wwPDB X-ray Structure Validation Summary Report (i)

Feb 12, 2024 – 03:12 pm GMT

PDB ID	:	8CHV
Title	:	Crystal structure of human PURA (fragment Glu57-Glu212, PUR repeat I and II) R140P mutant
Authors	:	Janowski, R.; Niessing, D.
Deposited on	:	2023-02-08
Resolution	:	2.15 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 (1)) were used in the production of this report:

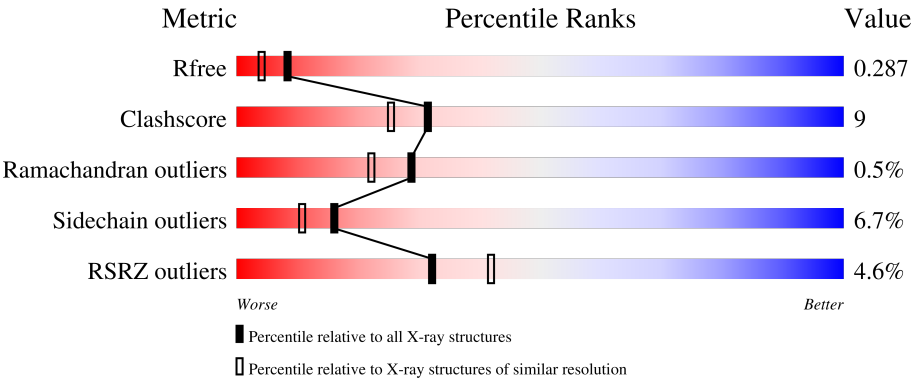
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.15 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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	156	5% 78% 15% • 6%
1	B	156	2% 76% 15% • 6%
1	C	156	6% 65% 18% • • 13%
1	D	156	3% 72% 15% • • 10%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4794 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 Transcriptional activator protein Pur-alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	146	Total	C	N	O	S	0	1	0
			1178	740	218	218	2			
1	B	147	Total	C	N	O	S	0	0	0
			1179	740	217	220	2			
1	C	135	Total	C	N	O	S	0	1	0
			1104	697	203	202	2			
1	D	141	Total	C	N	O	S	0	1	0
			1136	715	209	210	2			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	140	PRO	ARG	engineered mutation	UNP Q00577
B	140	PRO	ARG	engineered mutation	UNP Q00577
C	140	PRO	ARG	engineered mutation	UNP Q00577
D	140	PRO	ARG	engineered mutation	UNP Q00577

- Molecule 2 is water.

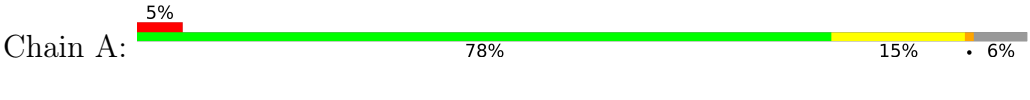
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	72	Total	O	0	0
			72	72		
2	B	58	Total	O	0	0
			58	58		
2	C	28	Total	O	0	0
			28	28		
2	D	39	Total	O	0	0
			39	39		

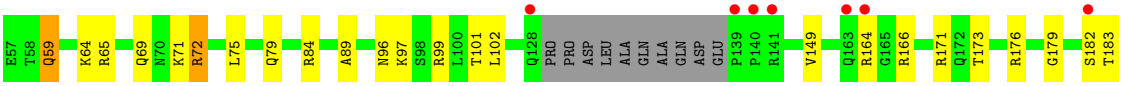
3 Residue-property plots (i)

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 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: Transcriptional activator protein Pur-alpha

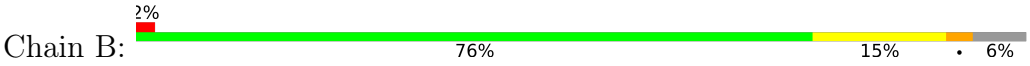
Chain A:

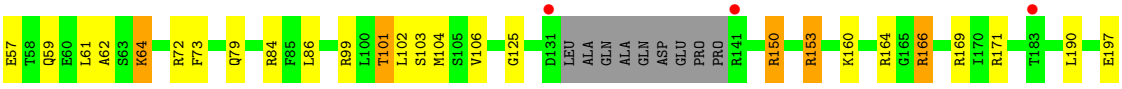




• Molecule 1: Transcriptional activator protein Pur-alpha

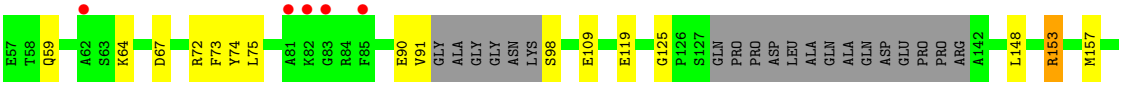
Chain B:

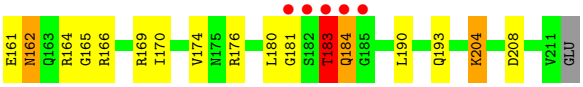




• Molecule 1: Transcriptional activator protein Pur-alpha

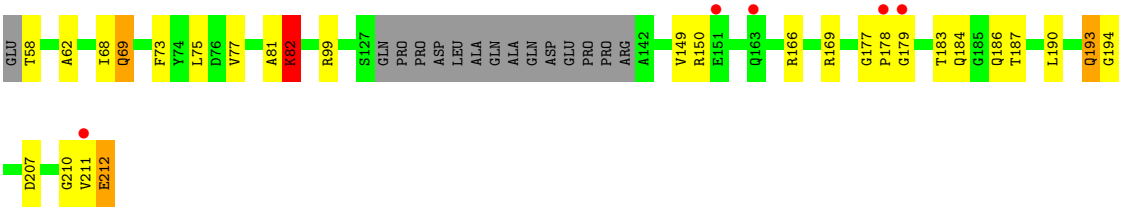
Chain C:





• Molecule 1: Transcriptional activator protein Pur-alpha

Chain D:


GLU T58 A62 I68 Q69 F73 Y74 L75 D76 V77 A81 K82 R99 S127 GLN PRO PRO ASP LEU ALA GLN ALA GLN ASP GLU PRO PRO ARG A142 V149 R150 E151 Q163 R166 R169 G177 P178 G179 T183 Q184 G185 Q186 T187 L190 Q193 G194
D207 G210 V211 E212

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	63.31Å 57.57Å 84.88Å 90.00° 102.33° 90.00°	Depositor
Resolution (Å)	47.29 – 2.15 47.29 – 2.15	Depositor EDS
% Data completeness (in resolution range)	99.4 (47.29-2.15) 99.4 (47.29-2.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.66 (at 2.16Å)	Xtriage
Refinement program	REFMAC 5.8.0352	Depositor
R, R_{free}	0.205 , 0.282 0.208 , 0.287	Depositor DCC
R_{free} test set	1621 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	38.7	Xtriage
Anisotropy	0.243	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 35.2	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	0.95	EDS
Total number of atoms	4794	wwPDB-VP
Average B, all atoms (Å^2)	52.0	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 31.69 % 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 1.0623e-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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

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	0.43	0/1195	0.82	0/1600
1	B	0.44	0/1196	0.87	2/1601 (0.1%)
1	C	0.37	0/1118	0.79	2/1495 (0.1%)
1	D	0.37	0/1151	0.77	1/1540 (0.1%)
All	All	0.40	0/4660	0.81	5/6236 (0.1%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	181	GLY	N-CA-C	-6.23	97.52	113.10
1	C	183	THR	N-CA-C	-5.81	95.31	111.00
1	B	72	ARG	NE-CZ-NH2	-5.27	117.67	120.30
1	B	171	ARG	NE-CZ-NH2	-5.19	117.70	120.30
1	D	166	ARG	NE-CZ-NH2	-5.04	117.78	120.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-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 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	1178	0	1190	25	0
1	B	1179	0	1188	16	0
1	C	1104	0	1121	22	0
1	D	1136	0	1150	28	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	72	0	0	5	0
2	B	58	0	0	1	0
2	C	28	0	0	3	0
2	D	39	0	0	3	0
All	All	4794	0	4649	83	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 9.

The worst 5 of 83 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:A:96:ASN:HD22	1:D:211:VAL:HG11	1.19	1.05
1:D:211:VAL:HG12	1:D:212:GLU:H	1.27	1.00
1:A:96:ASN:ND2	1:D:211:VAL:CG1	2.25	1.00
1:A:96:ASN:O	1:A:96:ASN:OD1	1.84	0.96
1:A:96:ASN:ND2	1:D:211:VAL:HG11	1.85	0.88

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	143/156 (92%)	135 (94%)	8 (6%)	0	100 100
1	B	143/156 (92%)	138 (96%)	5 (4%)	0	100 100
1	C	130/156 (83%)	121 (93%)	9 (7%)	0	100 100
1	D	138/156 (88%)	130 (94%)	5 (4%)	3 (2%)	6 2
All	All	554/624 (89%)	524 (95%)	27 (5%)	3 (0%)	29 22

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	178	PRO
1	D	82	LYS
1	D	210	GLY

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	123/130 (95%)	116 (94%)	7 (6%)	20 16
1	B	123/130 (95%)	115 (94%)	8 (6%)	17 12
1	C	116/130 (89%)	107 (92%)	9 (8%)	12 7
1	D	118/130 (91%)	109 (92%)	9 (8%)	13 8
All	All	480/520 (92%)	447 (93%)	33 (7%)	16 10

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	186	GLN
1	D	193[A]	GLN
1	D	212	GLU
1	B	153	ARG
1	B	150	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	79	GLN
1	B	163	GLN
1	D	184	GLN
1	D	96	ASN
1	D	120	HIS

5.3.3 RNA

There are no RNA molecules 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 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, 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	A	146/156 (93%)	0.08	8 (5%)	25	34	23, 39, 97, 117	0
1	B	147/156 (94%)	-0.06	3 (2%)	65	72	26, 44, 82, 98	3 (2%)
1	C	135/156 (86%)	0.33	10 (7%)	14	20	31, 59, 110, 133	8 (5%)
1	D	141/156 (90%)	0.04	5 (3%)	44	52	29, 51, 95, 114	6 (4%)
All	All	569/624 (91%)	0.09	26 (4%)	32	42	23, 49, 97, 133	17 (2%)

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	183	THR	6.4
1	A	182	SER	5.3
1	C	185	GLY	4.3
1	D	211	VAL	3.8
1	C	184	GLN	3.8

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 (i)

There are no such residues in this entry.