wwPDB X-ray Structure Validation Summary Report (i)

Nov 25, 2024 – 04:57 PM EST

PDB ID	:	2PQU
Title	:	Crystal structure of KH1 domain of human PCBP2 complexed to single-stranded 12-mer telomeric dna
Authors	:	James, T.L.; Lee, J.
Deposited on	:	2007-05-02
Resolution	:	2.12 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:

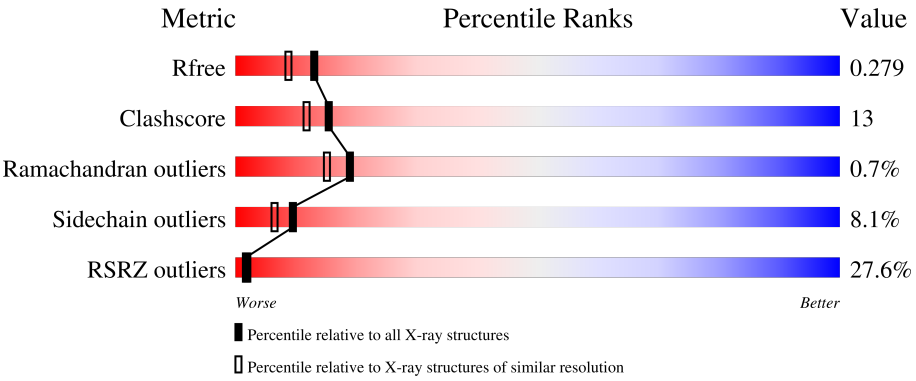
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 2.12 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}	164625	7689 (2.14-2.10)
Clashscore	180529	8431 (2.14-2.10)
Ramachandran outliers	177936	8366 (2.14-2.10)
Sidechain outliers	177891	8367 (2.14-2.10)
RSRZ outliers	164620	7689 (2.14-2.10)

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	E	12	8% 25% 58% 17%
1	G	12	17% 33% 50% 17%
2	A	73	38% 71% 23% • ••
2	B	73	38% 64% 18% 14% •
2	C	73	14% 71% 23% • •
2	D	73	16% 79% 12% • •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2752 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 DNA chain called 12-mer C-rich strand of human telomeric DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	12	Total	C	N	O	P	0	0	0
			235	114	42	68	11			
1	G	12	Total	C	N	O	P	0	0	0
			235	114	42	68	11			

- Molecule 2 is a protein called Poly(rC)-binding protein 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	A	72	Total	C	N	O	S	Se	0	0	0
			550	346	98	102	1	3			
2	B	70	Total	C	N	O	S	Se	0	0	0
			533	336	94	99	1	3			
2	C	70	Total	C	N	O	S	Se	0	0	0
			533	336	94	99	1	3			
2	D	70	Total	C	N	O	S	Se	0	0	0
			533	336	94	99	1	3			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	10	LYS	-	cloning artifact	UNP Q15366
A	20	MSE	MET	modified residue	UNP Q15366
A	39	MSE	MET	modified residue	UNP Q15366
A	74	MSE	MET	modified residue	UNP Q15366
B	10	LYS	-	cloning artifact	UNP Q15366
B	20	MSE	MET	modified residue	UNP Q15366
B	39	MSE	MET	modified residue	UNP Q15366
B	74	MSE	MET	modified residue	UNP Q15366
C	10	LYS	-	cloning artifact	UNP Q15366
C	20	MSE	MET	modified residue	UNP Q15366
C	39	MSE	MET	modified residue	UNP Q15366
C	74	MSE	MET	modified residue	UNP Q15366
D	10	LYS	-	cloning artifact	UNP Q15366

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	20	MSE	MET	modified residue	UNP Q15366
D	39	MSE	MET	modified residue	UNP Q15366
D	74	MSE	MET	modified residue	UNP Q15366

- Molecule 3 is water.

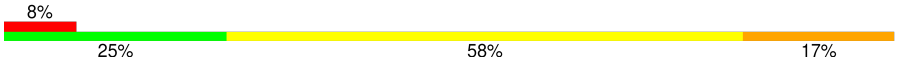
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	E	12	Total O 12 12	0	0
3	G	7	Total O 7 7	0	0
3	A	28	Total O 28 28	0	0
3	B	41	Total O 41 41	0	0
3	C	29	Total O 29 29	0	0
3	D	16	Total O 16 16	0	0

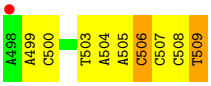
3 Residue-property plots (i)

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: 12-mer C-rich strand of human telomeric DNA

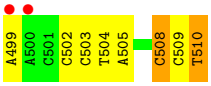
Chain E:





• Molecule 1: 12-mer C-rich strand of human telomeric DNA

Chain G:

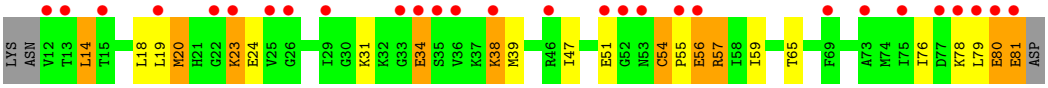


• Molecule 2: Poly(rC)-binding protein 2

Chain A:

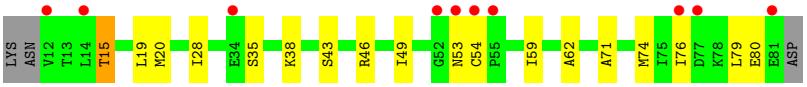
• Molecule 2: Poly(rC)-binding protein 2

Chain B:



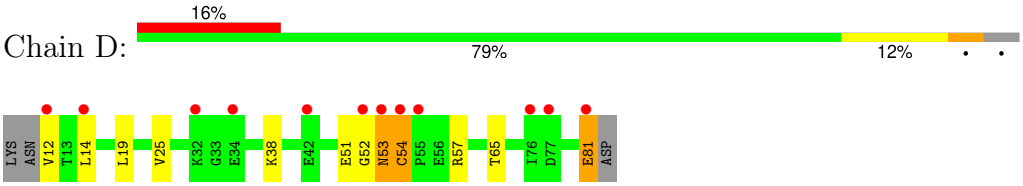
• Molecule 2: Poly(rC)-binding protein 2

Chain C:



• Molecule 2: Poly(rC)-binding protein 2

Chain D:


16%
79%
12%
LYS
ASN
V12
T13
L14
L19
V25
K32
G33
E34
K38
E42
E51
G52
N53
C54
P55
E56
R57
T65
I76
D77
E81
ASP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	92.23Å 58.61Å 71.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 2.12 40.00 – 2.12	Depositor EDS
% Data completeness (in resolution range)	99.4 (40.00-2.12) 99.4 (40.00-2.12)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.47 (at 2.12Å)	Xtriage
Refinement program	REFMAC refmac_5.2.0019, CNS	Depositor
R, R_{free}	0.215 , 0.268 0.227 , 0.279	Depositor DCC
R_{free} test set	1156 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å^2)	40.5	Xtriage
Anisotropy	0.316	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 41.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	2752	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 14.23% 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	E	1.38	1/262 (0.4%)	2.18	17/400 (4.2%)
1	G	1.65	4/262 (1.5%)	2.54	18/400 (4.5%)
2	A	1.16	4/551 (0.7%)	1.02	2/729 (0.3%)
2	B	1.25	4/534 (0.7%)	1.15	2/707 (0.3%)
2	C	0.96	1/534 (0.2%)	0.95	0/707
2	D	0.89	1/534 (0.2%)	0.88	0/707
All	All	1.18	15/2677 (0.6%)	1.42	39/3650 (1.1%)

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
2	A	0	1
2	D	0	1
All	All	0	2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	20	MSE	CG-SE	11.34	2.34	1.95
1	G	510	DT	C3'-O3'	9.14	1.55	1.44
2	B	20	MSE	SE-CE	6.35	2.33	1.95
2	A	57	ARG	CZ-NH1	6.30	1.41	1.33
2	A	69	PHE	CE2-CZ	6.20	1.49	1.37

The worst 5 of 39 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	57	ARG	NE-CZ-NH2	-15.91	112.34	120.30
1	G	509	DC	O4'-C1'-N1	14.08	117.85	108.00
1	G	510	DT	N3-C2-O2	-11.12	115.63	122.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	G	499	DA	P-O3'-C3'	10.14	131.86	119.70
1	G	504	DT	O4'-C1'-N1	9.68	114.78	108.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	A	54	CYS	Peptide
2	D	53	ASN	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	E	235	0	136	2	0
1	G	235	0	136	2	0
2	A	550	0	588	22	0
2	B	533	0	569	35	0
2	C	533	0	569	12	0
2	D	533	0	569	7	0
3	A	28	0	0	3	0
3	B	41	0	0	4	0
3	C	29	0	0	1	0
3	D	16	0	0	0	0
3	E	12	0	0	0	0
3	G	7	0	0	0	0
All	All	2752	0	2567	69	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 13.

The worst 5 of 69 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:20:MSE:SE	2:B:20:MSE:CE	2.32	1.26

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:20:MSE:SE	2:B:20:MSE:CG	2.34	1.26
2:C:43:SER:HB3	2:C:74:MSE:HE1	1.47	0.95
2:C:43:SER:HB3	2:C:74:MSE:CE	2.07	0.83
2:B:20:MSE:HE1	2:B:79:LEU:HD21	1.68	0.76

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
2	A	70/73 (96%)	68 (97%)	1 (1%)	1 (1%)	9 5
2	B	68/73 (93%)	67 (98%)	1 (2%)	0	100 100
2	C	68/73 (93%)	68 (100%)	0	0	100 100
2	D	68/73 (93%)	66 (97%)	1 (2%)	1 (2%)	8 4
All	All	274/292 (94%)	269 (98%)	3 (1%)	2 (1%)	19 15

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	53	ASN
2	A	54	CYS

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
2	A	60/58 (103%)	57 (95%)	3 (5%)	20 19
2	B	58/58 (100%)	51 (88%)	7 (12%)	4 2
2	C	58/58 (100%)	54 (93%)	4 (7%)	13 10
2	D	58/58 (100%)	53 (91%)	5 (9%)	8 5
All	All	234/232 (101%)	215 (92%)	19 (8%)	9 6

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	12	VAL
2	D	51	GLU
2	D	81	GLU
2	D	38	LYS
2	B	80	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	48	ASN
2	B	66	ASN
2	C	53	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 oligosaccharides 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

6.1 Protein, DNA and RNA chains

Warning: The R factor obtained from EDS is 0.3042, which does not match the depositor's R factor of 0.21506. Please interpret the results in this section carefully.

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	E	12/12 (100%)	0.99	1 (8%) 19 21	38, 44, 53, 55	0
1	G	12/12 (100%)	1.16	2 (16%) 5 6	38, 46, 54, 55	0
2	A	69/73 (94%)	1.91	28 (40%) 1 1	32, 39, 56, 72	0
2	B	67/73 (91%)	1.85	28 (41%) 1 1	31, 39, 56, 67	0
2	C	67/73 (91%)	1.19	10 (14%) 7 7	31, 39, 54, 61	0
2	D	67/73 (91%)	1.20	12 (17%) 4 5	32, 38, 57, 71	0
All	All	294/316 (93%)	1.50	81 (27%) 2 2	31, 40, 56, 72	0

The worst 5 of 81 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	12	VAL	4.7
2	A	54	CYS	4.6
2	B	81	GLU	4.4
2	B	55	PRO	4.3
2	D	54	CYS	4.2

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 monosaccharides 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.