Full wwPDB X-ray Structure Validation Report (i)

Jun 26, 2016 – 04:36 AM EDT

PDB ID	:	5K2Y
Title	:	Crystal structure of M. tuberculosis UspC (monoclinic crystal form)
Authors	:	Futterer, K.; Fullam, E.; Besra, G.S.
Deposited on	:	2016-05-19
Resolution	:	2.41 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

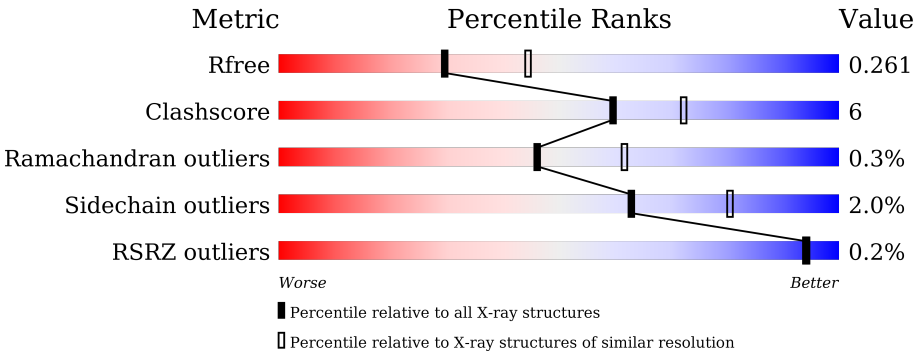
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027790
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027790

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 2.41 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}	91344	3386 (2.44-2.40)
Clashscore	102246	3897 (2.44-2.40)
Ramachandran outliers	100387	3837 (2.44-2.40)
Sidechain outliers	100360	3838 (2.44-2.40)
RSRZ outliers	91569	3396 (2.44-2.40)

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 <=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	415	83% 13% •
1	B	415	84% 11% • •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5930 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 Probable periplasmic sugar-binding lipoprotein UspC.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	400	Total	C	N	O	S	0	0	0
			2988	1902	513	567	6			
1	B	400	Total	C	N	O	S	0	0	0
			2890	1845	492	547	6			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	441	LYS	-	expression tag	UNP P71894
A	442	LEU	-	expression tag	UNP P71894
A	443	ALA	-	expression tag	UNP P71894
A	444	ALA	-	expression tag	UNP P71894
A	445	ALA	-	expression tag	UNP P71894
A	446	LEU	-	expression tag	UNP P71894
A	447	GLU	-	expression tag	UNP P71894
A	448	HIS	-	expression tag	UNP P71894
A	449	HIS	-	expression tag	UNP P71894
A	450	HIS	-	expression tag	UNP P71894
A	451	HIS	-	expression tag	UNP P71894
A	452	HIS	-	expression tag	UNP P71894
A	453	HIS	-	expression tag	UNP P71894
B	441	LYS	-	expression tag	UNP P71894
B	442	LEU	-	expression tag	UNP P71894
B	443	ALA	-	expression tag	UNP P71894
B	444	ALA	-	expression tag	UNP P71894
B	445	ALA	-	expression tag	UNP P71894
B	446	LEU	-	expression tag	UNP P71894
B	447	GLU	-	expression tag	UNP P71894
B	448	HIS	-	expression tag	UNP P71894
B	449	HIS	-	expression tag	UNP P71894
B	450	HIS	-	expression tag	UNP P71894
B	451	HIS	-	expression tag	UNP P71894
B	452	HIS	-	expression tag	UNP P71894

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	453	HIS	-	expression tag	UNP P71894

- Molecule 2 is water.

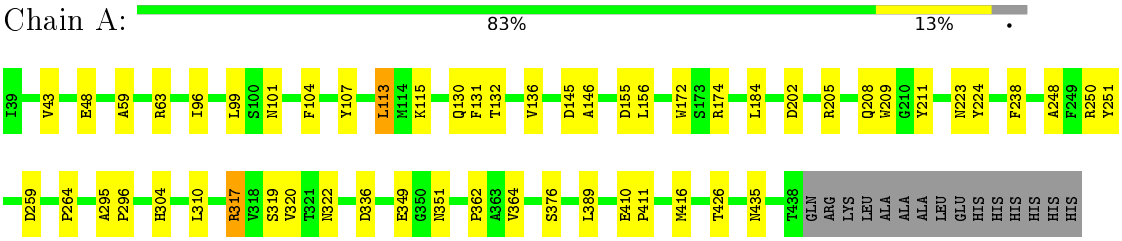
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	35	Total O 35 35	0	0
2	B	17	Total O 17 17	0	0

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 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.

