



# wwPDB NMR Structure Validation Summary Report ⓘ

Nov 9, 2024 – 11:22 am GMT

PDB ID : 2BZM  
Title : Solution structure of the primary host recognition region of complement factor H  
Authors : Herbert, A.P.; Uhrin, D.; Lyon, M.; Pangburn, M.K.; Barlow, P.N.  
Deposited on : 2005-08-18

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

We welcome your comments at [validation@mail.wwpdb.org](mailto: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	:	4.02b-467
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
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)	:	2.39

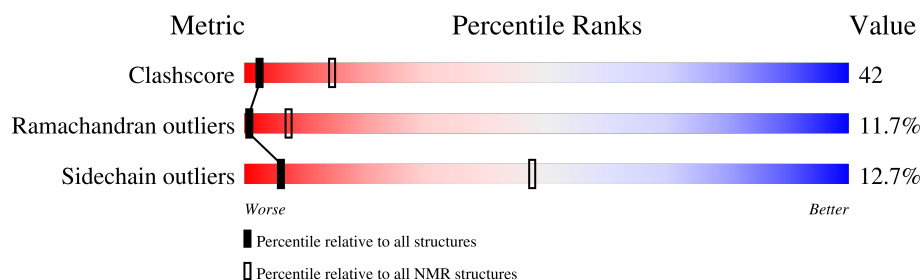
# 1 Overall quality at a glance

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	210492	14027
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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  $\geq 3$ , 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	129	<div> <div></div> <div>25%</div> <div>62%</div> <div>10%</div> <div>.</div> </div>

## 2 Ensemble composition and analysis

This entry contains 26 models. Model 17 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:1107-A:1231 (125)	1.41	17

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 5 clusters and 2 single-model clusters were found.

Cluster number	Models
1	8, 9, 10, 12, 17, 25, 26
2	3, 11, 14, 18, 20, 24
3	2, 5, 21, 22
4	7, 15, 16, 19
5	1, 6, 13
Single-model clusters	4; 23

### 3 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 2040 atoms, of which 1006 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called COMPLEMENT FACTOR H.

Mol	Chain	Residues	Atoms						Trace
1	A	129	Total	C	H	N	O	S	0
			2040	647	1006	185	193	9	



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *RESTRAINED MOLECULAR DYNAMICS*.

Of the ? calculated structures, 26 were deposited, based on the following criterion: *LOWEST OVERALL ENERGY*.

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

Software name	Classification	Version
CNS	refinement	
ANSIG	structure solution	

No chemical shift data was provided.

## 6 Model quality [i](#)

### 6.1 Standard geometry [i](#)

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 [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	1000	978	976	84±7
All	All	26000	25428	25376	2180

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 42.

5 of 905 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:1179:ILE:HA	1:A:1201:CYS:HA	0.89	1.42	19	7
1:A:1113:PRO:HD2	1:A:1157:TRP:HB2	0.88	1.45	12	23
1:A:1169:ILE:HG12	1:A:1223:LEU:HD23	0.88	1.45	14	1
1:A:1180:ALA:HB3	1:A:1200:VAL:HG23	0.84	1.48	8	1
1:A:1115:ILE:HG23	1:A:1160:PRO:HG2	0.84	1.50	5	10

### 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	124/129 (96%)	76±4 (61±4%)	34±4 (27±3%)	15±3 (12±2%)	1	7
All	All	3224/3354 (96%)	1966 (61%)	880 (27%)	378 (12%)	1	7

5 of 59 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	1161	PRO	25
1	A	1203	ARG	23
1	A	1230	LYS	23
1	A	1202	LYS	22
1	A	1187	GLN	20

### 6.3.2 Protein sidechains ⓘ

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	112/115 (97%)	98±3 (87±3%)	14±3 (13±3%)	6	47
All	All	2912/2990 (97%)	2543 (87%)	369 (13%)	6	47

5 of 62 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	1148	LYS	24
1	A	1153	ARG	23
1	A	1120	ILE	19
1	A	1128	TYR	18
1	A	1189	LEU	18

### 6.3.3 RNA ⓘ

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 oligosaccharides 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

No chemical shift data were provided