Full wwPDB X-ray Structure Validation Report (i)

Sep 29, 2024 – 12:41 AM EDT

PDB ID	:	2P0O
Title	:	Crystal structure of a conserved protein from locus EF_2437 in Enterococcus faecalis with an unknown function
Authors	:	Cuff, M.E.; Zhou, M.; Moy, S.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2007-02-28
Resolution	:	2.15 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 (i) 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 (i)) were used in the production of this report:

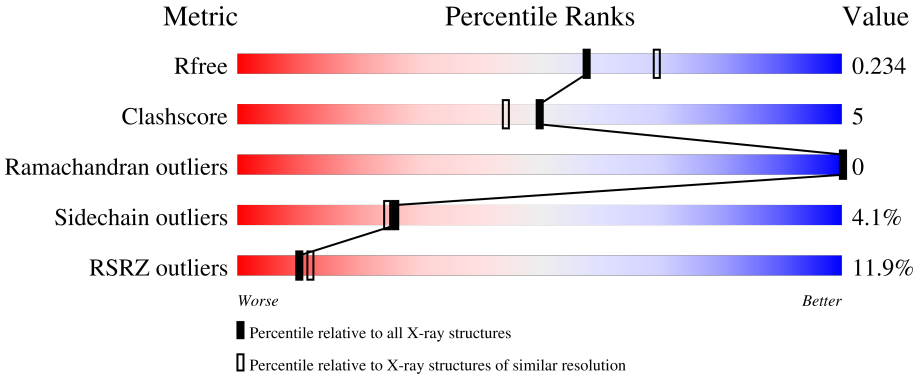
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1 Overall quality at a glance (i)

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

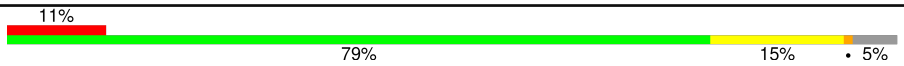
The reported resolution of this entry is 2.15 Å.

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}	164625	1881 (2.16-2.16)
Clashscore	180529	2047 (2.16-2.16)
Ramachandran outliers	177936	2027 (2.16-2.16)
Sidechain outliers	177891	2026 (2.16-2.16)
RSRZ outliers	164620	1882 (2.16-2.16)

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%

Mol	Chain	Length	Quality of chain
1	A	372	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2988 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 DUF871.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	353	2791	1776	475	531	9	0	2	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	cloning artifact	UNP Q831R3
A	-1	ASN	-	cloning artifact	UNP Q831R3
A	0	ALA	-	cloning artifact	UNP Q831R3
A	1	MSE	MET	modified residue	UNP Q831R3
A	23	MSE	MET	modified residue	UNP Q831R3
A	61	MSE	MET	modified residue	UNP Q831R3
A	64	MSE	MET	modified residue	UNP Q831R3
A	95	MSE	MET	modified residue	UNP Q831R3
A	104	MSE	MET	modified residue	UNP Q831R3
A	208	MSE	MET	modified residue	UNP Q831R3
A	228	MSE	MET	modified residue	UNP Q831R3
A	280	MSE	MET	modified residue	UNP Q831R3
A	313	MSE	MET	modified residue	UNP Q831R3
A	337	ASP	GLU	SEE REMARK 999	UNP Q831R3

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
2	A	1	1	1	0	0

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
3	A	1	1	1	0	0

- Molecule 4 is water.

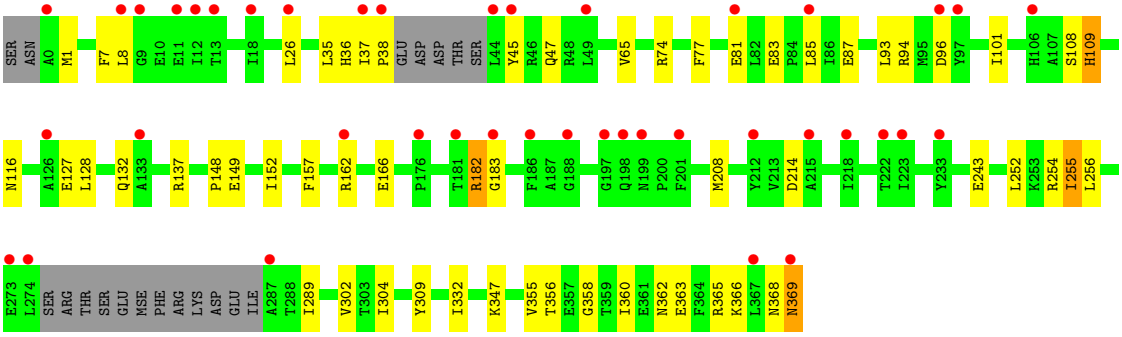
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	195	195	195	0	0

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. 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. 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 DUF871

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 63	Depositor
Cell constants a, b, c, α, β, γ	86.24Å 86.24Å 109.76Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	33.90 – 2.15 33.90 – 2.15	Depositor EDS
% Data completeness (in resolution range)	95.2 (33.90-2.15) 95.1 (33.90-2.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	2.70 (at 2.16Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.183 , 0.233 0.182 , 0.234	Depositor DCC
R_{free} test set	1224 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	39.3	Xtriage
Anisotropy	0.550	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 44.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.046 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2988	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.49% 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)

Bond lengths and bond angles in the following residue types are not validated in this section: NA, CL

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	0.93	9/2839 (0.3%)	0.73	2/3825 (0.1%)

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	83	GLU	CD-OE2	17.59	1.45	1.25
1	A	369	ASN	C-O	16.29	1.54	1.23
1	A	83	GLU	CG-CD	14.27	1.73	1.51
1	A	83	GLU	CD-OE1	9.63	1.36	1.25
1	A	109	HIS	CE1-NE2	7.63	1.50	1.32
1	A	87	GLU	CD-OE1	6.84	1.33	1.25
1	A	87	GLU	CD-OE2	5.92	1.32	1.25
1	A	109	HIS	CG-CD2	5.69	1.45	1.35
1	A	109	HIS	CG-ND1	5.13	1.50	1.38

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	83	GLU	OE1-CD-OE2	9.53	134.73	123.30
1	A	369	ASN	CA-C-O	-5.39	108.78	120.10

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	2791	0	2793	29	0
2	A	1	0	0	0	0
3	A	1	0	0	0	0
4	A	195	0	0	0	0
All	All	2988	0	2793	29	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 5.

All (29) 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:355:VAL:HG11	1:A:363:GLU:HG2	1.61	0.81
1:A:304:ILE:HG22	1:A:332:ILE:HD11	1.68	0.75
1:A:368:ASN:O	1:A:369:ASN:HB2	1.88	0.73
1:A:255:ILE:C	1:A:255:ILE:HD13	2.17	0.66
1:A:37:ILE:HD12	1:A:45:TYR:CG	2.36	0.61
1:A:35:LEU:HD23	1:A:37:ILE:HD11	1.85	0.58
1:A:255:ILE:HD13	1:A:256:LEU:N	2.18	0.58
1:A:254:ARG:CZ	1:A:289:ILE:HD11	2.38	0.54
1:A:81:GLU:HA	1:A:81:GLU:OE1	2.09	0.52
1:A:182:ARG:HG3	1:A:183:GLY:O	2.10	0.51
1:A:38:PRO:HA	1:A:74:ARG:HH12	1.75	0.51
1:A:362[B]:ASN:ND2	1:A:365:ARG:HH21	2.09	0.51
1:A:94:ARG:HH21	1:A:116:ASN:HA	1.78	0.49
1:A:7:PHE:CE2	1:A:182:ARG:HD2	2.48	0.48
1:A:8:LEU:HD12	1:A:35:LEU:HD11	1.94	0.48
1:A:109:HIS:NE2	1:A:132:GLN:O	2.46	0.48
1:A:208:MSE:HE1	1:A:214:ASP:O	2.16	0.46
1:A:243:GLU:CD	1:A:358:GLY:H	2.20	0.45
1:A:243:GLU:HG3	1:A:360:ILE:HG13	1.97	0.45
1:A:255:ILE:HG12	1:A:302:VAL:HG11	1.98	0.45
1:A:252:LEU:O	1:A:255:ILE:HD12	2.18	0.44
1:A:101:ILE:HG21	1:A:127:GLU:HB3	1.99	0.43
1:A:36:HIS:C	1:A:37:ILE:HD13	2.39	0.43
1:A:152:ILE:HD11	1:A:157:PHE:HD1	1.86	0.41
1:A:65:VAL:HB	1:A:93:LEU:HD23	2.03	0.41
1:A:148:PRO:O	1:A:149:GLU:HB2	2.20	0.41
1:A:77:PHE:CD1	1:A:85:LEU:HG	2.56	0.40
1:A:101:ILE:HG23	1:A:128:LEU:HG	2.03	0.40

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:108:SER:O	1:A:137:ARG:HD3	2.22	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	349/372 (94%)	340 (97%)	9 (3%)	0	100 100

There are no Ramachandran outliers to report.

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	298/307 (97%)	286 (96%)	12 (4%)	27 25

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1	MSE
1	A	26	LEU
1	A	47	GLN
1	A	96	ASP
1	A	162	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	166	GLU
1	A	182	ARG
1	A	255	ILE
1	A	309	TYR
1	A	347	LYS
1	A	356	THR
1	A	366	LYS

Sometimes 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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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.

5.7 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)

Warning: The R factor obtained from EDS is 0.2542,which does not match the depositor's R factor of 0.18295. Please interpret the results in this section carefully.

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	344/372 (92%)	0.97	41 (11%)	10	12	34, 57, 72, 84	2 (0%)

All (41) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	44	LEU	5.1
1	A	38	PRO	4.3
1	A	49	LEU	4.2
1	A	287	ALA	3.9
1	A	0	ALA	3.7
1	A	273	GLU	3.3
1	A	97	TYR	3.2
1	A	37	ILE	3.1
1	A	106	HIS	3.0
1	A	96	ASP	3.0
1	A	9	GLY	2.7
1	A	188	GLY	2.7
1	A	369	ASN	2.6
1	A	215	ALA	2.6
1	A	176	PRO	2.6
1	A	199	ASN	2.6
1	A	126	ALA	2.5
1	A	186	PHE	2.5
1	A	183	GLY	2.5
1	A	181	THR	2.4
1	A	18	ILE	2.4
1	A	197	GLY	2.4
1	A	222	THR	2.3
1	A	11	GLU	2.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	212	TYR	2.3
1	A	218	ILE	2.3
1	A	26	LEU	2.3
1	A	162	ARG	2.2
1	A	223	ILE	2.2
1	A	13	THR	2.2
1	A	274	LEU	2.2
1	A	201	PHE	2.2
1	A	8	LEU	2.2
1	A	233	TYR	2.2
1	A	12	ILE	2.1
1	A	45	TYR	2.1
1	A	133	ALA	2.1
1	A	198	GLN	2.1
1	A	81	GLU	2.1
1	A	367	LEU	2.1
1	A	85	LEU	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 monosaccharides 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(Å^2)	Q<0.9
3	NA	A	371	1/1	0.92	0.12	61,61,61,61	0
2	CL	A	370	1/1	0.98	0.08	44,44,44,44	0

6.5 Other polymers (i)

There are no such residues in this entry.