Full wwPDB NMR Structure Validation Report (i)

Jun 16, 2024 – 12:48 AM EDT

PDB ID	:	2RSY
BMRB ID	:	11508
Title	:	Solution structure of the SH2 domain of Csk in complex with a phosphopeptide from Cbp
Authors	:	Tanaka, H.; Akagi, K.; Oneyama, C.; Tanaka, M.; Sasaki, Y.; Kanou, T.; Lee, Y.; Yokogawa, D.; Debenecker, M.; Nakagawa, A.; Okada, M.; Ikegami, T.
Deposited on	:	2012-09-10

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 (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 (1)) were used in the production of this report:

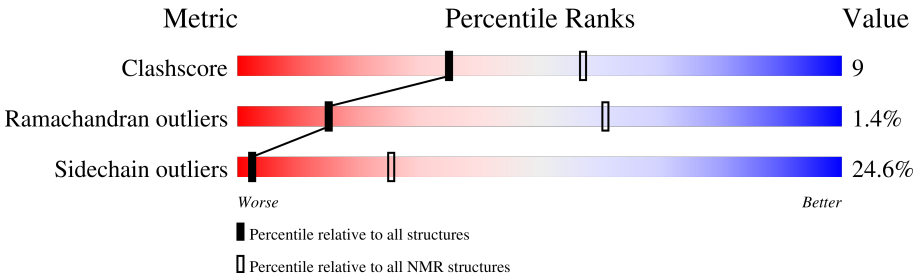
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 87%.

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	99	67% 22% 5% 6%
2	B	38	39% 24% · 34%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *target function*.

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:81-A:173, B:294-B:313, B:315-B:319 (118)	0.33	4

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 4 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2163 atoms, of which 1064 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Tyrosine-protein kinase CSK.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	99	1585	514	782	132	150	7	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	75	GLY	-	expression tag	UNP P32577
A	76	PRO	-	expression tag	UNP P32577
A	77	LEU	-	expression tag	UNP P32577
A	78	GLY	-	expression tag	UNP P32577
A	79	SER	-	expression tag	UNP P32577

- Molecule 2 is a protein called Phosphoprotein associated with glycosphingolipid-enriched microdomains 1.

Mol	Chain	Residues	Atoms							Trace
			Total	C	H	N	O	P	S	
2	B	38	578	180	282	48	66	1	1	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	284	GLY	-	expression tag	UNP Q9JM80
B	285	PRO	-	expression tag	UNP Q9JM80
B	286	LEU	-	expression tag	UNP Q9JM80
B	287	GLY	-	expression tag	UNP Q9JM80

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

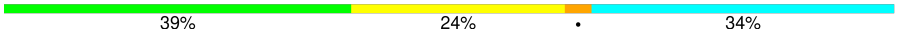
These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for 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.

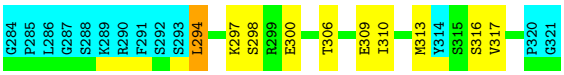
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:

• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

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: Tyrosine-protein kinase CSK

Chain A:

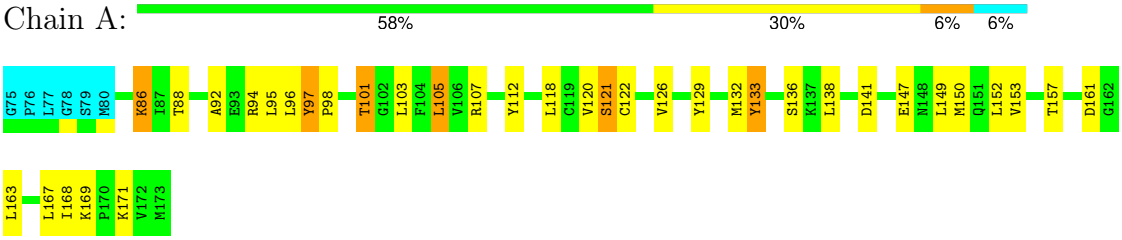
• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

Chain B:

4.2.2 Score per residue for model 2

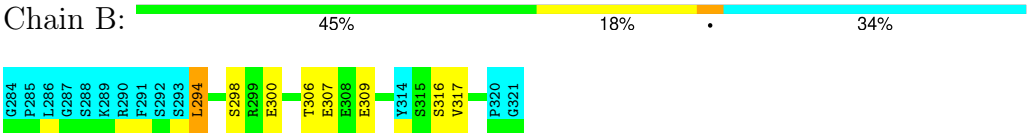
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

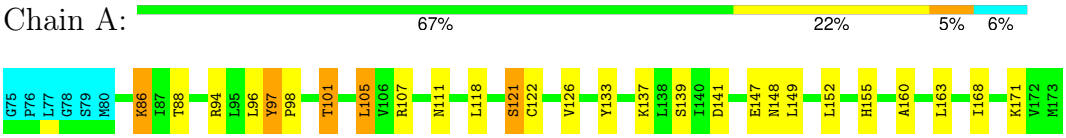
Chain B:



