



Full wwPDB NMR Structure Validation Report ⓘ

Mar 8, 2026 – 12:41 PM UTC

PDB ID : 1MSZ / pdb_00001msz
Title : Solution structure of the R3H domain from human Smubp-2
Authors : Liepinsh, E.; Leonchiks, A.; Sharipo, A.; Guignard, L.; Otting, G.
Deposited on : 2002-09-20

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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4-5-2 with Phenix2.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
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)	:	2.49

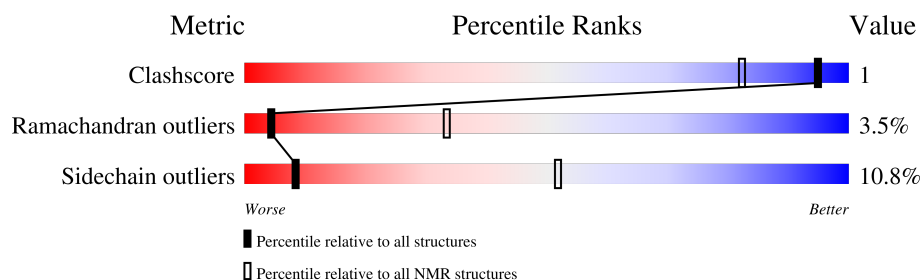
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment 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	229148	14424
Ramachandran outliers	224038	12848
Sidechain outliers	223484	12823

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	86	

2 Ensemble composition and analysis

This entry contains 20 models. Model 9 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:727-A:786 (60)	0.46	9

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 8 single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1000 atoms, of which 497 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called DNA-binding protein SMUBP-2.

Mol	Chain	Residues	Atoms						Trace
1	A	62	Total	C	H	N	O	S	0
			1000	312	497	100	88	3	

There are 10 discrepancies between the modelled and reference sequences:

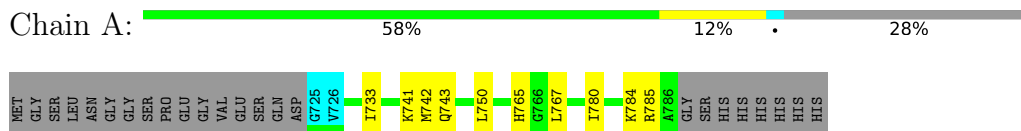
Chain	Residue	Modelled	Actual	Comment	Reference
A	709	MET	-	cloning artifact	UNP P38935
A	710	GLY	-	cloning artifact	UNP P38935
A	787	GLY	-	expression tag	UNP P38935
A	788	SER	-	expression tag	UNP P38935
A	789	HIS	-	expression tag	UNP P38935
A	790	HIS	-	expression tag	UNP P38935
A	791	HIS	-	expression tag	UNP P38935
A	792	HIS	-	expression tag	UNP P38935
A	793	HIS	-	expression tag	UNP P38935
A	794	HIS	-	expression tag	UNP P38935

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: DNA-binding protein SMUBP-2

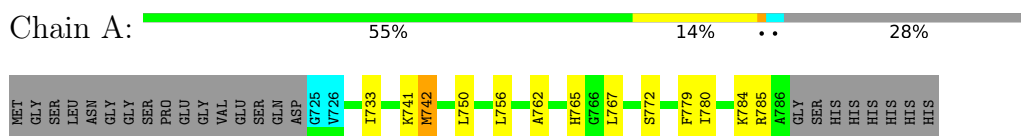


4.2 Scores per residue for each member of the ensemble

Colouring as in section [4.1](#) above.

