wwPDB NMR Structure Validation Summary Report

May 7, 2024 – 01:51 pm BST

PDB ID	:	1E0E
BMRB ID	:	4619
Title	:	N-terminal zinc-binding HHCC domain of HIV-2 integrase
Authors	:	Eijkelenboom, A.P.A.M.; Van Den ent, F.M.I.; Plasterk, R.H.A.; Kaptein, R.; Boelens, R.
Deposited on	:	2000-03-25

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:

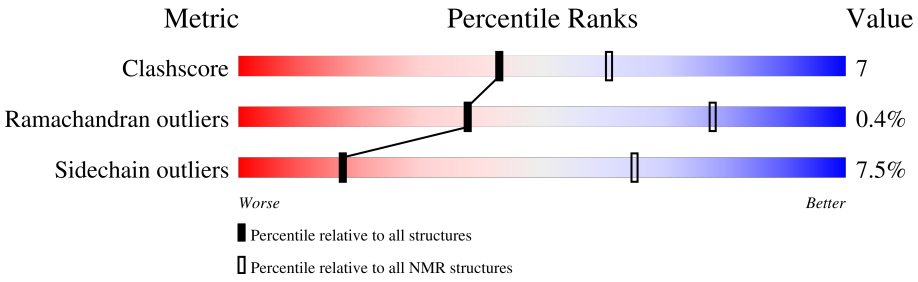
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 85%.

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	55	62% 9% 15% 15%
1	B	55	62% 9% 15% 15%

2 Ensemble composition and analysis

This entry contains 32 models. Model 27 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:6-A:44 (39)	0.27	27
2	B:6-B:44 (39)	0.27	27

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 6 clusters and 1 single-model cluster was found.

Cluster number	Models
1	5, 11, 13, 15, 16, 28, 31
2	1, 4, 9, 14, 17, 18, 24
3	7, 10, 20, 21, 27, 29, 32
4	2, 8, 19, 25
5	3, 6, 23
6	12, 22, 26
Single-model clusters	30

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1494 atoms, of which 740 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called HUMAN IMMUNODEFICIENCY VIRUS TYPE 2 INTEGRASE.

Mol	Chain	Residues	Atoms						Trace
1	A	47	Total	C	H	N	O	S	1
			746	235	370	69	70	2	
1	B	47	Total	C	H	N	O	S	1
			746	235	370	69	70	2	

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
2	A	1	Total	Zn
			1	1
2	B	1	Total	Zn
			1	1

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: HUMAN IMMUNODEFICIENCY VIRUS TYPE 2 INTEGRASE

Chain A:

• Molecule 1: HUMAN IMMUNODEFICIENCY VIRUS TYPE 2 INTEGRASE

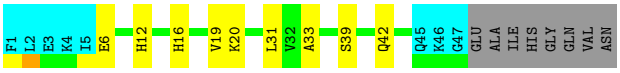
Chain B:

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

The representative model is number 27. Colouring as in section 4.1 above.

• Molecule 1: HUMAN IMMUNODEFICIENCY VIRUS TYPE 2 INTEGRASE

Chain A:



• Molecule 1: HUMAN IMMUNODEFICIENCY VIRUS TYPE 2 INTEGRASE

Chain B:

5 Refinement protocol and experimental data overview

The models were refined using the following method: *simulated annealing*.

Of the 50 calculated structures, 32 were deposited, based on the following criterion: *LOW OVERALL ENERGY*.

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

Software name	Classification	Version
X-PLOR	refinement	3.851
X-PLOR	structure solution	3.851

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	1288
Number of shifts mapped to atoms	1104
Number of unparsed shifts	0
Number of shifts with mapping errors	184
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	85%

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	312	297	297	4±1
1	B	312	297	297	4±1
All	All	20032	19008	19008	277

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 7.

5 of 33 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:B:19:VAL:HG13	1:B:33:ALA:HB3	0.94	1.40	15	32
1:A:19:VAL:HG13	1:A:33:ALA:HB3	0.93	1.40	15	32
1:B:19:VAL:HG13	1:B:33:ALA:CB	0.77	2.10	9	32
1:A:19:VAL:HG13	1:A:33:ALA:CB	0.77	2.10	13	32
1:B:19:VAL:CG1	1:B:33:ALA:HB3	0.67	2.20	25	27

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	39/55 (71%)	36±1 (93±3%)	3±1 (7±3%)	0±0 (0±1%)	38	78
1	B	39/55 (71%)	36±1 (93±3%)	3±1 (7±3%)	0±0 (0±1%)	38	78
All	All	2496/3520 (71%)	2321 (93%)	165 (7%)	10 (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	27	GLY	3
1	B	27	GLY	3
1	A	29	PRO	2
1	B	29	PRO	2

