wwPDB X-ray Structure Validation Summary Report (i)

Jul 4, 2022 – 01:11 pm BST

PDB ID	:	7OXX
Title	:	CrabP2 mutant R30AK31A
Authors	:	Tomlinson, C.W.E.; Basle, A.; Pohl, E.
Deposited on	:	2021-06-23
Resolution	:	1.33 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:

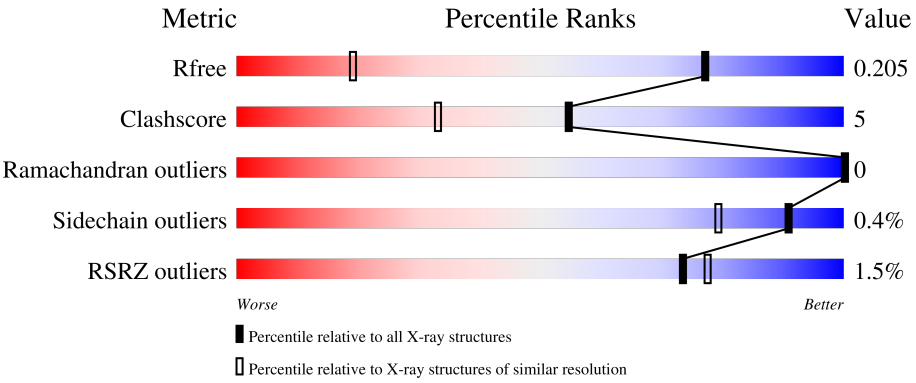
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.29
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 1.33 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	1385 (1.36-1.32)
Clashscore	141614	1417 (1.36-1.32)
Ramachandran outliers	138981	1397 (1.36-1.32)
Sidechain outliers	138945	1397 (1.36-1.32)
RSRZ outliers	127900	1369 (1.36-1.32)

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	138	92% 7%
1	B	138	89% 10%
1	C	138	92% 8%
1	D	138	3% 93% 6%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9389 atoms, of which 4541 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 Cellular retinoic acid-binding protein 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	136	Total	C	H	N	O	S	26	6	0
			2237	705	1130	180	214	8			
1	B	137	Total	C	H	N	O	S	27	6	0
			2240	706	1131	180	215	8			
1	C	138	Total	C	H	N	O	S	26	5	0
			2258	706	1142	186	216	8			
1	D	138	Total	C	H	N	O	S	26	5	0
			2251	706	1138	182	217	8			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	30	ALA	ARG	engineered mutation	UNP P29373
A	31	ALA	LYS	engineered mutation	UNP P29373
B	30	ALA	ARG	engineered mutation	UNP P29373
B	31	ALA	LYS	engineered mutation	UNP P29373
C	30	ALA	ARG	engineered mutation	UNP P29373
C	31	ALA	LYS	engineered mutation	UNP P29373
D	30	ALA	ARG	engineered mutation	UNP P29373
D	31	ALA	LYS	engineered mutation	UNP P29373

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

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

- Molecule 3 is water.

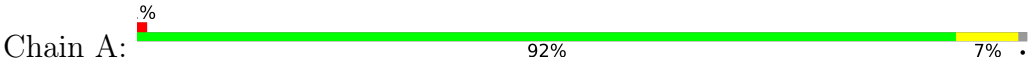
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	106	Total O 106 106	0	0
3	B	117	Total O 117 117	0	0
3	C	95	Total O 95 95	0	0
3	D	81	Total O 81 81	0	0

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 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: Cellular retinoic acid-binding protein 2

Chain A:

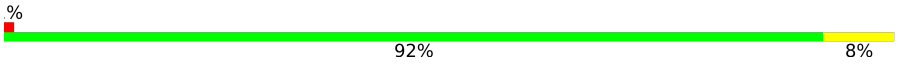


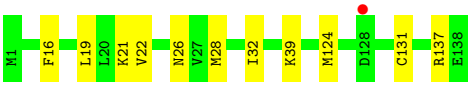
• Molecule 1: Cellular retinoic acid-binding protein 2

Chain B:

• Molecule 1: Cellular retinoic acid-binding protein 2

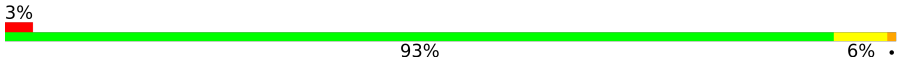
Chain C:

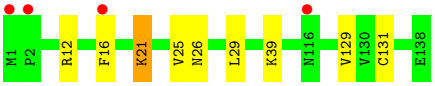




• Molecule 1: Cellular retinoic acid-binding protein 2

Chain D:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	51.17Å 106.02Å 55.54Å 90.00° 101.70° 90.00°	Depositor
Resolution (Å)	54.39 – 1.33 54.39 – 1.33	Depositor EDS
% Data completeness (in resolution range)	96.9 (54.39-1.33) 96.9 (54.39-1.33)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.65 (at 1.33Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.158 , 0.200 0.168 , 0.205	Depositor DCC
R_{free} test set	6503 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	18.2	Xtriage
Anisotropy	0.211	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	9389	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.75% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NA

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	0.82	0/1138	0.99	2/1535 (0.1%)
1	B	0.84	2/1144 (0.2%)	1.01	3/1544 (0.2%)
1	C	0.77	0/1147	0.99	1/1546 (0.1%)
1	D	0.78	0/1144	0.98	0/1543
All	All	0.80	2/4573 (0.0%)	0.99	6/6168 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	14	GLU	CD-OE1	8.28	1.34	1.25
1	B	17	GLU	CD-OE2	5.04	1.31	1.25

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	135	TYR	CB-CG-CD1	-8.44	115.94	121.00
1	B	135	TYR	CB-CG-CD2	7.87	125.72	121.00
1	A	135	TYR	CB-CG-CD1	-6.37	117.18	121.00
1	A	135	TYR	CB-CG-CD2	5.74	124.45	121.00
1	B	60	ARG	NE-CZ-NH1	5.20	122.90	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	1107	1130	1132	22	0
1	B	1109	1131	1135	19	0
1	C	1116	1142	1147	18	0
1	D	1113	1138	1143	22	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	106	0	0	1	0
3	B	117	0	0	2	0
3	C	95	0	0	1	0
3	D	81	0	0	2	0
All	All	4848	4541	4557	47	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 5.

The worst 5 of 47 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:39:LYS:H	1:D:21[B]:LYS:HD3	1.22	1.04
1:B:39:LYS:H	1:C:21[A]:LYS:HD3	1.30	0.94
1:A:21:LYS:HD3	1:D:39:LYS:H	1.34	0.89
1:B:39:LYS:HB3	1:C:21[A]:LYS:HE2	1.60	0.81
1:B:21:LYS:HE3	3:C:346:HOH:O	1.83	0.77

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	140/138 (101%)	139 (99%)	1 (1%)	0	100	100
1	B	141/138 (102%)	140 (99%)	1 (1%)	0	100	100
1	C	141/138 (102%)	140 (99%)	1 (1%)	0	100	100
1	D	141/138 (102%)	139 (99%)	2 (1%)	0	100	100
All	All	563/552 (102%)	558 (99%)	5 (1%)	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	126/122 (103%)	126 (100%)	0	100	100
1	B	127/122 (104%)	126 (99%)	1 (1%)	81	57
1	C	127/122 (104%)	127 (100%)	0	100	100
1	D	127/122 (104%)	125 (98%)	2 (2%)	62	29
All	All	507/488 (104%)	504 (99%)	3 (1%)	91	67

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	114	LEU
1	D	21[A]	LYS
1	D	21[B]	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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

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	136/138 (98%)	-0.13	1 (0%)	87	89	15, 22, 38, 56	0
1	B	137/138 (99%)	-0.08	2 (1%)	73	77	16, 22, 34, 52	0
1	C	138/138 (100%)	0.00	1 (0%)	87	89	16, 25, 44, 56	0
1	D	138/138 (100%)	0.14	4 (2%)	51	58	16, 27, 51, 67	0
All	All	549/552 (99%)	-0.02	8 (1%)	73	77	15, 24, 43, 67	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	1	MET	3.2
1	B	2	PRO	3.1
1	C	128	ASP	2.5
1	D	16	PHE	2.4
1	D	2	PRO	2.3

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

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	NA	A	201	1/1	1.00	0.17	16,16,16,16	0
2	NA	B	201	1/1	1.00	0.06	19,19,19,19	0
2	NA	C	201	1/1	1.00	0.05	19,19,19,19	0
2	NA	D	201	1/1	1.00	0.13	26,26,26,26	0

6.5 Other polymers (i)

There are no such residues in this entry.