Full wwPDB NMR Structure Validation Report (i)

Oct 10, 2021 – 02:50 PM EDT

PDB ID	:	2KA6
Title	:	NMR structure of the CBP-TAZ2/STAT1-TAD complex
Authors	:	Wojciak, J.M.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E.
Deposited on	:	2008-10-30

This is a Full wwPDB NMR 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/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

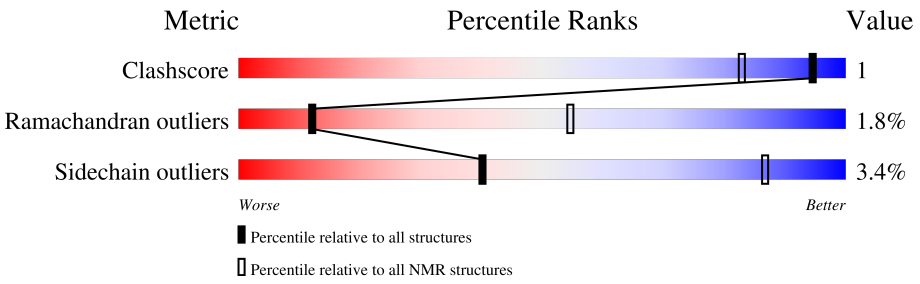
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 is 82%.

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	92	82% 10% • 8%
2	B	45	51% • • 44%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 18 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:4-A:88, B:722-B:746 (110)	0.22	18

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2146 atoms, of which 1071 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called CREB-binding protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	92	1466	439	743	149	121	14	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	87	LYS	ASN	engineered mutation	UNP P45481
A	88	LEU	VAL	engineered mutation	UNP P45481

- Molecule 2 is a protein called Signal transducer and activator of transcription 1-alpha/beta.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	45	677	213	328	58	74	4	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	706	GLY	-	expression tag	UNP P42224
B	707	SER	-	expression tag	UNP P42224
B	708	HIS	-	expression tag	UNP P42224
B	709	MET	-	expression tag	UNP P42224

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
3	A	3	3	3

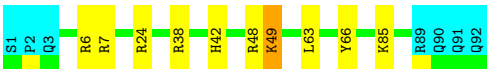
4 Residue-property plots (i)

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.

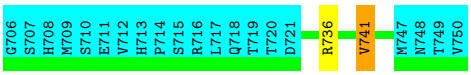
• Molecule 1: CREB-binding protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

Chain B:



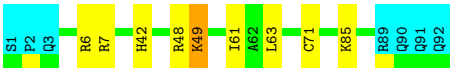
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

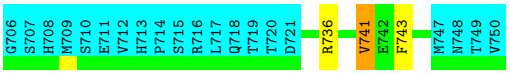
• Molecule 1: CREB-binding protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

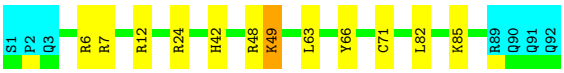
Chain B:



4.2.2 Score per residue for model 2

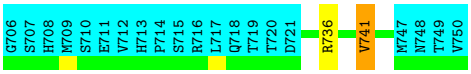
• Molecule 1: CREB-binding protein

Chain A: 79% 12% • 8%



• Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

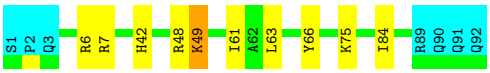
Chain B: 51% • • 44%



4.2.3 Score per residue for model 3

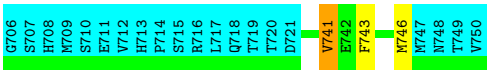
• Molecule 1: CREB-binding protein

Chain A: 82% 10% • 8%



• Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

Chain B: 49% • • 44%



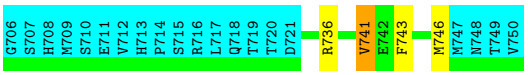
4.2.4 Score per residue for model 4

• Molecule 1: CREB-binding protein

Chain A: 77% 14% • 8%

• Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

Chain B: 47% 7% • 44%



4.2.5 Score per residue for model 5

• Molecule 1: CREB-binding protein

Chain A: 76% 15% • 8%

S1 P2 Q3 R6 R7 R24 R38 H42 R48 K49 I61 A62 L63 Y66 C71 K75 C81 I84 K85 R89 Q90 Q91 Q92

• Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

Chain B: 47% 7% • 44%

G706 S707 H708 M709 S710 E711 V712 H713 P714 S715 R716 L717 Q718 T719 T720 D721 R736 V741 E742 F743 M746 M747 N748 T749 V750

4.2.6 Score per residue for model 6

• Molecule 1: CREB-binding protein

Chain A: 79% 13% 8%

S1 P2 Q3 R6 R7 R12 R24 H42 R48 K49 I61 A62 L63 C71 C81 K85 R89 Q90 Q91 Q92

• Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

Chain B: 51% • 44%

G706 S707 H708 M709 S710 E711 V712 H713 P714 S715 R716 L717 Q718 T719 T720 D721 V741 E742 F743 M747 N748 T749 V750

4.2.7 Score per residue for model 7

• Molecule 1: CREB-binding protein

Chain A: 83% 8% • 8%

S1 P2 Q3 R7 R24 R38 H42 K47 R48 K49 C71 K85 R89 Q90 Q91 Q92

• Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

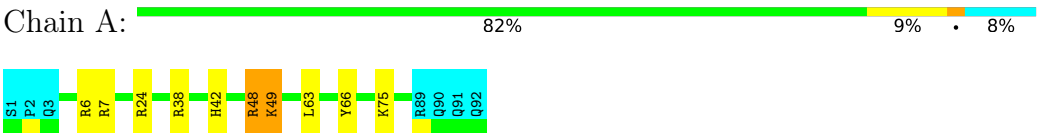
Chain B: 51% • • 44%

G706 S707 H708 M709 S710 E711 V712 H713 P714 S715 R716 L717 Q718 T719 T720 D721 R736 V741 M747 N748 T749 V750

4.2.8 Score per residue for model 8

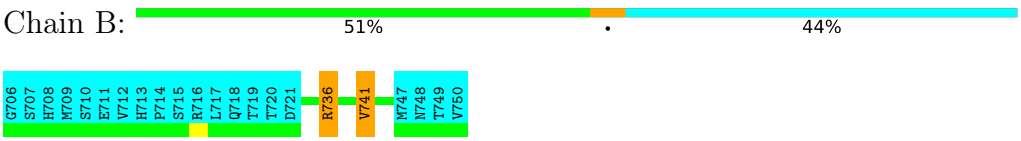
- Molecule 1: CREB-binding protein

Chain A:



- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

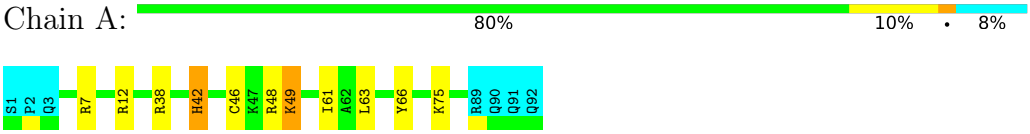
Chain B:



4.2.9 Score per residue for model 9

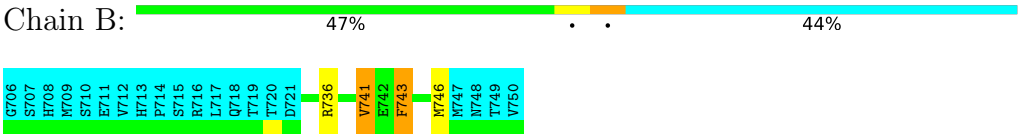
- Molecule 1: CREB-binding protein

Chain A:



- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

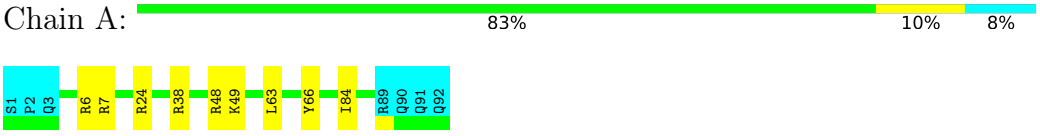
Chain B:



4.2.10 Score per residue for model 10

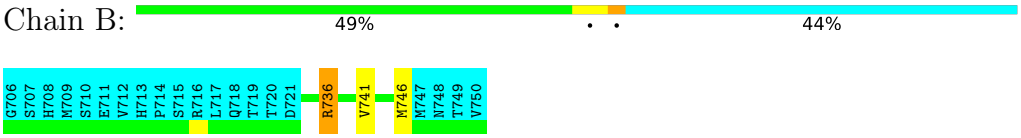
- Molecule 1: CREB-binding protein

Chain A:



- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

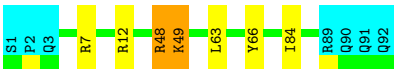
Chain B:



4.2.11 Score per residue for model 11

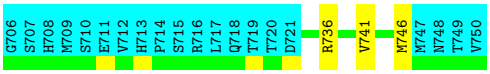
- Molecule 1: CREB-binding protein

Chain A: 85% 5% • 8%



- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

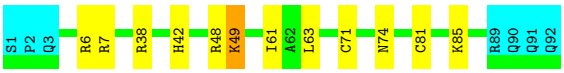
Chain B: 49% 7% 44%



4.2.12 Score per residue for model 12

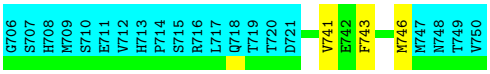
- Molecule 1: CREB-binding protein

Chain A: 79% 12% • 8%



- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

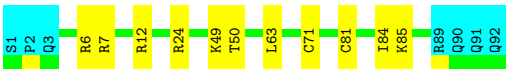
Chain B: 49% 7% 44%



4.2.13 Score per residue for model 13

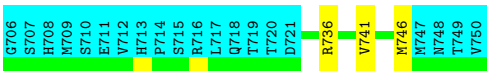
- Molecule 1: CREB-binding protein

Chain A: 80% 12% 8%



- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

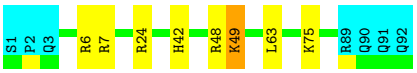
Chain B: 49% 7% 44%



4.2.14 Score per residue for model 14

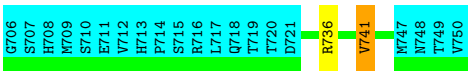
- Molecule 1: CREB-binding protein

Chain A: 84% 8% • 8%



- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

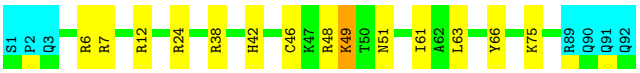
Chain B: 51% • • 44%



4.2.15 Score per residue for model 15

