wwPDB X-ray Structure Validation Summary Report (i)

May 23, 2022 – 01:14 pm BST

PDB ID	:	7OHZ
Title	:	Crystal structure of AP2 Mu2 - FCHO2 chimera (His6-tagged)
Authors	:	Zaccai, N.R.; Kelly, B.T.; Evans, P.R.; Owen, D.J.
Deposited on	:	2021-05-11
Resolution	:	2.27 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 following versions of software and data (see references (1)) were used in the production of this report:

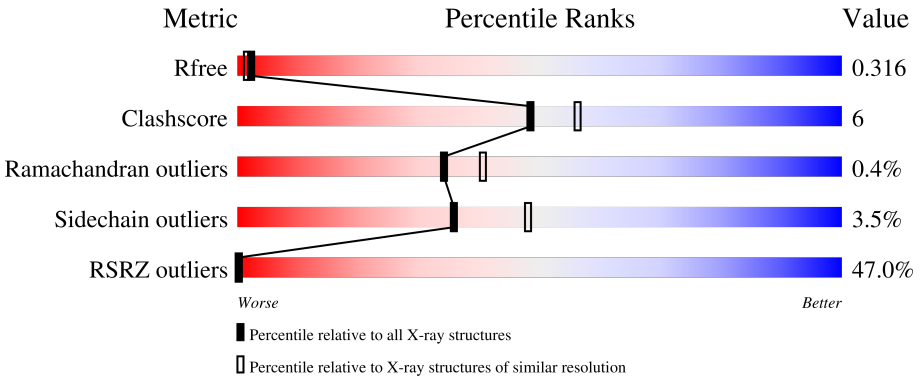
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.28.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

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 2.27 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	6980 (2.30-2.26)
Clashscore	141614	7711 (2.30-2.26)
Ramachandran outliers	138981	7597 (2.30-2.26)
Sidechain outliers	138945	7598 (2.30-2.26)
RSRZ outliers	127900	6849 (2.30-2.26)

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	361	36% 69% 9% · 21%
1	B	361	38% 66% 12% · 21%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4761 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 AP-2 complex subunit mu,F-BAR domain only protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	285	Total	C	N	O	S	0	0	0
			2307	1475	409	409	14			
1	A	285	Total	C	N	O	S	0	0	0
			2309	1476	410	409	14			

There are 94 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	151	MET	-	initiating methionine	UNP P84092
B	152	HIS	-	expression tag	UNP P84092
B	153	HIS	-	expression tag	UNP P84092
B	154	HIS	-	expression tag	UNP P84092
B	155	HIS	-	expression tag	UNP P84092
B	156	HIS	-	expression tag	UNP P84092
B	157	HIS	-	expression tag	UNP P84092
B	436	GLY	-	linker	UNP P84092
B	437	ALA	-	linker	UNP P84092
B	438	SER	-	linker	UNP P84092
B	439	GLY	-	linker	UNP P84092
B	440	SER	-	linker	UNP P84092
B	441	ALA	-	linker	UNP P84092
B	442	GLY	-	linker	UNP P84092
B	443	SER	-	linker	UNP P84092
B	444	ALA	-	linker	UNP P84092
B	445	GLY	-	linker	UNP P84092
B	446	PRO	-	linker	UNP P84092
B	447	SER	-	linker	UNP P84092
B	448	GLY	-	linker	UNP P84092
B	449	ALA	-	linker	UNP P84092
B	450	GLY	-	linker	UNP P84092
B	451	SER	-	linker	UNP P84092
B	452	ALA	-	linker	UNP P84092
B	453	GLY	-	linker	UNP P84092

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	454	SER	-	linker	UNP P84092
B	455	ALA	-	linker	UNP P84092
B	456	GLY	-	linker	UNP P84092
B	457	PRO	-	linker	UNP P84092
B	458	SER	-	linker	UNP P84092
B	459	ALA	-	linker	UNP P84092
B	460	GLY	-	linker	UNP P84092
B	461	SER	-	linker	UNP P84092
B	462	ALA	-	linker	UNP P84092
B	463	GLY	-	linker	UNP P84092
B	464	SER	-	linker	UNP P84092
B	465	ALA	-	linker	UNP P84092
B	466	GLY	-	linker	UNP P84092
B	467	SER	-	linker	UNP P84092
B	468	GLY	-	linker	UNP P84092
B	469	SER	-	linker	UNP P84092
B	470	ALA	-	linker	UNP P84092
B	471	GLY	-	linker	UNP P84092
B	472	SER	-	linker	UNP P84092
B	473	ALA	-	linker	UNP P84092
B	474	PRO	-	linker	UNP P84092
B	475	GLY	-	linker	UNP P84092
A	151	MET	-	initiating methionine	UNP P84092
A	152	HIS	-	expression tag	UNP P84092
A	153	HIS	-	expression tag	UNP P84092
A	154	HIS	-	expression tag	UNP P84092
A	155	HIS	-	expression tag	UNP P84092
A	156	HIS	-	expression tag	UNP P84092
A	157	HIS	-	expression tag	UNP P84092
A	436	GLY	-	linker	UNP P84092
A	437	ALA	-	linker	UNP P84092
A	438	SER	-	linker	UNP P84092
A	439	GLY	-	linker	UNP P84092
A	440	SER	-	linker	UNP P84092
A	441	ALA	-	linker	UNP P84092
A	442	GLY	-	linker	UNP P84092
A	443	SER	-	linker	UNP P84092
A	444	ALA	-	linker	UNP P84092
A	445	GLY	-	linker	UNP P84092
A	446	PRO	-	linker	UNP P84092
A	447	SER	-	linker	UNP P84092
A	448	GLY	-	linker	UNP P84092

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	449	ALA	-	linker	UNP P84092
A	450	GLY	-	linker	UNP P84092
A	451	SER	-	linker	UNP P84092
A	452	ALA	-	linker	UNP P84092
A	453	GLY	-	linker	UNP P84092
A	454	SER	-	linker	UNP P84092
A	455	ALA	-	linker	UNP P84092
A	456	GLY	-	linker	UNP P84092
A	457	PRO	-	linker	UNP P84092
A	458	SER	-	linker	UNP P84092
A	459	ALA	-	linker	UNP P84092
A	460	GLY	-	linker	UNP P84092
A	461	SER	-	linker	UNP P84092
A	462	ALA	-	linker	UNP P84092
A	463	GLY	-	linker	UNP P84092
A	464	SER	-	linker	UNP P84092
A	465	ALA	-	linker	UNP P84092
A	466	GLY	-	linker	UNP P84092
A	467	SER	-	linker	UNP P84092
A	468	GLY	-	linker	UNP P84092
A	469	SER	-	linker	UNP P84092
A	470	ALA	-	linker	UNP P84092
A	471	GLY	-	linker	UNP P84092
A	472	SER	-	linker	UNP P84092
A	473	ALA	-	linker	UNP P84092
A	474	PRO	-	linker	UNP P84092
A	475	GLY	-	linker	UNP P84092

- Molecule 2 is water.

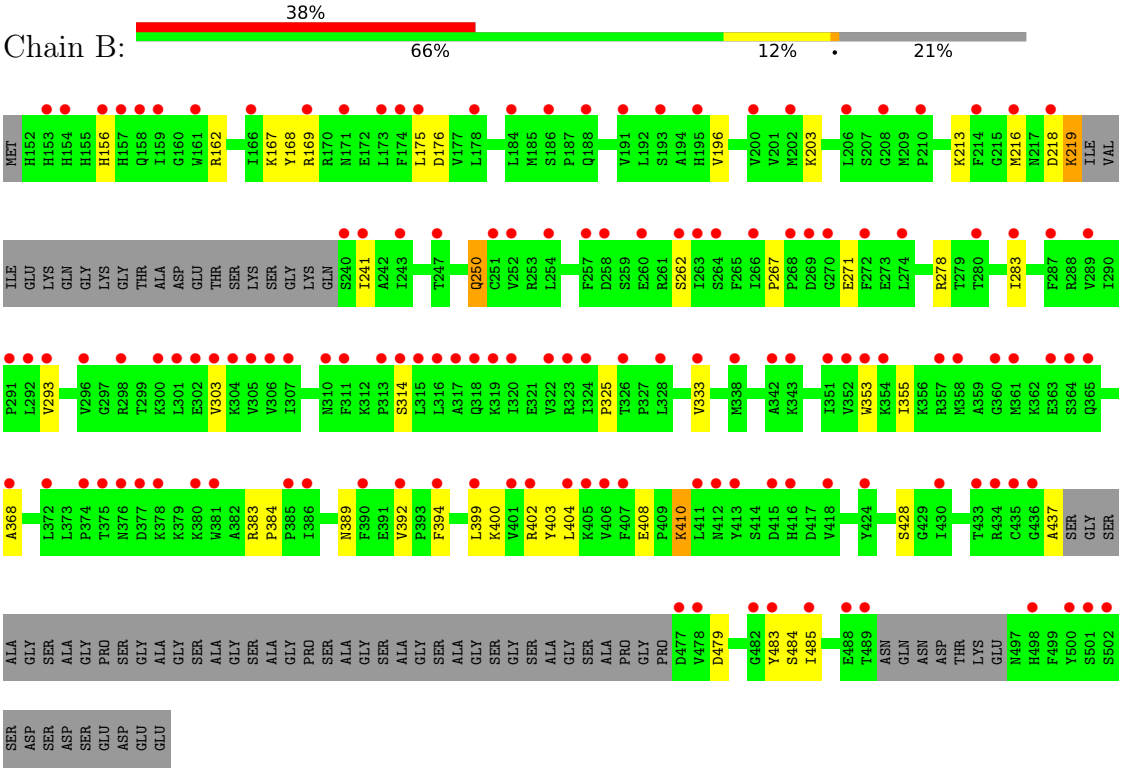
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	62	Total O 62 62	0	0
2	A	83	Total O 83 83	0	0

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.

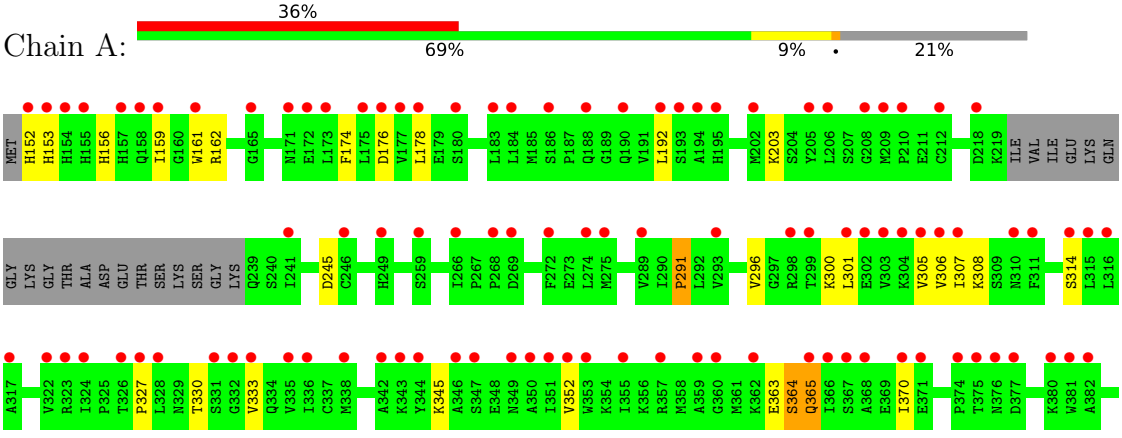
• Molecule 1: AP-2 complex subunit mu,F-BAR domain only protein 2

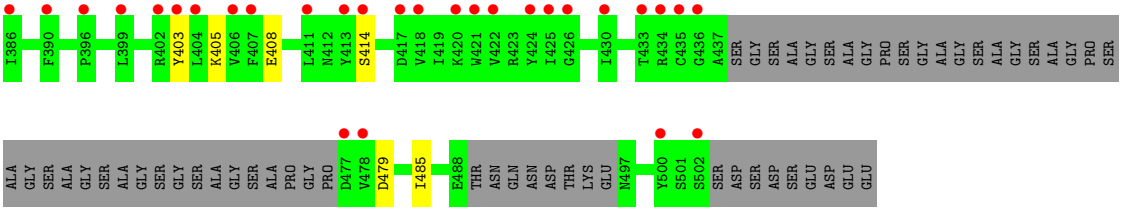
Chain B:



• Molecule 1: AP-2 complex subunit mu,F-BAR domain only protein 2

Chain A:




I386
F390
P396
L399
R402
Y403
L404
K405
V406
F407
E408
L411
N412
Y413
S414
D417
V418
I419
K420
W421
V422
R423
Y424
I425
G426
I430
T433
R434
C435
G436
A437
SER GLY SER ALA GLY SER ALA GLY PRO SER GLY ALA GLY SER ALA GLY SER ALA GLY PRO SER
ALA GLY SER ALA GLY SER ALA GLY SER GLY SER ALA GLY SER ALA PRO GLY PRO
D477
V478
D479
I485
E488
THR ASN GLN ASN ASP THR LYS GLU
N497
Y500
S501
S502
SER ASP SER ASP SER GLU ASP GLU GLU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	55.68Å 129.86Å 64.42Å 90.00° 102.47° 90.00°	Depositor
Resolution (Å)	62.90 – 2.27 62.90 – 2.27	Depositor EDS
% Data completeness (in resolution range)	95.9 (62.90-2.27) 97.0 (62.90-2.27)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.77 (at 2.27Å)	Xtriage
Refinement program	REFMAC 5.8.0257, PHENIX 1.19rc1_4016	Depositor
R, R_{free}	0.267 , 0.314 0.270 , 0.316	Depositor DCC
R_{free} test set	1989 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å^2)	38.1	Xtriage
Anisotropy	0.698	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	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	0.90	EDS
Total number of atoms	4761	wwPDB-VP
Average B, all atoms (Å^2)	53.0	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 32.09 % 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 9.9390e-04. 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	0.29	0/2361	0.54	0/3175
1	B	0.25	0/2359	0.53	0/3173
All	All	0.27	0/4720	0.54	0/6348

There are no bond length outliers.

There are no bond angle outliers.

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	2309	0	2347	27	0
1	B	2307	0	2346	30	0
2	A	83	0	0	0	0
2	B	62	0	0	0	0
All	All	4761	0	4693	56	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 6.

The worst 5 of 56 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:333:VAL:HG11	1:A:370:ILE:HG13	1.28	1.13

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:333:VAL:HG12	1:A:370:ILE:HA	1.05	1.05
1:A:333:VAL:CG1	1:A:370:ILE:HG13	1.92	0.98
1:A:333:VAL:CG1	1:A:370:ILE:HA	1.96	0.95
1:A:333:VAL:HG12	1:A:370:ILE:CA	1.97	0.92

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	277/361 (77%)	259 (94%)	16 (6%)	2 (1%)	22 25
1	B	277/361 (77%)	262 (95%)	15 (5%)	0	100 100
All	All	554/722 (77%)	521 (94%)	31 (6%)	2 (0%)	34 40

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	153	HIS
1	A	291	PRO

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.

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	258/307 (84%)	251 (97%)	7 (3%)	44	59
1	B	258/307 (84%)	247 (96%)	11 (4%)	29	38
All	All	516/614 (84%)	498 (96%)	18 (4%)	36	48

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	364	SER
1	A	414	SER
1	A	408	GLU
1	B	402	ARG
1	A	314	SER

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)

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

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	285/361 (78%)	2.21	130 (45%)	0	0	22, 48, 90, 135	0
1	B	285/361 (78%)	2.21	138 (48%)	0	0	22, 49, 85, 119	0
All	All	570/722 (78%)	2.21	268 (47%)	0	0	22, 49, 88, 135	0

The worst 5 of 268 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	328	LEU	11.6
1	A	377	ASP	10.6
1	B	351	ILE	10.2
1	B	343	LYS	8.9
1	A	352	VAL	8.0

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.