Full wwPDB NMR Structure Validation Report

Aug 15, 2022 – 12:06 PM EDT

PDB ID	:	2MH0
Title	:	Solution NMR structure of the p300 Taz2:ETAD1 complex
Authors	:	Langelaan, D.N.; Smith, S.P.; Chitayat, S.
Deposited on	:	2013-11-12

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 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:

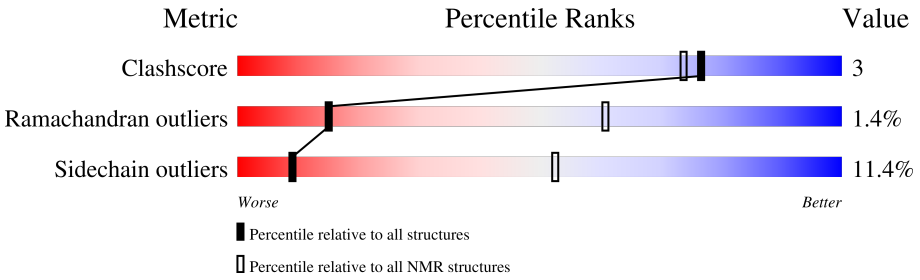
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.29
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 88%.

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	39	33% 8% 59%
2	B	92	77% 13% 10%

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: *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:13-A:28, B:1726-B:1774, B:1779-B:1812 (99)	0.34	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 7 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2011 atoms, of which 1017 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Transcription factor E2-alpha.

Mol	Chain	Residues	Atoms						Trace
1	A	39	Total	C	H	N	O	S	0
			592	184	297	51	56	4	

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP P15923
A	0	SER	-	expression tag	UNP P15923

- Molecule 2 is a protein called Histone acetyltransferase p300.

Mol	Chain	Residues	Atoms						Trace
2	B	92	Total	C	H	N	O	S	0
			1419	425	720	142	122	10	

There are 6 discrepancies between the modelled and reference sequences:

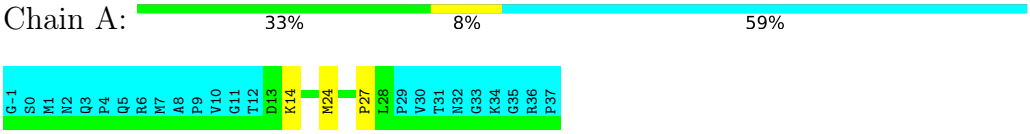
Chain	Residue	Modelled	Actual	Comment	Reference
B	1721	GLY	-	expression tag	UNP Q09472
B	1722	SER	-	expression tag	UNP Q09472
B	1738	ALA	CYS	engineered mutation	UNP Q09472
B	1746	ALA	CYS	engineered mutation	UNP Q09472
B	1789	ALA	CYS	engineered mutation	UNP Q09472
B	1790	ALA	CYS	engineered mutation	UNP Q09472

4 Residue-property plots

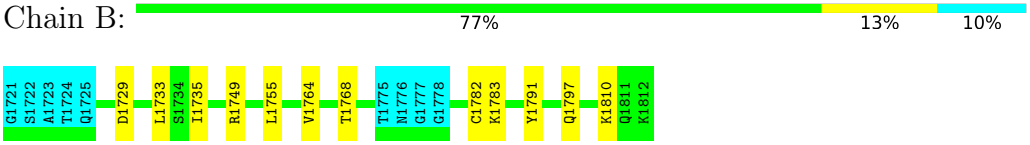
4.1 Average score per residue in the NMR ensemble

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: Transcription factor E2-alpha



• Molecule 2: Histone acetyltransferase p300

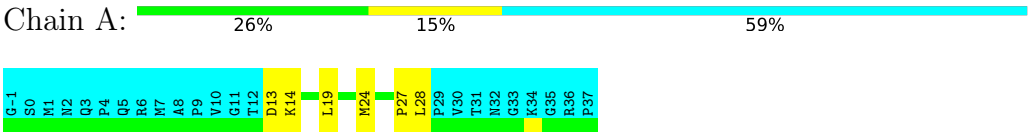


4.2 Scores per residue for each member of the ensemble

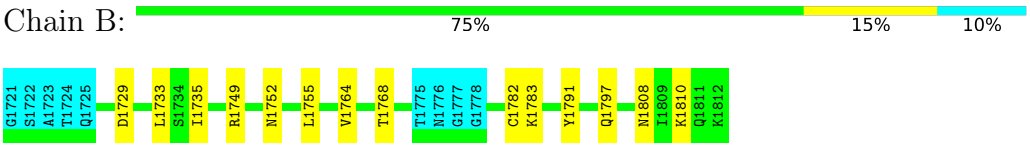
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Transcription factor E2-alpha



