Full wwPDB X-ray Structure Validation Report ⓘ

Oct 14, 2023 – 08:19 PM EDT

PDB ID	:	7UA2
Title	:	Pfs230 D1 domain in complex with 230AL-18
Authors	:	Tang, W.K.; Tolia, N.H.
Deposited on	:	2022-03-11
Resolution	:	2.19 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ 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 ⓘ) were used in the production of this report:

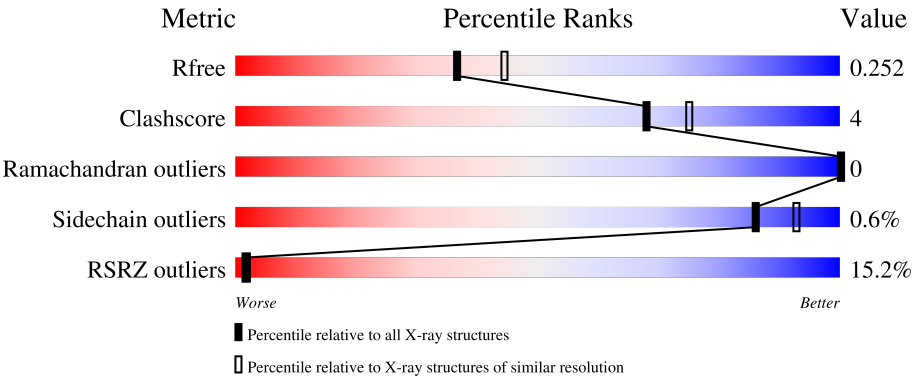
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.19 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	191	6% 84% 7% 9%
2	B	261	30% 80% 6% 14%
3	H	260	3% 77% 11% 12%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 9755 atoms, of which 4809 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 Gametocyte surface protein P230.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	174	Total	C	H	N	O	S	0	0	0
			2795	891	1404	217	279	4			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	585	GLN	ASN	engineered mutation	UNP P68874

- Molecule 2 is a protein called LMIV230-01.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	224	Total	C	H	N	O	S	0	0	0
			3413	1091	1692	290	333	7			

- Molecule 3 is a protein called 230AL-18.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	H	229	Total	C	H	N	O	S	0	0	0
			3478	1110	1713	307	340	8			

- Molecule 4 is water.

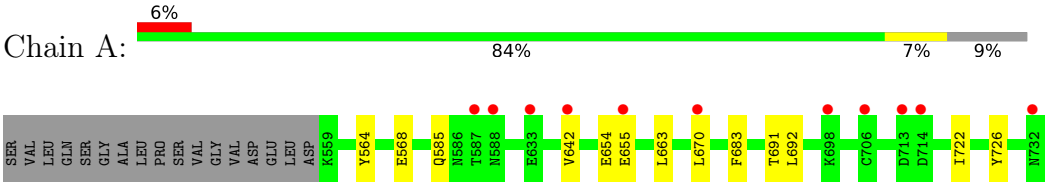
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	18	Total	O	0	0
			18	18		
4	B	9	Total	O	0	0
			9	9		
4	H	42	Total	O	0	0
			42	42		

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.

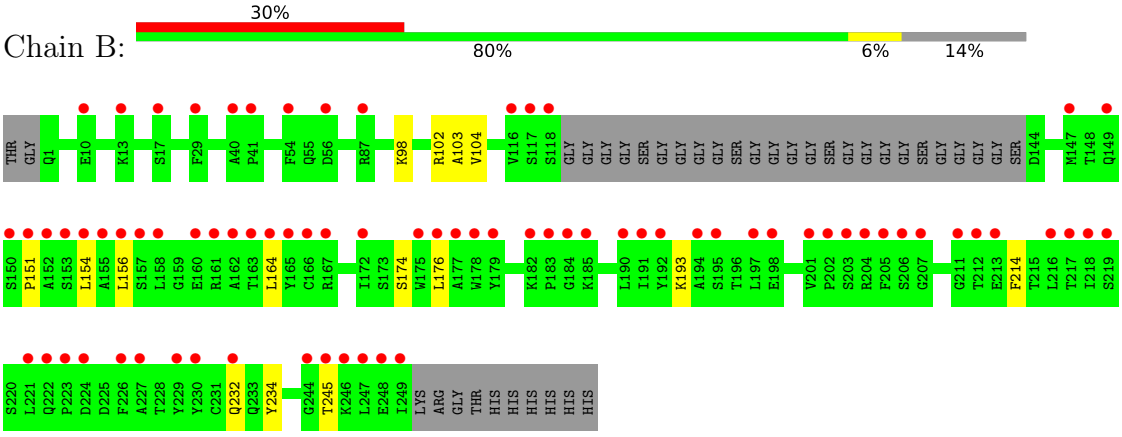
• Molecule 1: Gametocyte surface protein P230

Chain A:



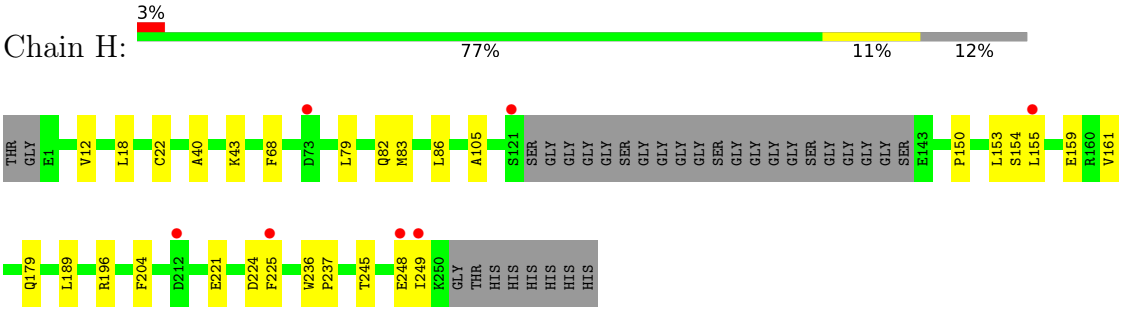
• Molecule 2: LMIV230-01

Chain B:



• Molecule 3: 230AL-18

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	152.65Å 77.93Å 58.78Å 90.00° 99.22° 90.00°	Depositor
Resolution (Å)	36.56 – 2.19 46.17 – 2.19	Depositor EDS
% Data completeness (in resolution range)	97.1 (36.56-2.19) 97.2 (46.17-2.19)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.16 (at 2.20Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.222 , 0.256 0.220 , 0.252	Depositor DCC
R_{free} test set	1703 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	43.7	Xtriage
Anisotropy	0.711	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 43.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	9755	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.05% 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

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	0.31	0/1417	0.52	0/1916
2	B	0.29	0/1760	0.52	0/2385
3	H	0.37	0/1805	0.61	0/2446
All	All	0.33	0/4982	0.55	0/6747

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	1391	1404	1403	7	0
2	B	1721	1692	1693	9	0
3	H	1765	1713	1715	22	0
4	A	18	0	0	0	0
4	B	9	0	0	0	0
4	H	42	0	0	0	0
All	All	4946	4809	4811	37	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 4.

All (37) 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)
3:H:153:LEU:HD11	3:H:161:VAL:HG21	1.73	0.69
1:A:568:GLU:O	1:A:722:ILE:HD11	1.97	0.65
3:H:40:ALA:HB3	3:H:43:LYS:HD3	1.82	0.61
2:B:154:LEU:HD22	2:B:156:LEU:HD11	1.83	0.61
3:H:155:LEU:CD1	3:H:161:VAL:HB	2.31	0.59
2:B:164:LEU:HG	2:B:245:THR:HG21	1.86	0.58
3:H:225:PHE:CD1	3:H:248:GLU:HA	2.40	0.57
1:A:663:LEU:HG	1:A:670:LEU:HD22	1.88	0.56
3:H:221:GLU:O	3:H:249:ILE:HD13	2.05	0.56
1:A:670:LEU:HD13	1:A:692:LEU:HD23	1.88	0.56
3:H:179:GLN:HB2	3:H:189:LEU:HD11	1.89	0.55
3:H:224:ASP:HB2	3:H:249:ILE:HD11	1.89	0.54
1:A:670:LEU:HD12	1:A:691:THR:O	2.09	0.53
3:H:224:ASP:HB2	3:H:249:ILE:CD1	2.40	0.51
2:B:174:SER:O	2:B:193:LYS:O	2.28	0.51
3:H:150:PRO:O	3:H:245:THR:HG23	2.11	0.51
3:H:83:MET:HB3	3:H:86:LEU:HD21	1.91	0.51
3:H:155:LEU:HD11	3:H:161:VAL:HB	1.91	0.51
2:B:176:LEU:HD22	2:B:214:PHE:CG	2.45	0.50
3:H:18:LEU:HB2	3:H:83:MET:HE3	1.93	0.49
3:H:12:VAL:HG11	3:H:86:LEU:CD1	2.43	0.48
3:H:22:CYS:HB3	3:H:79:LEU:HB3	1.95	0.48
3:H:12:VAL:HG11	3:H:86:LEU:HD13	1.95	0.47
2:B:176:LEU:HD12	2:B:232:GLN:O	2.15	0.47
3:H:155:LEU:HD22	3:H:159:GLU:OE1	2.17	0.45
2:B:151:PRO:O	2:B:245:THR:HG23	2.16	0.45
1:A:564:TYR:OH	1:A:642:VAL:HG11	2.18	0.44
3:H:196:ARG:HD3	3:H:204:PHE:O	2.18	0.44
2:B:102:ARG:O	2:B:103:ALA:HB3	2.19	0.43
2:B:176:LEU:O	2:B:193:LYS:O	2.36	0.43
1:A:585:GLN:HG3	3:H:105:ALA:HB2	2.01	0.43
2:B:104:VAL:HG13	2:B:234:TYR:CZ	2.55	0.42
1:A:654:GLU:O	1:A:655:GLU:CG	2.67	0.42
3:H:68:PHE:HA	3:H:82:GLN:O	2.20	0.42
3:H:225:PHE:HD1	3:H:248:GLU:HA	1.85	0.41
3:H:236:TRP:HB2	3:H:237:PRO:HA	2.01	0.41
3:H:154:SER:HA	3:H:248:GLU:O	2.22	0.40

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	172/191 (90%)	168 (98%)	4 (2%)	0	100	100
2	B	220/261 (84%)	212 (96%)	8 (4%)	0	100	100
3	H	225/260 (86%)	217 (96%)	8 (4%)	0	100	100
All	All	617/712 (87%)	597 (97%)	20 (3%)	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	163/177 (92%)	161 (99%)	2 (1%)	71	83
2	B	185/200 (92%)	184 (100%)	1 (0%)	88	94
3	H	190/203 (94%)	190 (100%)	0	100	100
All	All	538/580 (93%)	535 (99%)	3 (1%)	86	93

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

Mol	Chain	Res	Type
1	A	683	PHE
1	A	726	TYR
2	B	98	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are

no such sidechains identified.

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	174/191 (91%)	1.06	11 (6%) 20 19	39, 56, 90, 112	0
2	B	224/261 (85%)	1.90	77 (34%) 0 0	45, 87, 156, 200	0
3	H	229/260 (88%)	0.87	7 (3%) 49 47	34, 44, 62, 85	0
All	All	627/712 (88%)	1.29	95 (15%) 2 2	34, 56, 120, 200	0

All (95) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	226	PHE	12.8
2	B	161	ARG	7.7
2	B	221	LEU	7.7
2	B	165	TYR	7.5
2	B	163	THR	6.9
2	B	155	ALA	6.7
2	B	249	ILE	6.5
2	B	248	GLU	6.4
2	B	172	ILE	5.6
2	B	218	ILE	5.4
2	B	224	ASP	5.3
2	B	201	VAL	5.2
2	B	40	ALA	5.1
2	B	246	LYS	5.0
2	B	162	ALA	5.0
2	B	212	THR	4.9
2	B	118	SER	4.8
2	B	164	LEU	4.6
2	B	217	THR	4.4
2	B	229	TYR	4.4
2	B	150	SER	4.3
2	B	191	ILE	4.0
2	B	152	ALA	4.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	213	GLU	3.9
2	B	190	LEU	3.8
1	A	633	GLU	3.8
2	B	154	LEU	3.7
2	B	216	LEU	3.7
2	B	204	ARG	3.6
3	H	249	ILE	3.6
2	B	227	ALA	3.5
2	B	194	ALA	3.4
2	B	195	SER	3.4
2	B	223	PRO	3.4
2	B	156	LEU	3.4
2	B	207	GLY	3.3
2	B	178	TRP	3.3
2	B	202	PRO	3.3
2	B	222	GLN	3.2
2	B	158	LEU	3.1
2	B	197	LEU	3.1
2	B	247	LEU	3.1
3	H	73	ASP	3.0
2	B	149	GLN	3.0
2	B	117	SER	3.0
2	B	219	SER	2.9
2	B	176	LEU	2.9
3	H	248	GLU	2.9
1	A	714	ASP	2.8
2	B	205	PHE	2.8
2	B	153	SER	2.8
2	B	157	SER	2.8
2	B	166	CYS	2.8
3	H	121	SER	2.7
2	B	179	TYR	2.6
1	A	732	ASN	2.6
1	A	588	ASN	2.6
1	A	670	LEU	2.6
2	B	198	GLU	2.6
3	H	225	PHE	2.6
1	A	655	GLU	2.5
1	A	713	ASP	2.5
2	B	232	GLN	2.4
2	B	10	GLU	2.4
3	H	212	ASP	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
2	B	184	GLY	2.4
2	B	244	GLY	2.4
2	B	211	GLY	2.4
2	B	147	MET	2.4
2	B	230	TYR	2.4
2	B	182	LYS	2.3
2	B	41	PRO	2.3
2	B	87	ARG	2.3
2	B	175	TRP	2.3
2	B	183	PRO	2.3
2	B	17	SER	2.3
2	B	245	THR	2.3
2	B	13	LYS	2.2
2	B	167	ARG	2.2
2	B	203	SER	2.2
1	A	706	CYS	2.2
2	B	192	TYR	2.2
2	B	151	PRO	2.2
1	A	587	THR	2.2
2	B	206	SER	2.1
2	B	29	PHE	2.1
1	A	698	LYS	2.1
3	H	155	LEU	2.1
2	B	56	ASP	2.1
2	B	160	GLU	2.1
1	A	642	VAL	2.0
2	B	185	LYS	2.0
2	B	177	ALA	2.0
2	B	54	PHE	2.0
2	B	116	VAL	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

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.