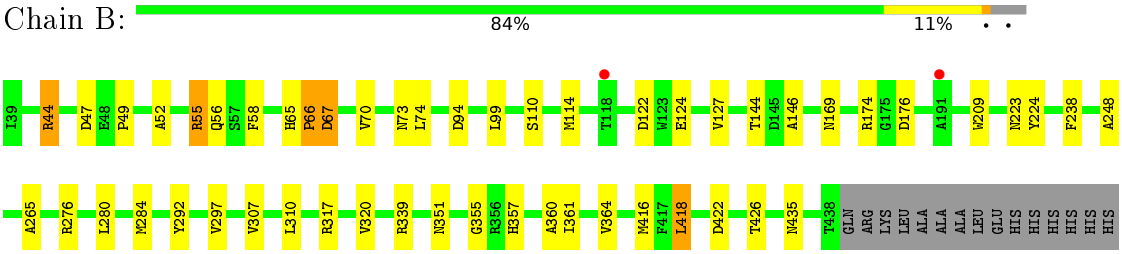
- Molecule 1: Probable periplasmic sugar-binding lipoprotein UspC

Chain A:



- Molecule 1: Probable periplasmic sugar-binding lipoprotein UspC

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	86.99Å 52.35Å 87.89Å 90.00° 90.88° 90.00°	Depositor
Resolution (Å)	87.88 – 2.41 87.88 – 2.41	Depositor EDS
% Data completeness (in resolution range)	99.5 (87.88-2.41) 99.5 (87.88-2.41)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	2.48 (at 2.42Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.203 , 0.261 0.203 , 0.261	Depositor DCC
R_{free} test set	1551 reflections (5.30%)	DCC
Wilson B-factor (Å^2)	31.5	Xtriage
Anisotropy	0.176	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 29.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.049 for l,k,-h 0.028 for h,-k,-l 0.023 for l,-k,h	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	5930	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.16% 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

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	0.89	9/4199 (0.2%)
1	B	0.77	0/2968	0.86	7/4076 (0.2%)
All	All	0.80	0/6036	0.87	16/8275 (0.2%)

There are no bond length outliers.

All (16) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	317	ARG	NE-CZ-NH1	8.79	124.69	120.30
1	A	202	ASP	CB-CG-OD1	7.30	124.87	118.30
1	A	205	ARG	NE-CZ-NH2	-7.09	116.75	120.30
1	B	44	ARG	NE-CZ-NH2	-6.95	116.83	120.30
1	A	250	ARG	NE-CZ-NH2	-6.55	117.02	120.30
1	A	259	ASP	CB-CG-OD1	6.41	124.07	118.30
1	B	66	PRO	N-CA-C	6.36	128.64	112.10
1	A	174	ARG	NE-CZ-NH2	-6.25	117.17	120.30
1	B	174	ARG	NE-CZ-NH2	-6.14	117.23	120.30
1	A	205	ARG	NE-CZ-NH1	6.09	123.35	120.30
1	B	122	ASP	CB-CG-OD1	5.56	123.31	118.30
1	B	44	ARG	NE-CZ-NH1	5.39	123.00	120.30
1	B	66	PRO	C-N-CA	5.28	134.89	121.70
1	A	317	ARG	NE-CZ-NH2	-5.27	117.66	120.30
1	A	250	ARG	NE-CZ-NH1	5.05	122.83	120.30
1	B	174	ARG	NE-CZ-NH1	5.03	122.81	120.30

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	2988	0	2806	32	0
1	B	2890	0	2640	33	0
2	A	35	0	0	0	0
2	B	17	0	0	2	0
All	All	5930	0	5446	64	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 6.

All (64) 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:146:ALA:HA	2:B:505:HOH:O	1.76	0.85
1:A:238:PHE:H	1:A:435:ASN:HD21	1.22	0.84
1:A:101:ASN:HB2	1:A:130:GLN:NE2	1.96	0.80
1:A:310:LEU:O	1:A:317:ARG:HD3	1.83	0.76
1:A:223:ASN:ND2	1:A:320:VAL:H	1.84	0.75
1:B:66:PRO:N	1:B:67:ASP:HB3	2.02	0.74
1:B:265:ALA:HB3	1:B:418:LEU:HG	1.79	0.65
1:A:223:ASN:HD22	1:A:319:SER:HA	1.63	0.63
1:A:101:ASN:HB2	1:A:130:GLN:HE21	1.64	0.63
1:A:101:ASN:HB2	1:A:130:GLN:HG3	1.82	0.61
1:B:47:ASP:OD2	1:B:49:PRO:HD2	2.02	0.58
1:A:156:LEU:HD22	1:A:208:GLN:OE1	2.04	0.58
1:B:114:MET:HE1	1:B:339:ARG:HG3	1.86	0.57
1:A:295:ALA:HB3	1:A:296:PRO:HD3	1.86	0.57
1:A:351:ASN:ND2	1:A:362:PRO:HB3	2.20	0.56
1:A:101:ASN:H	1:A:130:GLN:HE21	1.54	0.56
1:B:224:TYR:HB3	1:B:248:ALA:HB1	1.88	0.55
1:B:209:TRP:O	1:B:284:MET:HA	2.05	0.55
1:B:416:MET:HB2	1:B:426:THR:HG21	1.88	0.55
1:A:155:ASP:OD1	1:A:304:HIS:HD2	1.89	0.55
1:A:145:ASP:OD2	1:A:322:ASN:HB3	2.08	0.54
1:B:114:MET:CE	1:B:339:ARG:HG3	2.37	0.54
1:A:238:PHE:N	1:A:435:ASN:HD21	2.00	0.53
1:B:44:ARG:HD3	1:B:94:ASP:OD2	2.09	0.53
1:A:224:TYR:HB3	1:A:248:ALA:HB1	1.90	0.52
1:B:223:ASN:ND2	1:B:320:VAL:H	2.07	0.51
1:B:351:ASN:HD22	1:B:364:VAL:HB	1.76	0.51