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	35/48 (73%)	32±1 (92±4%)	3±1 (8±4%)	17	65
1	B	35/48 (73%)	32±1 (93±4%)	3±1 (7±4%)	17	65
All	All	2240/3072 (73%)	2073 (93%)	167 (7%)	17	65

5 of 28 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	20	LYS	19
1	B	20	LYS	19
1	A	6	GLU	12

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	B	6	GLU	12
1	A	14	LYS	9

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)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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 85% 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	1288
Number of shifts mapped to atoms	1104
Number of unparsed shifts	0
Number of shifts with mapping errors	184
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	8

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 184) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	1	PHE	H	8.3	.	.
1	B	1	PHE	H	8.3	.	.
1	A	47	GLY	H	8.202	.	.
1	B	47	GLY	H	8.202	.	.
1	A	47	GLY	HA3	3.981	.	.
1	B	47	GLY	HA3	3.981	.	.
1	A	47	GLY	HA2	3.981	.	.
1	B	47	GLY	HA2	3.981	.	.
1	A	47	GLY	CA	43.306	.	.
1	B	47	GLY	CA	43.306	.	.
1	A	48	GLU	H	8.224	.	.
1	B	48	GLU	H	8.224	.	.
1	A	48	GLU	HA	4.28	.	.
1	B	48	GLU	HA	4.28	.	.

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	48	GLU	HB3	2.064	.	.
1	B	48	GLU	HB3	2.064	.	.
1	A	48	GLU	HB2	1.948	.	.
1	B	48	GLU	HB2	1.948	.	.
1	A	48	GLU	HG3	2.315	.	.
1	B	48	GLU	HG3	2.315	.	.
1	A	48	GLU	HG2	2.277	.	.
1	B	48	GLU	HG2	2.277	.	.
1	A	48	GLU	N	121.446	.	.
1	B	48	GLU	N	121.446	.	.
1	A	48	GLU	CA	54.387	.	.
1	B	48	GLU	CA	54.387	.	.
1	A	48	GLU	CG	34.136	.	.
1	B	48	GLU	CG	34.136	.	.
1	A	48	GLU	CB	28.31	.	.
1	B	48	GLU	CB	28.31	.	.
1	A	49	ALA	H	8.264	.	.
1	B	49	ALA	H	8.264	.	.
1	A	49	ALA	HA	4.314	.	.
1	B	49	ALA	HA	4.314	.	.
1	A	49	ALA	HB1	1.367	.	.
1	B	49	ALA	HB1	1.367	.	.
1	A	49	ALA	HB2	1.367	.	.
1	B	49	ALA	HB2	1.367	.	.
1	A	49	ALA	HB3	1.367	.	.
1	B	49	ALA	HB3	1.367	.	.
1	A	49	ALA	N	125.657	.	.
1	B	49	ALA	N	125.657	.	.
1	A	49	ALA	CB	17.156	.	.
1	B	49	ALA	CB	17.156	.	.
1	A	49	ALA	CA	50.412	.	.
1	B	49	ALA	CA	50.412	.	.
1	A	50	ILE	H	7.963	.	.
1	B	50	ILE	H	7.963	.	.
1	A	50	ILE	HA	4.093	.	.
1	B	50	ILE	HA	4.093	.	.
1	A	50	ILE	HB	1.797	.	.
1	B	50	ILE	HB	1.797	.	.
1	A	50	ILE	HG13	1.366	.	.
1	B	50	ILE	HG13	1.366	.	.
1	A	50	ILE	HG12	1.134	.	.

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	B	50	ILE	HG12	1.134	.	.
1	A	50	ILE	HG21	0.805	.	.
1	B	50	ILE	HG21	0.805	.	.
1	A	50	ILE	HG22	0.805	.	.
1	B	50	ILE	HG22	0.805	.	.
1	A	50	ILE	HG23	0.805	.	.
1	B	50	ILE	HG23	0.805	.	.
1	A	50	ILE	HD11	0.805	.	.
1	B	50	ILE	HD11	0.805	.	.
1	A	50	ILE	HD12	0.805	.	.
1	B	50	ILE	HD12	0.805	.	.
1	A	50	ILE	HD13	0.805	.	.
1	B	50	ILE	HD13	0.805	.	.
1	A	50	ILE	N	120.168	.	.
1	B	50	ILE	N	120.168	.	.
1	A	50	ILE	CB	36.52	.	.
1	B	50	ILE	CB	36.52	.	.
1	A	50	ILE	CA	59.071	.	.
1	B	50	ILE	CA	59.071	.	.
1	A	50	ILE	CG1	25.196	.	.
1	B	50	ILE	CG1	25.196	.	.
1	A	50	ILE	CG2	15.545	.	.
1	B	50	ILE	CG2	15.545	.	.
1	A	50	ILE	CD1	10.99	.	.
1	B	50	ILE	CD1	10.99	.	.
1	A	51	HIS	H	8.354	.	.
1	B	51	HIS	H	8.354	.	.
1	A	51	HIS	HA	4.71	.	.
1	B	51	HIS	HA	4.71	.	.
1	A	51	HIS	HB3	3.234	.	.
1	B	51	HIS	HB3	3.234	.	.
1	A	51	HIS	HB2	3.137	.	.
1	B	51	HIS	HB2	3.137	.	.
1	A	51	HIS	HD2	7.172	.	.
1	B	51	HIS	HD2	7.172	.	.
1	A	51	HIS	HE1	8.19	.	.
1	B	51	HIS	HE1	8.19	.	.
1	A	51	HIS	N	123.39	.	.
1	B	51	HIS	N	123.39	.	.
1	A	51	HIS	CA	53.47	.	.
1	B	51	HIS	CA	53.47	.	.

