Full wwPDB NMR Structure Validation Report (i)

Oct 10, 2021 – 09:04 PM EDT

PDB ID	:	2KA4
Title	:	NMR structure of the CBP-TAZ1/STAT2-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:

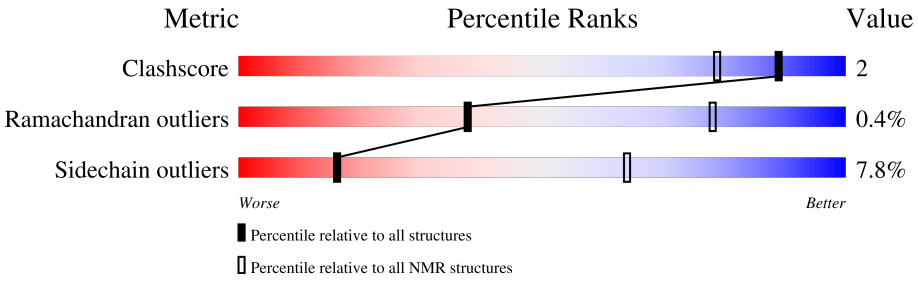
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al.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

1 Overall quality at a glance

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

The overall completeness of chemical shifts assignment is 81%.

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	100	58% 8% • 32%
2	B	57	44% 7% 49%

2 Ensemble composition and analysis

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: *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:346-A:367, A:381-A:395, A:405-A:435, B:788-B:816 (97)	0.16	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 3 clusters and 10 single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2447 atoms, of which 1215 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Crebbp protein.

Mol	Chain	Residues	Atoms						Trace
1	A	100	Total	C	H	N	O	S	0
			1563	472	783	162	135	11	

- Molecule 2 is a protein called Signal transducer and activator of transcription 2.

Mol	Chain	Residues	Atoms						Trace
2	B	57	Total	C	H	N	O	S	0
			881	279	432	78	89	3	

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	782	GLY	-	expression tag	UNP P52630
B	783	SER	-	expression tag	UNP P52630
B	784	HIS	-	expression tag	UNP P52630
B	785	MET	-	expression tag	UNP P52630
B	793	HIS	CYS	engineered mutation	UNP P52630
B	809	SER	CYS	engineered mutation	UNP P52630

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

Mol	Chain	Residues	Atoms	
3	A	3	Total	Zn
			3	3

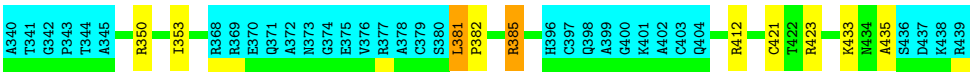
4 Residue-property plots (i)

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.

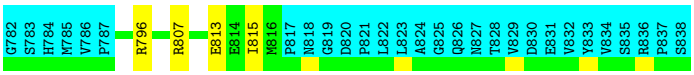
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

Chain B:



4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

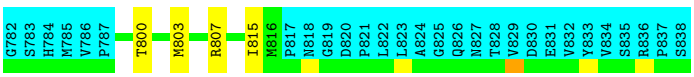
4.2.1 Score per residue for model 1

• Molecule 1: Crebbp protein

Chain A:

• Molecule 2: Signal transducer and activator of transcription 2

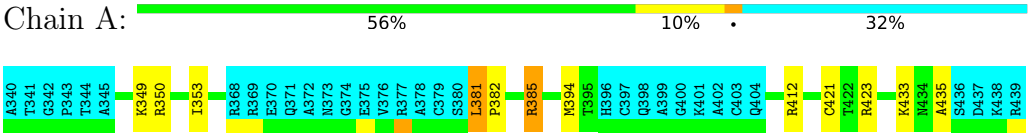
Chain B:



4.2.2 Score per residue for model 2

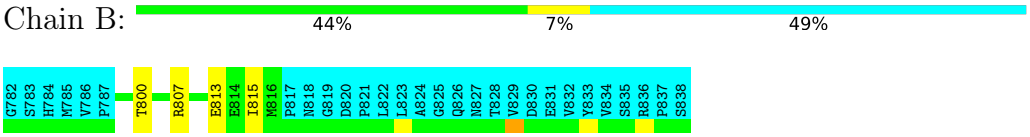
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