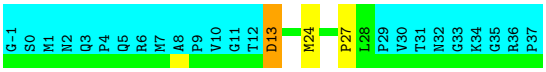
• Molecule 2: Histone acetyltransferase p300



4.2.2 Score per residue for model 2

• Molecule 1: Transcription factor E2-alpha

Chain A:



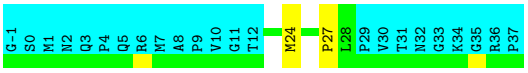
• Molecule 2: Histone acetyltransferase p300

Chain B:

4.2.3 Score per residue for model 3

• Molecule 1: Transcription factor E2-alpha

Chain A:



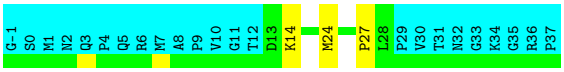
• Molecule 2: Histone acetyltransferase p300

Chain B:

4.2.4 Score per residue for model 4

• Molecule 1: Transcription factor E2-alpha

Chain A:



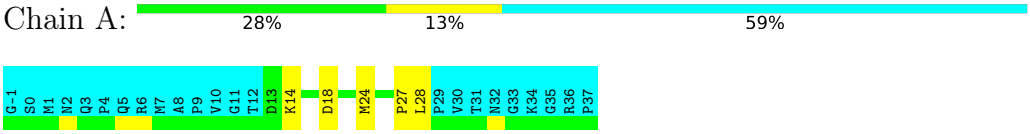
• Molecule 2: Histone acetyltransferase p300

Chain B:

4.2.5 Score per residue for model 5

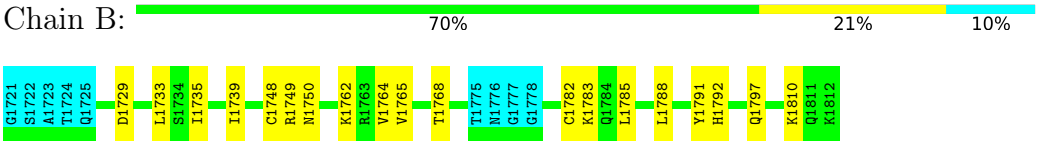
• Molecule 1: Transcription factor E2-alpha

Chain A:



• Molecule 2: Histone acetyltransferase p300

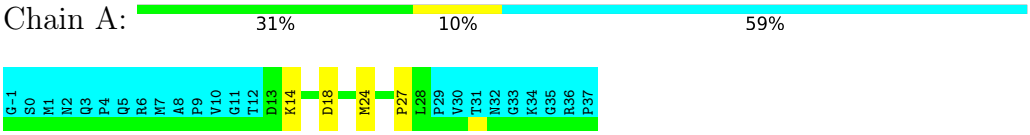
Chain B:



4.2.6 Score per residue for model 6 (medoid)

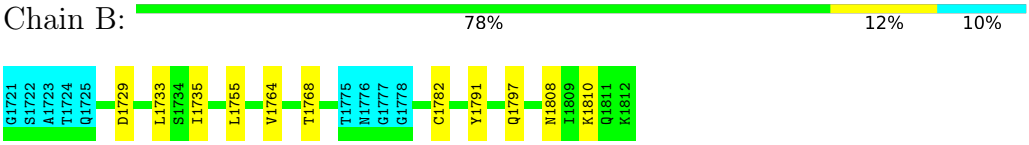
• Molecule 1: Transcription factor E2-alpha

Chain A:



• Molecule 2: Histone acetyltransferase p300

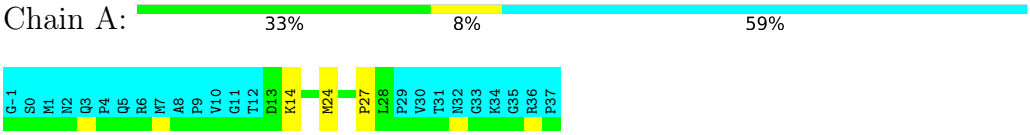
Chain B:



