wwPDB X-ray Structure Validation Summary Report (i)

Sep 11, 2023 – 04:18 AM EDT

PDB ID	:	4JGH
Title	:	Structure of the SOCS2-Elongin BC complex bound to an N-terminal fragment of Cullin5
Authors	:	Kim, Y.K.; Kwak, M.J.; Ku, B.; Suh, H.Y.; Joo, K.; Lee, J.; Jung, J.U.; Oh, B.H.
Deposited on	:	2013-03-01
Resolution	:	3.00 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 (i)) were used in the production of this report:

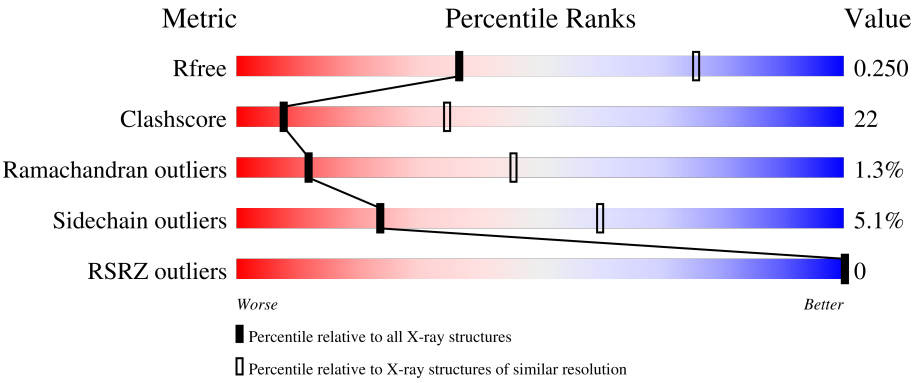
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	173	53% 37% · 6%
2	B	118	52% 36% · 12%
3	C	96	69% 27% ·
4	D	378	53% 41% · ·

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5902 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 Suppressor of cytokine signaling 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	163	Total	C	N	O	S	0	0	0
			1310	842	220	242	6			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	26	HIS	-	expression tag	UNP O14508
A	27	MET	-	expression tag	UNP O14508
A	28	ASP	-	expression tag	UNP O14508
A	29	PRO	-	expression tag	UNP O14508
A	30	GLU	-	expression tag	UNP O14508
A	31	PHE	-	expression tag	UNP O14508

- Molecule 2 is a protein called Transcription elongation factor B polypeptide 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	104	Total	C	N	O	S	0	0	0
			828	524	141	159	4			

- Molecule 3 is a protein called Transcription elongation factor B polypeptide 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	96	Total	C	N	O	S	0	0	0
			762	489	122	145	6			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	100	PRO	ALA	SEE REMARK 999	UNP P83940

- Molecule 4 is a protein called Cullin-5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	368	Total	C	N	O	S	0	0	0
			3002	1907	509	568	18			

There are 3 discrepancies between the modelled and reference sequences:

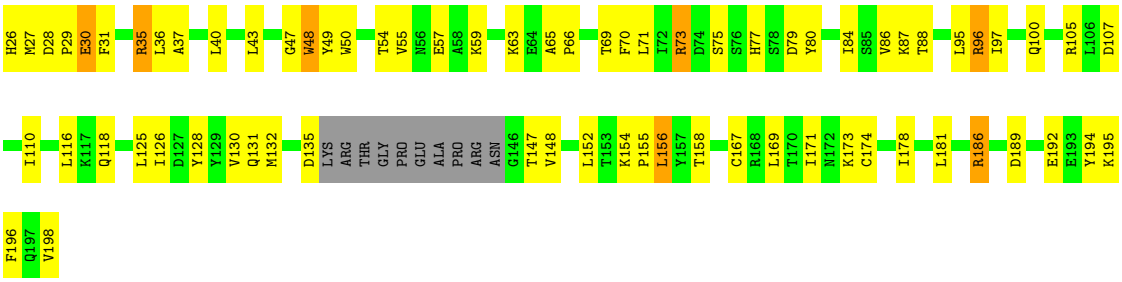
Chain	Residue	Modelled	Actual	Comment	Reference
D	341	ARG	VAL	engineered mutation	UNP Q93034
D	345	ASP	LEU	engineered mutation	UNP Q93034
D	387	VAL	-	expression tag	UNP Q93034

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.

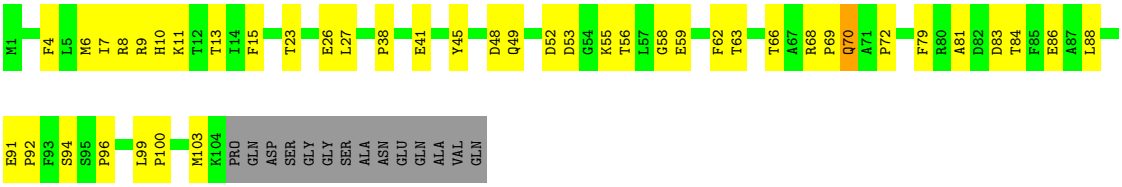
• Molecule 1: Suppressor of cytokine signaling 2

Chain A:



• Molecule 2: Transcription elongation factor B polypeptide 2

Chain B:

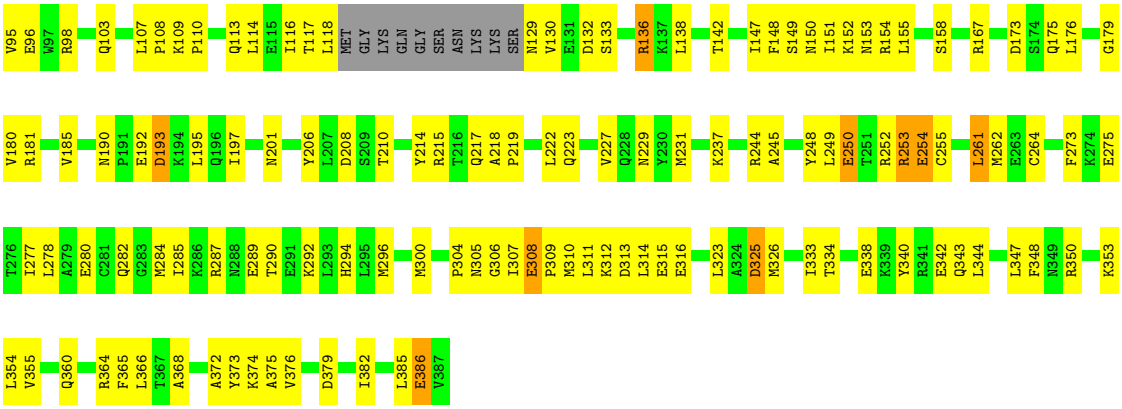


• Molecule 3: Transcription elongation factor B polypeptide 1

Chain C:

• Molecule 4: Cullin-5

Chain D:


V95 E96 W97 R98 Q103 L107 P108 K109 P110 Q113 L114 E115 I116 T117 L118 MET GLY LYS GLN GLY SER ASN LYS LYS SER N129 V130 E131 D132 S133 R136 K137 L138 T142 I147 F148 S149 N150 I151 K152 N153 R154 L155 S158 R167 D173 S174 Q175 L176 G179
V180 R181 V185 N190 P191 E192 D193 K194 L195 Q196 I197 N201 Y206 L207 D208 S209 T210 Y214 R215 T216 Q217 A218 P219 L222 Q223 V227 Q228 N229 Y230 M231 K237 R244 A245 Y248 L249 E250 T251 R252 R253 E254 C255 L261 M262 E263 C264 F273 K274 E275
T276 I277 L278 A279 E280 C281 Q282 G283 M284 I285 K286 R287 N288 E289 T290 E291 K292 L293 H294 L295 M296 M300 P304 N305 G306 I307 E308 P309 M310 L311 K312 D313 L314 E315 E316 L323 A324 D325 M326 I333 T334 E338 K339 Y340 R341 E342 Q343 L344 L347 F348 N349 R350 K353
L354 V355 Q360 R364 F365 L366 T367 A368 A372 Y373 K374 A375 V376 D379 I382 L385 E386 V387

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	138.85Å 141.46Å 182.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.00 49.54 – 3.01	Depositor EDS
% Data completeness (in resolution range)	(Not available) (30.00-3.00) 96.8 (49.54-3.01)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.82 (at 3.01Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.225 , 0.248 0.226 , 0.250	Depositor DCC
R_{free} test set	3438 reflections (9.96%)	wwPDB-VP
Wilson B-factor (Å^2)	58.1	Xtriage
Anisotropy	0.134	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 11.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	0.045 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	5902	wwPDB-VP
Average B, all atoms (Å^2)	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.31% 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	0.30	0/1340	0.60	1/1812 (0.1%)
2	B	0.30	0/844	0.59	0/1139
3	C	0.31	0/780	0.53	0/1055
4	D	0.29	0/3055	0.52	0/4116
All	All	0.29	0/6019	0.55	1/8122 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	186	ARG	NE-CZ-NH1	5.24	122.92	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	1310	0	1310	66	0
2	B	828	0	834	31	0
3	C	762	0	751	34	0
4	D	3002	0	2987	137	0
All	All	5902	0	5882	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 22.

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)
4:D:167:ARG:HH12	4:D:210:THR:HA	1.20	1.06
4:D:253:ARG:HB2	4:D:253:ARG:HH11	1.31	0.96
1:A:107:ASP:HB3	1:A:110:ILE:HD13	1.48	0.96
4:D:333:ILE:HG23	4:D:340:TYR:HA	1.56	0.87
4:D:109:LYS:HB2	4:D:110:PRO:HD3	1.54	0.87

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	159/173 (92%)	142 (89%)	16 (10%)	1 (1%)	25 64
2	B	102/118 (86%)	93 (91%)	8 (8%)	1 (1%)	15 53
3	C	94/96 (98%)	87 (93%)	3 (3%)	4 (4%)	2 15
4	D	364/378 (96%)	322 (88%)	39 (11%)	3 (1%)	19 57
All	All	719/765 (94%)	644 (90%)	66 (9%)	9 (1%)	12 45

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	149	SER
4	D	193	ASP
4	D	19	TRP
1	A	96	ARG
2	B	81	ALA

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	143/153 (94%)	136 (95%)	7 (5%)	25	61
2	B	92/102 (90%)	90 (98%)	2 (2%)	52	81
3	C	86/86 (100%)	83 (96%)	3 (4%)	36	71
4	D	329/340 (97%)	308 (94%)	21 (6%)	17	51
All	All	650/681 (95%)	617 (95%)	33 (5%)	24	60

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

Mol	Chain	Res	Type
4	D	308	GLU
4	D	313	ASP
4	D	386	GLU
4	D	13	LEU
3	C	98	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	288	ASN
4	D	229	ASN
4	D	48	HIS
4	D	223	GLN
3	C	108	ASN

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	163/173 (94%)	-0.20	0	100	100	49, 72, 101, 131	0
2	B	104/118 (88%)	-0.20	0	100	100	44, 62, 89, 98	0
3	C	96/96 (100%)	-0.22	0	100	100	45, 60, 89, 94	0
4	D	368/378 (97%)	-0.24	0	100	100	51, 78, 103, 124	0
All	All	731/765 (95%)	-0.22	0	100	100	44, 73, 100, 131	0

There are no RSRZ outliers to report.

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.