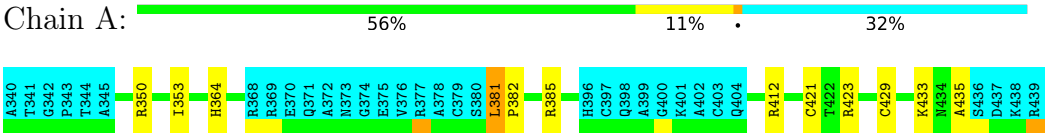
Chain B:



4.2.3 Score per residue for model 3

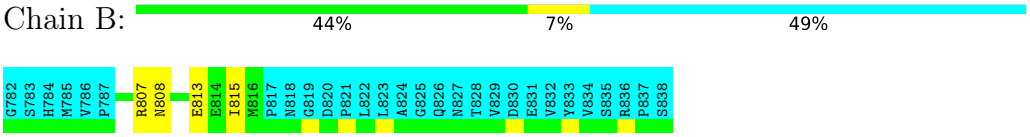
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

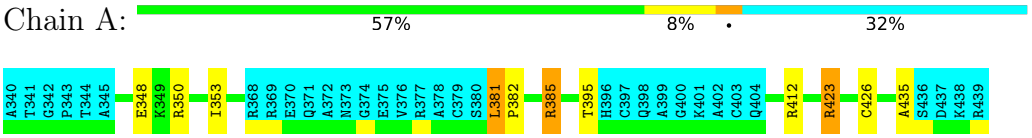
Chain B:



4.2.4 Score per residue for model 4

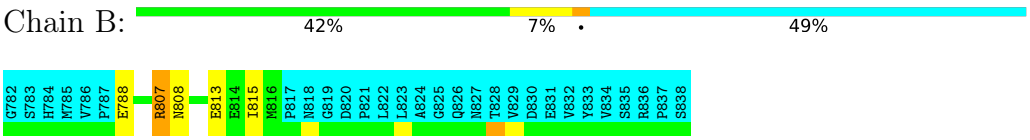
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

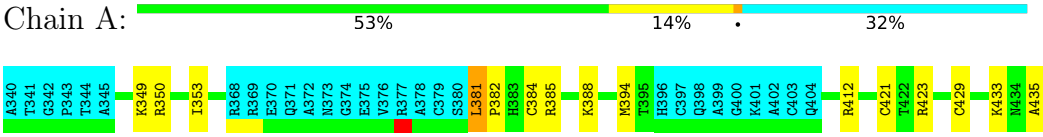
Chain B:



4.2.5 Score per residue for model 5

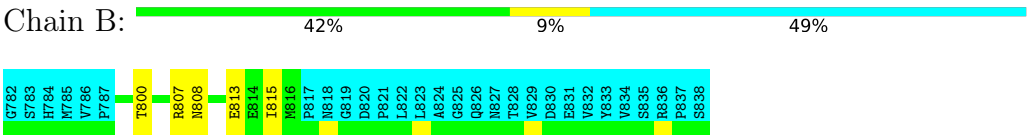
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

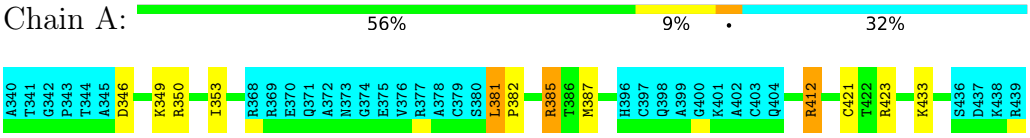
Chain B:



4.2.6 Score per residue for model 6 (medoid)

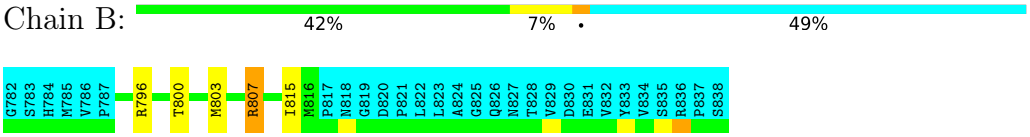
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

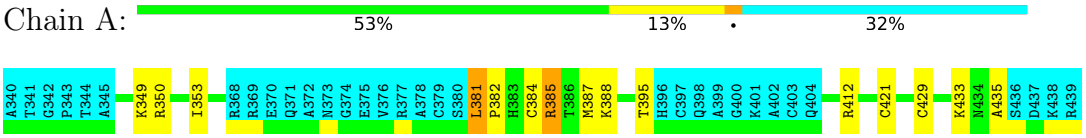
Chain B:



4.2.7 Score per residue for model 7

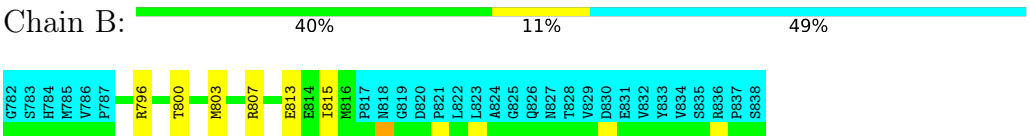
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

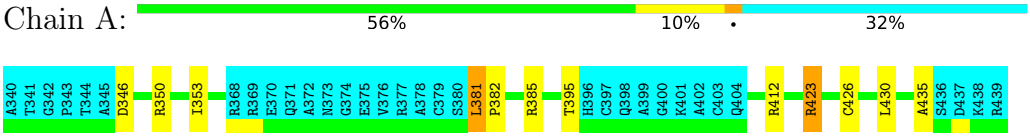
Chain B:



4.2.8 Score per residue for model 8

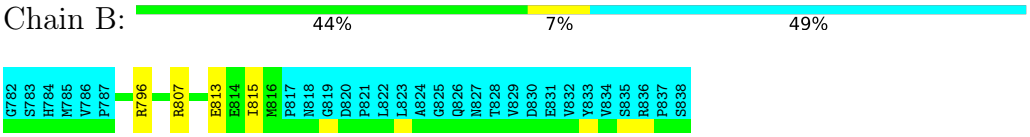
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

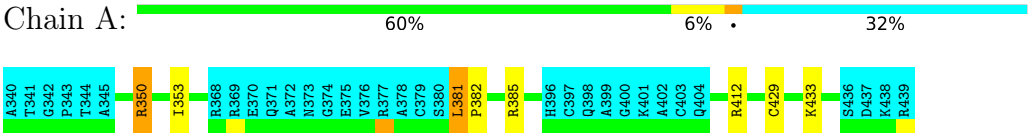
Chain B:



4.2.9 Score per residue for model 9

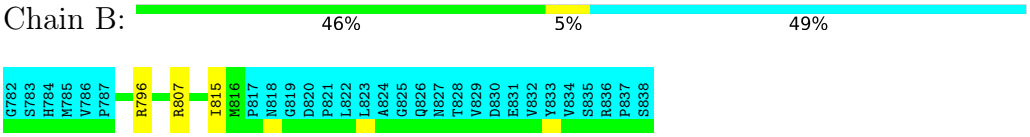
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

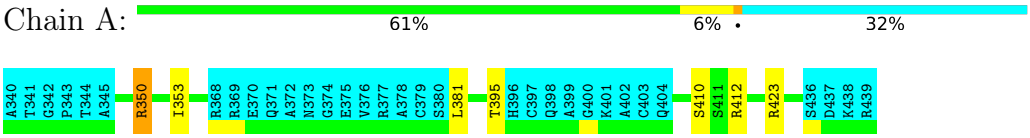
Chain B:



4.2.10 Score per residue for model 10

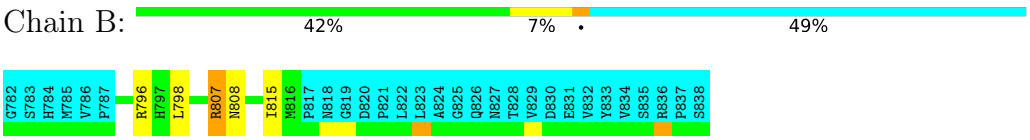
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

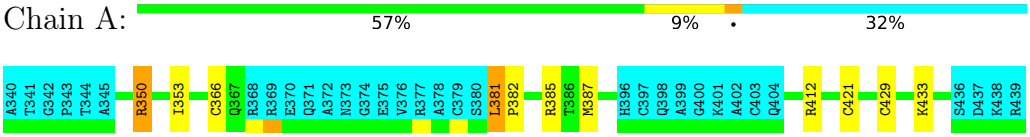
Chain B:



4.2.11 Score per residue for model 11

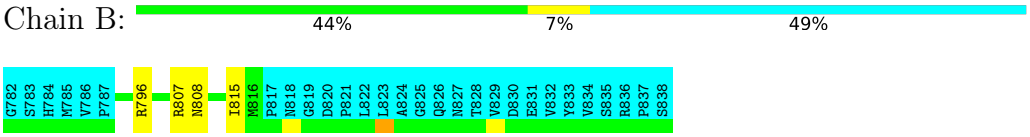
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

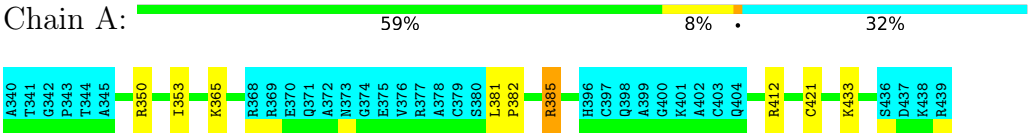
Chain B:



4.2.12 Score per residue for model 12

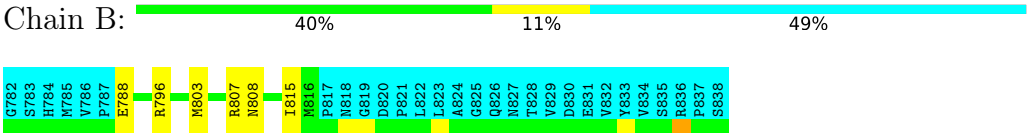
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

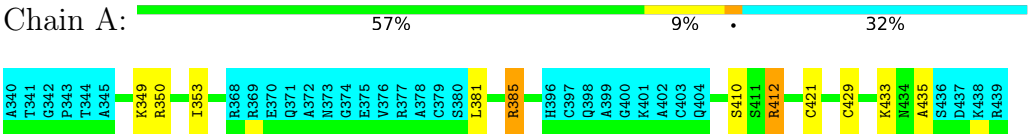
Chain B:



4.2.13 Score per residue for model 13

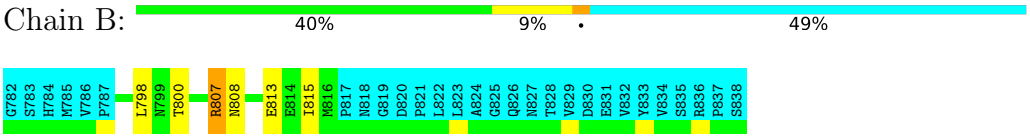
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

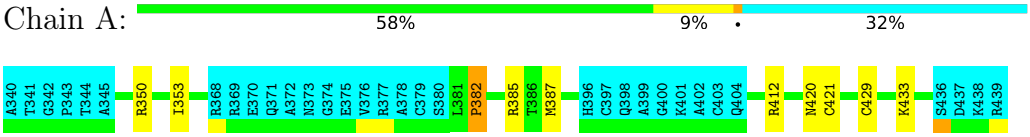
Chain B:



4.2.14 Score per residue for model 14

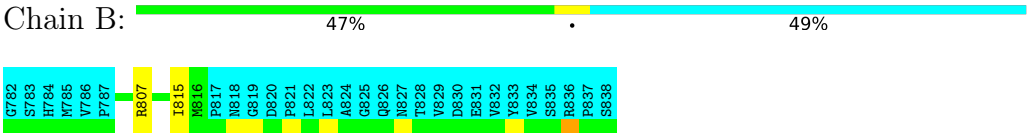
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

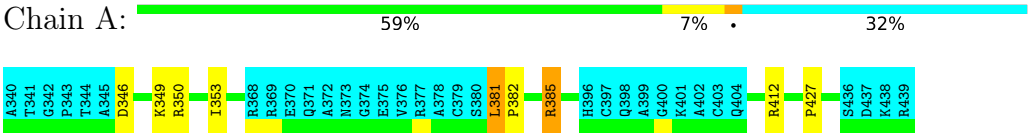
Chain B:



4.2.15 Score per residue for model 15

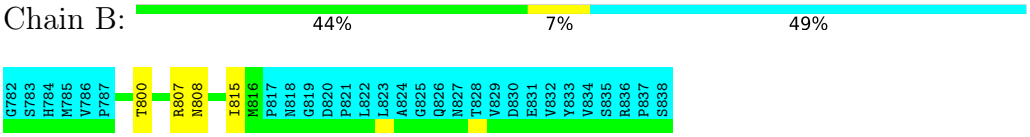
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

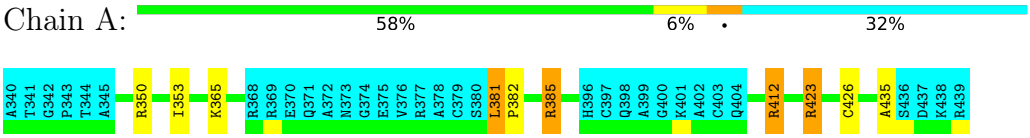
Chain B:



4.2.16 Score per residue for model 16

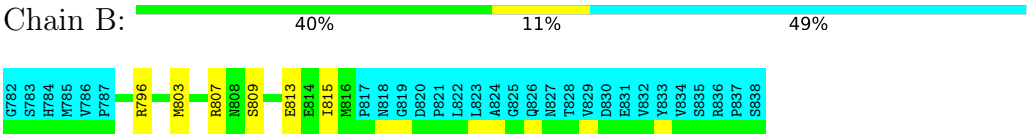
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

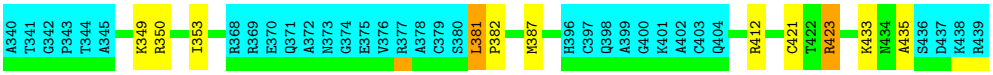
Chain B:



4.2.17 Score per residue for model 17

• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

Chain B:

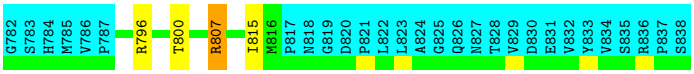
4.2.18 Score per residue for model 18

• Molecule 1: Crebbp protein

Chain A:

• Molecule 2: Signal transducer and activator of transcription 2

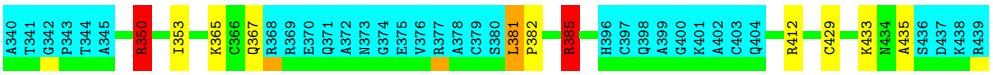
Chain B:



4.2.19 Score per residue for model 19

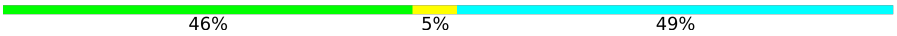
• Molecule 1: Crebbp protein

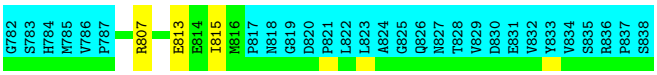
Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

Chain B:

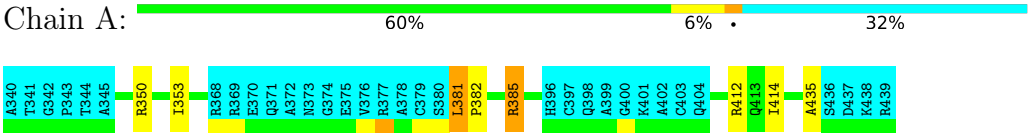




4.2.20 Score per residue for model 20

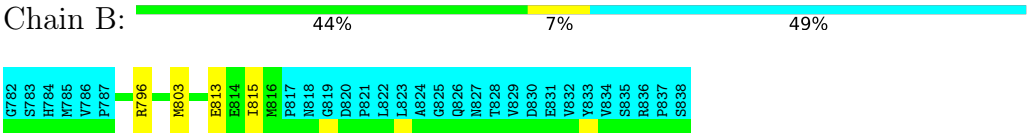
• Molecule 1: Crebbp protein

Chain A:



• Molecule 2: Signal transducer and activator of transcription 2

Chain B:



5 Refinement protocol and experimental data overview

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	1719
Number of shifts mapped to atoms	1719
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	81%

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	0.60±0.01	0±0/560 (0.0± 0.0%)	1.12±0.03	3±1/758 (0.4± 0.1%)
2	B	0.68±0.01	0±0/250 (0.0± 0.0%)	1.12±0.05	2±1/340 (0.5± 0.3%)
All	All	0.62	0/16200 (0.0%)	1.12	100/21960 (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	0.3±0.6
2	B	0.0±0.0	0.1±0.4
All	All	0	9