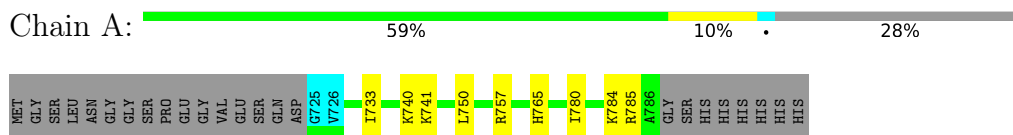
4.2.1 Score per residue for model 1

- Molecule 1: DNA-binding protein SMUBP-2



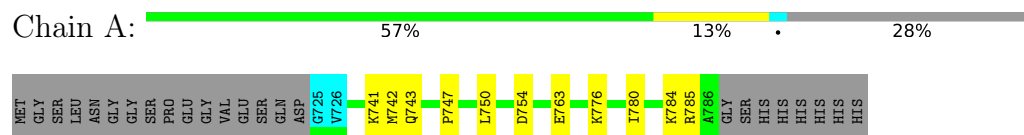
4.2.2 Score per residue for model 2

- Molecule 1: DNA-binding protein SMUBP-2



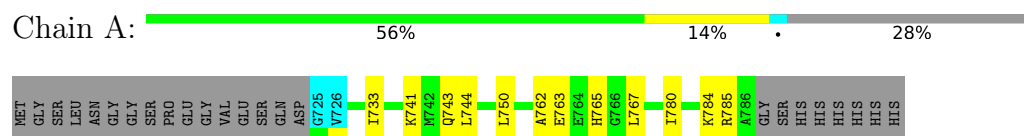
4.2.3 Score per residue for model 3

- Molecule 1: DNA-binding protein SMUBP-2



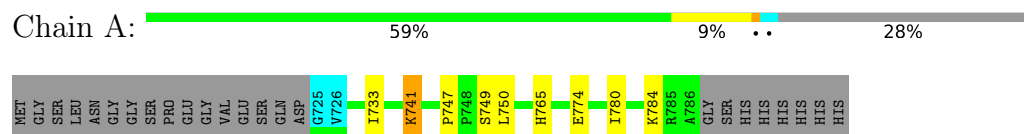
4.2.4 Score per residue for model 4

- Molecule 1: DNA-binding protein SMUBP-2



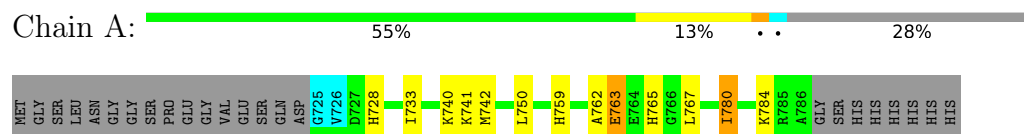
4.2.5 Score per residue for model 5

- Molecule 1: DNA-binding protein SMUBP-2



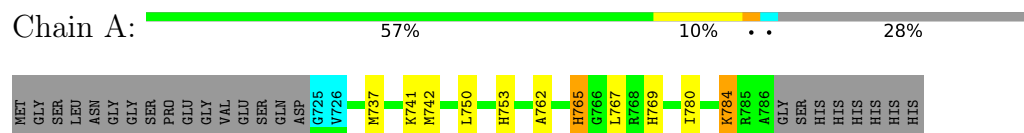
4.2.6 Score per residue for model 6

- Molecule 1: DNA-binding protein SMUBP-2



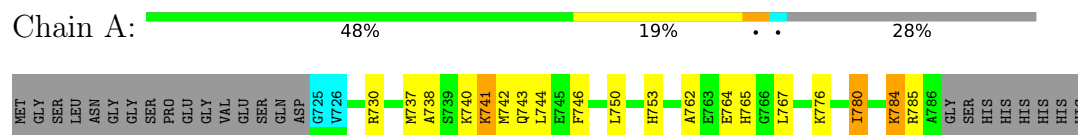
4.2.7 Score per residue for model 7

- Molecule 1: DNA-binding protein SMUBP-2



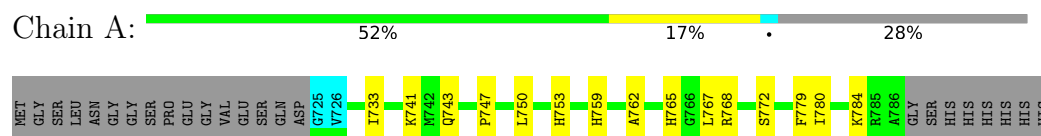
4.2.8 Score per residue for model 8

- Molecule 1: DNA-binding protein SMUBP-2



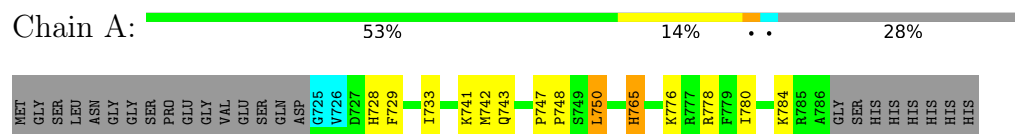
4.2.9 Score per residue for model 9 (medoid)

- Molecule 1: DNA-binding protein SMUBP-2



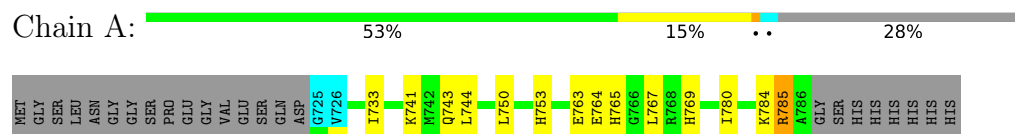
4.2.10 Score per residue for model 10

- Molecule 1: DNA-binding protein SMUBP-2



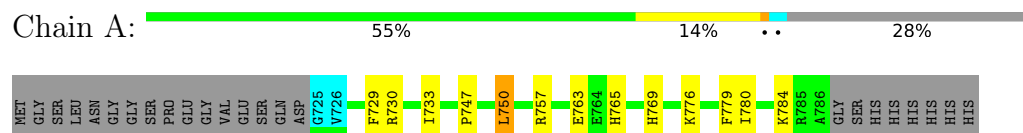
4.2.11 Score per residue for model 11

- Molecule 1: DNA-binding protein SMUBP-2



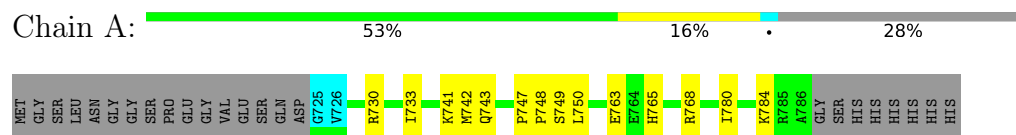
4.2.12 Score per residue for model 12

- Molecule 1: DNA-binding protein SMUBP-2



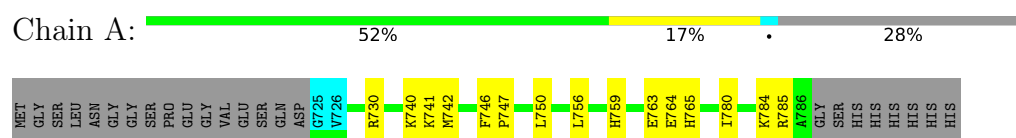
4.2.13 Score per residue for model 13

- Molecule 1: DNA-binding protein SMUBP-2



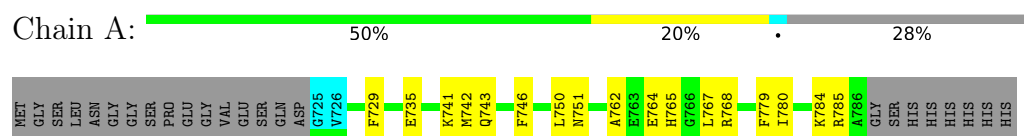
4.2.14 Score per residue for model 14

- Molecule 1: DNA-binding protein SMUBP-2



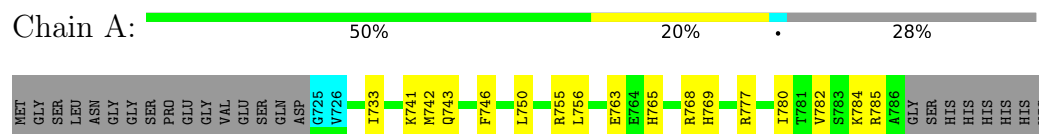
4.2.15 Score per residue for model 15

- Molecule 1: DNA-binding protein SMUBP-2



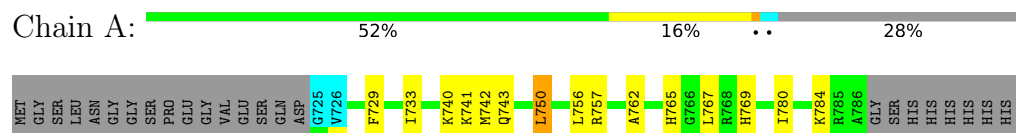
4.2.16 Score per residue for model 16

- Molecule 1: DNA-binding protein SMUBP-2



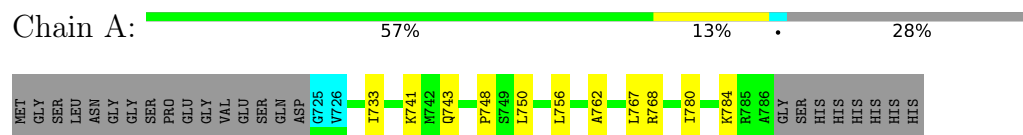
4.2.17 Score per residue for model 17

- Molecule 1: DNA-binding protein SMUBP-2



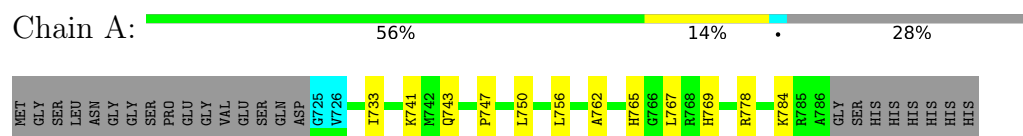
4.2.18 Score per residue for model 18

- Molecule 1: DNA-binding protein SMUBP-2



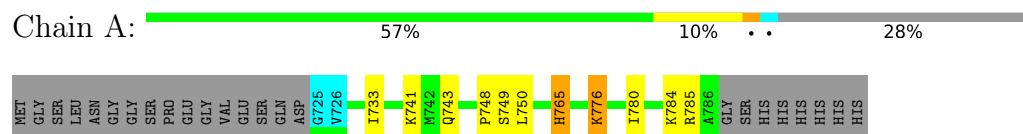
4.2.19 Score per residue for model 19

- Molecule 1: DNA-binding protein SMUBP-2



4.2.20 Score per residue for model 20

- Molecule 1: DNA-binding protein SMUBP-2



5 Refinement protocol and experimental data overview

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

Of the 80 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
DYANA	structure solution	1.5
OPAL	refinement	2.6

No chemical shift data was provided.

6 Model quality i

6.1 Standard geometry i

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	0.72±0.01	0±0/502 (0.0± 0.0%)	1.59±0.03	4±2/668 (0.7± 0.3%)
All	All	0.72	0/10040 (0.0%)	1.59	90/13360 (0.7%)

