wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 03:41 PM EDT

PDB ID	:	2JZ3
Title	:	SOCS box elonginBC ternary complex
Authors	:	Babon, J.J.; Sabo, J.; Soetopo, A.; Yao, S.; Bailey, M.F.; Zhang, J.; Nicola, N.A.; Norton, R.S.
Deposited on	:	2007-12-27

This is a wwPDB NMR 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/NMRValidationReportHelp
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:

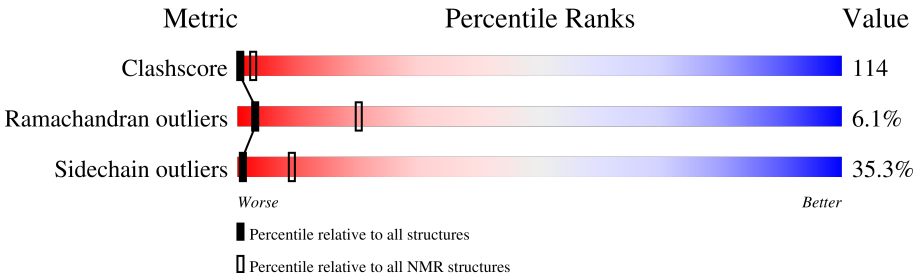
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment was not calculated.

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)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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%

Mol	Chain	Length	Quality of chain
1	A	40	• 30% 5% 62%
2	B	118	8% 51% 15% • 25%
3	C	96	5% 53% 21% 21%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 6 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:1-A:15, B:1-B:80, B:86-B:93, C:17-C:46, C:58-C:83, C:93-C:112 (179)	1.24	6

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 2 clusters. No single-model clusters were found.

Cluster number	Models
1	2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20
2	1, 7, 14

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3578 atoms, of which 1781 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Suppressor of cytokine signaling 3.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	15	239	71	123	25	19	1	0

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	118	1829	575	908	156	185	5	0

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
3	C	96	1510	487	750	122	145	6	0

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

• Molecule 1: Suppressor of cytokine signaling 3

Chain A: • 30% 5% 62%

V1 A2 T3 L4 Q5 H6 L7 C8 R9 K10 T11 V12 N13 G14 H15 LEU ASP SER TYR GLU LYS VAL THR GLN LEU PRO GLY PRO ILE ARG GLU PHE LEU ASP GLN TYR ASP ALA PRO LEU

• Molecule 2: Transcription elongation factor B polypeptide 2

Chain B: 8% 51% 15% • 25%

M1 D2 V3 F4 L5 M6 I7 R8 R9 H10 K11 T12 T13 I14 F15 T16 D17 A18 K19 E20 S21 S22 T23 V24 F25 E26 L27 K28 R29 I30 V31 E32 G33 I34 L35 K36 R37 D40 E41 Q42 R43 L44 Y45 K46 D47 D48 Q49 L50 L51 D52 D53 G54 K55 T56 L57 G58 E59 C60 G61

F62 T63 S64 Q65 T66 A67 R68 P69 Q70 A71 P72 A73 T74 V75 G76 L77 A78 F79 R80 A81 D82 D83 T84 F85 E86 A87 L88 C89 I90 E91 P92 F93 S94 S95 P96 P97 E98 L99 P100 D101 V102 M103 K104 P105 Q106 D107 S108 G109 S110 S111 A112 N113 E114 Q115 A116 V117 Q118

• Molecule 3: Transcription elongation factor B polypeptide 1

Chain C: 5% 53% 21% 21%

M17 Y18 V19 K20 L21 I22 S23 S24 D25 G26 H27 E28 F29 I30 V31 K32 R33 E34 H35 A36 L37 T38 S39 G40 T41 I42 K43 A44 M45 L46 S47 G48 P49 G50 Q51 F52 A53 E54 N55 E56 T57 N58 E59 V60 N61 F62 R63 E64 I65 P66 S67 H68 V69 L70 S71 K72 V73 C74 M75 Y76

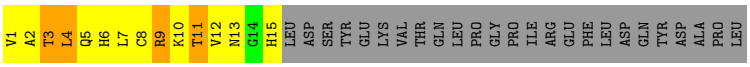
F77 T78 Y79 K80 V81 R82 Y83 T84 N85 S86 S87 T88 E89 I90 P91 E92 F93 P94 I95 A96 P97 E98 I99 A100 L101 E102 L103 L104 M105 A106 A107 N108 F109 L110 D111 C112

4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 6. Colouring as in section 4.1 above.