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	412	ARG	NE-CZ-NH1	11.17	125.89	120.30	12	18
2	B	807	ARG	NE-CZ-NH1	9.94	125.27	120.30	4	19
1	A	423	ARG	NE-CZ-NH1	8.00	124.30	120.30	17	10
1	A	350	ARG	NE-CZ-NH1	7.30	123.95	120.30	14	20
2	B	796	ARG	NE-CZ-NH2	-6.58	117.01	120.30	12	1
1	A	385	ARG	NE-CZ-NH1	6.41	123.50	120.30	2	12
2	B	796	ARG	NE-CZ-NH1	6.24	123.42	120.30	12	11
2	B	807	ARG	NE-CZ-NH2	-5.91	117.34	120.30	4	3
2	B	796	ARG	CD-NE-CZ	5.58	131.41	123.60	12	1
1	A	350	ARG	NE-CZ-NH2	-5.55	117.53	120.30	19	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	423	ARG	NH1-CZ-NH2	-5.14	113.75	119.40	17	1
2	B	807	ARG	CD-NE-CZ	5.02	130.63	123.60	4	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)
2	B	807	ARG	Sidechain	3
1	A	350	ARG	Sidechain	3
1	A	385	ARG	Sidechain	2
1	A	412	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	547	561	561	3±1
2	B	244	238	238	1±1
All	All	15880	15980	15980	57

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 2.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:381:LEU:HD13	1:A:382:PRO:HD2	0.54	1.79	4	16
1:A:429:CYS:CB	1:A:433:LYS:HZ2	0.50	2.20	7	10
1:A:435:ALA:HB2	2:B:813:GLU:HA	0.49	1.85	7	11
1:A:384:CYS:SG	1:A:388:LYS:NZ	0.46	2.89	5	2
1:A:421:CYS:SG	1:A:433:LYS:NZ	0.44	2.91	13	12
1:A:410:SER:HA	2:B:798:LEU:HD22	0.42	1.92	13	2
1:A:421:CYS:SG	1:A:423:ARG:NH1	0.42	2.93	17	1
1:A:423:ARG:NH2	1:A:426:CYS:SG	0.41	2.93	16	3

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	68/100 (68%)	64±1 (95±2%)	3±1 (5±2%)	0±0 (0±1%)	38	78
2	B	29/57 (51%)	26±1 (91±3%)	3±1 (9±3%)	0±0 (0±1%)	50	82
All	All	1940/3140 (62%)	1817 (94%)	116 (6%)	7 (0%)	38	78

All 4 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	346	ASP	4
2	B	807	ARG	1
1	A	382	PRO	1
1	A	422	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	64/87 (74%)	59±1 (92±1%)	5±1 (8±1%)	17	65
2	B	29/53 (55%)	26±1 (91±4%)	2±1 (9±4%)	14	61
All	All	1860/2800 (66%)	1714 (92%)	146 (8%)	16	64

All 26 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	353	ILE	20
2	B	815	ILE	20
1	A	381	LEU	19

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	385	ARG	18
2	B	800	THR	9
2	B	808	ASN	9
1	A	349	LYS	8
2	B	803	MET	7
1	A	387	MET	5
1	A	395	THR	4
1	A	412	ARG	4
1	A	394	MET	3
1	A	365	LYS	3
2	B	788	GLU	2
1	A	350	ARG	2
1	A	382	PRO	2
1	A	420	ASN	2
1	A	364	HIS	1
1	A	348	GLU	1
2	B	807	ARG	1
1	A	430	LEU	1
1	A	427	PRO	1
2	B	809	SER	1
2	B	793	HIS	1
1	A	367	GLN	1
1	A	414	ILE	1

6.3.3 RNA

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains

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

6.5 Carbohydrates

There are no monosaccharides in this entry.

