wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 05:55 PM GMT

PDB ID	:	4JW3
Title	:	Selection of specific protein binders for pre-defined targets from an optimized library of artificial helicoidal repeat proteins (alphaRep)
Authors	:	Guellouz, A.; Valerio-Lepiniec, M.; Urvoas, A.; Chevrel, A.; Graille, M.; Fourati-Kammoun, Z.; Desmadril, M.; van Tilbeurgh, H.; Minard, P.
Deposited on	:	2013-03-27
Resolution	:	2.60 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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:

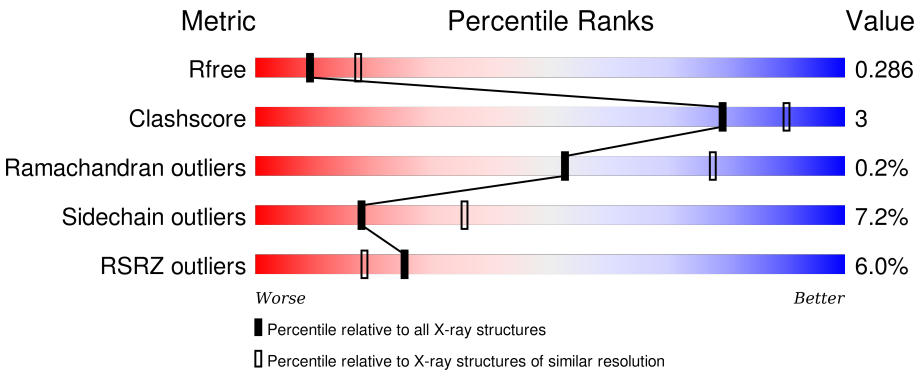
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 2.60 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	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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	120	3%, 81%, 11%, • 8%
1	B	120	8%, 84%, 9%, 7%
2	C	139	4%, 60%, 9%, • 30%
2	D	139	4%, 57%, 12%, • 31%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3095 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 Neocarzinostatin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	111	785	486	135	162	2	0	0	0
1	B	112	790	489	136	163	2	0	0	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	33	TRP	ASP	ENGINEERED MUTATION	UNP P0A3S0
A	35	ALA	GLY	ENGINEERED MUTATION	UNP P0A3S0
A	37	TRP	CYS	ENGINEERED MUTATION	UNP P0A3S0
A	39	ARG	TRP	ENGINEERED MUTATION	UNP P0A3S0
A	45	TRP	LEU	ENGINEERED MUTATION	UNP P0A3S0
A	47	TYR	CYS	ENGINEERED MUTATION	UNP P0A3S0
A	52	ASN	PHE	ENGINEERED MUTATION	UNP P0A3S0
A	113	ALA	-	EXPRESSION TAG	UNP P0A3S0
A	114	ALA	-	EXPRESSION TAG	UNP P0A3S0
A	115	HIS	-	EXPRESSION TAG	UNP P0A3S0
A	116	HIS	-	EXPRESSION TAG	UNP P0A3S0
A	117	HIS	-	EXPRESSION TAG	UNP P0A3S0
A	118	HIS	-	EXPRESSION TAG	UNP P0A3S0
A	119	HIS	-	EXPRESSION TAG	UNP P0A3S0
A	120	HIS	-	EXPRESSION TAG	UNP P0A3S0
B	33	TRP	ASP	ENGINEERED MUTATION	UNP P0A3S0
B	35	ALA	GLY	ENGINEERED MUTATION	UNP P0A3S0
B	37	TRP	CYS	ENGINEERED MUTATION	UNP P0A3S0
B	39	ARG	TRP	ENGINEERED MUTATION	UNP P0A3S0
B	45	TRP	LEU	ENGINEERED MUTATION	UNP P0A3S0
B	47	TYR	CYS	ENGINEERED MUTATION	UNP P0A3S0
B	52	ASN	PHE	ENGINEERED MUTATION	UNP P0A3S0
B	113	ALA	-	EXPRESSION TAG	UNP P0A3S0
B	114	ALA	-	EXPRESSION TAG	UNP P0A3S0
B	115	HIS	-	EXPRESSION TAG	UNP P0A3S0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	116	HIS	-	EXPRESSION TAG	UNP P0A3S0
B	117	HIS	-	EXPRESSION TAG	UNP P0A3S0
B	118	HIS	-	EXPRESSION TAG	UNP P0A3S0
B	119	HIS	-	EXPRESSION TAG	UNP P0A3S0
B	120	HIS	-	EXPRESSION TAG	UNP P0A3S0

- Molecule 2 is a protein called Alpha-helical artificial proteins.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	97	762	480	137	144	1	0	0	0
			Total	C	N	O	S			
2	D	96	752	474	134	143	1	0	0	0

- Molecule 3 is water.

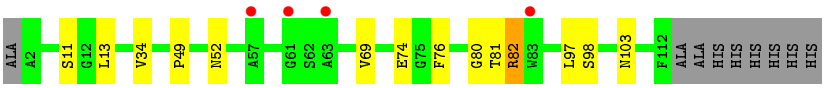
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	1	1	1	0	0
			Total	O		
3	B	2	2	2	0	0
			Total	O		
3	C	1	1	1	0	0
			Total	O		
3	D	2	2	2	0	0

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.

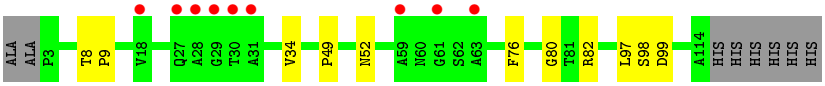
- Molecule 1: Neocarzinostatin

Chain A:



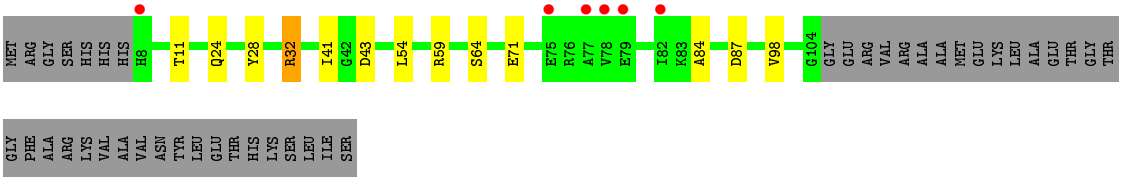
- Molecule 1: Neocarzinostatin

Chain B:



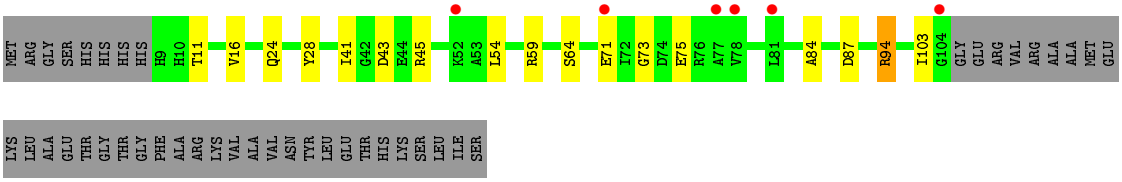
- Molecule 2: Alpha-helical artificial proteins

Chain C:



- Molecule 2: Alpha-helical artificial proteins

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	118.21Å 59.21Å 65.53Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.77 – 2.60 39.40 – 2.60	Depositor EDS
% Data completeness (in resolution range)	95.5 (33.77-2.60) 95.9 (39.40-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	2.12 (at 2.61Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.224 , 0.264 0.232 , 0.286	Depositor DCC
R_{free} test set	704 reflections (5.27%)	DCC
Wilson B-factor (Å^2)	60.0	Xtriage
Anisotropy	0.731	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 52.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Outliers	0 of 14076 reflections	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3095	wwPDB-VP
Average B, all atoms (Å^2)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 31.64 % of the origin peak, indicating pseudo translational symmetry.The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.0708e-03.The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

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	0.48	0/802	0.71	0/1101
1	B	0.44	0/807	0.68	0/1107
2	C	0.56	0/775	0.70	0/1044
2	D	0.54	0/764	0.70	0/1029
All	All	0.51	0/3148	0.70	0/4281

There are no bond length outliers.

There are no bond angle outliers.

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	785	0	734	6	0
1	B	790	0	740	5	0
2	C	762	0	756	6	0
2	D	752	0	749	8	0
3	A	1	0	0	0	0
3	B	2	0	0	0	0
3	C	1	0	0	0	0
3	D	2	0	0	1	0
All	All	3095	0	2979	21	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 3.

The worst 5 of 21 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:D:11:THR:HG21	2:D:41:ILE:HG23	1.55	0.87
2:C:32:ARG:HH11	2:C:32:ARG:HG2	1.42	0.84
2:D:73:GLY:HA2	2:D:103:ILE:HG12	1.86	0.58
1:A:13:LEU:HD13	1:A:69:VAL:HG21	1.88	0.56
2:C:11:THR:HG21	2:C:41:ILE:HG12	1.90	0.53

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	109/120 (91%)	105 (96%)	3 (3%)	1 (1%)	21	42
1	B	110/120 (92%)	106 (96%)	4 (4%)	0	100	100
2	C	95/139 (68%)	93 (98%)	2 (2%)	0	100	100
2	D	94/139 (68%)	92 (98%)	2 (2%)	0	100	100
All	All	408/518 (79%)	396 (97%)	11 (3%)	1 (0%)	52	77

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	103	ASN

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	79/85 (93%)	75 (95%)	4 (5%)	29	55
1	B	79/85 (93%)	76 (96%)	3 (4%)	40	68
2	C	75/108 (69%)	68 (91%)	7 (9%)	11	21
2	D	74/108 (68%)	66 (89%)	8 (11%)	8	15
All	All	307/386 (80%)	285 (93%)	22 (7%)	18	35

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	32	ARG
2	C	59	ARG
2	D	75	GLU
2	C	43	ASP
2	C	54	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	52	ASN
1	B	103	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 (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)

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	111/120 (92%)	0.34	4 (3%)	46	38	45, 69, 88, 97	0
1	B	112/120 (93%)	0.80	9 (8%)	15	10	44, 92, 120, 143	0
2	C	97/139 (69%)	0.29	6 (6%)	24	18	51, 72, 108, 111	0
2	D	96/139 (69%)	0.33	6 (6%)	23	17	51, 74, 103, 122	0
All	All	416/518 (80%)	0.45	25 (6%)	25	18	44, 75, 109, 143	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	77	ALA	4.2
1	B	28	ALA	3.8
1	B	61	GLY	3.6
2	C	77	ALA	3.3
1	B	30	THR	3.2

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 carbohydrates in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.