W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Oct 9, 2023 – 02:15 AM EDT

PDB ID	:	6WJL
Title	:	Crystal structure of Glypican-2 core protein in complex with D3 Fab
Authors	:	Raman, S.; Maris, J.M.; Bosse, K.R.; Julien, J.P.
Deposited on	:	2020-04-14
Resolution	:	3.30 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

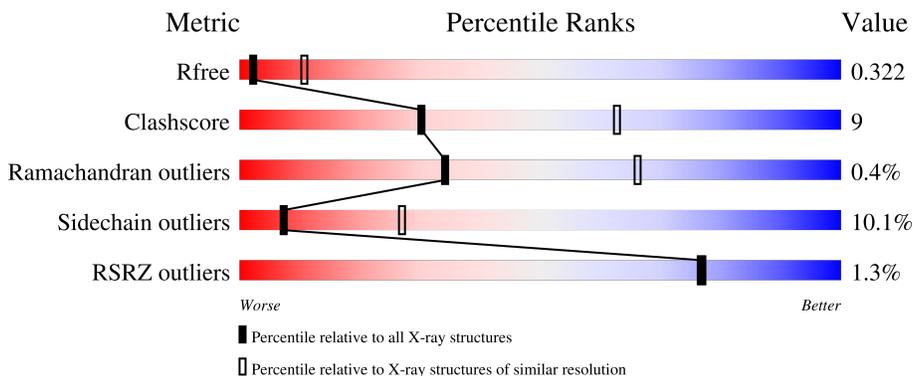|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1   Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.30 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	Percentile Ranks	Value
Rfree		0.322
Clashscore		9
Ramachandran outliers		0.4%
Sidechain outliers		10.1%
RSRZ outliers		1.3%

Worse   *Better*

▊ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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	E	482	2% / 60% 20% 5% 16%
1	G	482	.% / 59% 22% 6% 13%
2	F	220	75% 23% ••
2	H	220	78% 20% ••
3	I	121	6% / 81% 13% 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	K	121	2% 72% 20% 5%
4	J	214	86% 14%
4	L	214	82% 16%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 14241 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 Glypican-2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	G	419	Total 3209	C 2021	N 576	O 591	S 21	0	0	0
1	E	406	Total 3111	C 1958	N 559	O 573	S 21	0	0	0

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	21	GLU	-	expression tag	UNP Q8N158
G	22	THR	-	expression tag	UNP Q8N158
G	55	THR	SER	engineered mutation	UNP Q8N158
G	92	THR	SER	engineered mutation	UNP Q8N158
G	155	THR	SER	engineered mutation	UNP Q8N158
G	494	GLY	-	expression tag	UNP Q8N158
G	495	THR	-	expression tag	UNP Q8N158
G	496	LYS	-	expression tag	UNP Q8N158
G	497	HIS	-	expression tag	UNP Q8N158
G	498	HIS	-	expression tag	UNP Q8N158
G	499	HIS	-	expression tag	UNP Q8N158
G	500	HIS	-	expression tag	UNP Q8N158
G	501	HIS	-	expression tag	UNP Q8N158
G	502	HIS	-	expression tag	UNP Q8N158
E	21	GLU	-	expression tag	UNP Q8N158
E	22	THR	-	expression tag	UNP Q8N158
E	55	THR	SER	engineered mutation	UNP Q8N158
E	92	THR	SER	engineered mutation	UNP Q8N158
E	155	THR	SER	engineered mutation	UNP Q8N158
E	494	GLY	-	expression tag	UNP Q8N158
E	495	THR	-	expression tag	UNP Q8N158
E	496	LYS	-	expression tag	UNP Q8N158
E	497	HIS	-	expression tag	UNP Q8N158
E	498	HIS	-	expression tag	UNP Q8N158
E	499	HIS	-	expression tag	UNP Q8N158

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	500	HIS	-	expression tag	UNP Q8N158
E	501	HIS	-	expression tag	UNP Q8N158
E	502	HIS	-	expression tag	UNP Q8N158

- Molecule 2 is a protein called D3 Fab Heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	218	Total 1639	C 1039	N 268	O 326	S 6	0	0	0
2	F	218	Total 1639	C 1039	N 268	O 326	S 6	0	0	0

- Molecule 3 is a protein called VHH domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	K	115	Total 670	C 407	N 126	O 135	S 2	0	0	0
3	I	114	Total 687	C 420	N 129	O 136	S 2	0	0	0