4.2.3 Score per residue for model 3

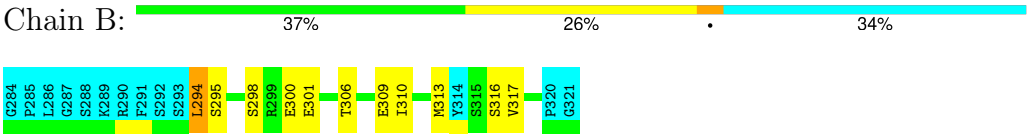
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

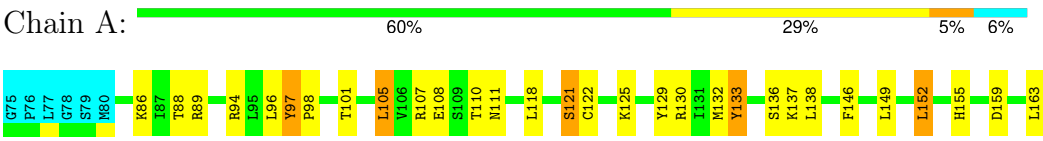
Chain B:



4.2.4 Score per residue for model 4 (medoid)

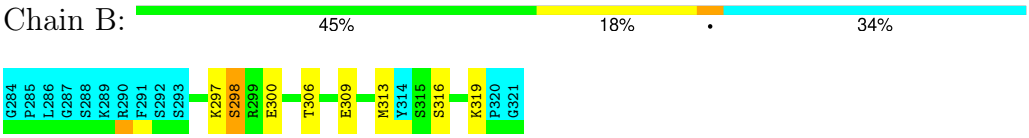
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

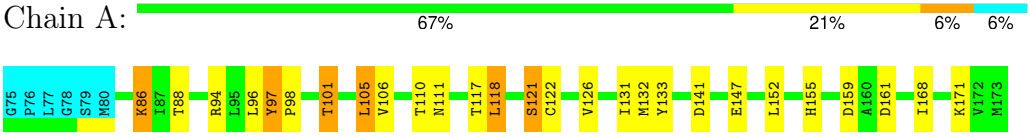
Chain B:



4.2.5 Score per residue for model 5

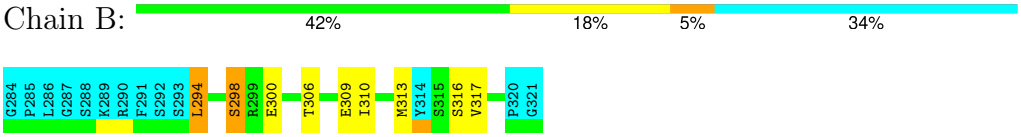
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

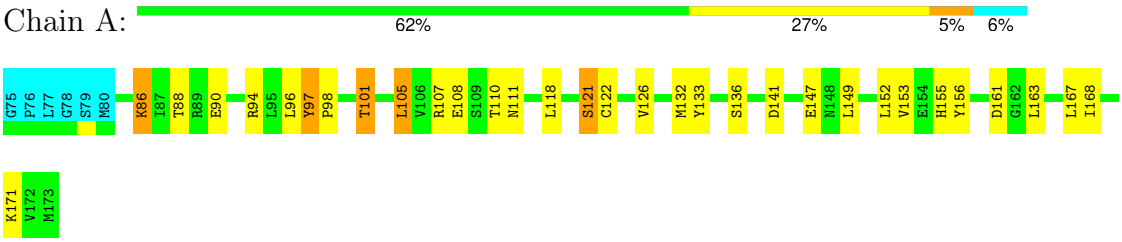
Chain B:



4.2.6 Score per residue for model 6

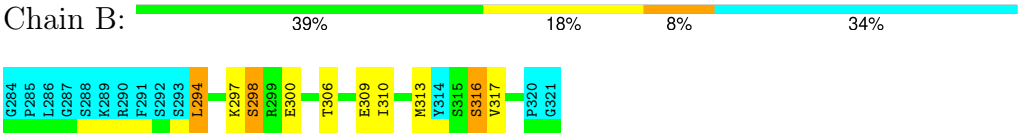
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

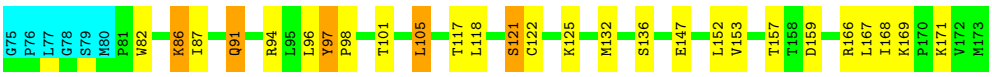
Chain B:



4.2.7 Score per residue for model 7

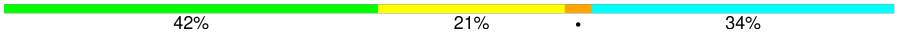
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

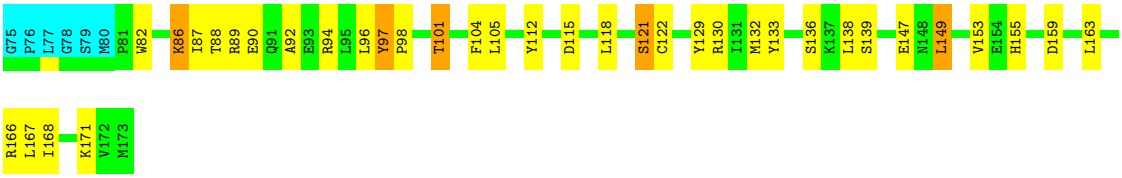
Chain B:



4.2.8 Score per residue for model 8

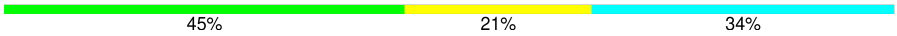
• Molecule 1: Tyrosine-protein kinase CSK

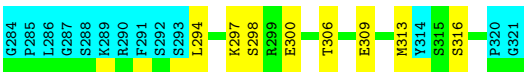
Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

Chain B:

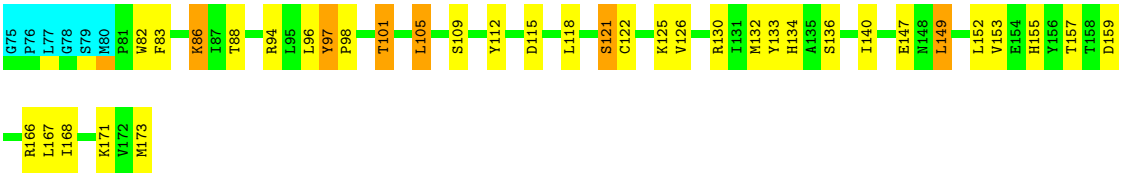




4.2.9 Score per residue for model 9

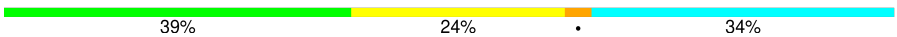
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

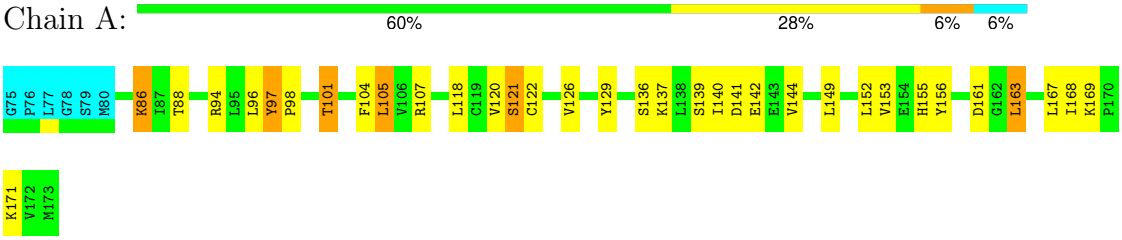
Chain B:



4.2.10 Score per residue for model 10

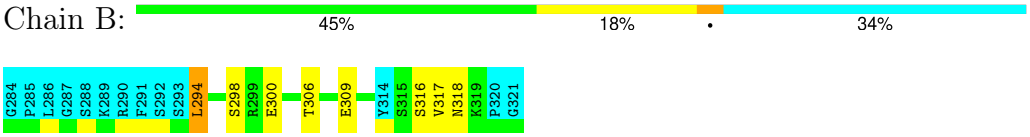
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

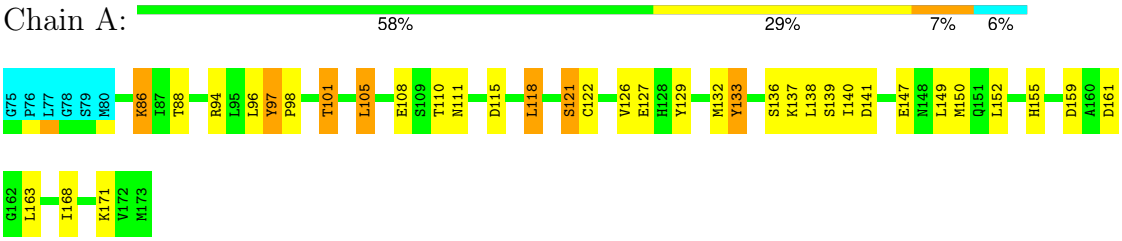
Chain B:



4.2.11 Score per residue for model 11

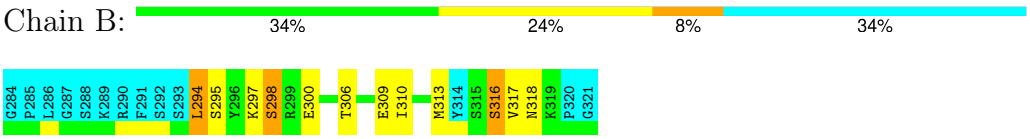
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

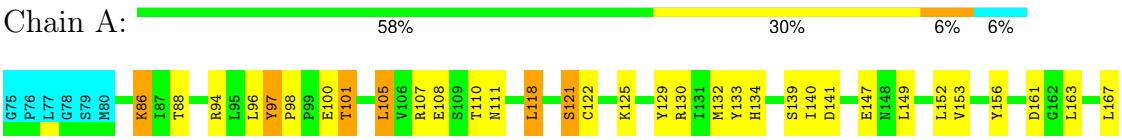
Chain B:

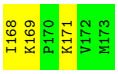


4.2.12 Score per residue for model 12

• Molecule 1: Tyrosine-protein kinase CSK

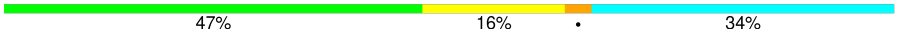
Chain A:

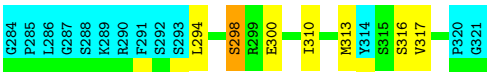




• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

Chain B:





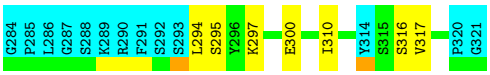
4.2.13 Score per residue for model 13

• Molecule 1: Tyrosine-protein kinase CSK

Chain A:

• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

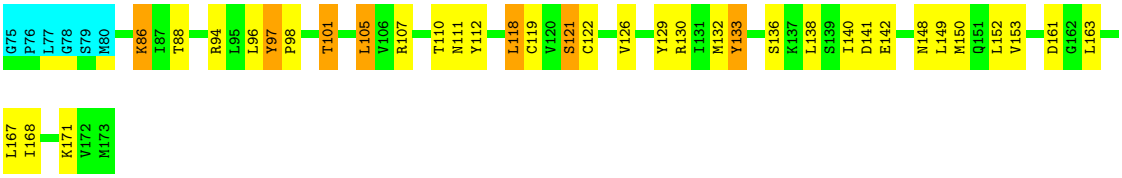
Chain B:



4.2.14 Score per residue for model 14

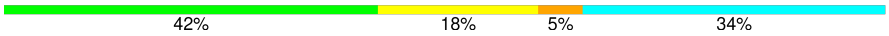
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

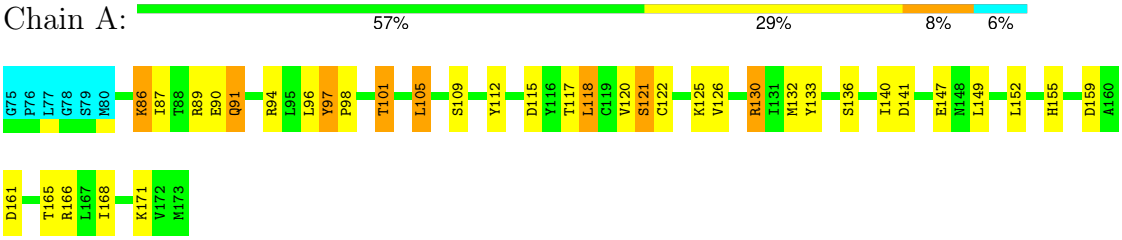
Chain B:



4.2.15 Score per residue for model 15

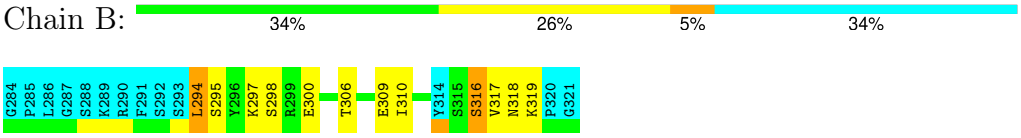
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

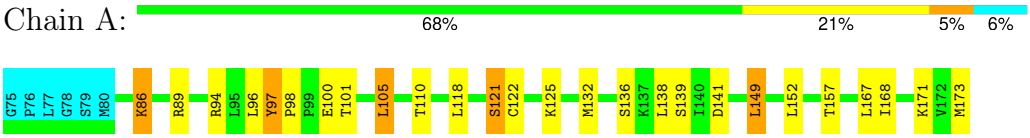
Chain B:



4.2.16 Score per residue for model 16

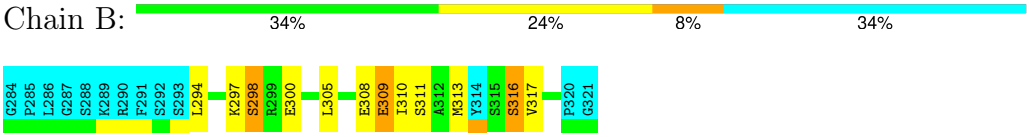
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

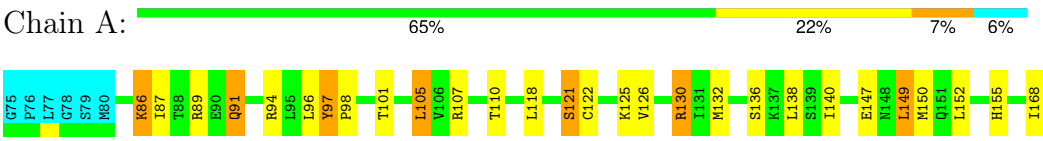
Chain B:



4.2.17 Score per residue for model 17

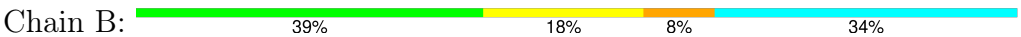
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

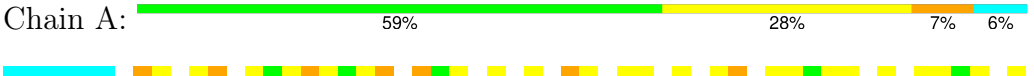
Chain B:

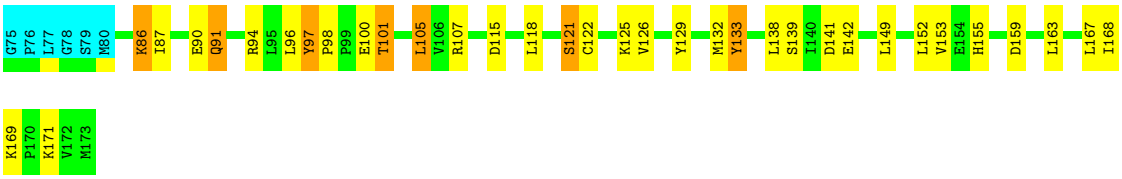


4.2.18 Score per residue for model 18

• Molecule 1: Tyrosine-protein kinase CSK

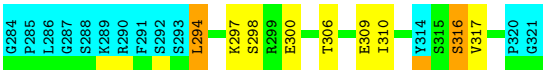
Chain A:





• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

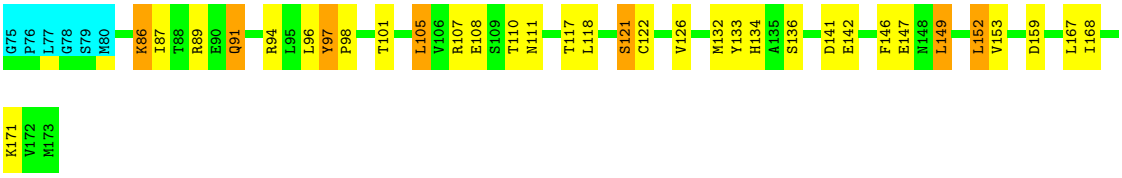
Chain B:



4.2.19 Score per residue for model 19

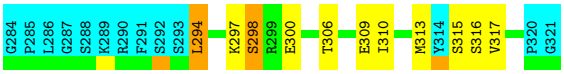
• Molecule 1: Tyrosine-protein kinase CSK

Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

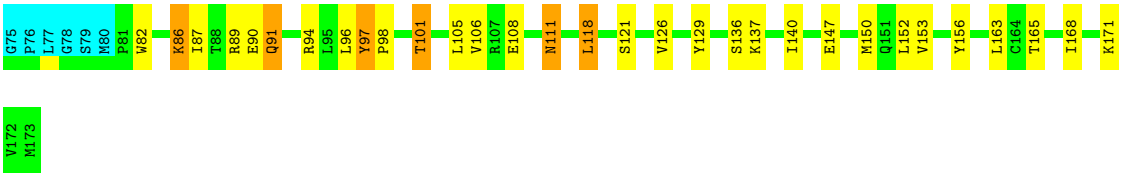
Chain B:



4.2.20 Score per residue for model 20

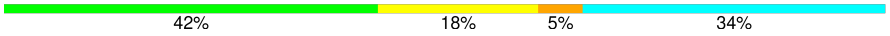
• Molecule 1: Tyrosine-protein kinase CSK

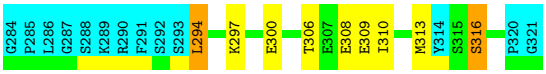
Chain A:



• Molecule 2: Phosphoprotein associated with glycosphingolipid-enriched microdomains 1

Chain B:





5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function.*

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

Software name	Classification	Version
CYANA	structure solution	
MOLMOL	geometry optimization	
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	1605
Number of shifts mapped to atoms	1605
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	87%

6 Model quality

6.1 Standard geometry

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

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	766	744	744	15±3
2	B	197	190	190	4±2
All	All	19260	18680	18680	330

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 9.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:106:VAL:HG22	1:A:118:LEU:HD22	0.66	1.66	5	1
1:A:118:LEU:HD23	1:A:131:ILE:HD11	0.65	1.68	5	1
1:A:162:GLY:HA3	2:B:317:VAL:HG23	0.62	1.69	13	1
1:A:156:TYR:CE1	1:A:163:LEU:HD23	0.62	2.30	10	1
1:A:153:VAL:HG13	1:A:167:LEU:HD12	0.62	1.72	12	7
1:A:141:ASP:HB3	2:B:317:VAL:HG21	0.61	1.73	10	9
1:A:96:LEU:HD21	1:A:105:LEU:HD23	0.61	1.73	2	19
1:A:141:ASP:CB	2:B:317:VAL:HG21	0.60	2.26	18	8
1:A:96:LEU:HD12	1:A:121:SER:CB	0.59	2.28	12	19
1:A:101:THR:O	1:A:168:ILE:HD12	0.59	1.98	12	20
1:A:129:TYR:CZ	1:A:163:LEU:HD13	0.57	2.34	2	5

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:156:TYR:CD1	1:A:163:LEU:HD13	0.57	2.34	13	3
1:A:126:VAL:O	2:B:294:LEU:HD13	0.56	2.01	9	11
1:A:129:TYR:CE1	1:A:163:LEU:HD13	0.55	2.36	12	7
2:B:305:LEU:HD13	2:B:309:GLU:OE2	0.55	2.02	17	2
1:A:157:THR:OG1	1:A:167:LEU:HD12	0.54	2.01	7	3
1:A:130:ARG:O	1:A:140:ILE:HG23	0.54	2.02	17	3
1:A:111:ASN:CG	2:B:310:ILE:HD12	0.54	2.24	1	8
1:A:118:LEU:CD2	1:A:131:ILE:HD11	0.54	2.32	5	1
1:A:96:LEU:HD12	1:A:121:SER:HB2	0.54	1.80	14	8
1:A:116:TYR:CE2	1:A:149:LEU:HD21	0.53	2.39	1	1
1:A:96:LEU:HD12	1:A:121:SER:OG	0.52	2.04	10	17
1:A:96:LEU:CD1	1:A:105:LEU:HD23	0.51	2.34	20	1
1:A:96:LEU:HD13	1:A:105:LEU:HD23	0.51	1.81	20	1
1:A:140:ILE:CG2	2:B:317:VAL:HG13	0.50	2.37	10	4
1:A:101:THR:HG22	1:A:165:THR:CB	0.50	2.37	15	1
1:A:140:ILE:HG22	2:B:317:VAL:HG13	0.49	1.83	14	4
1:A:156:TYR:CE1	1:A:163:LEU:HD13	0.49	2.41	13	2
1:A:153:VAL:HG13	1:A:167:LEU:CD1	0.49	2.38	18	7
1:A:110:THR:HG21	2:B:298:SER:HB2	0.49	1.85	19	9
1:A:118:LEU:CD2	1:A:163:LEU:HD11	0.48	2.38	11	3
1:A:104:PHE:HB3	1:A:167:LEU:HD22	0.48	1.85	13	2
1:A:96:LEU:HD12	1:A:121:SER:HB3	0.47	1.85	19	1
1:A:126:VAL:O	2:B:294:LEU:HD22	0.47	2.10	15	3
1:A:101:THR:HG22	1:A:165:THR:HB	0.47	1.85	15	2
1:A:82:TRP:CH2	1:A:149:LEU:HD12	0.46	2.45	8	2
2:B:310:ILE:HD13	2:B:313:MET:HE2	0.46	1.87	11	2
1:A:119:CYS:SG	1:A:126:VAL:HG13	0.46	2.50	14	1
1:A:97:TYR:CB	1:A:98:PRO:CD	0.46	2.94	1	20
1:A:96:LEU:CD2	1:A:105:LEU:HD23	0.46	2.39	2	17
1:A:87:ILE:HG23	1:A:91:GLN:HB3	0.46	1.87	17	8
1:A:146:PHE:CD1	1:A:152:LEU:HD12	0.46	2.45	19	3
1:A:104:PHE:CG	1:A:167:LEU:HD13	0.46	2.46	13	1
1:A:149:LEU:HD13	1:A:152:LEU:HD22	0.46	1.86	19	1
2:B:306:THR:HG22	2:B:309:GLU:HG2	0.45	1.88	5	15
1:A:138:LEU:HD22	1:A:149:LEU:HD22	0.45	1.88	8	1
1:A:156:TYR:CD1	1:A:163:LEU:HD23	0.45	2.46	12	1
1:A:87:ILE:CG2	1:A:92:ALA:HB2	0.45	2.41	8	1
1:A:111:ASN:HD22	2:B:310:ILE:HD12	0.45	1.72	20	1
1:A:97:TYR:N	1:A:98:PRO:HD2	0.45	2.26	2	20
1:A:104:PHE:HB2	1:A:118:LEU:HD11	0.45	1.89	10	1
1:A:160:ALA:HB1	1:A:163:LEU:O	0.44	2.11	3	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:141:ASP:HA	2:B:317:VAL:HG21	0.44	1.89	1	4
1:A:141:ASP:O	1:A:142:GLU:CG	0.44	2.66	19	4
1:A:104:PHE:CD1	1:A:167:LEU:HD13	0.44	2.46	13	1
1:A:118:LEU:HD21	1:A:120:VAL:CG2	0.44	2.43	15	1
1:A:127:GLU:OE2	2:B:294:LEU:HD11	0.44	2.12	11	1
1:A:92:ALA:HA	1:A:95:LEU:HD12	0.43	1.90	2	1
1:A:126:VAL:HG13	1:A:126:VAL:O	0.43	2.13	2	1
1:A:82:TRP:CH2	1:A:153:VAL:HG21	0.43	2.48	9	4
1:A:141:ASP:OD1	1:A:144:VAL:HG12	0.43	2.14	10	1
1:A:103:LEU:HD12	1:A:169:LYS:O	0.43	2.13	2	1
1:A:138:LEU:HD21	1:A:149:LEU:HG	0.43	1.89	16	3
1:A:120:VAL:HG21	1:A:163:LEU:HD12	0.43	1.89	2	1
1:A:133:TYR:CZ	1:A:138:LEU:HD11	0.42	2.50	2	4
2:B:310:ILE:HD13	2:B:313:MET:CE	0.42	2.45	6	2
1:A:120:VAL:HG21	1:A:163:LEU:CD1	0.42	2.44	10	1
1:A:106:VAL:HG22	1:A:118:LEU:CD2	0.42	2.41	5	1
1:A:157:THR:HG23	1:A:167:LEU:O	0.42	2.15	2	1
1:A:87:ILE:HG21	1:A:92:ALA:HB2	0.41	1.91	8	1
1:A:133:TYR:CE2	1:A:138:LEU:HD21	0.41	2.49	4	1
2:B:310:ILE:HA	2:B:313:MET:HE2	0.41	1.91	16	1
1:A:106:VAL:CG2	1:A:118:LEU:HD12	0.41	2.46	20	1
1:A:129:TYR:HB3	1:A:140:ILE:HD13	0.41	1.92	20	1
1:A:141:ASP:CA	2:B:317:VAL:HG21	0.41	2.46	14	1
1:A:138:LEU:HD13	1:A:148:ASN:C	0.40	2.37	14	1

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	92/99 (93%)	86±1 (93±1%)	5±1 (6±1%)	1±0 (1±0%)	18 66
2	B	25/38 (66%)	22±1 (87±3%)	3±1 (10±4%)	1±0 (2±2%)	9 46
All	All	2340/2740 (85%)	2148 (92%)	160 (7%)	32 (1%)	15 61

All 2 unique Ramachandran outliers are listed below. They are sorted by the frequency of occur-

rence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	86	LYS	20
2	B	316	SER	12

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	85/89 (96%)	64±3 (76±3%)	21±3 (24±3%)	2	26
2	B	24/33 (73%)	18±1 (75±6%)	6±1 (25±6%)	2	24
All	All	2180/2440 (89%)	1644 (75%)	536 (25%)	2	25

All 55 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	94	ARG	20
1	A	97	TYR	20
1	A	121	SER	20
1	A	171	LYS	20
2	B	300	GLU	20
1	A	86	LYS	19
1	A	118	LEU	19
1	A	122	CYS	19
1	A	152	LEU	19
2	B	294	LEU	19
1	A	105	LEU	18
2	B	298	SER	18
1	A	149	LEU	17
1	A	132	MET	16
1	A	133	TYR	15
1	A	136	SER	15
2	B	316	SER	15
1	A	147	GLU	14
1	A	101	THR	14
1	A	155	HIS	13
2	B	297	LYS	13