4.2.7 Score per residue for model 7

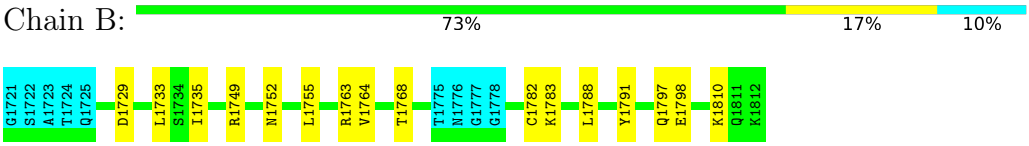
• Molecule 1: Transcription factor E2-alpha

Chain A:



• Molecule 2: Histone acetyltransferase p300

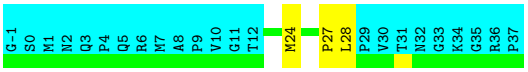
Chain B:



4.2.8 Score per residue for model 8

• Molecule 1: Transcription factor E2-alpha

Chain A:



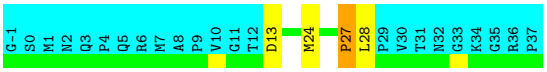
• Molecule 2: Histone acetyltransferase p300

Chain B:

4.2.9 Score per residue for model 9

• Molecule 1: Transcription factor E2-alpha

Chain A:



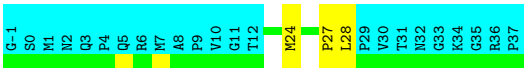
• Molecule 2: Histone acetyltransferase p300

Chain B:

4.2.10 Score per residue for model 10

• Molecule 1: Transcription factor E2-alpha

Chain A:



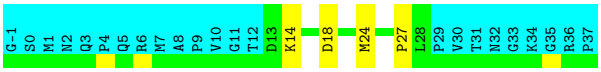
• Molecule 2: Histone acetyltransferase p300

Chain B:

4.2.11 Score per residue for model 11

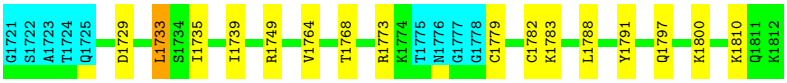
• Molecule 1: Transcription factor E2-alpha

Chain A:



• Molecule 2: Histone acetyltransferase p300

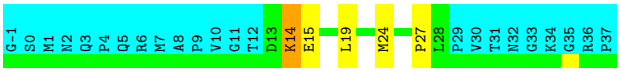
Chain B:



4.2.12 Score per residue for model 12

• Molecule 1: Transcription factor E2-alpha

Chain A:



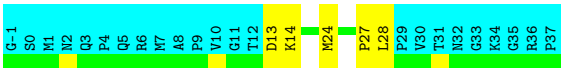
• Molecule 2: Histone acetyltransferase p300

Chain B:

4.2.13 Score per residue for model 13

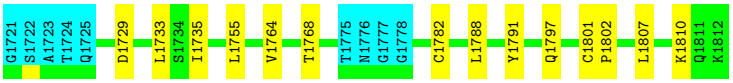
• Molecule 1: Transcription factor E2-alpha

Chain A:



• Molecule 2: Histone acetyltransferase p300

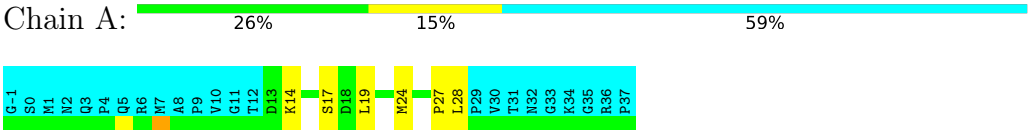
Chain B:



4.2.14 Score per residue for model 14

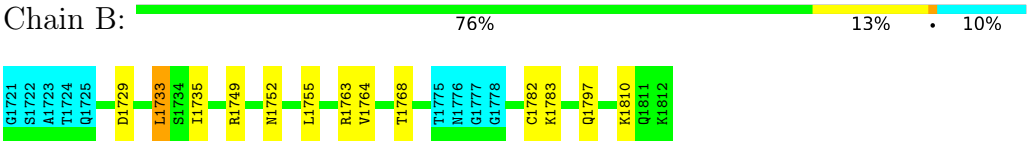
• Molecule 1: Transcription factor E2-alpha

Chain A:



• Molecule 2: Histone acetyltransferase p300

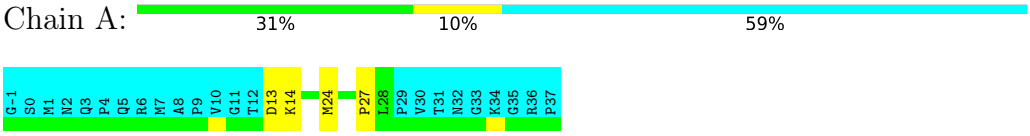
Chain B:



4.2.15 Score per residue for model 15

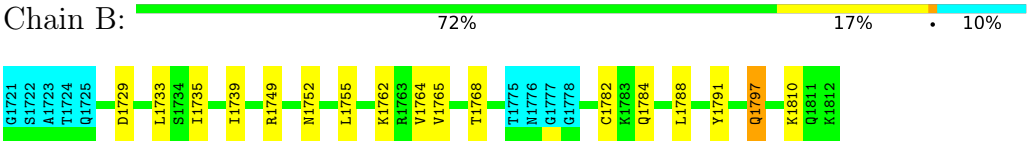
• Molecule 1: Transcription factor E2-alpha

Chain A:



• Molecule 2: Histone acetyltransferase p300

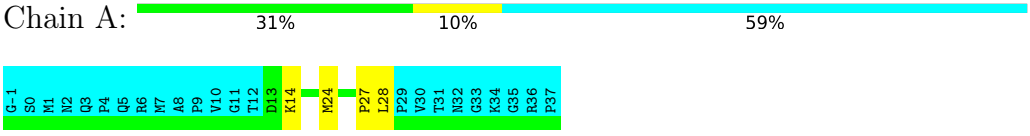
Chain B:



4.2.16 Score per residue for model 16

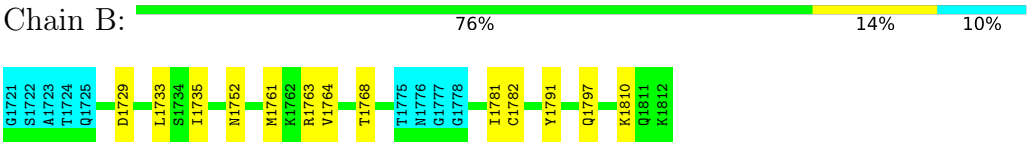
• Molecule 1: Transcription factor E2-alpha

Chain A:



• Molecule 2: Histone acetyltransferase p300

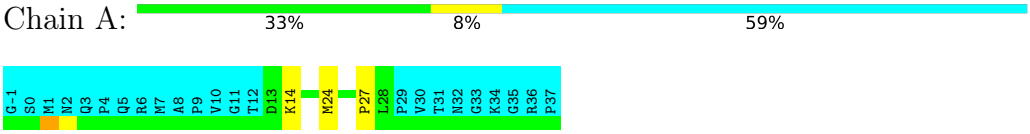
Chain B:



4.2.17 Score per residue for model 17

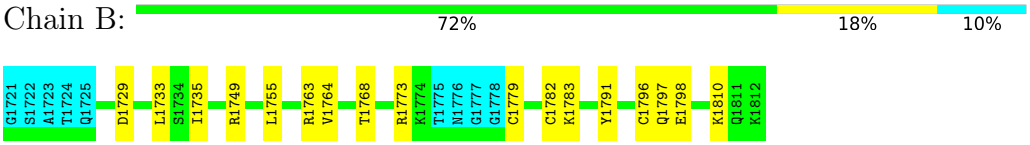
• Molecule 1: Transcription factor E2-alpha

Chain A:



• Molecule 2: Histone acetyltransferase p300

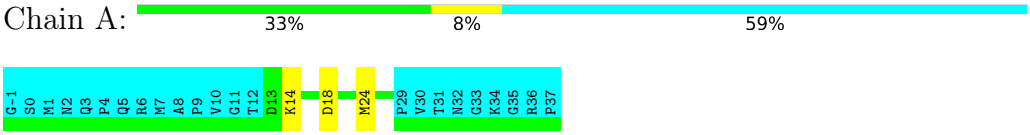
Chain B:



4.2.18 Score per residue for model 18

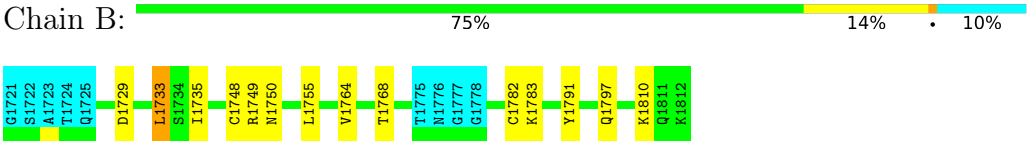
• Molecule 1: Transcription factor E2-alpha

Chain A:



• Molecule 2: Histone acetyltransferase p300

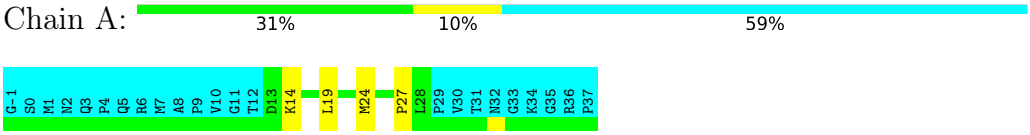
Chain B:



4.2.19 Score per residue for model 19

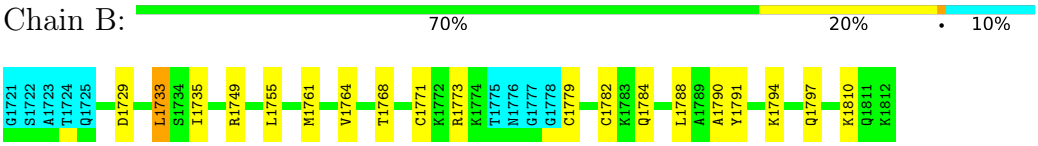
• Molecule 1: Transcription factor E2-alpha

Chain A:



• Molecule 2: Histone acetyltransferase p300

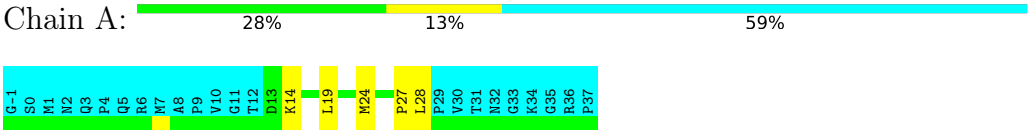
Chain B:



4.2.20 Score per residue for model 20

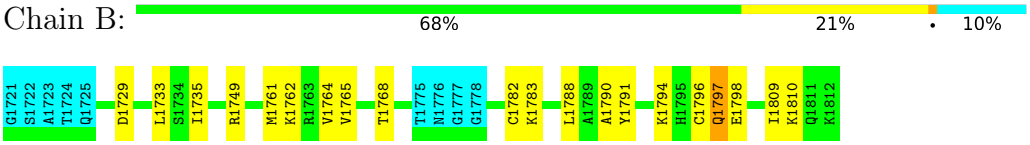
• Molecule 1: Transcription factor E2-alpha

Chain A:



• Molecule 2: Histone acetyltransferase p300

Chain B:



5 Refinement protocol and experimental data overview

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

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
CYANA	structure solution	2.1
CNS	refinement	
CYANA	refinement	2.1

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	1533
Number of shifts mapped to atoms	1533
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	88%

6 Model quality

6.1 Standard geometry

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

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	131	128	128	0±0
2	B	645	673	681	5±2
All	All	15520	16020	16180	94

