wwPDB NMR Structure Validation Summary Report

Apr 16, 2023 – 08:35 AM EDT

PDB ID	:	8GBC
BMRB ID	:	51833
Title	:	Homo sapiens Zalpha mutant - N173S
Authors	:	Langeberg, C.J.; Nichols, P.J.; Henen, M.; Vicens, Q.; Vogeli, B.
Deposited on	:	2023-02-25

This is a wwPDB NMR Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

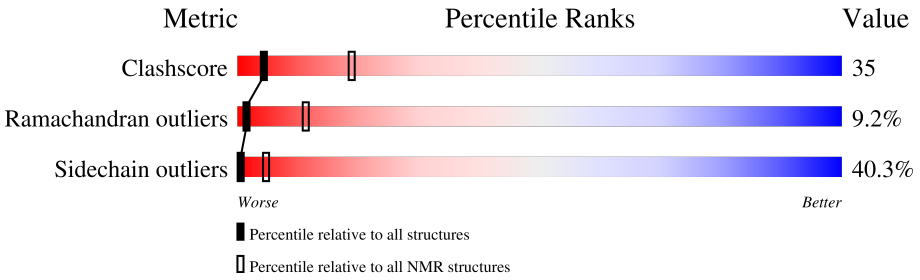
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.2

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 59%.

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	84	21% 26% 18% 10% 25%

2 Ensemble composition and analysis

This entry contains 20 models. Model 5 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:140-A:187, A:192-A:198 (55)	0.22	5

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 4 clusters and 5 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1011 atoms, of which 524 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Double-stranded RNA-specific adenosine deaminase.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	63	1011	312	524	85	90	0

There are 22 discrepancies between the modelled and reference sequences:

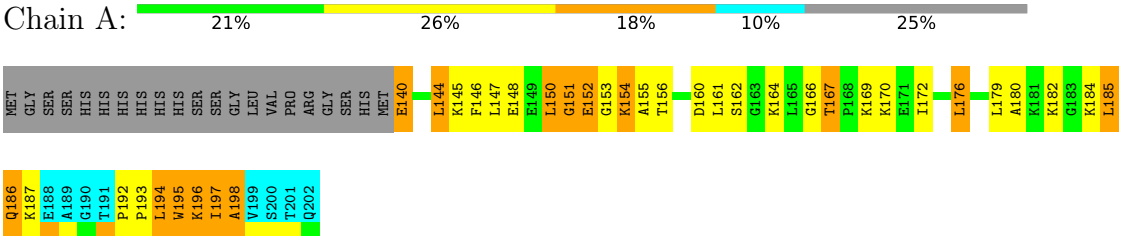
Chain	Residue	Modelled	Actual	Comment	Reference
A	119	MET	-	initiating methionine	UNP P55265
A	120	GLY	-	expression tag	UNP P55265
A	121	SER	-	expression tag	UNP P55265
A	122	SER	-	expression tag	UNP P55265
A	123	HIS	-	expression tag	UNP P55265
A	124	HIS	-	expression tag	UNP P55265
A	125	HIS	-	expression tag	UNP P55265
A	126	HIS	-	expression tag	UNP P55265
A	127	HIS	-	expression tag	UNP P55265
A	128	HIS	-	expression tag	UNP P55265
A	129	SER	-	expression tag	UNP P55265
A	130	SER	-	expression tag	UNP P55265
A	131	GLY	-	expression tag	UNP P55265
A	132	LEU	-	expression tag	UNP P55265
A	133	VAL	-	expression tag	UNP P55265
A	134	PRO	-	expression tag	UNP P55265
A	135	ARG	-	expression tag	UNP P55265
A	136	GLY	-	expression tag	UNP P55265
A	137	SER	-	expression tag	UNP P55265
A	138	HIS	-	expression tag	UNP P55265
A	139	MET	-	expression tag	UNP P55265
A	173	SER	ASN	engineered mutation	UNP P55265

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.

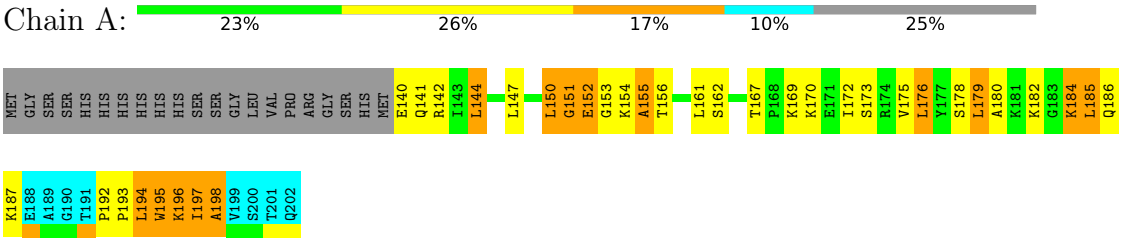
• Molecule 1: Double-stranded RNA-specific adenosine deaminase



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

The representative model is number 5. Colouring as in section 4.1 above.

• Molecule 1: Double-stranded RNA-specific adenosine deaminase



5 Refinement protocol and experimental data overview

The models were refined using the following method: *na*.

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 calculation	3.98.13

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	528
Number of shifts mapped to atoms	518
Number of unparsed shifts	0
Number of shifts with mapping errors	10
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	59%

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	433	474	473	32±6
All	All	8660	9480	9460	640

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 35.

5 of 260 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:194:LEU:C	1:A:194:LEU:HD13	0.99	1.75	6	1
1:A:194:LEU:HD12	1:A:195:TRP:N	0.91	1.80	17	5
1:A:194:LEU:C	1:A:194:LEU:HD23	0.90	1.88	10	7
1:A:172:ILE:HD13	1:A:172:ILE:O	0.88	1.68	16	2
1:A:179:LEU:HD21	1:A:185:LEU:HD22	0.86	1.47	9	2

6.3 Torsion angles

6.3.1 Protein backbone

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	54/84 (64%)	43±2 (80±3%)	6±2 (11±3%)	5±1 (9±2%)	1 11
All	All	1080/1680 (64%)	859 (80%)	122 (11%)	99 (9%)	1 11

5 of 10 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	152	GLU	20
1	A	197	ILE	19
1	A	151	GLY	15
1	A	198	ALA	12
1	A	150	LEU	11

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	46/70 (66%)	27±3 (60±6%)	19±3 (40±6%)	0 5
All	All	920/1400 (66%)	549 (60%)	371 (40%)	0 5

5 of 40 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	194	LEU	19
1	A	156	THR	18
1	A	160	ASP	18
1	A	144	LEU	17
1	A	167	THR	17

6.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 59% for the well-defined parts and 59% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chemical_shifts_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	528
Number of shifts mapped to atoms	518
Number of unparsed shifts	0
Number of shifts with mapping errors	10
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. First 5 (of 10) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	138	HIS	HA	3.302	0.000	.
1	A	138	HIS	HB2	3.302	0.000	.
1	A	138	HIS	CA	59.784	0.000	.
1	A	138	HIS	CB	30.101	0.000	.
1	A	139	MET	H	8.469	0.000	.
1	A	139	MET	HA	4.227	0.000	.
1	A	139	MET	HB2	1.938	0.000	.
1	A	139	MET	CA	57.815	0.000	.
1	A	139	MET	CB	31.861	0.000	.
1	A	139	MET	N	118.493	0.000	.

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	65	-0.43 ± 0.21	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	57	-0.36 ± 0.24	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	61	0.74 ± 0.64	None needed (imprecise)

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 59%, i.e. 467 atoms were assigned a chemical shift out of a possible 785. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	214/274 (78%)	107/112 (96%)	55/110 (50%)	52/52 (100%)
Sidechain	251/473 (53%)	166/307 (54%)	83/148 (56%)	2/18 (11%)
Aromatic	2/38 (5%)	1/19 (5%)	0/17 (0%)	1/2 (50%)
Overall	467/785 (59%)	274/438 (63%)	138/275 (50%)	55/72 (76%)

7.1.4 Statistically unusual chemical shifts

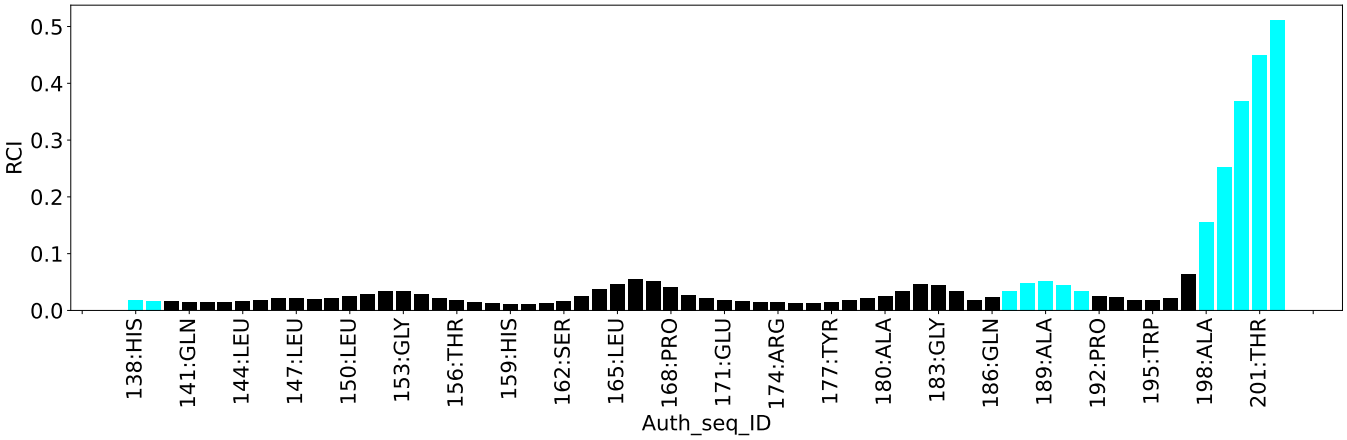
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	140	GLU	HG2	3.62	1.24 – 3.30	6.6
1	A	140	GLU	HG3	3.62	1.20 – 3.30	6.5
1	A	175	VAL	CG1	28.58	14.71 – 28.29	5.2

7.1.5 Random Coil Index (RCI) plots

The image below reports *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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:


RCI
0.0
0.1
0.2
0.3
0.4
0.5
138:HIS
141:GLN
144:LEU
147:LEU
150:LEU
153:GLY
156:THR
159:HIS
162:SER
165:LEU
168:PRO
171:GLU
174:ARG
177:TYR
180:ALA
183:GLY
186:GLN
189:ALA
192:PRO
195:TRP
198:ALA
201:THR
Auth_seq_ID