wwPDB X-ray Structure Validation Summary Report (i)

Apr 21, 2024 – 11:02 am BST

PDB ID	:	4D49
Title	:	Crystal structure of computationally designed armadillo repeat proteins for modular peptide recognition.
Authors	:	Reichen, C.; Forzani, C.; Zhou, T.; Parmeggiani, F.; Fleishman, S.J.; Mittl, P.R.E.; Madhurantakam, C.; Honegger, A.; Ewald, C.; Zerbe, O.; Baker, D.; Caflisch, A.; Pluckthun, A.
Deposited on	:	2014-10-27
Resolution	:	2.09 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
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:

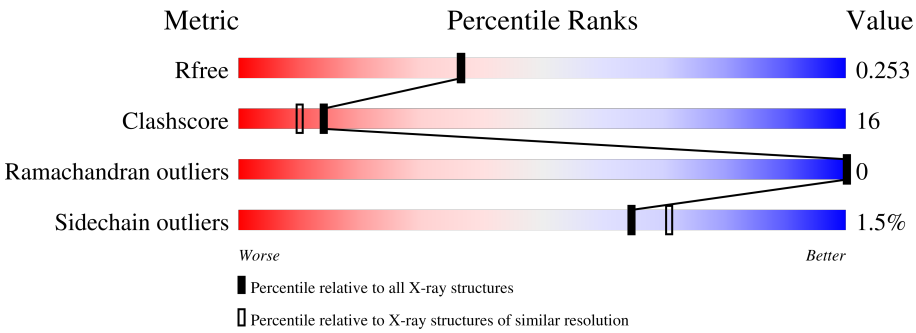
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 2.09 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}	130704	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)

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	243	77% 20% • •
1	B	243	75% 22% • •
1	E	243	77% 21% •
1	F	243	77% 21% • •
2	C	10	10% 60% 20% 10%
2	D	10	20% 40% 30% 10%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	10	10% 30% 50% 10%
2	H	10	20% 70% 10%

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:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	ARG	E	1011	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7834 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.

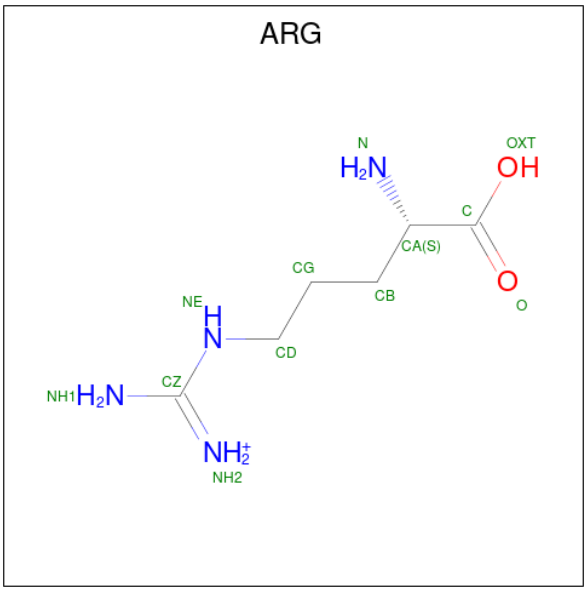
- Molecule 1 is a protein called ARMADILLO REPEAT PROTEIN ARM00027.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	239	1789	1119	301	367	2	0	0	1
1	B	239	1795	1123	301	369	2	0	1	1
1	E	239	1789	1119	301	367	2	0	0	1
1	F	239	1797	1124	304	367	2	0	1	1

- Molecule 2 is a protein called POLY ARG DECAPEPTIDE.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	C	9	107	59	39	9	0	1	0
2	D	9	99	54	36	9	0	0	0
2	G	9	99	54	36	9	0	0	0
2	H	9	99	54	36	9	0	0	0

- Molecule 3 is ARGININE (three-letter code: ARG) (formula: $C_6H_{15}N_4O_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	D	1	11	6	4	1	0	0
3	E	1	11	6	4	1	0	0

- Molecule 4 is water.

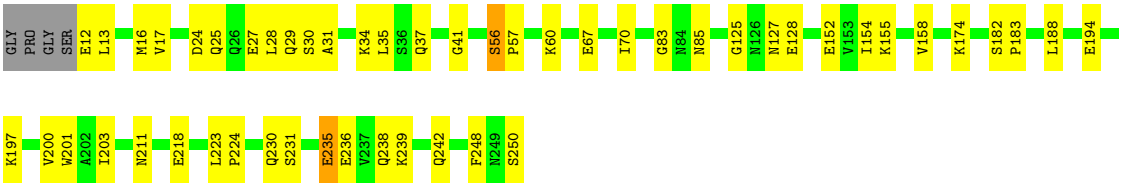
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	50	50	50	0	0
4	B	53	53	53	0	0
4	C	3	3	3	0	0
4	D	3	3	3	0	0
4	E	70	70	70	0	0
4	F	51	51	51	0	0
4	G	3	3	3	0	0
4	H	5	5	5	0	0

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.

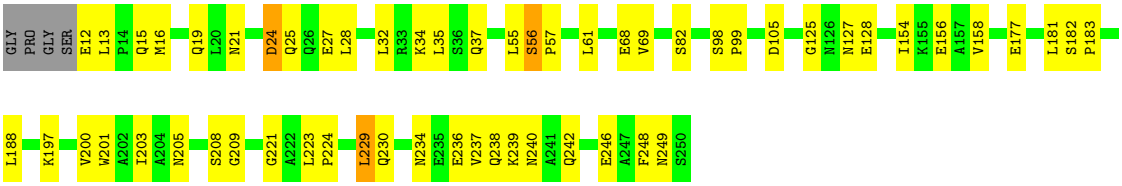
• Molecule 1: ARMADILLO REPEAT PROTEIN ARM00027

Chain A:



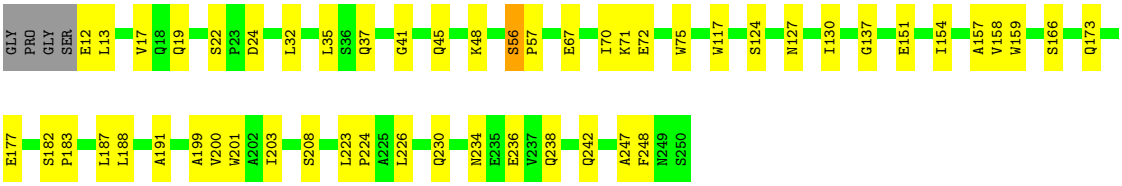
• Molecule 1: ARMADILLO REPEAT PROTEIN ARM00027

Chain B:



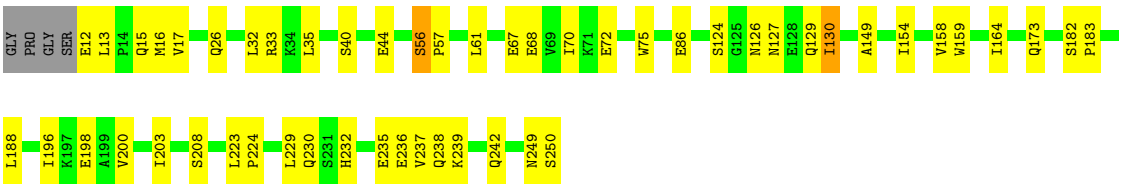
• Molecule 1: ARMADILLO REPEAT PROTEIN ARM00027

Chain E:



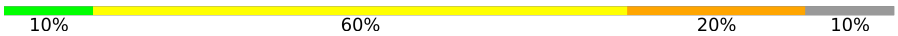
• Molecule 1: ARMADILLO REPEAT PROTEIN ARM00027

Chain F:



- Molecule 2: POLY ARG DECAPEPTIDE

Chain C:



- Molecule 2: POLY ARG DECAPEPTIDE

Chain D:

- Molecule 2: POLY ARG DECAPEPTIDE

Chain G:

- Molecule 2: POLY ARG DECAPEPTIDE

Chain H:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	88.57Å 51.73Å 107.61Å 90.00° 90.16° 90.00°	Depositor
Resolution (Å)	20.00 – 2.09 46.62 – 1.69	Depositor EDS
% Data completeness (in resolution range)	98.1 (20.00-2.09) 97.3 (46.62-1.69)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.30 (at 1.69Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.201 , 0.255 0.205 , 0.253	Depositor DCC
R_{free} test set	5257 reflections (4.89%)	wwPDB-VP
Wilson B-factor (Å^2)	20.6	Xtriage
Anisotropy	0.296	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 25.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.41$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	0.459 for h,-k,-l	Xtriage
Reported twinning fraction	0.505 for H, K, L 0.495 for -h,-k,l	Depositor
Outliers	1 of 107543 reflections (0.001%)	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7834	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 16.28% 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)

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	0.74	1/1805 (0.1%)	0.84	2/2447 (0.1%)
1	B	0.82	0/1814	0.85	1/2459 (0.0%)
1	E	0.89	1/1805 (0.1%)	0.87	0/2447
1	F	0.68	0/1816	0.78	0/2461
2	C	0.78	0/109	1.63	1/137 (0.7%)
2	D	1.05	1/98 (1.0%)	1.72	2/123 (1.6%)
2	G	1.27	0/98	2.33	7/123 (5.7%)
2	H	0.75	0/98	1.46	2/123 (1.6%)
All	All	0.80	3/7643 (0.0%)	0.91	15/10320 (0.1%)

