Full wwPDB NMR Structure Validation Report ⓘ

Apr 20, 2024 – 11:32 AM EDT

PDB ID	:	2N3A
BMRB ID	:	25639
Title	:	Solution structure of LEDGF/p75 IBD in complex with POGZ peptide (1389-1404)
Authors	:	Tesina, P.; Cermakova, K.; Horejsi, M.; Prochazkova, K.; Fabry, M.; Sharma, S.; Christ, F.; Demeulemeester, J.; Debyser, Z.; De Rijck, J.; Veverka, V.; Rezacova, P.
Deposited on	:	2015-05-26

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

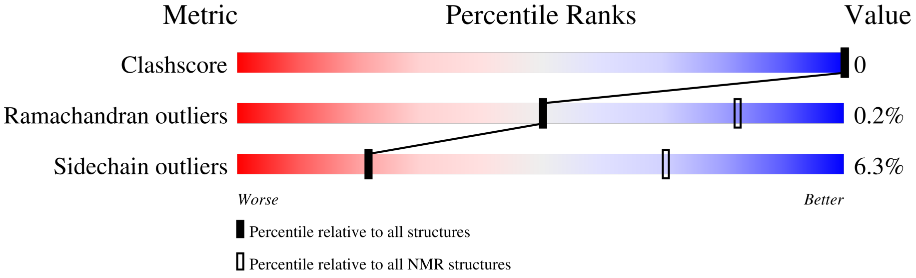
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson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

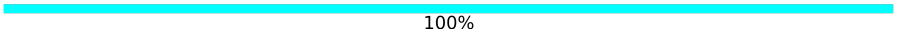
The overall completeness of chemical shifts assignment is 86%.

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	16	
2	B	79	

2 Ensemble composition and analysis ⓘ

This entry contains 29 models. Model 28 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	B:348-B:423 (76)	0.69	28

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 2 clusters and 1 single-model cluster was found.

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

3 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 1564 atoms, of which 786 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Pogo transposable element with ZNF domain.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	16	226	78	97	16	35	0

- Molecule 2 is a protein called PC4 and SFRS1-interacting protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	79	1338	404	689	118	120	7	0

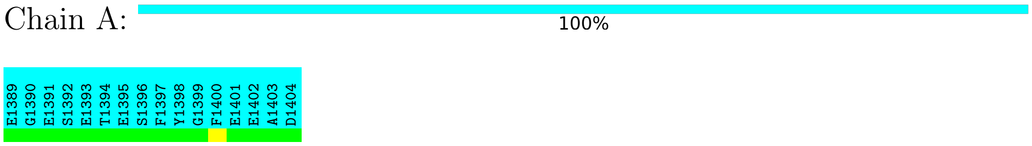
4 Residue-property plots (i)

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.

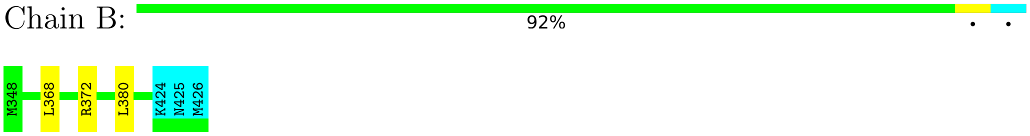
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

Chain B:



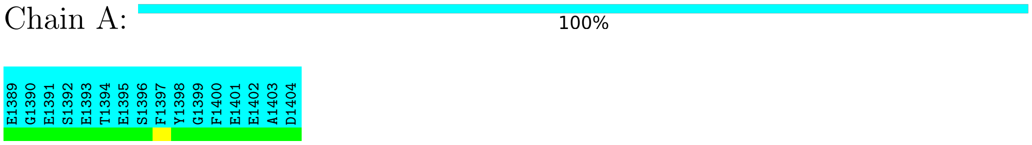
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

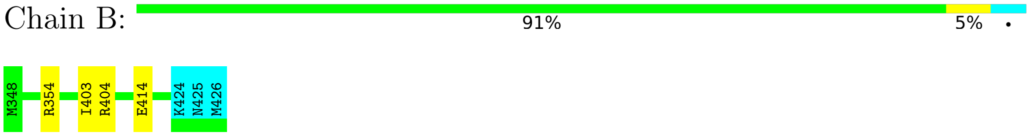
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

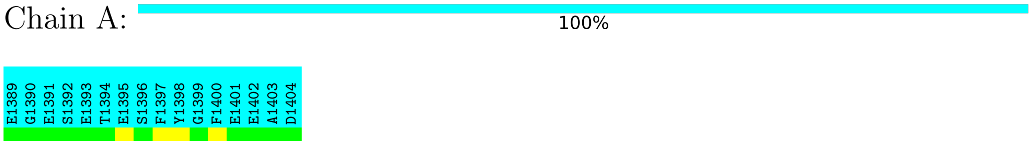
Chain B:



4.2.2 Score per residue for model 2

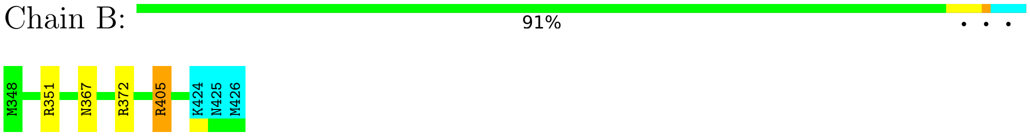
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

