WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

Nov 30, 2020 – 06:06 PM EST

PDB ID	:	6UZ5
Title	:	Solution structure of KTI55-Kringle 2 complex
Authors	:	Qiu, C.; Yuan, Y.; Castellino, F.J.
Deposited on	:	2019-11-14

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

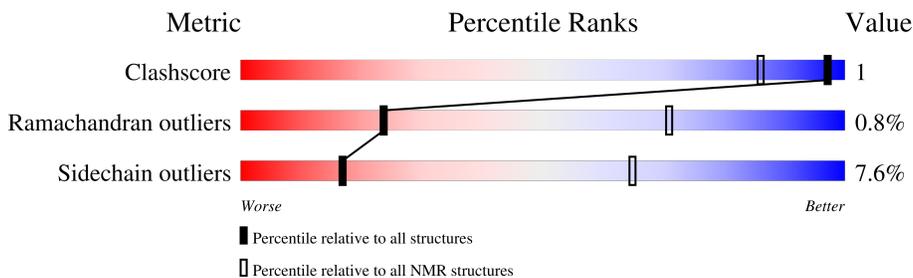|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.14.6
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.14.6

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 30%.

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	Percentile Ranks	Value
Clashscore		1
Ramachandran outliers		0.8%
Sidechain outliers		7.6%

Worse   *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all NMR structures

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	87	85% 7% 8%
2	B	57	95% 5%

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:-2-A:-1, A:166-A:243, B:-2-B:-1, B:85-B:139 (137)	0.93	1

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

NmrClust was unable to cluster the ensemble.

Error message: 1 Unexpected character

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2198 atoms, of which 1067 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Plasminogen.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	80	1255	405	600	119	123	8	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-7	TYR	-	expression tag	UNP P00747
A	-6	VAL	-	expression tag	UNP P00747
A	-5	GLU	-	expression tag	UNP P00747
A	-4	PHE	-	expression tag	UNP P00747
A	-3	SER	-	expression tag	UNP P00747
A	-2	GLU	-	expression tag	UNP P00747
A	-1	GLU	-	expression tag	UNP P00747
A	169	GLY	CYS	engineered mutation	UNP P00747
A	221	ASP	GLU	engineered mutation	UNP P00747
A	237	TYR	LEU	engineered mutation	UNP P00747
A	244	ALA	-	expression tag	UNP P00747
A	245	ALA	-	expression tag	UNP P00747

- Molecule 2 is a protein called M protein.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	B	57	943	287	467	89	100	0

There are 3 discrepancies between the modelled and reference sequences:

| Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	GLY	-	expression tag	UNP M4I070
B	-1	SER	-	expression tag	UNP M4I070
B	105	ARG	GLN	variant	UNP M4I070

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: Plasminogen

Chain A:

- Molecule 2: M protein

Chain B:
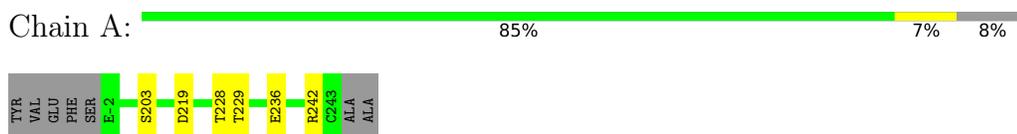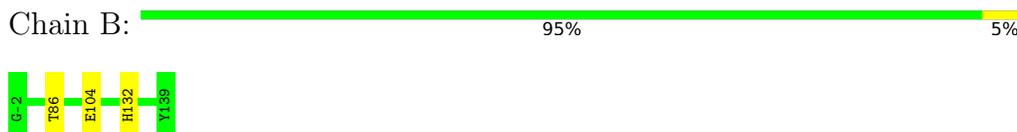
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

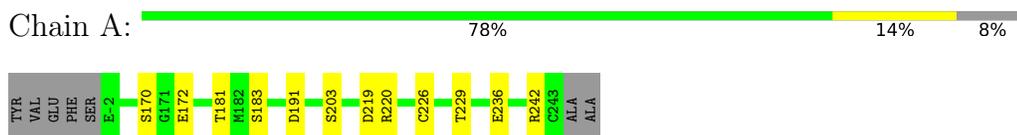### 4.2.1 Score per residue for model 1 (medoid)

