



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 14, 2018 – 09:33 am GMT

PDB ID : 5Y1I
Title : The crystal structure of GfsF
Authors : Miyanaga, A.; Kudo, F.; Eguchi, T.
Deposited on : 2017-07-20
Resolution : 2.00 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix) : 1.13
EDS : trunk31020
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk31020

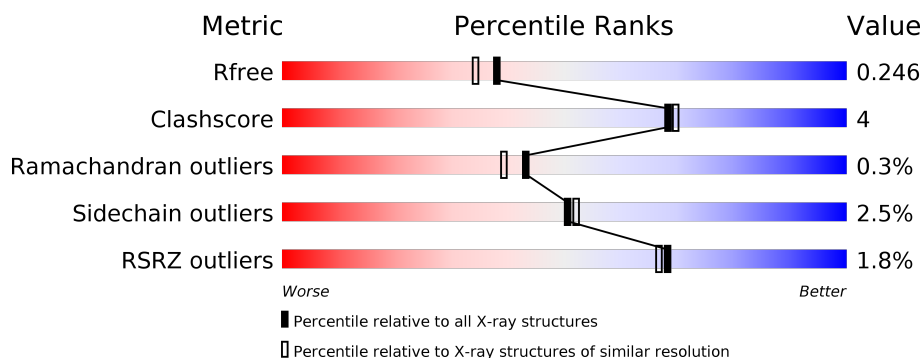
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 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}	111664	7193 (2.00-2.00)
Clashscore	122126	8267 (2.00-2.00)
Ramachandran outliers	120053	8166 (2.00-2.00)
Sidechain outliers	120020	8165 (2.00-2.00)
RSRZ outliers	108989	7011 (2.00-2.00)

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. 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	403	<div> <div>2%</div> <div> <div></div> <div>85%</div> <div>9%</div> <div>• •</div> </div> </div>
1	B	403	<div> <div>%</div> <div> <div></div> <div>85%</div> <div>12%</div> <div>• •</div> </div> </div>

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6561 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.

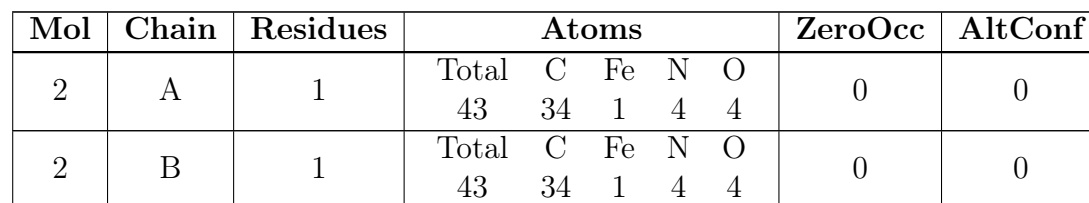
- Molecule 1 is a protein called Cytochrome P450.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	388	Total	C	N	O	S	0	0	0
			3007	1903	529	562	13			
1	B	394	Total	C	N	O	S	0	0	0
			3052	1929	539	571	13			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	12	GLY	-	expression tag	UNP E0D207
A	13	SER	-	expression tag	UNP E0D207
A	14	HIS	-	expression tag	UNP E0D207
A	15	MET	-	expression tag	UNP E0D207
B	12	GLY	-	expression tag	UNP E0D207
B	13	SER	-	expression tag	UNP E0D207
B	14	HIS	-	expression tag	UNP E0D207
B	15	MET	-	expression tag	UNP E0D207

- Molecule 2 is PROTOPORPHYRIN IX CONTAINING FE (three-letter code: HEM) (formula: $C_{34}H_{32}FeN_4O_4$).



- PEG
-
- The diagram illustrates the chemical structure of Polyethylene Glycol (PEG). It shows a chain of carbon (C) and oxygen (O) atoms. The terminal groups are hydroxyl (OH) groups, labeled O1 and O4. The repeating unit consists of a carbon atom bonded to two hydrogen atoms (H) and an oxygen atom, which is then bonded to another carbon atom. The carbon atoms are labeled C1, C2, C3, and C4. The oxygen atoms are labeled O2 and O4. The hydroxyl groups are labeled HO and OH.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 7 4 3	0	0
3	B	1	Total C O 7 4 3	0	0

- Molecule 4 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	C	O	0	0
			10	6	4		

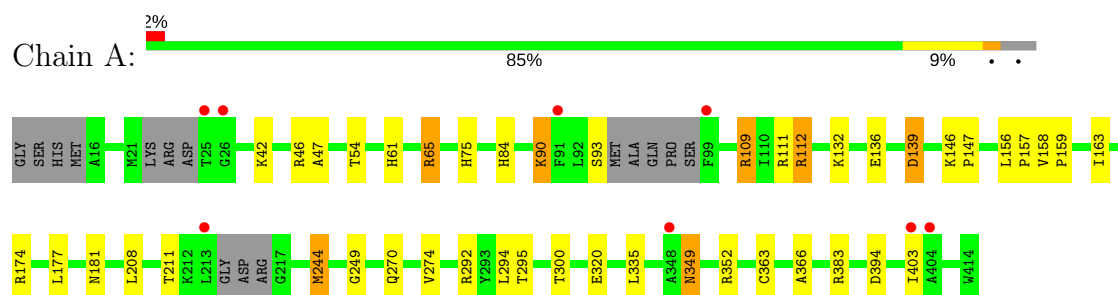
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	185	Total	O	0	0
			185	185		
5	B	207	Total	O	0	0
			207	207		

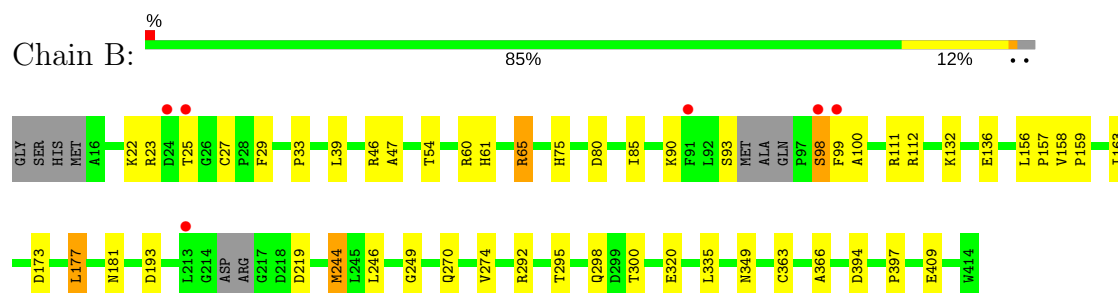
3 Residue-property plots [i](#)

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: Cytochrome P450



• Molecule 1: Cytochrome P450



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	67.08Å 102.87Å 114.43Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.49 – 2.00 50.00 – 2.00	Depositor EDS
% Data completeness (in resolution range)	93.5 (50.49-2.00) 93.5 (50.00-2.00)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.95 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.194 , 0.236 0.201 , 0.246	Depositor DCC
R_{free} test set	2589 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å ²)	23.8	Xtriage
Anisotropy	0.170	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 39.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	6561	wwPDB-VP
Average B, all atoms (Å ²)	27.0	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 43.02 % 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 1.8727e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: HEM, PEG, PGE

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	0.83	0/3068	0.96	10/4175 (0.2%)
1	B	0.83	0/3115	0.93	7/4238 (0.2%)
All	All	0.83	0/6183	0.95	17/8413 (0.2%)

There are no bond length outliers.

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	111	ARG	NE-CZ-NH1	7.66	124.13	120.30
1	A	112	ARG	NE-CZ-NH2	-7.25	116.68	120.30
1	A	139	ASP	CB-CG-OD1	7.18	124.76	118.30
1	A	65	ARG	NE-CZ-NH2	-7.04	116.78	120.30
1	A	174	ARG	CG-CD-NE	6.51	125.46	111.80

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	3007	0	3004	21	0
1	B	3052	0	3051	26	0
2	A	43	0	30	2	0
2	B	43	0	30	3	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	7	0	10	0	0
3	B	7	0	10	0	0
4	B	10	0	14	0	0
5	A	185	0	0	0	0
5	B	207	0	0	2	0
All	All	6561	0	6149	51	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 4.

The worst 5 of 51 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:B:25:THR:HG22	1:B:33:PRO:HG2	1.64	0.78
2:A:501:HEM:HMB2	2:A:501:HEM:HBB2	1.79	0.63
1:A:270:GLN:O	1:A:274:VAL:HG23	2.01	0.60
1:B:25:THR:HG22	1:B:33:PRO:CG	2.33	0.58
1:B:158:VAL:HB	1:B:159:PRO:HD3	1.88	0.55

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	380/403 (94%)	368 (97%)	12 (3%)	0	100	100
1	B	388/403 (96%)	373 (96%)	13 (3%)	2 (0%)	31	25
All	All	768/806 (95%)	741 (96%)	25 (3%)	2 (0%)	43	39

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	98	SER
1	B	295	THR

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	316/328 (96%)	308 (98%)	8 (2%)	50	52
1	B	321/328 (98%)	313 (98%)	8 (2%)	50	52
All	All	637/656 (97%)	621 (98%)	16 (2%)	50	52

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	394	ASP
1	B	22	LYS
1	B	181	ASN
1	A	244	MET
1	B	219	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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

5 ligands are modelled in this entry.

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$
2	HEM	A	501	1,5	27,50,50	1.33	4 (14%)	17,82,82	1.44	2 (11%)
3	PEG	A	502	-	6,6,6	0.58	0	5,5,5	0.23	0
2	HEM	B	501	1,5	27,50,50	1.66	5 (18%)	17,82,82	2.54	6 (35%)
3	PEG	B	502	-	6,6,6	0.40	0	5,5,5	0.42	0
4	PGE	B	503	-	9,9,9	0.69	0	8,8,8	0.61	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
2	HEM	A	501	1,5	-	0/6/54/54	0/0/8/8
3	PEG	A	502	-	-	0/4/4/4	0/0/0/0
2	HEM	B	501	1,5	-	0/6/54/54	0/0/8/8
3	PEG	B	502	-	-	0/4/4/4	0/0/0/0
4	PGE	B	503	-	-	0/7/7/7	0/0/0/0

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	501	HEM	C3B-C2B	-4.10	1.34	1.40
2	A	501	HEM	C1D-ND	-3.25	1.29	1.36
2	B	501	HEM	C1D-ND	-2.63	1.30	1.36
2	A	501	HEM	C3B-C2B	-2.34	1.37	1.40
2	A	501	HEM	C1A-CHA	-2.32	1.34	1.40

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	501	HEM	CMA-C3A-C4A	-4.42	121.67	128.46

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	501	HEM	C4C-C3C-C2C	-3.68	104.33	106.90
2	B	501	HEM	CAA-CBA-CGA	-3.01	107.51	112.66
2	A	501	HEM	C3B-C4B-NB	-2.56	105.90	109.21
2	B	501	HEM	C4A-C3A-C2A	2.38	108.65	107.00

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	501	HEM	2	0
2	B	501	HEM	3	0

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	388/403 (96%)	-0.14	8 (2%) 63 62	14, 24, 49, 76	0
1	B	394/403 (97%)	-0.13	6 (1%) 73 72	13, 25, 49, 63	0
All	All	782/806 (97%)	-0.14	14 (1%) 68 67	13, 25, 49, 76	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	348	ALA	4.1
1	A	25	THR	3.9
1	B	24	ASP	3.5
1	A	99	PHE	3.3
1	A	26	GLY	3.3

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](#)

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, 95th 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(\AA^2)	Q<0.9
3	PEG	A	502	7/7	0.79	0.17	39,52,57,58	0
4	PGE	B	503	10/10	0.84	0.21	45,51,58,62	0
3	PEG	B	502	7/7	0.94	0.13	33,35,37,38	0
2	HEM	B	501	43/43	0.98	0.12	11,13,19,21	0
2	HEM	A	501	43/43	0.98	0.10	13,15,18,22	0

6.5 Other polymers [i](#)

There are no such residues in this entry.