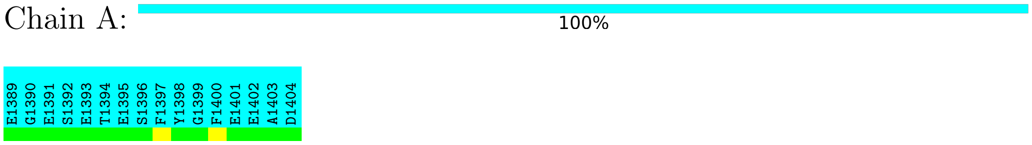
Chain B:



4.2.3 Score per residue for model 3

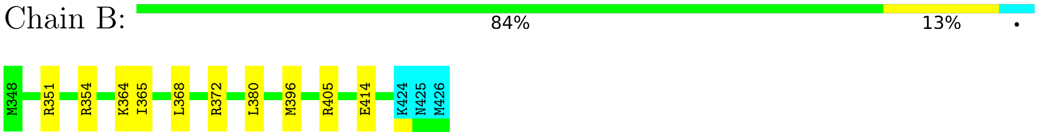
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

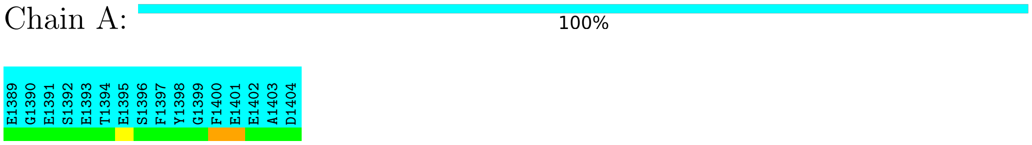
Chain B:



4.2.4 Score per residue for model 4

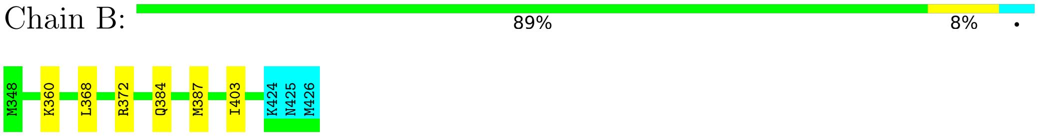
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



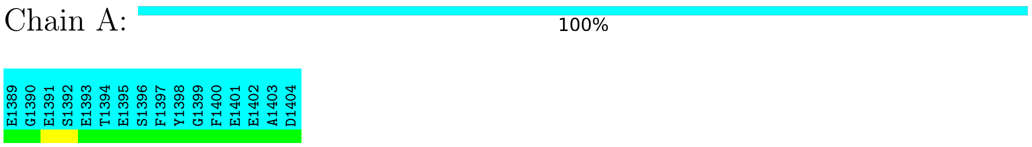
- Molecule 2: PC4 and SFRS1-interacting protein

Chain B:

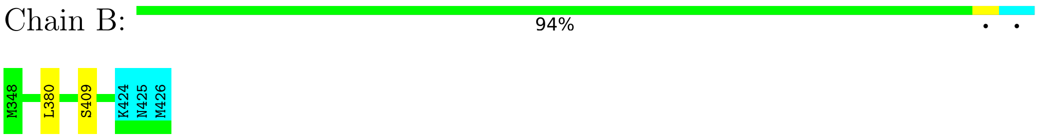


4.2.5 Score per residue for model 5

- Molecule 1: Pogo transposable element with ZNF domain

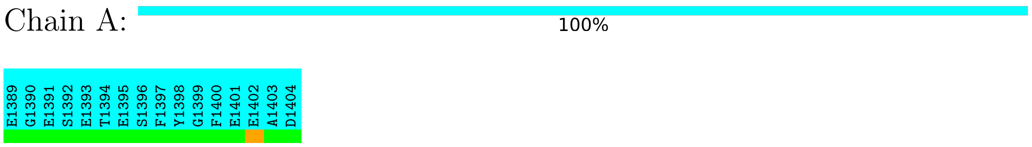


- Molecule 2: PC4 and SFRS1-interacting protein

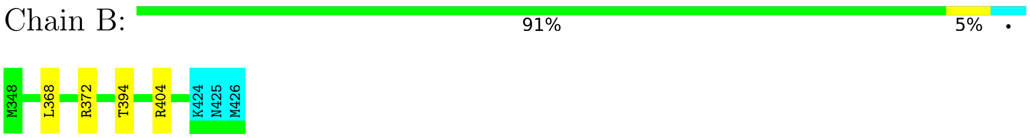


4.2.6 Score per residue for model 6

- Molecule 1: Pogo transposable element with ZNF domain

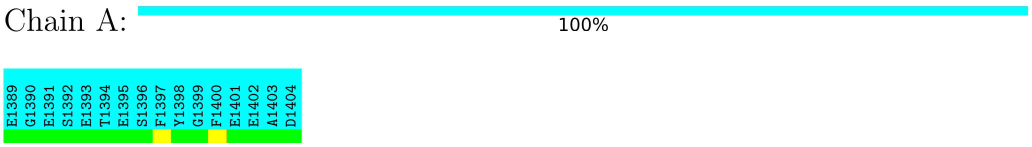


- Molecule 2: PC4 and SFRS1-interacting protein

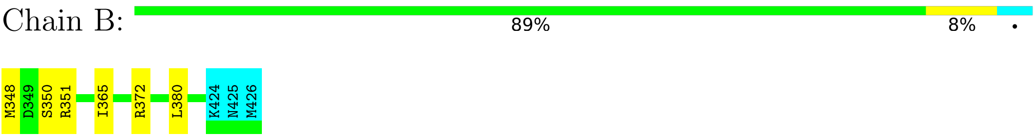


4.2.7 Score per residue for model 7

- Molecule 1: Pogo transposable element with ZNF domain



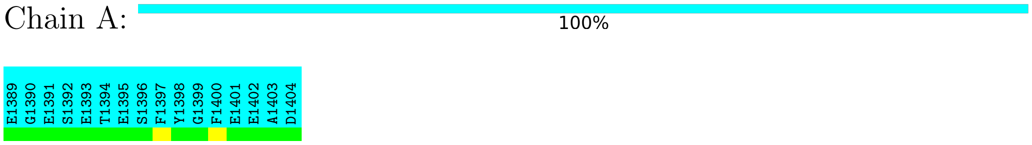
- Molecule 2: PC4 and SFRS1-interacting protein



4.2.8 Score per residue for model 8

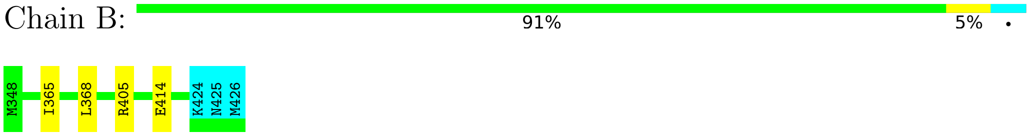
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

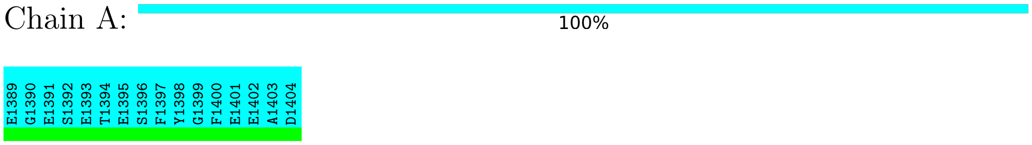
Chain B:



4.2.9 Score per residue for model 9

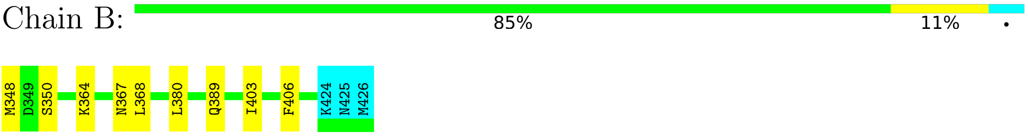
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

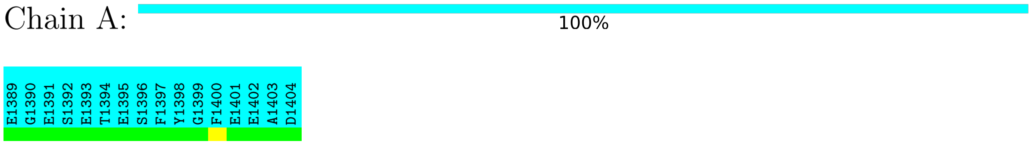
Chain B:



4.2.10 Score per residue for model 10

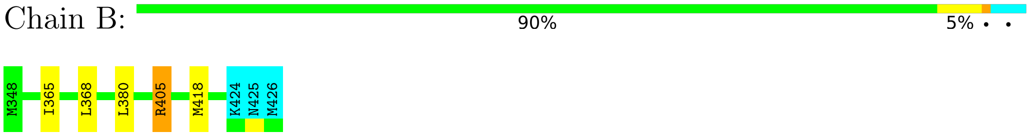
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

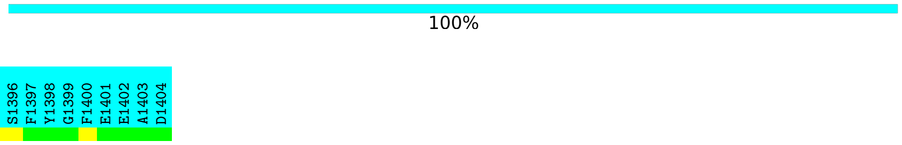
Chain B:



4.2.11 Score per residue for model 11

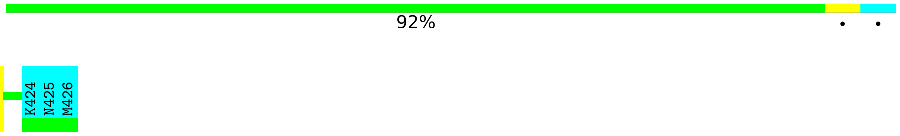
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

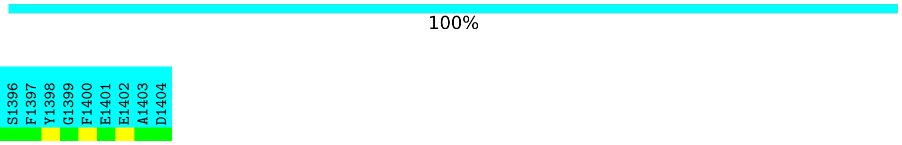
Chain B:



4.2.12 Score per residue for model 12

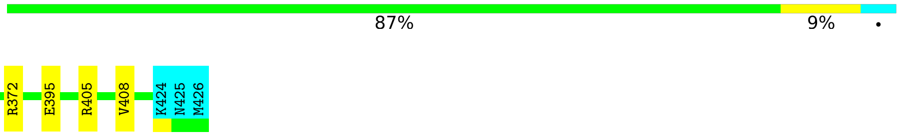
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

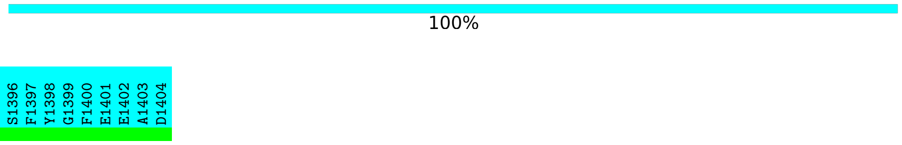
Chain B:



4.2.13 Score per residue for model 13

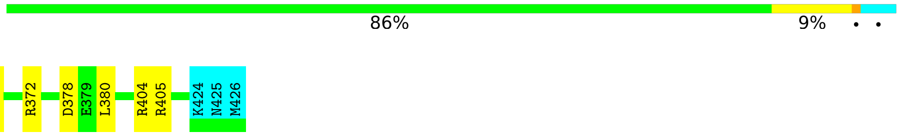
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

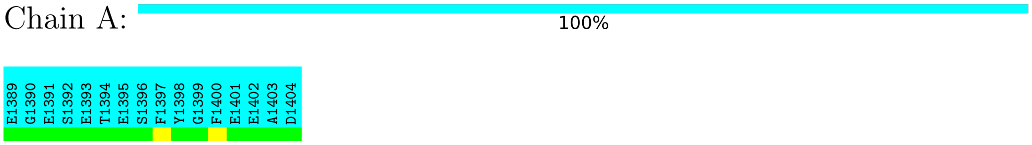
Chain B:



4.2.14 Score per residue for model 14

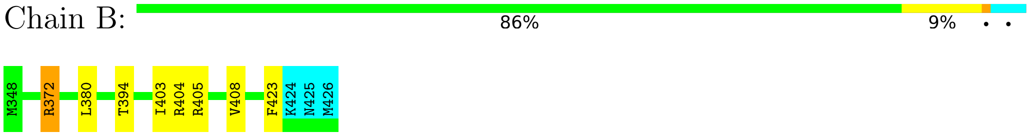
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

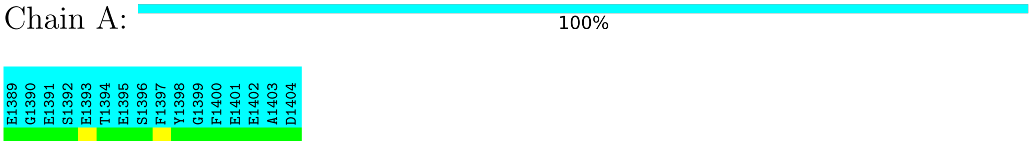
Chain B:



4.2.15 Score per residue for model 15

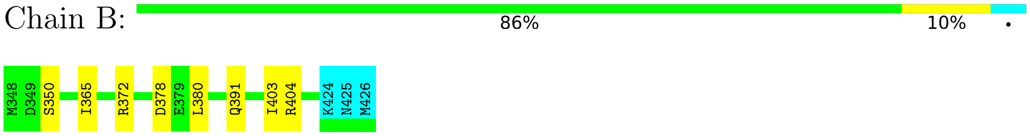
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

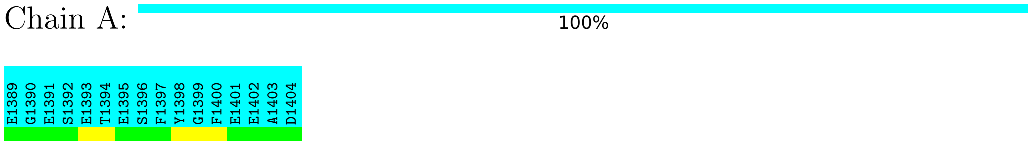
Chain B:



4.2.16 Score per residue for model 16

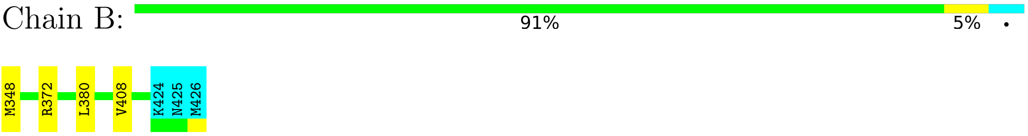
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



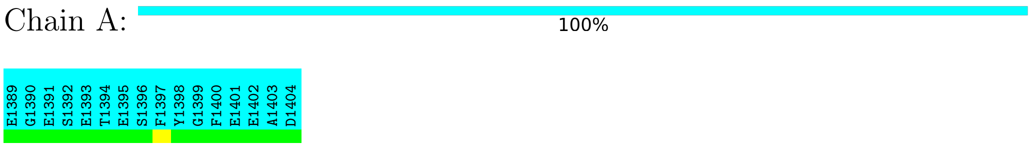
- Molecule 2: PC4 and SFRS1-interacting protein

Chain B:

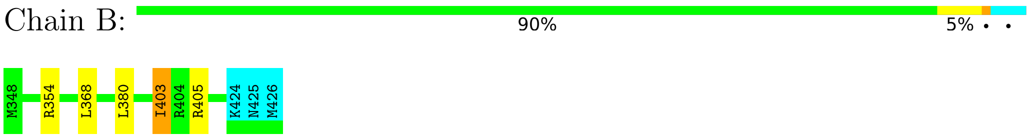


4.2.17 Score per residue for model 17

- Molecule 1: Pogo transposable element with ZNF domain

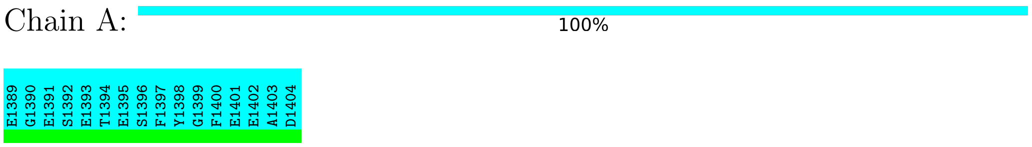


- Molecule 2: PC4 and SFRS1-interacting protein

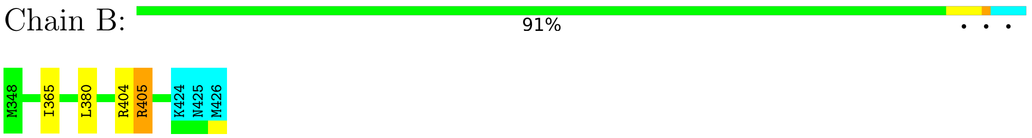


4.2.18 Score per residue for model 18

- Molecule 1: Pogo transposable element with ZNF domain

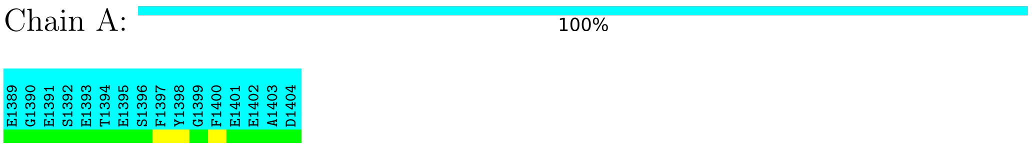


- Molecule 2: PC4 and SFRS1-interacting protein

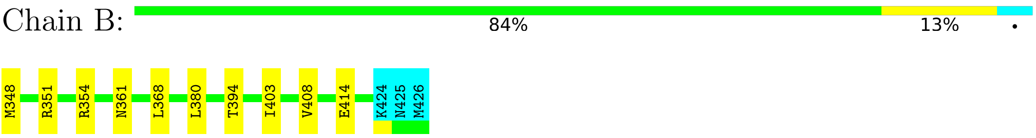


4.2.19 Score per residue for model 19

- Molecule 1: Pogo transposable element with ZNF domain



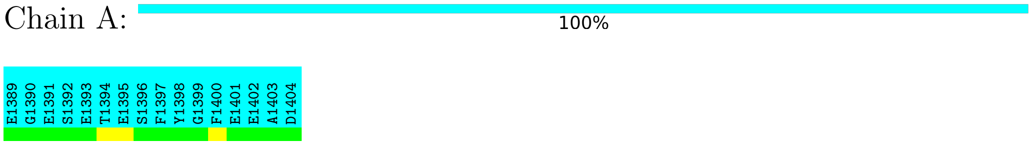
- Molecule 2: PC4 and SFRS1-interacting protein



4.2.20 Score per residue for model 20

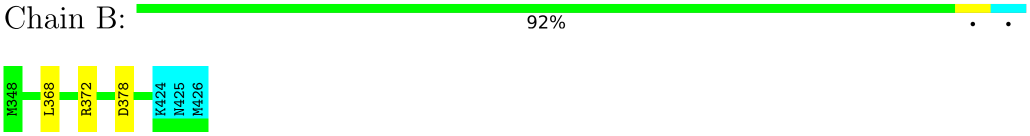
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

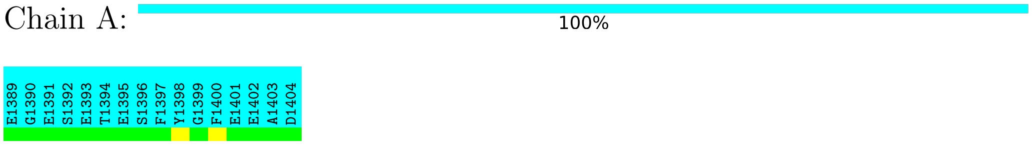
Chain B:



4.2.21 Score per residue for model 21

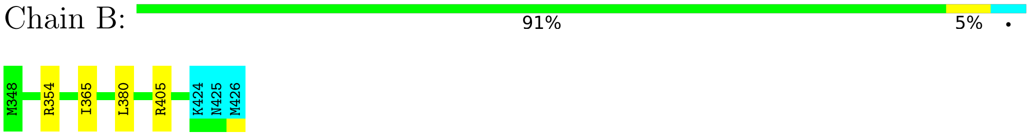
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

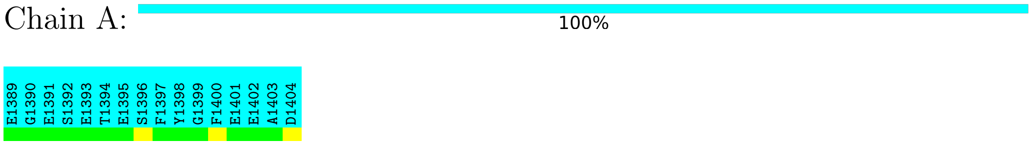
Chain B:



4.2.22 Score per residue for model 22

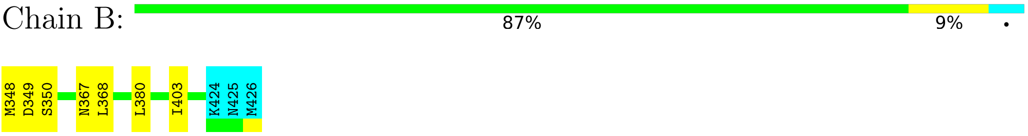
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

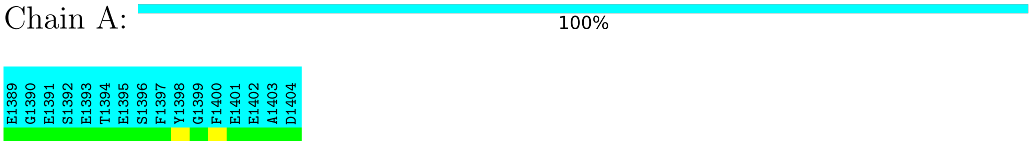
Chain B:



4.2.23 Score per residue for model 23

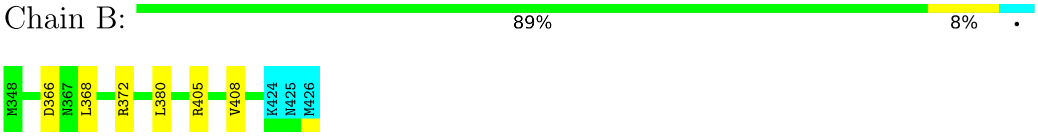
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

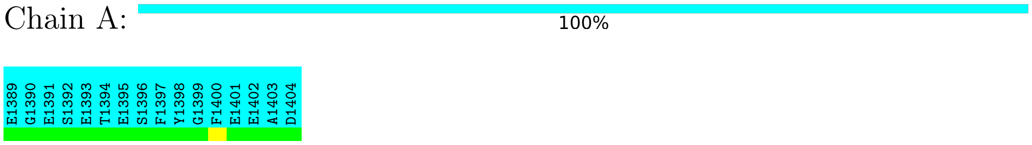
Chain B:



4.2.24 Score per residue for model 24

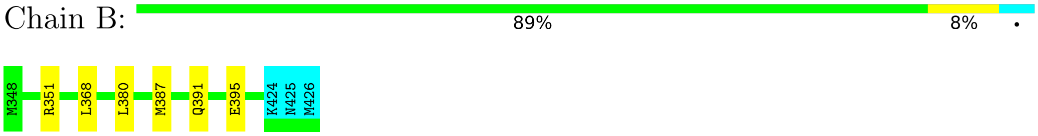
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

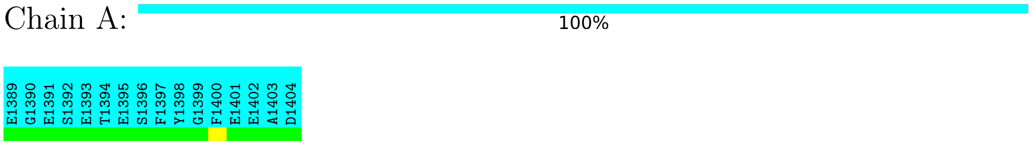
Chain B:



4.2.25 Score per residue for model 25

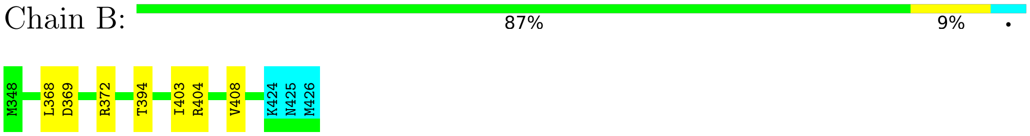
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

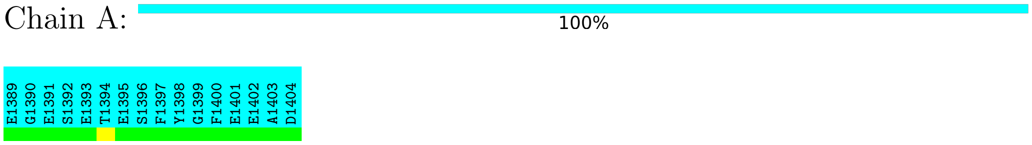
Chain B:



4.2.26 Score per residue for model 26

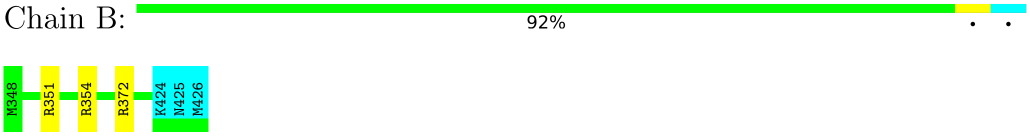
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

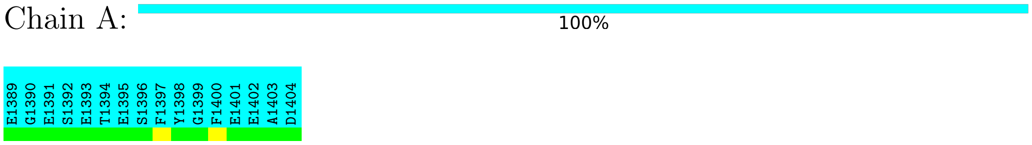
Chain B:



4.2.27 Score per residue for model 27

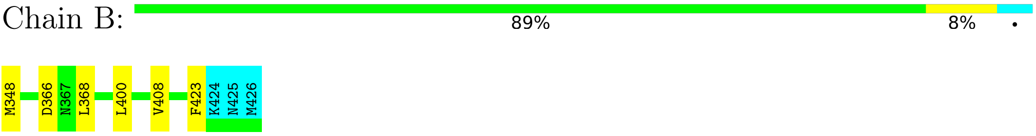
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

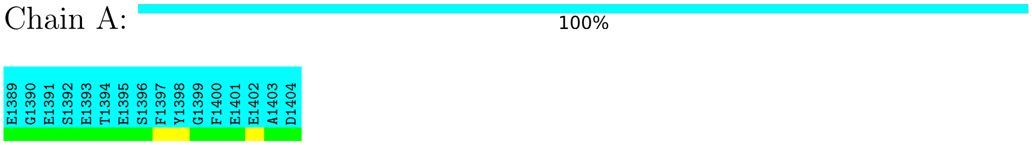
Chain B:



4.2.28 Score per residue for model 28 (medoid)

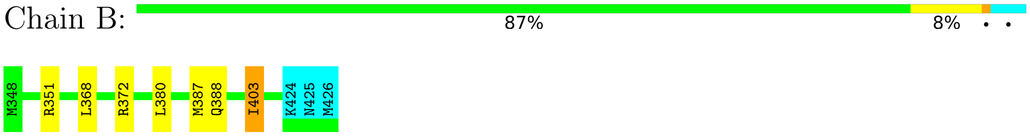
- Molecule 1: Pogo transposable element with ZNF domain

Chain A:



- Molecule 2: PC4 and SFRS1-interacting protein

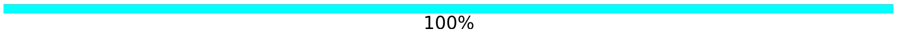
Chain B:

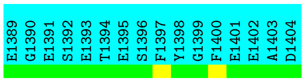


4.2.29 Score per residue for model 29

- Molecule 1: Pogo transposable element with ZNF domain

Chain A:





- Molecule 2: PC4 and SFRS1-interacting protein

Chain B:

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 29 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
CYANA	structure solution	
YASARA	refinement	
CYANA	refinement	

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	1036
Number of shifts mapped to atoms	1036
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	86%

6 Model quality ⓘ

6.1 Standard geometry ⓘ

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
2	B	0.60±0.03	0±0/628 (0.0± 0.0%)	0.70±0.04	1±1/839 (0.2± 0.2%)
All	All	0.60	0/18212 (0.0%)	0.70	37/24331 (0.2%)

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
2	B	405	ARG	NE-CZ-NH1	7.17	123.88	120.30	13	8
2	B	372	ARG	NE-CZ-NH1	6.71	123.66	120.30	13	15
2	B	404	ARG	NE-CZ-NH1	6.54	123.57	120.30	29	2
2	B	354	ARG	NE-CZ-NH1	6.31	123.46	120.30	13	5
2	B	351	ARG	NE-CZ-NH1	5.93	123.27	120.30	26	4
2	B	405	ARG	NE-CZ-NH2	-5.60	117.50	120.30	13	2
2	B	372	ARG	NE-CZ-NH2	-5.14	117.73	120.30	12	1

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	0	0	0	0±0
2	B	623	660	658	0±0
All	All	18067	19140	19082	1

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:417:THR:HA	2:B:420:TYR:CD2	0.47	2.45	29	1

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	0	-	-	-	-	
2	B	75/79 (95%)	72±1 (97±1%)	2±1 (3±1%)	0±0 (0±1%)	50	82
All	All	2175/2755 (79%)	2102 (97%)	68 (3%)	5 (0%)	50	82

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)
2	B	367	ASN	3
2	B	406	PHE	2

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	0	-	-	-	
2	B	72/75 (96%)	67±2 (94±3%)	5±2 (6±3%)	21	70
All	All	2088/2552 (82%)	1956 (94%)	132 (6%)	21	70

All 33 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)
2	B	368	LEU	19
2	B	380	LEU	18
2	B	403	ILE	10
2	B	365	ILE	8
2	B	408	VAL	8
2	B	348	MET	7
2	B	404	ARG	6
2	B	405	ARG	6
2	B	351	ARG	5
2	B	414	GLU	4
2	B	394	THR	4
2	B	350	SER	4
2	B	354	ARG	3
2	B	387	MET	3
2	B	378	ASP	3
2	B	364	LYS	2
2	B	395	GLU	2
2	B	372	ARG	2
2	B	423	PHE	2
2	B	391	GLN	2
2	B	366	ASP	2
2	B	396	MET	1
2	B	384	GLN	1
2	B	409	SER	1
2	B	389	GLN	1
2	B	418	MET	1
2	B	361	ASN	1
2	B	349	ASP	1
2	B	367	ASN	1
2	B	369	ASP	1
2	B	400	LEU	1
2	B	388	GLN	1
2	B	401	LYS	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 (i)

The completeness of assignment taking into account all chemical shift lists is 86% for the well-defined parts and 77% 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 (i)

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	1036
Number of shifts mapped to atoms	1036
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 (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	**# values**	**Correction ± precision, *ppm***	**Suggested action**
$^{13}C_{\alpha}$	79	-0.68 ± 0.19	Should be checked
$^{13}C_{\beta}$	79	0.36 ± 0.15	None needed (< 0.5 ppm)
$^{13}C'$	74	2.46 ± 0.09	Should be applied
^{15}N	75	0.05 ± 0.24	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 86%, i.e. 965 atoms were assigned a chemical shift out of a possible 1120. 0 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	367/380 (97%)	147/152 (97%)	148/152 (97%)	72/76 (95%)
Sidechain	585/697 (84%)	398/451 (88%)	182/213 (85%)	5/33 (15%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	13/43 (30%)	11/22 (50%)	2/19 (11%)	0/2 (0%)
Overall	965/1120 (86%)	556/625 (89%)	332/384 (86%)	77/111 (69%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 77%, i.e. 1034 atoms were assigned a chemical shift out of a possible 1347. 0 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	391/477 (82%)	163/192 (85%)	153/190 (81%)	75/95 (79%)
Sidechain	623/798 (78%)	427/513 (83%)	190/250 (76%)	6/35 (17%)
Aromatic	20/72 (28%)	18/36 (50%)	2/34 (6%)	0/2 (0%)
Overall	1034/1347 (77%)	608/741 (82%)	345/474 (73%)	81/132 (61%)

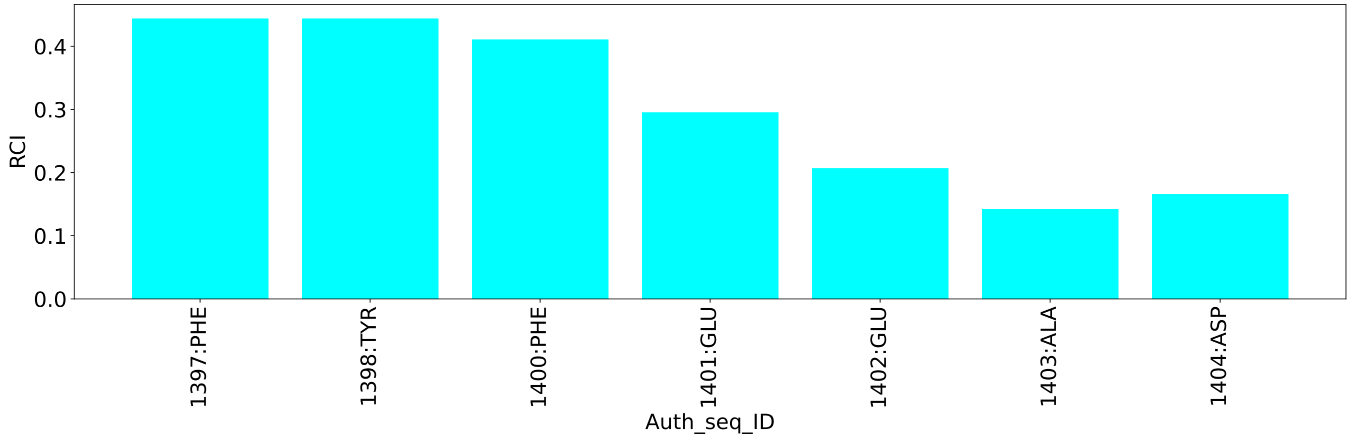
7.1.4 Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

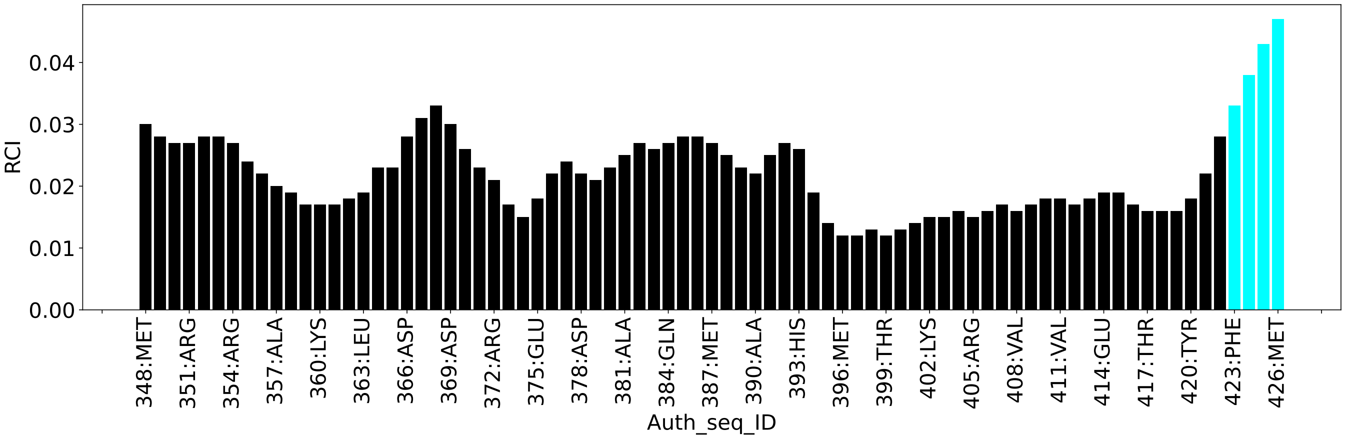
7.1.5 Random Coil Index (RCI) plots ⓘ

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:



Random coil index (RCI) for chain B:


RCI
0.00
0.01
0.02
0.03
0.04
348:MET
351:ARG
354:ARG
357:ALA
360:LYS
363:LEU
366:ASP
369:ASP
372:ARG
375:GLU
378:ASP
381:ALA
384:GLN
387:MET
390:ALA
393:HIS
396:MET
399:THR
402:LYS
405:ARG
408:VAL
411:VAL
414:GLU
417:THR
420:TYR
423:PHE
426:MET
Auth_seq_ID