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	88	THR	11
2	B	313	MET	10
1	A	107	ARG	10
1	A	159	ASP	10
1	A	161	ASP	9
1	A	91	GLN	8
1	A	139	SER	8
1	A	125	LYS	8
1	A	130	ARG	7
1	A	89	ARG	7
1	A	108	GLU	7
1	A	112	TYR	6
2	B	295	SER	6
1	A	90	GLU	6
1	A	150	MET	5
1	A	137	LYS	5
1	A	166	ARG	5
1	A	115	ASP	5
1	A	100	GLU	4
1	A	169	LYS	4
2	B	308	GLU	4
2	B	315	SER	4
1	A	173	MET	3
2	B	319	LYS	3
1	A	134	HIS	3
2	B	318	ASN	3
2	B	301	GLU	2
1	A	111	ASN	2
1	A	109	SER	2
2	B	309	GLU	2
2	B	307	GLU	1
1	A	148	ASN	1
1	A	163	LEU	1
2	B	311	SER	1

6.3.3 RNA

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	PTR	B	314	2	15,16,17	0.90±0.01	1±0 (6±2%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	PTR	B	314	2	17,22,24	0.84±0.00	1±0 (5±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	PTR	B	314	2	-	0±0,10,11,13	0±0,1,1,1

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
2	B	314	PTR	P-OH	2.09	1.63	1.59	7	18

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	B	314	PTR	P-OH-CZ	2.34	115.57	123.88	18	20

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 87% for the well-defined parts and 87% 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	1605
Number of shifts mapped to atoms	1605
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	133	-0.27 ± 0.06	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	125	0.03 ± 0.07	None needed (< 0.5 ppm)
$^{13}C'$	124	-0.02 ± 0.12	None needed (< 0.5 ppm)
^{15}N	125	-0.03 ± 0.46	None needed (< 0.5 ppm)

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 87%, i.e. 1419 atoms were assigned a chemical shift out of a possible 1629. 0 out of 19 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	573/583 (98%)	234/235 (100%)	227/236 (96%)	112/112 (100%)
Sidechain	775/904 (86%)	515/586 (88%)	251/288 (87%)	9/30 (30%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	71/142 (50%)	36/69 (52%)	34/68 (50%)	1/5 (20%)
Overall	1419/1629 (87%)	785/890 (88%)	512/592 (86%)	122/147 (83%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 87%, i.e. 1593 atoms were assigned a chemical shift out of a possible 1837. 0 out of 21 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	647/672 (96%)	265/273 (97%)	257/272 (94%)	125/127 (98%)
Sidechain	869/1013 (86%)	577/658 (88%)	282/321 (88%)	10/34 (29%)
Aromatic	77/152 (51%)	39/74 (53%)	37/73 (51%)	1/5 (20%)
Overall	1593/1837 (87%)	881/1005 (88%)	576/666 (86%)	136/166 (82%)

7.1.4 Statistically unusual chemical shifts (i)

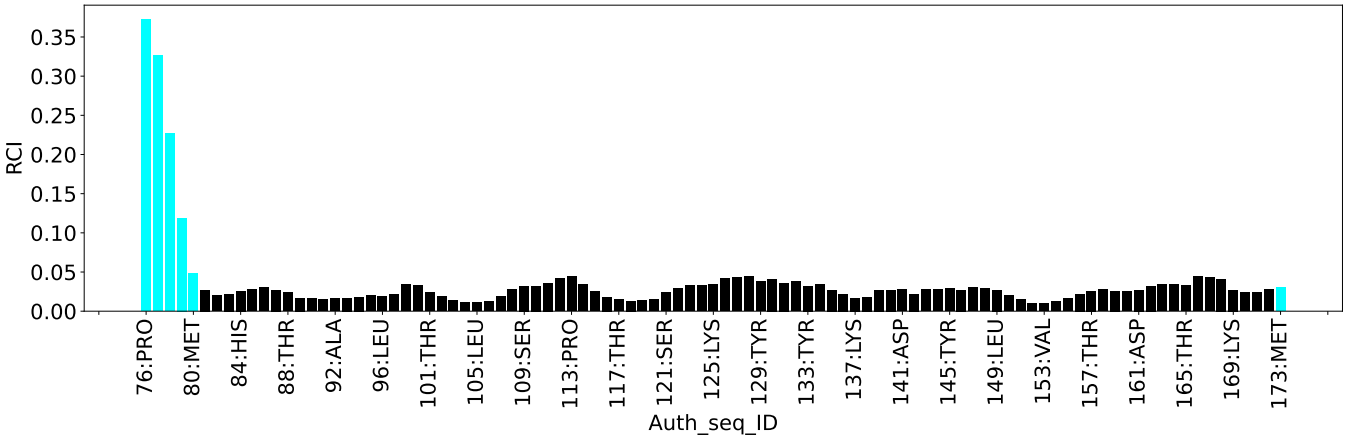
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	B	315	SER	HB2	2.39	2.61 – 5.13	-5.9
1	A	80	MET	HB2	0.21	0.42 – 3.63	-5.7
1	A	152	LEU	HA	1.98	2.04 – 6.55	-5.1
1	A	85	GLY	H	5.20	5.23 – 11.42	-5.0

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: