Full wwPDB X-ray Structure Validation Report (i)

Nov 7, 2023 – 12:52 pm GMT

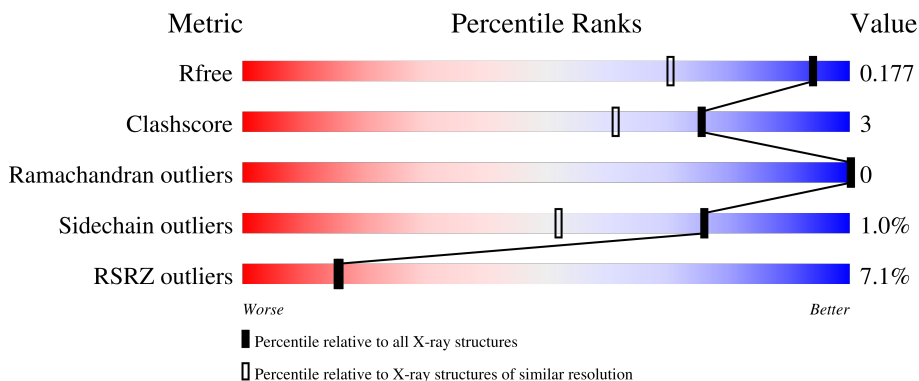|		
PDB ID	:	5AQB
Title	:	DARPin-based Crystallization Chaperones exploit Molecular Geometry as a Screening Dimension in Protein Crystallography
Authors	:	Batyuk, A.; Wu, Y.; Honegger, A.; Heberling, M.; Plueckthun, A.
Deposited on	:	2015-09-21
Resolution	:	1.37 Å(reported)

1   Overall quality at a glance ⓘ

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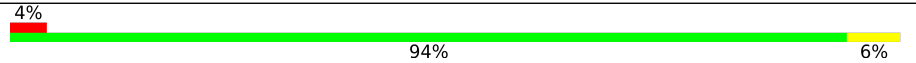*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.37 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		0.177
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		1.0%
RSRZ outliers		7.1%

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	2907 (1.40-1.36)
Clashscore	141614	3037 (1.40-1.36)
Ramachandran outliers	138981	2970 (1.40-1.36)
Sidechain outliers	138945	2969 (1.40-1.36)
RSRZ outliers	127900	2846 (1.40-1.36)

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	428	8% ... 89% ... 7% 5%
2	B	229	4% ... 94% ... 6%

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:

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	CRO	B	66	X	-	-	-

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 11102 atoms, of which 5230 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 3G61_DB15V4.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	408	Total 6498	C 2033	H 3283	N 553	O 615	S 14	0	23	0

- Molecule 2 is a protein called GREEN FLUORESCENT PROTEIN.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	229	Total 3887	C 1245	H 1947	N 319	O 370	S 6	0	21	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	SER	-	expression tag	UNP P42212
B	80	ARG	GLN	conflict	UNP P42212
B	66	CRO	THR	chromophore	UNP P42212
B	66	CRO	TYR	chromophore	UNP P42212
B	66	CRO	GLY	chromophore	UNP P42212

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	448	Total 448	O 448	0	0
3	B	269	Total 269	O 269	0	0

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: 3G61_DB15V4
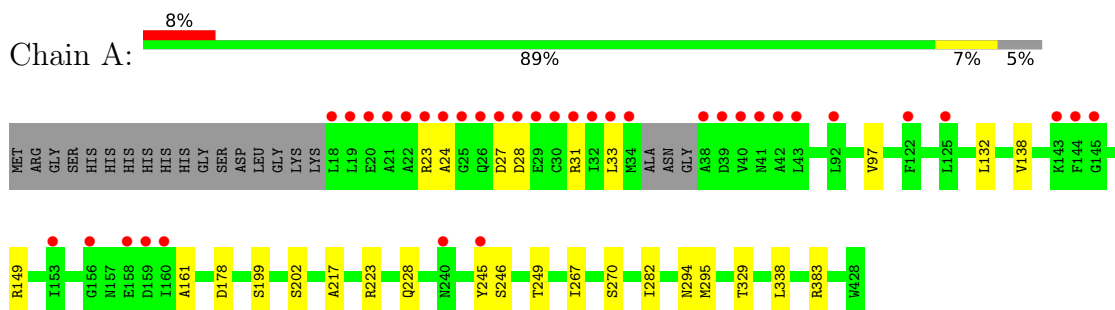
• Molecule 2: GREEN FLUORESCENT PROTEIN
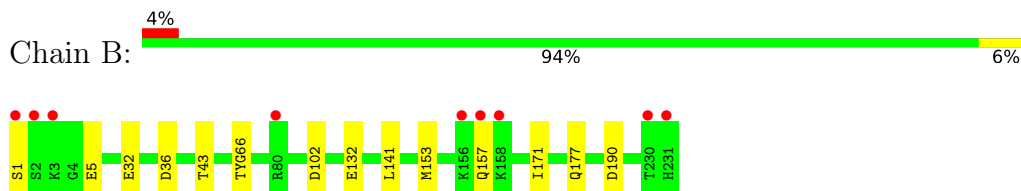
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants	90.15Å 96.17Å 92.78Å	Depositor
a, b, c, α, β, γ	90.00° 118.99° 90.00°	
Resolution (Å)	48.09 – 1.37	Depositor
	48.09 – 1.37	EDS
% Data completeness	99.1 (48.09-1.37)	Depositor
(in resolution range)	99.1 (48.09-1.37)	EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.33 (at 1.37Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.156 , 0.177	Depositor
	0.157 , 0.177	DCC
R_{free} test set	7180 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	20.2	Xtriage
Anisotropy	0.170	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 48.2	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.015 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	11102	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.48% 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: CRO

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	0.39	0/3337	0.65	1/4526 (0.0%)
2	B	0.41	0/2020	0.63	0/2726
All	All	0.40	0/5357	0.65	1/7252 (0.0%)

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	B	1	0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	383	ARG	NE-CZ-NH1	5.48	123.04	120.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	66	CRO	CB1

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	3215	3283	3274	20	0
2	B	1940	1947	1943	11	0
3	A	448	0	0	7	2
3	B	269	0	0	6	1
All	All	5872	5230	5217	32	2

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 3.

All (32) 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)
2:B:102:ASP:OD1	3:B:2141:HOH:O	1.96	0.84
2:B:190[A]:ASP:OD1	3:B:2226:HOH:O	2.00	0.80
2:B:32[B]:GLU:OE2	3:B:2048:HOH:O	2.02	0.76
1:A:28:ASP:OD1	1:A:31:ARG:NH2	2.23	0.71
1:A:228:GLN:OE1	3:A:2259:HOH:O	2.13	0.67
2:B:171[B]:ILE:HD11	2:B:177:GLN:HB2	1.82	0.61
1:A:294[B]:ASN:OD1	3:A:2319:HOH:O	2.17	0.56
1:A:149[B]:ARG:NH1	1:A:161:ALA:O	2.40	0.55
1:A:23:ARG:CB	1:A:24:ALA:HB3	2.39	0.53
2:B:141:LEU:CD2	2:B:171[B]:ILE:CD1	2.87	0.53
2:B:36:ASP:OD1	3:B:2011:HOH:O	2.18	0.51
1:A:217:ALA:CB	1:A:267[B]:ILE:HD11	2.41	0.51
1:A:295[B]:MET:CE	1:A:329:THR:HG23	2.44	0.48
1:A:223:ARG:NH2	3:A:2246:HOH:O	2.35	0.48
2:B:141:LEU:HD23	2:B:171[B]:ILE:CD1	2.44	0.48
2:B:132:GLU:OE1	3:B:2171:HOH:O	2.20	0.47
1:A:223:ARG:CD	1:A:282:ILE:HG22	2.45	0.46
2:B:43[B]:THR:HG23	3:B:2066:HOH:O	2.15	0.46
1:A:246:SER:HB3	1:A:249:THR:OG1	2.16	0.46
1:A:295[B]:MET:HE2	1:A:329:THR:HG23	1.98	0.45
1:A:97[B]:VAL:HG23	3:A:2065:HOH:O	2.17	0.45
2:B:141:LEU:CD2	2:B:171[B]:ILE:HD13	2.47	0.44
1:A:245:TYR:CD2	1:A:245:TYR:O	2.71	0.43
2:B:1:SER:N	2:B:5:GLU:OE1	2.44	0.43
1:A:149[B]:ARG:NH1	3:A:2163:HOH:O	2.33	0.42
1:A:132:LEU:HD21	1:A:138[B]:VAL:HG22	2.01	0.42
1:A:295[B]:MET:HE3	1:A:338:LEU:HD11	2.01	0.42

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:178[B]:ASP:OD2	3:A:2204:HOH:O	2.21	0.41
1:A:23:ARG:HB3	1:A:24:ALA:HB3	2.01	0.41
1:A:23:ARG:N	1:A:24:ALA:O	2.49	0.41
1:A:199:SER:HB2	1:A:202:SER:HB3	2.02	0.40

All (2) 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)
3:A:2141:HOH:O	3:B:2073:HOH:O[1_656]	2.09	0.11
3:A:2323:HOH:O	3:A:2323:HOH:O[2_756]	2.11	0.09

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	427/428 (100%)	417 (98%)	10 (2%)	0	100	100
2	B	245/229 (107%)	242 (99%)	3 (1%)	0	100	100
All	All	672/657 (102%)	659 (98%)	13 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	346/338 (102%)	343 (99%)	3 (1%)	78	56
2	B	221/200 (110%)	219 (99%)	2 (1%)	78	56
All	All	567/538 (105%)	562 (99%)	5 (1%)	76	56

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	27	ASP
1	A	33	LEU
1	A	270	SER
2	B	153	MET
2	B	157	GLN

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

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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	CRO	B	66	2	23,23,24	3.33	4 (17%)	30,32,34	2.87	9 (30%)

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	CRO	B	66	2	1/1/7/8	3/12/31/32	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	66	CRO	CA2-C2	-9.54	1.39	1.48
2	B	66	CRO	CB2-CA2	9.50	1.43	1.35
2	B	66	CRO	OG1-CB1	-6.48	1.25	1.43
2	B	66	CRO	O2-C2	3.95	1.31	1.23

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	66	CRO	CA2-C2-N3	10.23	108.21	103.37
2	B	66	CRO	OG1-CB1-CA1	5.66	121.17	109.04
2	B	66	CRO	C2-N3-C1	-4.71	105.58	107.97
2	B	66	CRO	CA1-C1-N3	-4.25	119.66	124.75
2	B	66	CRO	CA1-C1-N2	3.70	129.06	123.89
2	B	66	CRO	CG2-CB2-CA2	-3.39	125.78	129.94
2	B	66	CRO	OG1-CB1-CG1	3.29	119.48	109.74
2	B	66	CRO	CA3-N3-C1	2.90	130.65	127.16
2	B	66	CRO	CG1-CB1-CA1	2.01	116.90	112.16

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	66	CRO	CB1

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	66	CRO	N2-C1-CA1-CB1
2	B	66	CRO	N3-C1-CA1-CB1
2	B	66	CRO	C1-CA1-CB1-OG1

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides 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	408/428 (95%)	0.64	36 (8%)	10	10	15, 24, 57, 130	0
2	B	228/229 (99%)	0.28	9 (3%)	39	41	16, 23, 45, 82	0
All	All	636/657 (96%)	0.51	45 (7%)	16	16	15, 24, 53, 130	0

All (45) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	24	ALA	22.2
1	A	22	ALA	16.4
1	A	23	ARG	10.2
1	A	38	ALA	10.0
1	A	18	LEU	9.7
1	A	32	ILE	9.0
1	A	19	LEU	8.6
1	A	40	VAL	8.5
1	A	30	CYS	7.4
1	A	144	PHE	7.3
1	A	20	GLU	7.1
1	A	28	ASP	6.1
1	A	34	MET	6.0
1	A	21	ALA	5.8
1	A	31	ARG	5.6
2	B	1	SER	5.5
1	A	33	LEU	5.4
1	A	29	GLU	5.4
1	A	26	GLN	5.2
1	A	27	ASP	5.2
2	B	231	HIS	4.7
1	A	25	GLY	4.6
1	A	153[A]	ILE	4.0
1	A	41	ASN	4.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	157	GLN	3.8
2	B	230	THR	3.6
1	A	245	TYR	3.6
1	A	143	LYS	3.2
2	B	3	LYS	3.1
1	A	42	ALA	3.1
1	A	43	LEU	3.1
2	B	80	ARG	3.0
1	A	159	ASP	3.0
1	A	158	GLU	2.8
1	A	39	ASP	2.7
1	A	92	LEU	2.7
1	A	125	LEU	2.7
2	B	158	LYS	2.5
2	B	2	SER	2.4
1	A	122	PHE	2.4
1	A	240	ASN	2.2
2	B	156	LYS	2.2
1	A	145	GLY	2.2
1	A	156	GLY	2.1
1	A	160	ILE	2.1

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(Å^2)	Q<0.9
2	CRO	B	66	22/23	0.97	0.09	16,18,22,22	0

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

6.5 Other polymers ⓘ

There are no such residues in this entry.