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 3.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:1768:THR:HG22	2:B:1782:CYS:SG	0.79	2.18	2	20
2:B:1796:CYS:SG	2:B:1798:GLU:HG2	0.61	2.36	17	1
2:B:1748:CYS:SG	2:B:1750:ASN:OD1	0.56	2.64	8	4
2:B:1750:ASN:ND2	2:B:1753:CYS:SG	0.54	2.80	4	2
2:B:1748:CYS:SG	2:B:1762:LYS:HD3	0.54	2.43	4	1
2:B:1764:VAL:O	2:B:1768:THR:HG23	0.52	2.04	10	20
2:B:1771:CYS:SG	2:B:1773:ARG:HG2	0.52	2.44	19	1
2:B:1739:ILE:HG12	2:B:1788:LEU:HB3	0.52	1.80	11	2
2:B:1729:ASP:O	2:B:1733:LEU:HD22	0.51	2.06	17	18
1:A:27:PRO:HG2	2:B:1731:ARG:HB2	0.51	1.83	9	1
2:B:1796:CYS:SG	2:B:1798:GLU:HB3	0.47	2.49	20	1
2:B:1762:LYS:HA	2:B:1765:VAL:HG12	0.45	1.88	15	5
2:B:1773:ARG:HG3	2:B:1779:CYS:HB2	0.45	1.87	19	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:1797:GLN:H	2:B:1797:GLN:NE2	0.44	2.10	20	2
2:B:1773:ARG:HG2	2:B:1779:CYS:HB2	0.44	1.89	11	3
2:B:1801:CYS:SG	2:B:1802:PRO:HD2	0.43	2.53	13	1
2:B:1729:ASP:O	2:B:1733:LEU:HD13	0.43	2.13	8	1
2:B:1768:THR:HG21	2:B:1785:LEU:HD12	0.42	1.91	5	1
2:B:1807:LEU:HA	2:B:1810:LYS:HG2	0.42	1.91	3	1
2:B:1790:ALA:HA	2:B:1809:ILE:HD13	0.42	1.90	20	1
2:B:1739:ILE:HD13	2:B:1792:HIS:HB2	0.42	1.90	5	1
2:B:1784:GLN:O	2:B:1788:LEU:HD22	0.42	2.15	15	2
2:B:1790:ALA:O	2:B:1794:LYS:HG2	0.41	2.16	20	2
2:B:1797:GLN:HE21	2:B:1797:GLN:N	0.41	2.13	15	1
1:A:14:LYS:HD3	1:A:15:GLU:N	0.41	2.30	12	1

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	16/39 (41%)	14±0 (88±1%)	1±0 (3±3%)	1±1 (9±4%)	1 12
2	B	82/92 (89%)	81±1 (99±1%)	1±1 (1±1%)	0±0 (0±0%)	100 100
All	All	1960/2620 (75%)	1898 (97%)	34 (2%)	28 (1%)	15 61

All 2 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	27	PRO	19
1	A	28	LEU	9

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	16/34 (47%)	14±1 (86±5%)	2±1 (14±5%)	6 45
2	B	73/78 (94%)	65±1 (89±1%)	8±1 (11±1%)	10 54
All	All	1780/2240 (79%)	1577 (89%)	203 (11%)	9 52

All 27 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	24	MET	20
2	B	1735	ILE	20
2	B	1797	GLN	19
2	B	1810	LYS	19
2	B	1791	TYR	16
1	A	14	LYS	15
2	B	1749	ARG	14
2	B	1755	LEU	14
2	B	1783	LYS	13
2	B	1733	LEU	8
2	B	1752	ASN	7
2	B	1788	LEU	6
1	A	19	LEU	5
1	A	18	ASP	4
2	B	1763	ARG	3
2	B	1800	LYS	3
2	B	1761	MET	3
1	A	13	ASP	2
2	B	1808	ASN	2
2	B	1798	GLU	2
2	B	1781	ILE	2
2	B	1801	CYS	1
2	B	1753	CYS	1
2	B	1784	GLN	1
2	B	1799	ASN	1
2	B	1812	LYS	1
2	B	1807	LEU	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

There are no ligands in this entry.

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 88% for the well-defined parts and 86% 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	1533
Number of shifts mapped to atoms	1533
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
$^{13}C_{\alpha}$	130	-0.11 ± 0.08	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	122	0.31 ± 0.12	None needed (< 0.5 ppm)
$^{13}C'$	120	-0.14 ± 0.09	None needed (< 0.5 ppm)
^{15}N	116	-0.49 ± 0.36	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 88%, i.e. 1112 atoms were assigned a chemical shift out of a possible 1268. 14 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	471/483 (98%)	190/192 (99%)	189/198 (95%)	92/93 (99%)
Sidechain	590/721 (82%)	355/432 (82%)	223/248 (90%)	12/41 (29%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	51/64 (80%)	27/35 (77%)	24/24 (100%)	0/5 (0%)
Overall	1112/1268 (88%)	572/659 (87%)	436/470 (93%)	104/139 (75%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 86%, i.e. 1390 atoms were assigned a chemical shift out of a possible 1610. 16 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	606/635 (95%)	240/252 (95%)	250/262 (95%)	116/121 (96%)
Sidechain	733/911 (80%)	439/546 (80%)	276/311 (89%)	18/54 (33%)
Aromatic	51/64 (80%)	27/35 (77%)	24/24 (100%)	0/5 (0%)
Overall	1390/1610 (86%)	706/833 (85%)	550/597 (92%)	134/180 (74%)

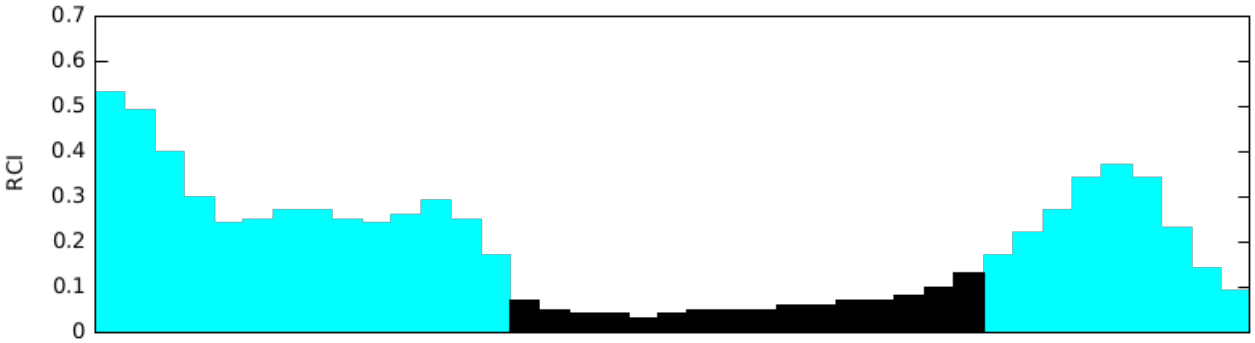
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

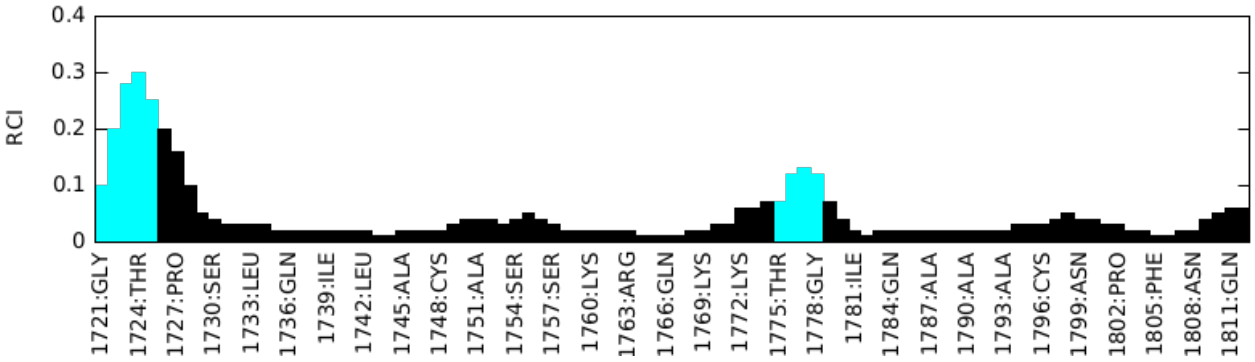
7.1.5 Random Coil Index (RCI) plots (i)

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
0.4
0.3
0.2
0.1
0
1721:GLY
1724:THR
1727:PRO
1730:SER
1733:LEU
1736:GLN
1739:ILE
1742:LEU
1745:ALA
1748:CYS
1751:ALA
1754:SER
1757:SER
1760:LYS
1763:ARG
1766:GLN
1769:LYS
1772:LYS
1775:THR
1778:GLY
1781:ILE
1784:GLN
1787:ALA
1790:ALA
1793:ALA
1796:CYS
1799:ASN
1802:PRO
1805:PHE
1808:ASN
1811:GLN