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	51	HIS	CB	27.625	.	.
1	B	51	HIS	CB	27.625	.	.
1	A	51	HIS	CD2	134.871	.	.
1	B	51	HIS	CD2	134.871	.	.
1	A	52	GLY	HA3	3.957	.	.
1	B	52	GLY	HA3	3.957	.	.
1	A	52	GLY	HA2	3.957	.	.
1	B	52	GLY	HA2	3.957	.	.
1	A	52	GLY	CA	43.095	.	.
1	B	52	GLY	CA	43.095	.	.
1	A	52	GLY	H	8.34	.	.
1	B	52	GLY	H	8.34	.	.
1	A	52	GLY	N	110.96	.	.
1	B	52	GLY	N	110.96	.	.
1	A	53	GLN	H	8.248	.	.
1	B	53	GLN	H	8.248	.	.
1	A	53	GLN	HA	4.394	.	.
1	B	53	GLN	HA	4.394	.	.
1	A	53	GLN	HB3	2.108	.	.
1	B	53	GLN	HB3	2.108	.	.
1	A	53	GLN	HB2	1.995	.	.
1	B	53	GLN	HB2	1.995	.	.
1	A	53	GLN	HG3	2.36	.	.
1	B	53	GLN	HG3	2.36	.	.
1	A	53	GLN	HG2	2.335	.	.
1	B	53	GLN	HG2	2.335	.	.
1	A	53	GLN	HE21	7.584	.	.
1	B	53	GLN	HE21	7.584	.	.
1	A	53	GLN	HE22	6.851	.	.
1	B	53	GLN	HE22	6.851	.	.
1	A	53	GLN	N	121.184	.	.
1	B	53	GLN	N	121.184	.	.
1	A	53	GLN	CA	53.66	.	.
1	B	53	GLN	CA	53.66	.	.
1	A	53	GLN	CB	27.567	.	.
1	B	53	GLN	CB	27.567	.	.
1	A	53	GLN	CG	31.7	.	.
1	B	53	GLN	CG	31.7	.	.
1	A	53	GLN	NE2	113.586	.	.
1	B	53	GLN	NE2	113.586	.	.
1	A	54	VAL	H	8.248	.	.

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	B	54	VAL	H	8.248	.	.
1	A	54	VAL	HA	4.165	.	.
1	B	54	VAL	HA	4.165	.	.
1	A	54	VAL	HB	2.096	.	.
1	B	54	VAL	HB	2.096	.	.
1	A	54	VAL	HG11	0.931	.	.
1	B	54	VAL	HG11	0.931	.	.
1	A	54	VAL	HG12	0.931	.	.
1	B	54	VAL	HG12	0.931	.	.
1	A	54	VAL	HG13	0.931	.	.
1	B	54	VAL	HG13	0.931	.	.
1	A	54	VAL	HG21	0.911	.	.
1	B	54	VAL	HG21	0.911	.	.
1	A	54	VAL	HG22	0.911	.	.
1	B	54	VAL	HG22	0.911	.	.
1	A	54	VAL	HG23	0.911	.	.
1	B	54	VAL	HG23	0.911	.	.
1	A	54	VAL	N	122.516	.	.
1	B	54	VAL	N	122.516	.	.
1	A	54	VAL	CA	60.109	.	.
1	B	54	VAL	CA	60.109	.	.
1	A	54	VAL	CB	30.81	.	.
1	B	54	VAL	CB	30.81	.	.
1	A	54	VAL	CG1	19.385	.	.
1	B	54	VAL	CG1	19.385	.	.
1	A	54	VAL	CG2	18.363	.	.
1	B	54	VAL	CG2	18.363	.	.
1	A	55	ASN	H	8.073	.	.
1	B	55	ASN	H	8.073	.	.
1	A	55	ASN	HA	4.488	.	.
1	B	55	ASN	HA	4.488	.	.
1	A	55	ASN	HB3	2.774	.	.
1	B	55	ASN	HB3	2.774	.	.
1	A	55	ASN	HB2	2.669	.	.
1	B	55	ASN	HB2	2.669	.	.
1	A	55	ASN	HD21	7.5	.	.
1	B	55	ASN	HD21	7.5	.	.
1	A	55	ASN	HD22	6.802	.	.
1	B	55	ASN	HD22	6.802	.	.
1	A	55	ASN	N	128.509	.	.
1	B	55	ASN	N	128.509	.	.

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	55	ASN	CA	52.658	.	.
1	B	55	ASN	CA	52.658	.	.
1	A	55	ASN	CB	38.427	.	.
1	B	55	ASN	CB	38.427	.	.
1	A	55	ASN	ND2	113.752	.	.
1	B	55	ASN	ND2	113.752	.	.

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	110	1.54 ± 0.12	Should be checked
$^{13}C_{\beta}$	104	2.06 ± 0.11	Should be checked
$^{13}C'$	0	—	None (insufficient data)
^{15}N	104	-0.37 ± 0.37	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 85%, i.e. 894 atoms were assigned a chemical shift out of a possible 1054. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	302/384 (79%)	152/154 (99%)	78/156 (50%)	72/74 (97%)
Sidechain	542/590 (92%)	368/378 (97%)	160/186 (86%)	14/26 (54%)
Aromatic	50/80 (62%)	32/42 (76%)	18/32 (56%)	0/6 (0%)
Overall	894/1054 (85%)	552/574 (96%)	256/374 (68%)	86/106 (81%)

7.1.4 Statistically unusual chemical shifts (i)

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	16	HIS	HA	1.95	2.49 – 6.71	-6.3
1	B	16	HIS	HA	1.95	2.49 – 6.71	-6.3
1	A	12	HIS	CD2	138.85	103.95 – 136.66	5.7

Continued on next page...

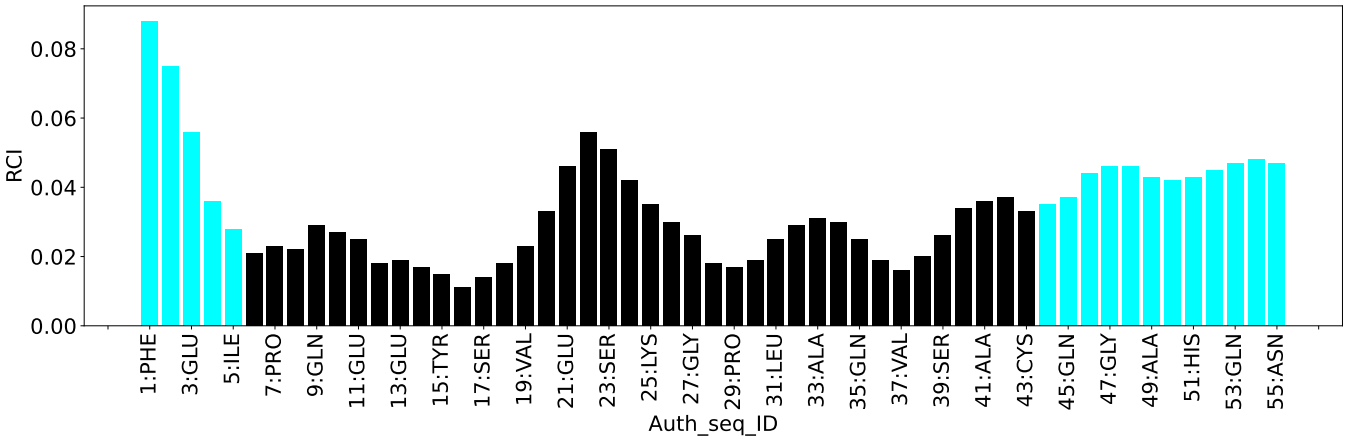
Continued from previous page...

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	B	12	HIS	CD2	138.85	103.95 – 136.66	5.7
1	A	25	LYS	HG2	0.08	0.13 – 2.61	-5.2
1	B	25	LYS	HG2	0.08	0.13 – 2.61	-5.2
1	A	16	HIS	CD2	136.86	103.95 – 136.66	5.1
1	B	16	HIS	CD2	136.86	103.95 – 136.66	5.1

7.1.5 Random Coil Index (RCI) plots (i)

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:



Random coil index (RCI) for chain B: