



Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 09:38 PM GMT

PDB ID : 2PGX
Title : Crystal structure of UPF0341 protein yhiQ from E. coli, Northeast Structural Genomics Target ER585
Authors : Forouhar, F.; Su, M.; Seetharaman, J.; Janjua, H.; Fang, Y.; Xiao, R.; Liu, J.; Baran, M.C.; Acton, T.B.; Montelione, G.T.; Tong, L.; Hunt, J.F.; Northeast Structural Genomics Consortium (NESG)
Deposited on : 2007-04-10
Resolution : 2.00 Å(reported)

This is a full wwPDB validation report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

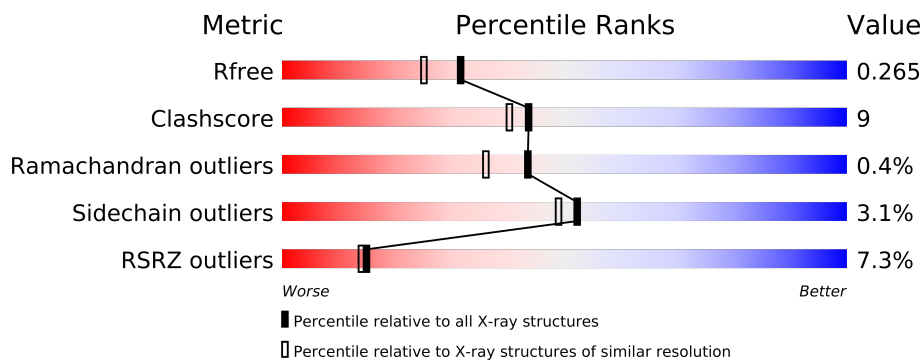
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

1 Overall quality at a glance

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}	66092	4888 (2.00-2.00)
Clashscore	79885	6188 (2.00-2.00)
Ramachandran outliers	78287	6102 (2.00-2.00)
Sidechain outliers	78261	6100 (2.00-2.00)
RSRZ outliers	66119	4890 (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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	258	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2054 atoms, of which 0 are hydrogen and 0 are deuterium.

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 UPF0341 protein yhiQ.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	247	Total	C	N	O	S	Se	0	0	0
			1875	1190	336	341	2	6			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP P68568
A	33	MSE	MET	MODIFIED RESIDUE	UNP P68568
A	64	MSE	MET	MODIFIED RESIDUE	UNP P68568
A	115	MSE	MET	MODIFIED RESIDUE	UNP P68568
A	173	MSE	MET	MODIFIED RESIDUE	UNP P68568
A	187	MSE	MET	MODIFIED RESIDUE	UNP P68568
A	251	LEU	-	CLONING ARTIFACT	UNP P68568
A	252	GLU	-	CLONING ARTIFACT	UNP P68568
A	253	HIS	-	CLONING ARTIFACT	UNP P68568
A	254	HIS	-	CLONING ARTIFACT	UNP P68568
A	255	HIS	-	CLONING ARTIFACT	UNP P68568
A	256	HIS	-	CLONING ARTIFACT	UNP P68568
A	257	HIS	-	CLONING ARTIFACT	UNP P68568
A	258	HIS	-	CLONING ARTIFACT	UNP P68568

- Molecule 2 is water.

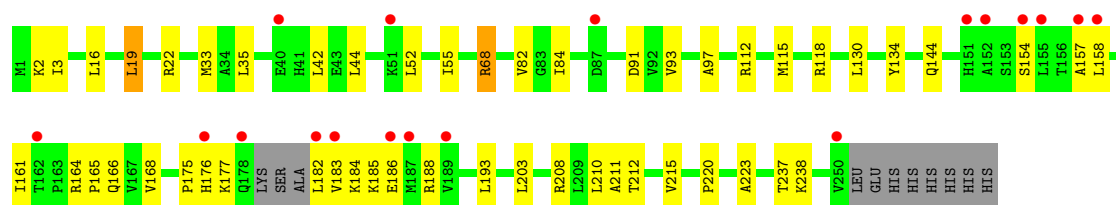
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	179	Total	O	0	0
			179	179		

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: UPF0341 protein yhiQ

Chain A: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	40.99Å 74.41Å 78.12Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	28.28 – 2.00 28.28 – 1.90	Depositor EDS
% Data completeness (in resolution range)	89.7 (28.28-2.00) 97.6 (28.28-1.90)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.20 (at 1.91Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.228 , 0.243 0.248 , 0.265	Depositor DCC
R_{free} test set	1587 reflections (9.66%)	DCC
Wilson B-factor (Å ²)	20.8	Xtriage
Anisotropy	0.484	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 30.2	EDS
Estimated twinning fraction	0.023 for -h,l,k	Xtriage
L-test for twinning	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 36231 reflections	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	2054	wwPDB-VP
Average B, all atoms (Å ²)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.86% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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	0.32	0/1905	0.59	0/2575

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1875	0	1917	33	0
2	A	179	0	0	5	0
All	All	2054	0	1917	33	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 9.

All (33) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:237:THR:HG22	1:A:238:LYS:H	1.51	0.73
1:A:118:ARG:HD3	2:A:310:HOH:O	1.89	0.72
1:A:182:LEU:HG	1:A:183:VAL:H	1.57	0.68
1:A:44:LEU:HD23	1:A:55:ILE:HD11	1.78	0.66

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:91:ASP:H	1:A:166:GLN:HE21	1.46	0.62
1:A:220:PRO:HG2	1:A:223:ALA:HB2	1.83	0.61
1:A:97:ALA:HB2	1:A:115:MSE:HB3	1.85	0.59
1:A:157:ALA:O	1:A:161:ILE:HG12	2.02	0.59
1:A:168:VAL:HG23	1:A:211:ALA:HB2	1.85	0.59
1:A:134:TYR:CE1	1:A:144:GLN:HA	2.38	0.59
1:A:3:ILE:HD13	1:A:35:LEU:HB2	1.85	0.57
1:A:91:ASP:H	1:A:166:GLN:NE2	2.02	0.56
1:A:154:SER:O	1:A:158:LEU:HB2	2.07	0.54
1:A:186:GLU:CD	1:A:186:GLU:H	2.15	0.49
1:A:82:VAL:HG23	1:A:84:ILE:HG13	1.94	0.48
1:A:93:VAL:HG23	1:A:165:PRO:HB3	1.95	0.48
1:A:68:ARG:HG2	2:A:422:HOH:O	2.13	0.47
1:A:158:LEU:O	1:A:210:LEU:HD22	2.15	0.46
1:A:182:LEU:HG	1:A:183:VAL:N	2.29	0.46
1:A:184:LYS:HE3	1:A:185:LYS:H	1.82	0.45
1:A:22:ARG:HH21	1:A:22:ARG:HG3	1.81	0.45
1:A:33:MSE:HE2	1:A:193:LEU:HB2	2.00	0.44
1:A:2:LYS:HE3	2:A:385:HOH:O	2.18	0.43
1:A:188:ARG:HH11	1:A:188:ARG:HG3	1.82	0.43
1:A:68:ARG:CG	2:A:422:HOH:O	2.67	0.43
1:A:164:ARG:CZ	1:A:212:THR:HG22	2.49	0.43
1:A:93:VAL:CG2	1:A:165:PRO:HG3	2.49	0.42
1:A:19:LEU:HD11	1:A:42:LEU:HD22	2.02	0.42
1:A:238:LYS:HE2	2:A:369:HOH:O	2.18	0.42
1:A:208:ARG:HA	1:A:215:VAL:HG21	2.02	0.41
1:A:176:HIS:C	1:A:177:LYS:HD2	2.41	0.40
1:A:52:LEU:O	1:A:185:LYS:HD2	2.21	0.40
1:A:188:ARG:HG3	1:A:188:ARG:NH1	2.37	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	243/258 (94%)	232 (96%)	10 (4%)	1 (0%)	43	36

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	175	PRO

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	194/198 (98%)	188 (97%)	6 (3%)	52	49

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	16	LEU
1	A	19	LEU
1	A	68	ARG
1	A	112	ARG
1	A	130	LEU
1	A	203	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	166	GLN
1	A	228	ASN

5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates 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	247/258 (95%)	0.46	18 (7%) 15 14	11, 23, 57, 62	0

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	183	VAL	7.8
1	A	182	LEU	7.3
1	A	154	SER	6.7
1	A	176	HIS	4.2
1	A	152	ALA	3.7
1	A	162	THR	3.4
1	A	87	ASP	3.3
1	A	51	LYS	2.9
1	A	250	VAL	2.9
1	A	157	ALA	2.9
1	A	40	GLU	2.9
1	A	155	LEU	2.7
1	A	158	LEU	2.5
1	A	187	MSE	2.3
1	A	186	GLU	2.1
1	A	151	HIS	2.1
1	A	178	GLN	2.1
1	A	189	VAL	2.0

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 carbohydrates 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.