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	0.6±0.7
All	All	0	11

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(°)	Ideal(°)	Models	
								Worst	Total
1	A	747	PRO	N-CA-C	8.27	120.79	110.70	14	5
1	A	785	ARG	N-CA-C	-7.13	97.62	108.67	1	9
1	A	748	PRO	CA-C-N	6.67	134.29	121.54	20	4
1	A	748	PRO	C-N-CA	6.67	134.29	121.54	20	4
1	A	768	ARG	N-CA-C	-5.90	101.75	110.48	9	4
1	A	765	HIS	CB-CG-CD2	-5.89	123.54	131.20	10	17
1	A	754	ASP	CA-CB-CG	5.88	118.48	112.60	3	1
1	A	779	PHE	CA-CB-CG	5.88	119.68	113.80	15	1
1	A	729	PHE	CA-CB-CG	5.86	119.66	113.80	15	1
1	A	769	HIS	CB-CG-CD2	-5.72	123.77	131.20	17	5
1	A	763	GLU	CA-C-N	5.67	128.45	120.28	4	6
1	A	763	GLU	C-N-CA	5.67	128.45	120.28	4	6
1	A	747	PRO	CA-C-N	5.64	125.32	119.56	13	4
1	A	747	PRO	C-N-CA	5.64	125.32	119.56	13	4
1	A	753	HIS	CA-CB-CG	5.64	119.44	113.80	8	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	767	LEU	CA-C-N	5.61	131.36	122.94	11	1
1	A	767	LEU	C-N-CA	5.61	131.36	122.94	11	1
1	A	753	HIS	CB-CG-CD2	-5.31	124.29	131.20	8	3
1	A	728	HIS	CA-CB-CG	5.29	119.09	113.80	10	1
1	A	779	PHE	CA-C-N	5.28	129.91	123.10	12	1
1	A	779	PHE	C-N-CA	5.28	129.91	123.10	12	1
1	A	751	ASN	CA-CB-CG	5.18	117.78	112.60	15	1
1	A	728	HIS	CB-CG-CD2	-5.18	124.47	131.20	6	1
1	A	746	PHE	CB-CA-C	5.17	115.87	108.68	15	2
1	A	746	PHE	CA-C-N	5.13	125.67	120.38	14	1
1	A	746	PHE	C-N-CA	5.13	125.67	120.38	14	1
1	A	780	ILE	CB-CA-C	5.08	117.75	110.33	8	1
1	A	738	ALA	CA-C-N	5.02	127.01	120.28	8	1
1	A	738	ALA	C-N-CA	5.02	127.01	120.28	8	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	730	ARG	Sidechain	4
1	A	757	ARG	Sidechain	2
1	A	785	ARG	Sidechain	2
1	A	768	ARG	Sidechain	1
1	A	755	ARG	Sidechain	1
1	A	777	ARG	Sidechain	1

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	492	485	489	1±1
All	All	9840	9700	9780	22

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:762:ALA:HB1	1:A:767:LEU:HB2	0.50	1.81	15	10
1:A:759:HIS:CE1	1:A:780:ILE:CD1	0.47	2.97	6	1
1:A:769:HIS:CB	1:A:782:VAL:HG12	0.47	2.39	16	1
1:A:759:HIS:CE1	1:A:780:ILE:HD13	0.46	2.45	9	2
1:A:772:SER:HB2	1:A:779:PHE:CE2	0.46	2.46	1	2
1:A:737:MET:SD	1:A:784:LYS:HE3	0.45	2.51	8	2
1:A:744:LEU:HD22	1:A:746:PHE:CE1	0.45	2.46	8	1
1:A:729:PHE:CZ	1:A:750:LEU:HD11	0.44	2.47	17	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	59/86 (69%)	49±1 (84±2%)	8±1 (13±2%)	2±1 (3±2%)	4	33
All	All	1180/1720 (69%)	987 (84%)	152 (13%)	41 (3%)	4	33

All 5 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	741	LYS	19
1	A	742	MET	11
1	A	776	LYS	4
1	A	765	HIS	4
1	A	749	SER	3

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	53/73 (73%)	47±1 (89±2%)	6±1 (11±2%)	8	52
All	All	1060/1460 (73%)	945 (89%)	115 (11%)	8	52

All 17 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	750	LEU	20
1	A	784	LYS	20
1	A	780	ILE	17
1	A	733	ILE	15
1	A	743	GLN	13
1	A	756	LEU	6
1	A	740	LYS	5
1	A	764	GLU	4
1	A	763	GLU	3
1	A	744	LEU	2
1	A	741	LYS	2
1	A	778	ARG	2
1	A	776	LYS	2
1	A	742	MET	1
1	A	774	GLU	1
1	A	757	ARG	1
1	A	735	GLU	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 oligosaccharides in this entry.

6.6 Ligand geometry ⓘ

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

No chemical shift data were provided