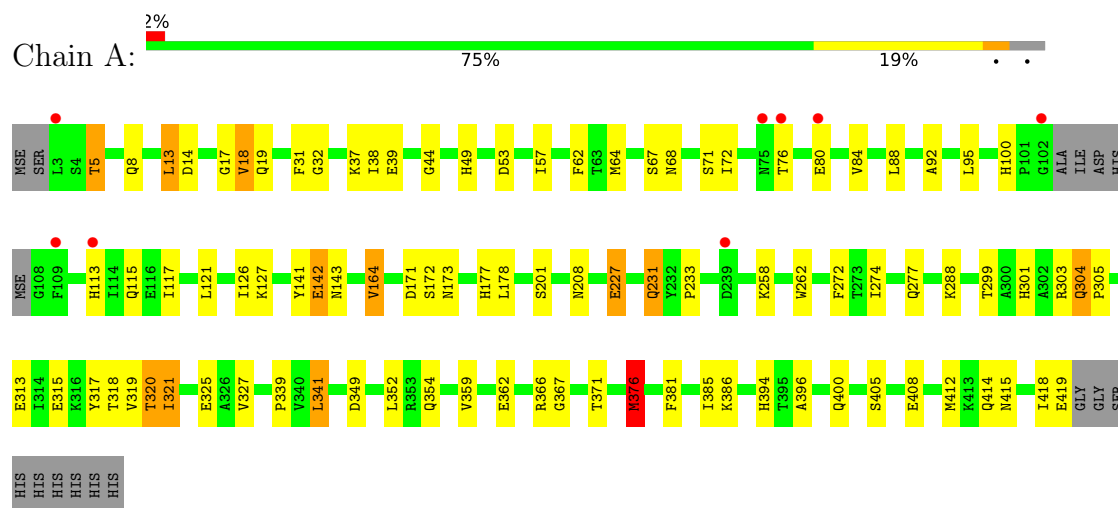
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	260	Total 260	O 260	0	0

3 Residue-property plots [i](#)

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: Hypothetical protein yfdW



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	92.26Å 69.54Å 73.11Å 90.00° 108.64° 90.00°	Depositor
Resolution (Å)	27.99 – 2.35 27.99 – 2.35	Depositor EDS
% Data completeness (in resolution range)	98.3 (27.99-2.35) 98.2 (27.99-2.35)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	0.09	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.69 (at 2.34Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.177 , 0.211 0.177 , 0.210	Depositor DCC
R_{free} test set	877 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å ²)	18.7	Xtriage
Anisotropy	0.360	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 51.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	3423	wwPDB-VP
Average B, all atoms (Å ²)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.41% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	0.52	1/3227 (0.0%)	1.01	13/4370 (0.3%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	376	MSE	SE-CE	-10.55	1.63	1.95

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	304	GLN	OE1-CD-NE2	-10.25	112.35	122.60
1	A	414	GLN	OE1-CD-NE2	-9.88	112.72	122.60
1	A	115	GLN	OE1-CD-NE2	-9.86	112.74	122.60
1	A	354	GLN	OE1-CD-NE2	-9.80	112.80	122.60
1	A	304	GLN	CG-CD-NE2	7.37	127.45	116.40
1	A	354	GLN	CG-CD-NE2	7.33	127.40	116.40
1	A	115	GLN	CG-CD-NE2	6.68	126.42	116.40
1	A	414	GLN	CG-CD-NE2	6.12	125.58	116.40
1	A	53	ASP	N-CA-C	-5.43	106.66	113.72
1	A	92	ALA	N-CA-C	5.34	118.39	110.48
1	A	14	ASP	N-CA-C	5.27	117.32	107.99
1	A	227	GLU	N-CA-C	5.22	117.64	111.33
1	A	415	ASN	CA-CB-CG	5.15	117.75	112.60

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	3163	0	3078	73	0
2	A	260	0	0	1	0
All	All	3423	0	3078	73	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 12.

All (73) 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:113:HIS:CE1	1:A:117:ILE:HD11	1.81	1.15
1:A:5:THR:HG21	1:A:32:GLY:HA3	1.33	1.05
1:A:113:HIS:NE2	1:A:117:ILE:HD11	1.80	0.95
1:A:80:GLU:HB3	1:A:412:MSE:HE2	1.54	0.89
1:A:5:THR:CG2	1:A:32:GLY:HA3	2.03	0.88
1:A:231:GLN:H	1:A:231:GLN:HE21	1.22	0.83
1:A:113:HIS:CE1	1:A:117:ILE:CD1	2.65	0.77
1:A:177:HIS:HE1	1:A:208:ASN:HD22	1.33	0.76
1:A:231:GLN:H	1:A:231:GLN:NE2	1.87	0.73
1:A:327:VAL:HG21	1:A:339:PRO:HD3	1.70	0.72
1:A:5:THR:HG23	1:A:31:PHE:O	1.91	0.70
1:A:68:ASN:HD21	1:A:386:LYS:H	1.37	0.70
1:A:13:LEU:HD21	1:A:88:LEU:HB3	1.76	0.68
1:A:339:PRO:HG2	1:A:341:LEU:HD13	1.74	0.68
1:A:304:GLN:HB3	1:A:305:PRO:HD3	1.76	0.67
1:A:64:MSE:HE3	1:A:371:THR:HA	1.78	0.65
1:A:80:GLU:HB3	1:A:412:MSE:CE	2.27	0.65
1:A:321:ILE:H	1:A:321:ILE:HD13	1.61	0.65
1:A:5:THR:HG22	1:A:8:GLN:N	2.12	0.64
1:A:320:THR:HG22	1:A:321:ILE:HG23	1.80	0.63
1:A:272:PHE:HE2	1:A:274:ILE:HD13	1.64	0.62
1:A:18:VAL:HG12	1:A:19:GLN:N	2.15	0.60
1:A:18:VAL:HG12	1:A:19:GLN:H	1.66	0.60
1:A:71:SER:H	1:A:394:HIS:HD2	1.47	0.60
1:A:39:GLU:HB3	1:A:44:GLY:HA2	1.82	0.60
1:A:317:TYR:O	1:A:320:THR:HB	2.00	0.60
1:A:177:HIS:HE1	1:A:208:ASN:ND2	2.02	0.57
1:A:349:ASP:OD2	1:A:352:LEU:HG	2.04	0.57
1:A:177:HIS:CE1	1:A:208:ASN:HD22	2.21	0.56

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:80:GLU:OE1	1:A:80:GLU:N	2.35	0.56
1:A:173:ASN:O	1:A:177:HIS:HD2	1.89	0.56
1:A:68:ASN:ND2	1:A:386:LYS:H	2.03	0.56
1:A:126:ILE:HG13	1:A:178:LEU:HD22	1.88	0.54
1:A:100:HIS:HB3	2:A:678:HOH:O	2.09	0.53
1:A:5:THR:HG21	1:A:32:GLY:CA	2.23	0.53
1:A:5:THR:HG22	1:A:8:GLN:H	1.74	0.52
1:A:299:THR:HG22	1:A:301:HIS:H	1.73	0.52
1:A:376:MSE:HE3	1:A:376:MSE:H	1.75	0.52
1:A:71:SER:H	1:A:394:HIS:CD2	2.27	0.51
1:A:37:LYS:HE2	1:A:62:PHE:HE1	1.75	0.51
1:A:272:PHE:CE2	1:A:274:ILE:HD13	2.45	0.51
1:A:408:GLU:O	1:A:412:MSE:HG3	2.11	0.49
1:A:113:HIS:NE2	1:A:117:ILE:CD1	2.66	0.49
1:A:231:GLN:HE21	1:A:231:GLN:N	2.00	0.49
1:A:321:ILE:HD13	1:A:321:ILE:N	2.28	0.48
1:A:299:THR:HG22	1:A:301:HIS:N	2.28	0.47
1:A:405:SER:OG	1:A:408:GLU:HG3	2.14	0.47
1:A:17:GLY:O	1:A:18:VAL:HB	2.13	0.47
1:A:339:PRO:HG2	1:A:341:LEU:CD1	2.43	0.47
1:A:143:ASN:CG	1:A:164:VAL:HG21	2.40	0.47
1:A:231:GLN:NE2	1:A:231:GLN:N	2.60	0.46
1:A:95:LEU:C	1:A:95:LEU:HD23	2.41	0.46
1:A:171:ASP:O	1:A:172:SER:C	2.57	0.44
1:A:67:SER:C	1:A:68:ASN:HD22	2.25	0.44
1:A:127:LYS:O	1:A:201:SER:HA	2.18	0.44
1:A:299:THR:O	1:A:303:ARG:HG3	2.18	0.44
1:A:80:GLU:O	1:A:84:VAL:HG23	2.17	0.44
1:A:315:GLU:HA	1:A:318:THR:OG1	2.17	0.44
1:A:258:LYS:HB2	1:A:319:VAL:HG12	2.00	0.43
1:A:57:ILE:HD12	1:A:366:ARG:HB2	2.00	0.43
1:A:396:ALA:O	1:A:400:GLN:HG3	2.19	0.42
1:A:84:VAL:CG2	1:A:412:MSE:HE1	2.49	0.42
1:A:359:VAL:HG21	1:A:385:ILE:HD12	2.01	0.42
1:A:141:TYR:O	1:A:142:GLU:C	2.62	0.42
1:A:38:ILE:HG12	1:A:72:ILE:HD11	2.02	0.42
1:A:121:LEU:C	1:A:121:LEU:HD23	2.45	0.42
1:A:362:GLU:OE2	1:A:367:GLY:HA2	2.20	0.41
1:A:67:SER:HB3	1:A:68:ASN:HD22	1.85	0.41
1:A:288:LYS:HE3	1:A:313:GLU:HG2	2.01	0.41
1:A:49:HIS:CD2	1:A:49:HIS:C	2.99	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:418:ILE:O	1:A:419:GLU:C	2.64	0.41
1:A:258:LYS:HD3	1:A:262:TRP:CE2	2.56	0.41
1:A:84:VAL:HG23	1:A:412:MSE:HE1	2.03	0.40

There are no symmetry-related clashes.

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	408/428 (95%)	399 (98%)	7 (2%)	2 (0%)	24 27

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	18	VAL
1	A	142	GLU

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	330/341 (97%)	316 (96%)	14 (4%)	26 35

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	5	THR
1	A	13	LEU
1	A	76	THR
1	A	164	VAL
1	A	227	GLU
1	A	231	GLN
1	A	233	PRO
1	A	277	GLN
1	A	320	THR
1	A	321	ILE
1	A	325	GLU
1	A	341	LEU
1	A	376	MSE
1	A	381	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	50	GLN
1	A	68	ASN
1	A	177	HIS
1	A	188	HIS
1	A	196	GLN
1	A	231	GLN
1	A	277	GLN
1	A	281	ASN
1	A	332	GLN
1	A	394	HIS

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 oligosaccharides 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 [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	403/428 (94%)	-0.23	8 (1%) 65 70	11, 18, 35, 56	0

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	3	LEU	4.7
1	A	80	GLU	3.0
1	A	76	THR	2.9
1	A	102	GLY	2.6
1	A	113	HIS	2.6
1	A	109	PHE	2.6
1	A	239	ASP	2.3
1	A	75	ASN	2.0

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 oligosaccharides 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.