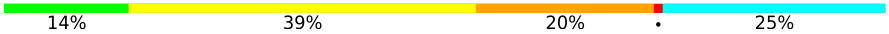
• Molecule 1: Suppressor of cytokine signaling 3

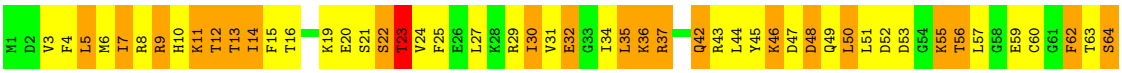
Chain A: • 25% 10% 62%

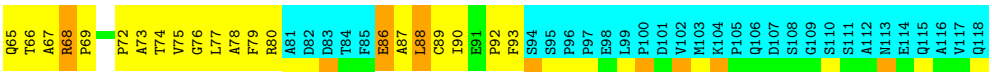


- Molecule 2: Transcription elongation factor B polypeptide 2

Chain B:

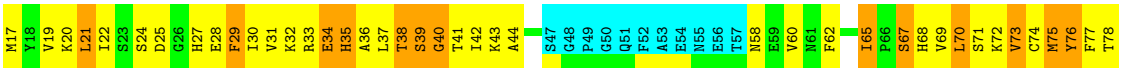






- Molecule 3: Transcription elongation factor B polypeptide 1

Chain C:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
TALOS	structure solution	
TALOS	refinement	

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	116	123	123	42±7
2	B	701	712	712	160±12
3	C	614	625	624	175±12
All	All	28620	29200	29180	6579

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 114.

5 of 2314 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
3:C:70:LEU:HD13	3:C:71:SER:N	1.16	1.55	4	14
3:C:70:LEU:HD12	3:C:71:SER:N	1.16	1.54	3	6
2:B:10:HIS:O	2:B:12:THR:N	1.15	1.79	1	20
3:C:36:ALA:HB1	3:C:42:ILE:HG21	1.10	1.18	14	20
1:A:4:LEU:HD13	3:C:103:LEU:HD22	1.09	1.09	16	2

6.3 Torsion angles (i)

6.3.1 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 PDB entries followed by that with respect to all NMR

entries. 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	13/40 (32%)	12±1 (90±5%)	1±1 (9±6%)	0±0 (2±3%)	14 59
2	B	87/118 (74%)	70±2 (81±2%)	10±2 (12±3%)	6±3 (7±3%)	2 15
3	C	74/96 (77%)	64±2 (87±3%)	6±1 (8±2%)	4±1 (5±1%)	3 23
All	All	3480/5080 (69%)	2925 (84%)	343 (10%)	212 (6%)	3 20

5 of 32 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	11	LYS	20
3	C	98	GLU	20
3	C	100	ALA	20
2	B	23	THR	14
2	B	68	ARG	14

6.3.2 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 PDB entries followed by that with respect to all NMR entries. 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	13/36 (36%)	9±1 (68±11%)	4±1 (32±11%)	1 14
2	B	77/103 (75%)	50±4 (65±5%)	27±4 (35±5%)	1 9
3	C	68/85 (80%)	43±3 (64±5%)	25±3 (36±5%)	1 8
All	All	3160/4480 (71%)	2043 (65%)	1117 (35%)	1 9

5 of 137 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	7	ILE	20
2	B	12	THR	20
2	B	14	ILE	20
2	B	88	LEU	20
3	C	73	VAL	20

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

No chemical shift data were provided