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
1	E	0	1
2	C	0	1
All	All	0	2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	152	GLU	CD-OE2	9.37	1.35	1.25
1	E	41	GLY	N-CA	5.94	1.54	1.46
2	D	4	ARG	CZ-NH2	-5.14	1.26	1.33

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	G	7	ARG	NE-CZ-NH1	9.25	124.92	120.30
2	G	1	ARG	NE-CZ-NH1	8.93	124.77	120.30
2	C	2	ARG	NE-CZ-NH1	8.33	124.46	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	41	GLY	N-CA-C	-7.60	94.11	113.10
2	G	1	ARG	NE-CZ-NH2	-6.67	116.97	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	8	ARG	Peptide
1	E	247	ALA	Peptide

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	1789	0	1815	41	0
1	B	1795	0	1821	68	0
1	E	1789	0	1815	53	0
1	F	1797	0	1828	57	1
2	C	107	0	132	8	0
2	D	99	0	116	17	0
2	G	99	0	119	13	0
2	H	99	0	116	12	0
3	D	11	0	12	0	0
3	E	11	0	12	6	0
4	A	50	0	0	8	1
4	B	53	0	0	19	2
4	C	3	0	0	0	1
4	D	3	0	0	4	0
4	E	70	0	0	28	2
4	F	51	0	0	23	1
4	G	3	0	0	0	0
4	H	5	0	0	0	0
All	All	7834	0	7786	242	4

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 16.

The worst 5 of 242 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:E:48:LYS:HD2	4:E:2018:HOH:O	1.22	1.36
1:E:19:GLN:NE2	4:E:2006:HOH:O	1.80	1.14
1:E:72:GLU:HG3	4:E:2010:HOH:O	1.49	1.11
1:A:83:GLY:O	2:C:6:ARG:NH1	1.83	1.10
1:B:201:TRP:CZ3	4:B:2042:HOH:O	2.14	1.00

All (4) 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)
4:B:2040:HOH:O	4:E:2022:HOH:O[2_555]	1.56	0.64
1:F:208:SER:O	4:A:2043:HOH:O[2_556]	1.99	0.21
4:C:2003:HOH:O	4:F:2048:HOH:O[2_546]	2.06	0.14
4:B:2020:HOH:O	4:E:2039:HOH:O[2_545]	2.19	0.01

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	237/243 (98%)	231 (98%)	6 (2%)	0	100	100
1	B	238/243 (98%)	234 (98%)	4 (2%)	0	100	100
1	E	237/243 (98%)	230 (97%)	7 (3%)	0	100	100
1	F	238/243 (98%)	234 (98%)	4 (2%)	0	100	100
2	C	8/10 (80%)	7 (88%)	1 (12%)	0	100	100
2	D	7/10 (70%)	7 (100%)	0	0	100	100
2	G	7/10 (70%)	7 (100%)	0	0	100	100
2	H	7/10 (70%)	5 (71%)	2 (29%)	0	100	100
All	All	979/1012 (97%)	955 (98%)	24 (2%)	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	189/192 (98%)	187 (99%)	2 (1%)	73	79
1	B	190/192 (99%)	184 (97%)	6 (3%)	39	41
1	E	189/192 (98%)	188 (100%)	1 (0%)	88	92
1	F	190/192 (99%)	188 (99%)	2 (1%)	73	79
2	C	10/10 (100%)	10 (100%)	0	100	100
2	D	9/10 (90%)	9 (100%)	0	100	100
2	G	9/10 (90%)	9 (100%)	0	100	100
2	H	9/10 (90%)	8 (89%)	1 (11%)	6	3
All	All	795/808 (98%)	783 (98%)	12 (2%)	65	71

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	246[B]	GLU
1	E	56	SER
2	H	10	ARG
1	F	56	SER
1	B	56	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	230	GLN
1	E	45	GLN
1	F	249	ASN
1	F	45	GLN
1	F	232	HIS

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

2 ligands are 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	$\#\|Z\| > 2$
3	ARG	D	1011	-	9,10,11	0.63	0	5,11,13	0.57	0
3	ARG	E	1011	-	9,10,11	0.75	0	5,11,13	0.33	0

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
3	ARG	D	1011	-	-	1/8/9/11	-
3	ARG	E	1011	-	-	3/8/9/11	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	D	1011	ARG	NE-CD-CG-CB
3	E	1011	ARG	NE-CD-CG-CB
3	E	1011	ARG	N-CA-CB-CG
3	E	1011	ARG	CA-CB-CG-CD

There are no ring outliers.

1 monomer is involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	E	1011	ARG	6	0

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)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.