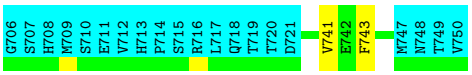
- Molecule 1: CREB-binding protein

Chain A: 77% 14% • 8%



- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

Chain B: 51% • 44%



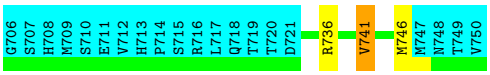
4.2.16 Score per residue for model 16

- Molecule 1: CREB-binding protein

Chain A: 76% 15% • 8%

- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

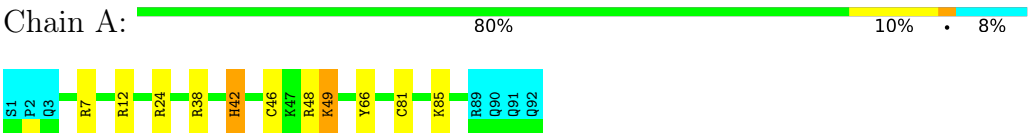
Chain B: 49% • • 44%



4.2.17 Score per residue for model 17

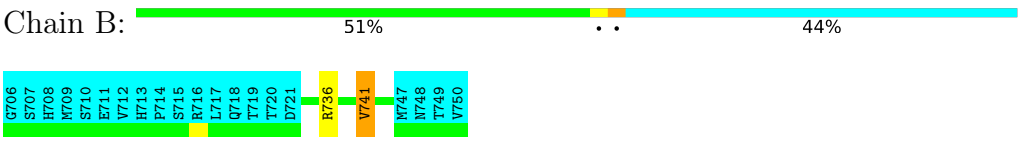
- Molecule 1: CREB-binding protein

Chain A:



- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

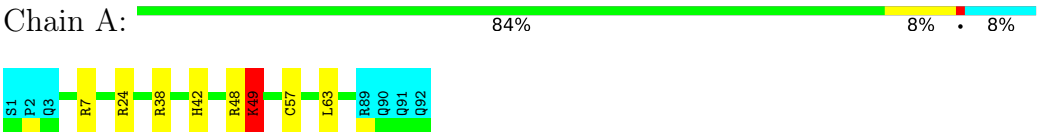
Chain B:



4.2.18 Score per residue for model 18 (medoid)

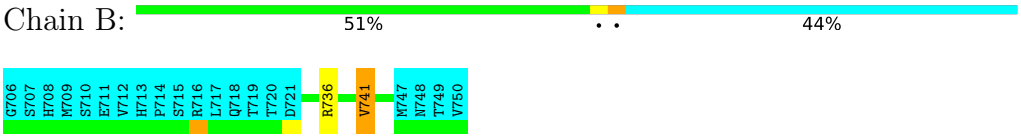
- Molecule 1: CREB-binding protein

Chain A:



- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

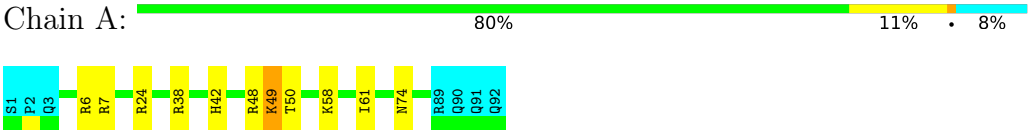
Chain B:



4.2.19 Score per residue for model 19

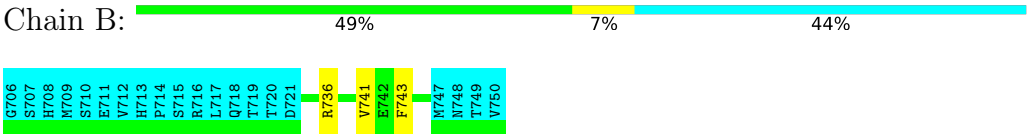
- Molecule 1: CREB-binding protein

Chain A:



- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

Chain B:



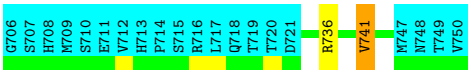
4.2.20 Score per residue for model 20

- Molecule 1: CREB-binding protein

Chain A:

- Molecule 2: Signal transducer and activator of transcription 1-alpha/beta

Chain B:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *molecular dynamics*.

Of the 200 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
Amber	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	1509
Number of shifts mapped to atoms	1509
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	82%

6 Model quality (i)

6.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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 (average) 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	0.62±0.00	0±0/674 (0.0± 0.0%)	1.10±0.03	6±1/902 (0.6± 0.1%)
2	B	0.72±0.01	0±0/200 (0.0± 0.0%)	1.07±0.03	1±0/271 (0.3± 0.1%)
All	All	0.64	0/17480 (0.0%)	1.10	129/23460 (0.5%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	1.5±0.6
2	B	0.0±0.0	0.1±0.3
All	All	0	32

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	7	ARG	NE-CZ-NH1	9.46	125.03	120.30	2	20
1	A	48	ARG	NE-CZ-NH1	9.29	124.95	120.30	14	19
1	A	6	ARG	NE-CZ-NH1	8.47	124.53	120.30	1	13
2	B	736	ARG	NE-CZ-NH1	8.10	124.35	120.30	2	16
1	A	7	ARG	NE-CZ-NH2	-7.70	116.45	120.30	19	20
1	A	38	ARG	NE-CZ-NH1	6.99	123.79	120.30	18	11
1	A	24	ARG	NE-CZ-NH1	6.66	123.63	120.30	4	13
1	A	7	ARG	CD-NE-CZ	6.02	132.03	123.60	7	6
1	A	12	ARG	NE-CZ-NH1	5.83	123.21	120.30	4	8
1	A	48	ARG	CD-NE-CZ	5.44	131.22	123.60	20	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	12	ARG	NE-CZ-NH2	5.30	122.95	120.30	13	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	42	HIS	Sidechain	16
1	A	66	TYR	Sidechain	12
2	B	736	ARG	Sidechain	2
1	A	38	ARG	Sidechain	1
1	A	48	ARG	Sidechain	1

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	663	686	686	2±1
2	B	196	184	184	1±1
All	All	17240	17400	17400	39

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 1.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:61:ILE:HG22	2:B:743:PHE:CE1	0.79	2.12	12	9
1:A:84:ILE:HG23	2:B:746:MET:HE2	0.57	1.75	10	7
1:A:50:THR:HG22	1:A:58:LYS:HG2	0.46	1.85	4	2
1:A:81:CYS:SG	1:A:85:LYS:NZ	0.46	2.89	13	7
1:A:71:CYS:SG	1:A:85:LYS:NZ	0.45	2.89	2	10
1:A:49:LYS:NZ	1:A:57:CYS:SG	0.41	2.93	16	2
1:A:42:HIS:CE1	1:A:46:CYS:HB2	0.41	2.51	9	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	85/92 (92%)	80±1 (94±2%)	4±1 (5±2%)	1±0 (1±0%)	17 64
2	B	25/45 (56%)	21±1 (85±3%)	3±1 (11±3%)	1±0 (4±0%)	5 31
All	All	2200/2740 (80%)	2015 (92%)	145 (7%)	40 (2%)	12 54

All 3 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	741	VAL	20
1	A	49	LYS	19
1	A	50	THR	1

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	78/85 (92%)	75±1 (97±1%)	3±1 (3±1%)	40 87
2	B	24/43 (56%)	23±1 (97±3%)	1±1 (3±3%)	41 87
All	All	2040/2560 (80%)	1971 (97%)	69 (3%)	40 87

All 13 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)
1	A	49	LYS	18
1	A	63	LEU	17
2	B	741	VAL	13
1	A	75	LYS	7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	24	ARG	3
1	A	48	ARG	2
2	B	746	MET	2
1	A	74	ASN	2
1	A	82	LEU	1
1	A	47	LYS	1
2	B	743	PHE	1
1	A	46	CYS	1
1	A	51	ASN	1

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)

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

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)