- Molecule 4 is a protein called D3 Fab Light Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	L	213	Total 1643	C 1032	N 275	O 331	S 5	0	0	0
4	J	213	Total 1643	C 1032	N 275	O 331	S 5	0	0	0

3   Residue-property plots ⓘ

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: Glypican-2

Chain G:

- Molecule 1: Glypican-2

Chain E:

L439 V443 P447 Q450 V451 N452 N453 L456 K457 V458 G462 P463 D464 V465 P466 T467 R468 R469 R470 R471 L474 R475 T478 A479 R480 M481 K482 L490 ASP GLY GLN GLY THR LYS HIS HIS HIS HIS HIS

● Molecule 2: D3 Fab Heavy chain

Chain H:
78%   20%

E1 C22 D31 Y32 M34 S35 Q39 L45 Y50 I51 S52 A60 D61 S62 V63 F67 R71 L78 Y79 L80 Q81 M82 L82C D86 T87 Y90 A93 R94 E95 S96 G97 F100B L108 V109 A110 P119 S120 V121 L124 S128 LYS

SER T131 L138 L141 V142 P147 V163 S179 V182 T183 L189 T193 P202 K206 K209 K210 V211 E212 P213 K214

● Molecule 2: D3 Fab Heavy chain

Chain F:
75%   23%

E1 V2 V5 V11 V12 C22 A23 F27 T28 Y33 W36 Q39 L45 Y50 I51 S52 T56 Y59 R66 R71 L78 Y79 L80 Q81 M82 N82A S82B L82C V89 A93 R94 E95 S96 G97 Y98 D99 Y100 V100A F100B D101 L108 V111

P119 L124 S128 LYS SER T131 L141 V142 F146 V150 T151 V152 N155 L159 V163 L175 S180 V181 V182 T183 T193 Y194 I195 H200 K210 K214

● Molecule 3: VHH domain

Chain K:
2%
72%   20%   5%

UNK VAL GLN LEU UNK X6 S7 G8 G9 G10 X11 V12 Q13 X16 S17 L18 S21 X33 W36 V48 X51 X52 X61 F68 T69 X70 D73 X76 X80 L81 Q82 X83 X84 X85 L86 X87 X88 X106 V117 T118 V119 S120 SER

● Molecule 3: VHH domain

Chain I:
6%
81%   13%   6%
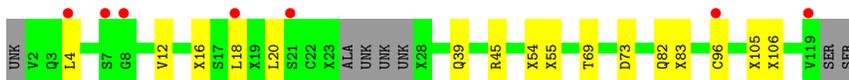
UNK V2 Q3 L4 S7 G8 V12 X16 S17 L18 X19 L20 S21 C22 X23 ALA UNK UNK X28 Q39 R45 X54 X55 T69 D73 Q82 X83 C96 X105 X106 V119 SER SER

● Molecule 4: D3 Fab Light Chain

Chain L:
82%   16%

D1 F14 V15 G16 D17 R24 I29 S30 S31 W32 L33 Q37 L47 A50 A51 S52 G64 E70 F71 L78 Q89 Q90 L91 Y94 P95 F118 E123 K145 N152 Q155 N158 H189 K190 T197 H198 Q199 P204 V205 R211

G212 E213 CYS

● Molecule 4: D3 Fab Light Chain

Chain J:

86%　　　　　　　14%

D1 I2 Q3 G16 T22 S26 W32 L46 Y49 T72 L78 Q90 L91 Y94 P95 I96 P113 S114 V115 V132 F139 V150 D151 Q155 N158 L179 H189 K190 T197 H198 Q199 P204 K207 R211 G212 E213 CYS

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	54.94Å 225.65Å 107.25Å 90.00° 104.75° 90.00°	Depositor
Resolution (Å)	29.48 – 3.30 29.48 – 3.30	Depositor EDS
% Data completeness (in resolution range)	99.9 (29.48-3.30) 95.2 (29.48-3.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.59 (at 3.31Å)	Xtriage
Refinement program	PHENIX 1.16_3549	Depositor
R, R_{free}	0.278 , 0.317 0.284 , 0.322	Depositor DCC
R_{free} test set	1833 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å^2)	72.4	Xtriage
Anisotropy	0.359	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 20.3	EDS
L-test for twinning[2]	$< |L| > = 0.45, < L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.390 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	14241	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 3.63% 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	E	0.29	0/3169	0.48	0/4287
1	G	0.34	1/3271 (0.0%)	0.54	0/4431
2	F	0.25	0/1679	0.47	0/2286
2	H	0.25	0/1679	0.47	0/2286
3	I	0.28	0/332	0.46	0/413
3	K	0.42	0/304	0.57	0/376
4	J	0.25	0/1679	0.45	0/2281
4	L	0.26	0/1679	0.46	0/2281
All	All	0.29	1/13792 (0.0%)	0.49	0/18641

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	175	PHE	C-N	8.79	1.50	1.34

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	E	3111	0	3077	55	0
1	G	3209	0	3171	81	0
2	F	1639	0	1591	29	0
2	H	1639	0	1591	28	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	I	687	0	400	16	0
3	K	670	0	366	22	0
4	J	1643	0	1604	22	0
4	L	1643	0	1604	23	0
All	All	14241	0	13404	257	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 257 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:I:12:VAL:HG21	3:I:16:UNK:CB	1.56	1.35
3:I:12:VAL:CG2	3:I:16:UNK:CB	2.11	1.29
3:I:106:UNK:CB	4:J:198:HIS:C	2.08	1.20
1:G:465:VAL:HG13	1:G:466:PRO:HD3	1.32	1.11
3:K:106:UNK:CB	4:L:198:HIS:C	2.29	0.99

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	E	392/482 (81%)	355 (91%)	34 (9%)	3 (1%)	19	51
1	G	409/482 (85%)	367 (90%)	39 (10%)	3 (1%)	22	54
2	F	214/220 (97%)	195 (91%)	19 (9%)	0	100	100
2	H	214/220 (97%)	203 (95%)	10 (5%)	1 (0%)	29	61
3	I	44/121 (36%)	37 (84%)	7 (16%)	0	100	100
3	K	44/121 (36%)	39 (89%)	5 (11%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	J	211/214 (99%)	203 (96%)	8 (4%)	0	100	100
4	L	211/214 (99%)	202 (96%)	9 (4%)	0	100	100
All	All	1739/2074 (84%)	1601 (92%)	131 (8%)	7 (0%)	34	66

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	442	VAL
1	E	140	ILE
1	G	247	SER
1	E	344	GLN
1	G	51	PRO

5.3.2   Protein sidechains   ⓘ

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	E	332/400 (83%)	279 (84%)	53 (16%)	2	11
1	G	341/400 (85%)	288 (84%)	53 (16%)	2	12
2	F	183/185 (99%)	171 (93%)	12 (7%)	16	46
2	H	183/185 (99%)	174 (95%)	9 (5%)	25	56
3	I	37/39 (95%)	34 (92%)	3 (8%)	11	36
3	K	32/39 (82%)	26 (81%)	6 (19%)	1	6
4	J	187/188 (100%)	182 (97%)	5 (3%)	44	71
4	L	187/188 (100%)	178 (95%)	9 (5%)	25	56
All	All	1482/1624 (91%)	1332 (90%)	150 (10%)	7	27

5 of 150 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	413	CYS
3	I	69	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	435	ARG
2	F	5	VAL
1	G	465	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
3	K	82	GLN

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 (i)

6.1 Protein, DNA and RNA chains (i)

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, 95th 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(Å²)	Q<0.9
1	E	406/482 (84%)	-0.14	8 (1%)	65	64	34, 87, 116, 149	0
1	G	419/482 (86%)	-0.16	6 (1%)	75	75	42, 86, 115, 136	0
2	F	218/220 (99%)	-0.32	0	100	100	33, 64, 97, 120	0
2	H	218/220 (99%)	-0.28	1 (0%)	91	91	32, 63, 87, 106	0
3	I	46/121 (38%)	0.45	7 (15%)	2	2	76, 96, 115, 128	0
3	K	45/121 (37%)	0.15	2 (4%)	34	33	74, 92, 109, 116	0
4	J	213/214 (99%)	-0.30	0	100	100	36, 55, 76, 90	0
4	L	213/214 (99%)	-0.31	0	100	100	32, 56, 76, 94	0
All	All	1778/2074 (85%)	-0.20	24 (1%)	77	77	32, 69, 111, 149	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	I	96	CYS	3.9
3	K	8	GLY	3.8
1	E	157	GLU	3.4
1	E	155	THR	3.4
1	G	441	PRO	3.3

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.