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:107:TYR:HB3	1:A:113:LEU:HD13	1.93	0.51
1:B:238:PHE:H	1:B:435:ASN:HD21	1.58	0.51
1:A:101:ASN:HB2	1:A:130:GLN:CG	2.40	0.50
1:B:65:HIS:C	1:B:67:ASP:HB3	2.32	0.50
1:A:209:TRP:CH2	1:A:264:PRO:HD3	2.47	0.50
1:B:238:PHE:N	1:B:435:ASN:HD21	2.10	0.49
1:A:59:ALA:O	1:A:63:ARG:HG3	2.13	0.49
1:B:310:LEU:O	1:B:317:ARG:HD3	2.12	0.49
1:A:101:ASN:HB2	1:A:130:GLN:CD	2.33	0.48
1:B:44:ARG:HD2	1:B:73:ASN:ND2	2.29	0.48
1:B:49:PRO:HB2	2:B:511:HOH:O	2.14	0.47
1:B:280:LEU:HD21	1:B:297:VAL:HA	1.97	0.47
1:B:58:PHE:HB3	1:B:70:VAL:HB	1.97	0.47
1:B:66:PRO:CA	1:B:67:ASP:HB3	2.44	0.47
1:A:416:MET:HB2	1:A:426:THR:HG21	1.97	0.46
1:B:44:ARG:CD	1:B:94:ASP:OD2	2.63	0.46
1:A:172:TRP:HB2	1:A:251:TYR:CZ	2.50	0.46
1:A:43:VAL:HG22	1:A:96:ILE:HB	1.98	0.46
1:B:276:ARG:O	1:B:280:LEU:HG	2.16	0.46
1:A:146:ALA:HA	1:A:389:LEU:HD11	1.97	0.46
1:B:124:GLU:HG3	1:B:127:VAL:HG23	1.98	0.45
1:B:56:GLN:HE22	1:B:357:HIS:CE1	2.35	0.45
1:B:66:PRO:HB2	1:B:67:ASP:HB2	1.98	0.45
1:A:130:GLN:HG2	1:A:131:PHE:CE2	2.51	0.45
1:A:99:LEU:HD13	1:A:104:PHE:HB2	1.99	0.44
1:A:410:GLU:HB3	1:A:411:PRO:CD	2.47	0.44
1:B:52:ALA:HA	1:B:55:ARG:HH11	1.82	0.44
1:A:48:GLU:HG3	1:B:110:SER:CB	2.48	0.44
1:B:99:LEU:HD23	1:B:99:LEU:HA	1.89	0.44
1:B:44:ARG:HD2	1:B:73:ASN:HD22	1.82	0.43
1:A:351:ASN:ND2	1:A:364:VAL:H	2.17	0.43
1:B:144:THR:O	1:B:361:ILE:HB	2.19	0.43
1:A:132:THR:HA	1:A:136:VAL:O	2.19	0.42
1:A:184:LEU:HD22	1:A:211:TYR:HB2	2.00	0.42
1:B:56:GLN:HE22	1:B:357:HIS:HE1	1.67	0.42
1:B:355:GLY:HA2	1:B:360:ALA:O	2.21	0.41
1:A:115:LYS:HG2	1:A:136:VAL:HG11	2.02	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	398/415 (96%)	392 (98%)	6 (2%)	0	100 100
1	B	398/415 (96%)	376 (94%)	20 (5%)	2 (0%)	34 47
All	All	796/830 (96%)	768 (96%)	26 (3%)	2 (0%)	46 62

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	67	ASP
1	B	292	TYR

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	284/309 (92%)	280 (99%)	4 (1%)	74 88
1	B	260/309 (84%)	253 (97%)	7 (3%)	52 72
All	All	544/618 (88%)	533 (98%)	11 (2%)	63 80

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	113	LEU
1	A	336	ASP
1	A	349	GLU
1	A	376	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	55	ARG
1	B	74	LEU
1	B	169	ASN
1	B	176	ASP
1	B	307	VAL
1	B	418	LEU
1	B	422	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	142	GLN
1	A	223	ASN
1	A	304	HIS
1	A	351	ASN
1	A	435	ASN
1	B	56	GLN
1	B	223	ASN
1	B	351	ASN
1	B	435	ASN

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 (i)

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, 95^{th} 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	400/415 (96%)	-0.57	0 100 100	15, 27, 42, 56	0
1	B	400/415 (96%)	-0.26	2 (0%) 91 91	20, 41, 63, 83	0
All	All	800/830 (96%)	-0.41	2 (0%) 94 94	15, 33, 56, 83	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	191	ALA	3.2
1	B	118	THR	2.1

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.