The completeness of assignment taking into account all chemical shift lists is 82% for the well-defined parts and 80% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	1509
Number of shifts mapped to atoms	1509
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	134	-0.18 ± 0.11	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	129	0.58 ± 0.09	Should be applied
$^{13}C'$	0	—	None (insufficient data)
^{15}N	122	-0.44 ± 0.23	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 82%, i.e. 1136 atoms were assigned a chemical shift out of a possible 1387. 0 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	424/538 (79%)	213/214 (100%)	109/220 (50%)	102/104 (98%)
Sidechain	663/779 (85%)	418/466 (90%)	231/269 (86%)	14/44 (32%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	49/70 (70%)	27/39 (69%)	22/26 (85%)	0/5 (0%)
Overall	1136/1387 (82%)	658/719 (92%)	362/515 (70%)	116/153 (76%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 80%, i.e. 1380 atoms were assigned a chemical shift out of a possible 1721. 0 out of 20 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	514/669 (77%)	258/266 (97%)	134/274 (49%)	122/129 (95%)
Sidechain	814/968 (84%)	514/579 (89%)	280/333 (84%)	20/56 (36%)
Aromatic	52/84 (62%)	29/47 (62%)	23/30 (77%)	0/7 (0%)
Overall	1380/1721 (80%)	801/892 (90%)	437/637 (69%)	142/192 (74%)

7.1.4 Statistically unusual chemical shifts (i)

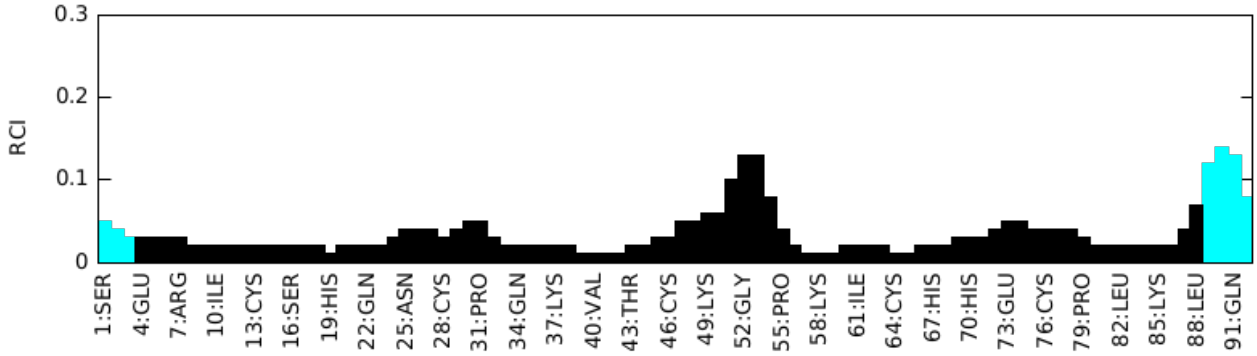
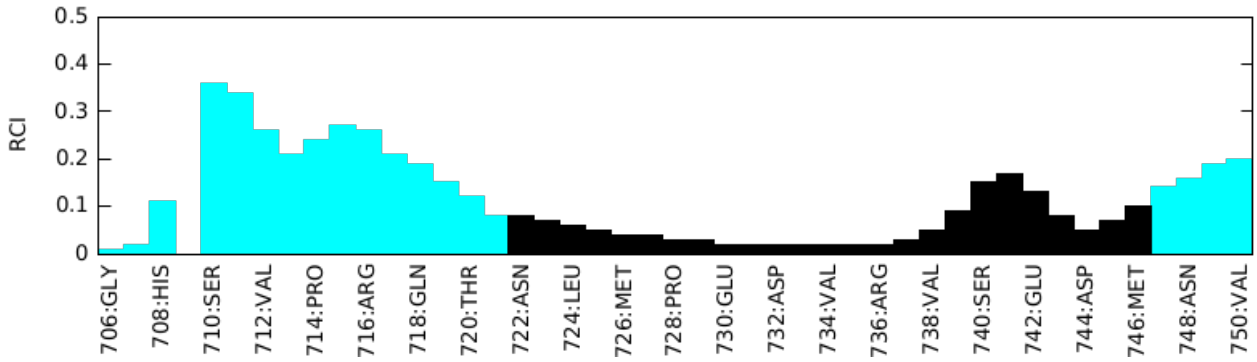
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
2	B	706	GLY	CA	57.88	51.81 – 38.91	9.7

7.1.5 Random Coil Index (RCI) plots (i)

The images below report *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A:



Random coil index (RCI) for chain B: