



wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 11:58 PM GMT

PDB ID : 4J2J
Title : Crystal structure of AXH domain complex with Capicua
Authors : Song, J.-J.; Kim, E.
Deposited on : 2013-02-04
Resolution : 2.50 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

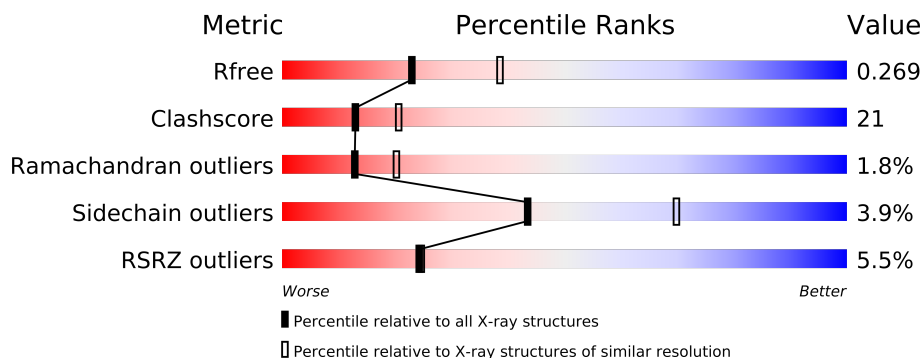
The following versions of software and data (see [references](#)) were used in the production of this report:

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 2.50 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	2784 (2.50-2.50)
Clashscore	79885	3562 (2.50-2.50)
Ramachandran outliers	78287	3480 (2.50-2.50)
Sidechain outliers	78261	3482 (2.50-2.50)
RSRZ outliers	66119	2785 (2.50-2.50)

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	129	
1	B	129	
1	C	129	
2	D	21	
2	E	21	
2	F	21	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3161 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 Ataxin-1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	121	Total	C	N	O	S	Se	0	0	0
			937	595	150	185	4	3			
1	B	118	Total	C	N	O	S	Se	0	0	0
			920	585	147	181	4	3			
1	C	115	Total	C	N	O	S	Se	0	0	0
			890	566	143	174	4	3			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	561	GLU	-	EXPRESSION TAG	UNP P54253
A	562	PHE	-	EXPRESSION TAG	UNP P54253
A	580	MSE	ILE	ENGINEERED MUTATION	UNP P54253
A	605	MSE	ILE	ENGINEERED MUTATION	UNP P54253
B	561	GLU	-	EXPRESSION TAG	UNP P54253
B	562	PHE	-	EXPRESSION TAG	UNP P54253
B	580	MSE	ILE	ENGINEERED MUTATION	UNP P54253
B	605	MSE	ILE	ENGINEERED MUTATION	UNP P54253
C	561	GLU	-	EXPRESSION TAG	UNP P54253
C	562	PHE	-	EXPRESSION TAG	UNP P54253
C	580	MSE	ILE	ENGINEERED MUTATION	UNP P54253
C	605	MSE	ILE	ENGINEERED MUTATION	UNP P54253

- Molecule 2 is a protein called Protein capicua homolog.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	D	15	Total	C	N	O	0	0	0
			125	87	21	17			
2	E	17	Total	C	N	O	0	0	0
			137	95	23	19			
2	F	15	Total	C	N	O	0	0	0
			124	86	21	17			

- Molecule 3 is water.

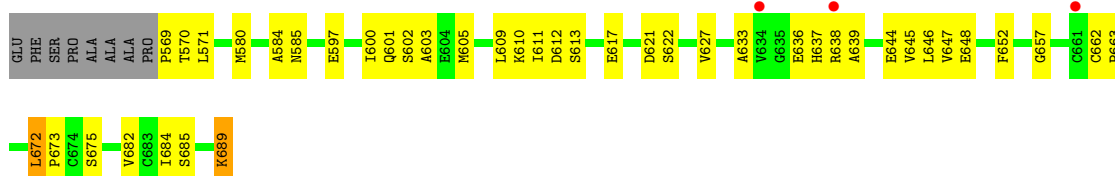
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	8	Total 8	O 8	0	0
3	B	5	Total 5	O 5	0	0
3	C	11	Total 11	O 11	0	0
3	E	2	Total 2	O 2	0	0
3	F	2	Total 2	O 2	0	0

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.

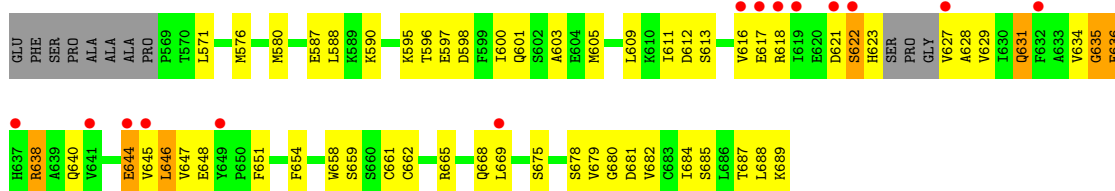
- Molecule 1: Ataxin-1

Chain A: 



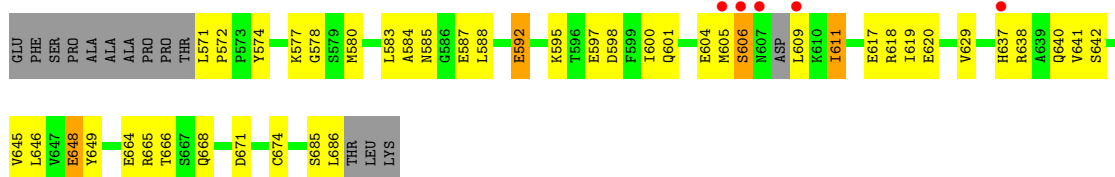
- Molecule 1: Ataxin-1

Chain B: 



- Molecule 1: Ataxin-1

Chain C: 



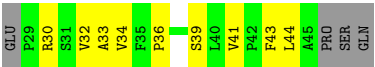
- Molecule 2: Protein capicua homolog

Chain D: 



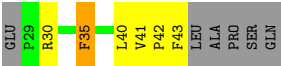
- Molecule 2: Protein capicua homolog

Chain E: 



● Molecule 2: Protein capicua homolog

Chain F: 



4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α , β , γ	77.72Å 89.10Å 132.66Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.90 – 2.50 43.90 – 2.50	Depositor EDS
% Data completeness (in resolution range)	97.1 (43.90-2.50) 98.7 (43.90-2.50)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.22 (at 2.51Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.223 , 0.270 0.230 , 0.269	Depositor DCC
R_{free} test set	790 reflections (4.93%)	DCC
Wilson B-factor (Å ²)	46.7	Xtriage
Anisotropy	0.234	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 32.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 30627 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	3161	wwPDB-VP
Average B, all atoms (Å ²)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.65% 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	0.42	0/953	0.69	0/1283
1	B	0.36	0/934	0.64	0/1255
1	C	0.40	0/904	0.61	0/1216
2	D	0.47	0/131	0.67	0/180
2	E	0.50	0/144	0.71	0/198
2	F	0.38	0/131	0.62	0/180
All	All	0.40	0/3197	0.65	0/4312

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	937	0	925	28	0
1	B	920	0	909	61	0
1	C	890	0	874	39	0
2	D	125	0	125	7	0
2	E	137	0	138	9	0
2	F	124	0	122	8	0
3	A	8	0	0	0	0
3	B	5	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	11	0	0	1	0
3	E	2	0	0	0	0
3	F	2	0	0	0	0
All	All	3161	0	3093	129	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 21.

The worst 5 of 129 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:580:MSE:HE1	1:B:590:LYS:HG2	1.14	1.11
1:B:646:LEU:HB3	1:B:648:GLU:OE2	1.77	0.84
1:B:685:SER:HA	2:E:41:VAL:HG13	1.65	0.78
1:A:647:VAL:HB	1:A:662:CYS:SG	2.25	0.77
1:B:617:GLU:HA	1:B:679:VAL:HG13	1.68	0.76

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	119/129 (92%)	109 (92%)	9 (8%)	1 (1%)	27	46
1	B	114/129 (88%)	97 (85%)	13 (11%)	4 (4%)	6	7
1	C	111/129 (86%)	104 (94%)	5 (4%)	2 (2%)	13	20
2	D	13/21 (62%)	12 (92%)	1 (8%)	0	100	100
2	E	15/21 (71%)	13 (87%)	2 (13%)	0	100	100
2	F	13/21 (62%)	12 (92%)	1 (8%)	0	100	100
All	All	385/450 (86%)	347 (90%)	31 (8%)	7 (2%)	13	20

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	583	LEU
1	B	622	SER
1	B	638	ARG
1	A	636	GLU
1	C	606	SER

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	109/111 (98%)	106 (97%)	3 (3%)	56	82
1	B	107/111 (96%)	103 (96%)	4 (4%)	45	72
1	C	103/111 (93%)	98 (95%)	5 (5%)	35	59
2	D	14/19 (74%)	14 (100%)	0	100	100
2	E	15/19 (79%)	14 (93%)	1 (7%)	23	40
2	F	14/19 (74%)	13 (93%)	1 (7%)	21	37
All	All	362/390 (93%)	348 (96%)	14 (4%)	43	70

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	644	GLU
1	C	592	GLU
1	C	686	LEU
1	B	636	GLU
1	C	671	ASP

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	601	GLN
1	A	640	GLN

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	121/129 (93%)	0.17	3 (2%) 54 57	29, 45, 83, 163	0
1	B	118/129 (91%)	0.52	14 (11%) 5 5	32, 57, 120, 150	0
1	C	115/129 (89%)	0.06	5 (4%) 34 35	30, 53, 94, 124	1 (0%)
2	D	15/21 (71%)	0.10	0 100 100	33, 40, 68, 74	0
2	E	17/21 (80%)	0.13	0 100 100	29, 45, 99, 107	0
2	F	15/21 (71%)	-0.03	0 100 100	48, 53, 78, 80	0
All	All	401/450 (89%)	0.23	22 (5%) 24 24	29, 51, 101, 163	1 (0%)

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	619	ILE	6.7
1	B	645	VAL	4.2
1	B	627	VAL	3.9
1	B	644	GLU	3.7
1	C	606	SER	3.6

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.