6.6 Ligand geometry

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

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 81% for the well-defined parts and 78% 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

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	1719
Number of shifts mapped to atoms	1719
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	156	-0.39 ± 0.19	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	150	0.35 ± 0.18	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	130	0.01 ± 0.34	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 81%, i.e. 1031 atoms were assigned a chemical shift out of a possible 1278. 17 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	363/471 (77%)	182/187 (97%)	96/194 (49%)	85/90 (94%)
Sidechain	614/723 (85%)	382/429 (89%)	221/258 (86%)	11/36 (31%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	54/84 (64%)	28/47 (60%)	25/27 (93%)	1/10 (10%)
Overall	1031/1278 (81%)	592/663 (89%)	342/479 (71%)	97/136 (71%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 78%, i.e. 1532 atoms were assigned a chemical shift out of a possible 1960. 24 out of 24 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	573/761 (75%)	287/302 (95%)	156/314 (50%)	130/145 (90%)
Sidechain	891/1093 (82%)	556/649 (86%)	321/384 (84%)	14/60 (23%)
Aromatic	68/106 (64%)	36/59 (61%)	31/35 (89%)	1/12 (8%)
Overall	1532/1960 (78%)	879/1010 (87%)	508/733 (69%)	145/217 (67%)

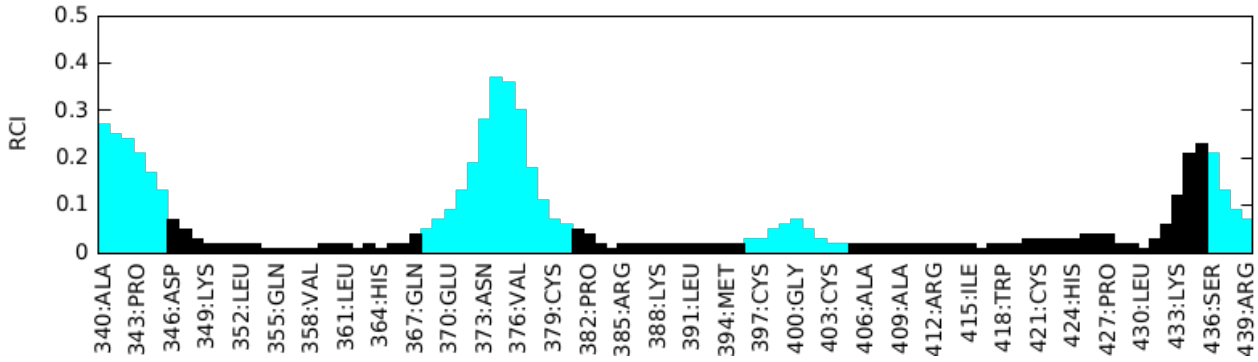
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

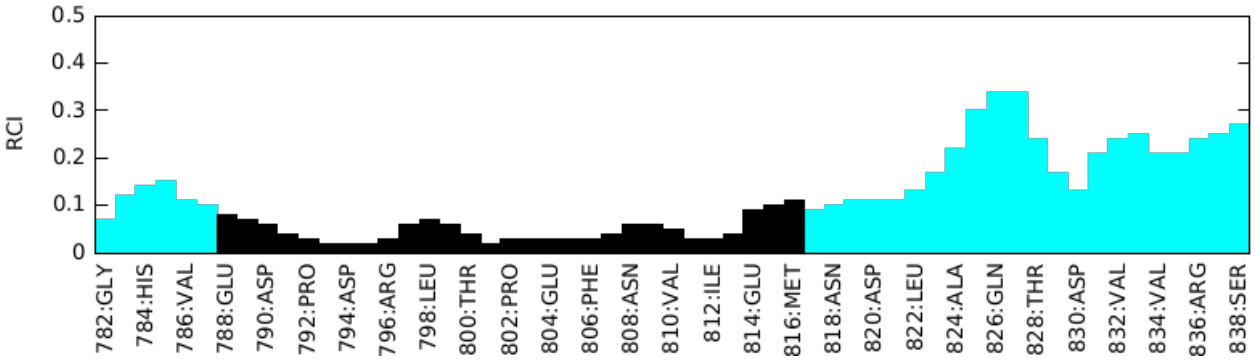
7.1.5 Random Coil Index (RCI) plots (i)

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:


RCI
0.5
0.4
0.3
0.2
0.1
0
782:GLY
784:HIS
786:VAL
788:GLU
790:ASP
792:PRO
794:ASP
796:ARG
798:LEU
800:THR
802:PRO
804:GLU
806:PHE
808:ASN
810:VAL
812:ILE
814:GLU
816:MET
818:ASN
820:ASP
822:LEU
824:ALA
826:GLN
828:THR
830:ASP
832:VAL
834:VAL
836:ARG
838:SER