WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Oct 26, 2023 – 09:05 AM EDT

PDB ID	:	3FHV
Title	:	Structural basis of Salmonella typhi type IVb PilS and cystic fibrosis transmembrane conductance regulator (CFTR) interaction
Authors	:	Balakrishna, A.M.; Swaminathan, K.
Deposited on	:	2008-12-10
Resolution	:	1.90 Å(reported)

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

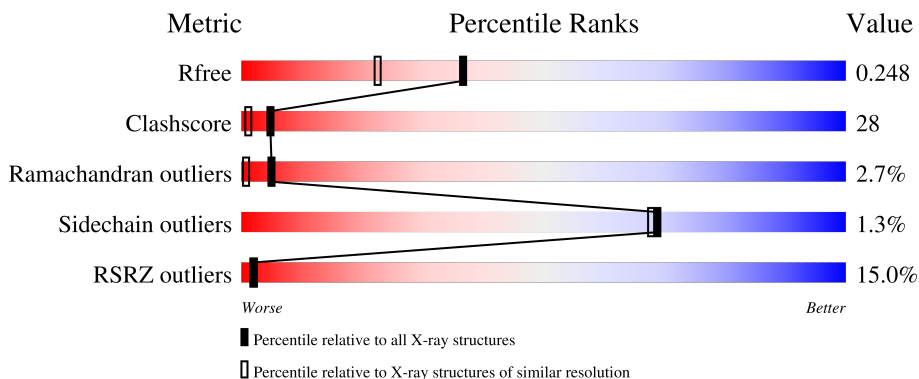|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 1.90 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		0.248
Clashscore		28
Ramachandran outliers		2.7%
Sidechain outliers		1.3%
RSRZ outliers		15.0%

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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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	156	14% · 72% · 22% · · ·
1	B	156	12% · 76% · 19% · ·
2	C	10	60% · 40% · 40% · 20%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	DTY	C	109	-	-	X	X
2	DPR	C	111	-	-	X	-
2	DGL	C	116	-	-	X	X

2 Entity composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 2474 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 Prepilin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	150	Total	C	N	O	S	0	0	0
			1054	639	183	225	7			
1	B	150	Total	C	N	O	S	0	0	0
			1054	639	183	225	7			

- Molecule 2 is a protein called cftr peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	C	10	Total	C	N	O	0	0	0
			88	51	15	22			

- Molecule 3 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	142	Total	O	0	0
			142	142		
3	B	130	Total	O	0	0
			130	130		
3	C	6	Total	O	0	0
			6	6		

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.
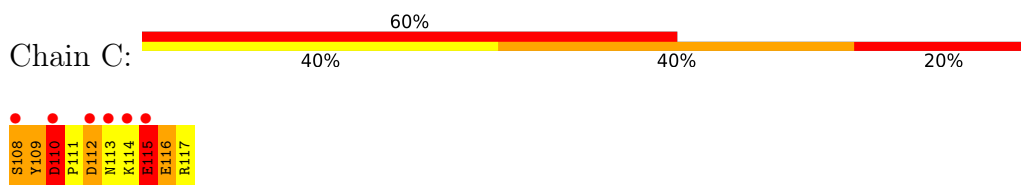
- Molecule 1: Prepilin
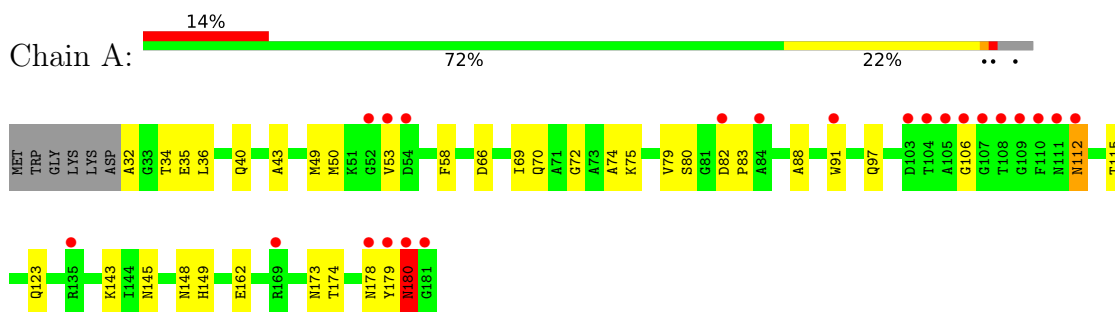
- Molecule 1: Prepilin
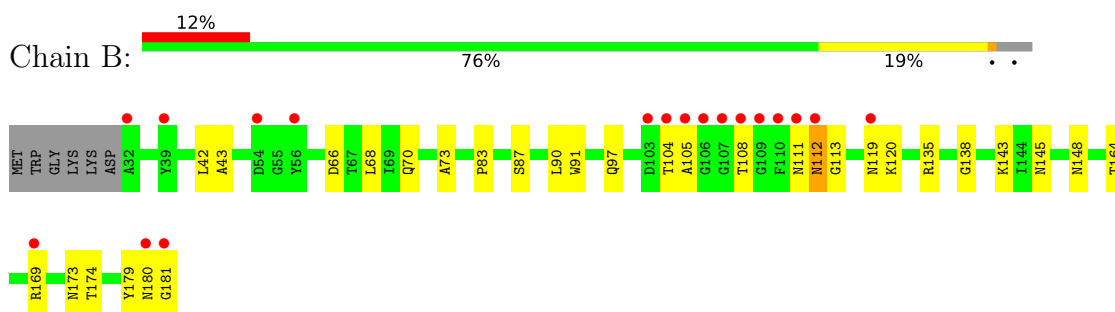
- Molecule 2: cftr peptide

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	77.80Å 114.37Å 31.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 1.90 29.41 – 1.70	Depositor EDS
% Data completeness (in resolution range)	88.6 (8.00-1.90) 96.6 (29.41-1.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	4.88 (at 1.70Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.207 , 0.248 0.214 , 0.248	Depositor DCC
R_{free} test set	2252 reflections (7.25%)	wwPDB-VP
Wilson B-factor (Å^2)	12.8	Xtriage
Anisotropy	0.220	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 51.9	EDS
L-test for twinning[2]	$< |L| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	2474	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 7.59% 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 ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: DGL, DTY, DAR, DPR

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	0.33	0/1068	0.59	0/1449
1	B	0.33	0/1068	0.62	0/1449
2	C	0.18	0/45	0.25	0/55
All	All	0.33	0/2181	0.60	0/2953

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
2	C	0	5

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	108	SER	Peptide
2	C	109	DTY	Peptide
2	C	110	ASP	Peptide
2	C	115	GLU	Peptide
2	C	116	DGL	Peptide

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	1054	0	1009	49	0
1	B	1054	0	1009	52	6
2	C	88	0	70	65	6
3	A	142	0	0	1	0
3	B	130	0	0	2	0
3	C	6	0	0	1	0
All	All	2474	0	2088	119	6

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 28.

The worst 5 of 119 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:120:LYS:NZ	2:C:115:GLU:HG3	1.21	1.44
1:B:120:LYS:NZ	2:C:115:GLU:CG	1.98	1.24
1:A:75:LYS:CD	2:C:112:ASP:OD2	1.87	1.21
1:A:75:LYS:HD3	2:C:112:ASP:OD2	1.05	1.20
1:A:75:LYS:HD3	2:C:112:ASP:CG	1.76	1.04

The worst 5 of 6 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)
1:B:138:GLY:CA	2:C:115:GLU:OE2[1_556]	1.61	0.59
1:B:138:GLY:C	2:C:115:GLU:OE2[1_556]	2.02	0.18
1:B:179:TYR:CZ	2:C:114:LYS:NZ[1_556]	2.02	0.18
1:B:179:TYR:CE1	2:C:114:LYS:NZ[1_556]	2.03	0.17
1:B:43:ALA:CB	2:C:114:LYS:NZ[1_556]	2.12	0.08

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	148/156 (95%)	141 (95%)	5 (3%)	2 (1%)	11	3
1	B	148/156 (95%)	143 (97%)	2 (1%)	3 (2%)	7	1
2	C	5/10 (50%)	1 (20%)	1 (20%)	3 (60%)	0	0
All	All	301/322 (94%)	285 (95%)	8 (3%)	8 (3%)	5	1

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	180	ASN
1	B	112	ASN
2	C	110	ASP
2	C	112	ASP
1	B	180	ASN

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	111/116 (96%)	108 (97%)	3 (3%)	44	38
1	B	111/116 (96%)	111 (100%)	0	100	100
2	C	6/6 (100%)	6 (100%)	0	100	100
All	All	228/238 (96%)	225 (99%)	3 (1%)	69	68

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	82	ASP
1	A	123	GLN
1	A	180	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	70	GLN
1	B	111	ASN
2	C	113	ASN
1	B	173	ASN
1	A	149	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

4 non-standard protein/DNA/RNA residues are modelled in this entry.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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 (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 ⓘ

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(Å²)	Q<0.9
1	A	150/156 (96%)	0.86	22 (14%)	2	2	6, 12, 56, 90	0
1	B	150/156 (96%)	0.97	18 (12%)	4	4	7, 13, 54, 91	0
2	C	6/10 (60%)	7.97	6 (100%)	0	0	45, 56, 62, 63	0
All	All	306/322 (95%)	1.05	46 (15%)	2	2	6, 12, 58, 91	0

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	108	THR	20.6
1	B	107	GLY	18.6
1	B	110	PHE	16.7
1	A	108	THR	15.7
1	A	107	GLY	15.4

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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(Å²)	Q<0.9
2	DTY	C	109	12/13	0.24	0.65	56,60,61,63	0
2	DGL	C	116	9/10	0.33	0.71	54,57,59,60	0
2	DAR	C	117	12/12	0.49	0.37	59,62,63,63	0
2	DPR	C	111	7/8	0.57	0.30	49,51,52,52	0

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.