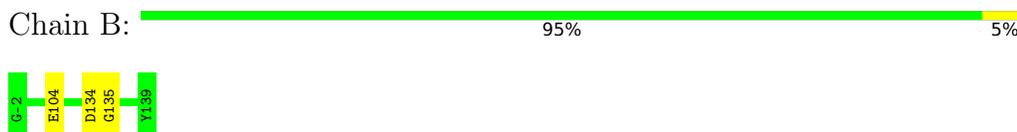- Molecule 1: Plasminogen

Chain A:

- Molecule 2: M protein

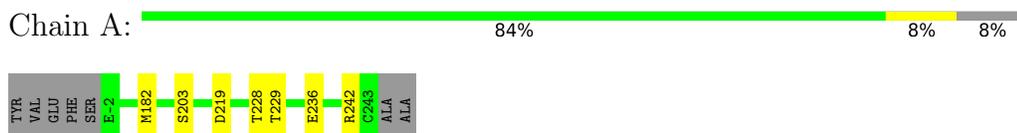Chain B:
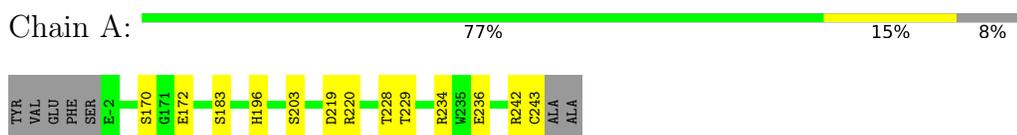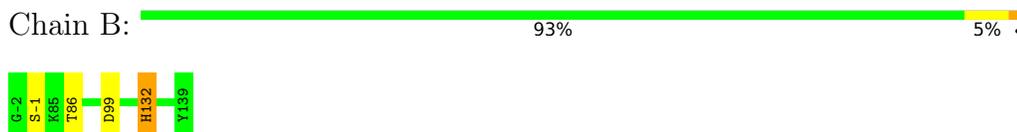
4.2.2 Score per residue for model 2

• Molecule 1: Plasminogen

Chain A:

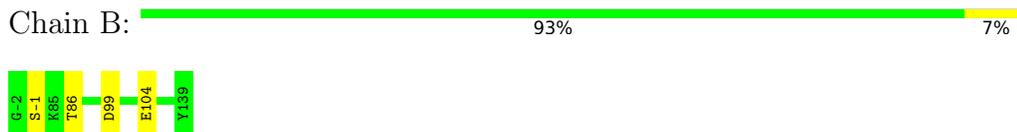• Molecule 2: M protein

Chain B:
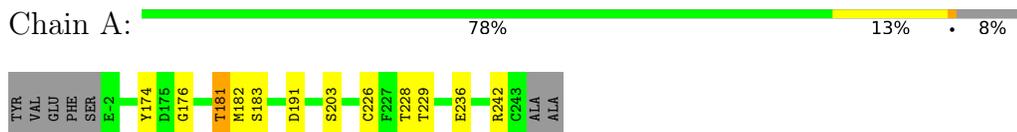

4.2.3 Score per residue for model 3

• Molecule 1: Plasminogen

Chain A:

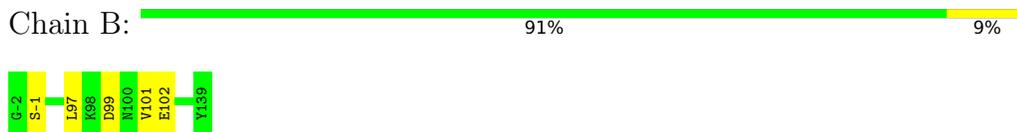• Molecule 2: M protein

Chain B:

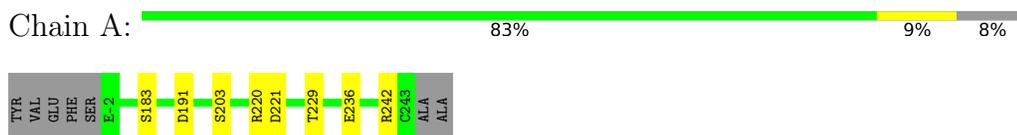### 4.2.4 Score per residue for model 4

• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.5 Score per residue for model 5

• Molecule 1: Plasminogen

Chain A:

83%   9%   8%

TYR VAL GLU PHE SER E-2 S183 D191 S203 R220 D221 T229 E236 R242 C243 ALA ALA

• Molecule 2: M protein

Chain B:

96%   .

G-2 E104 H132 Y139
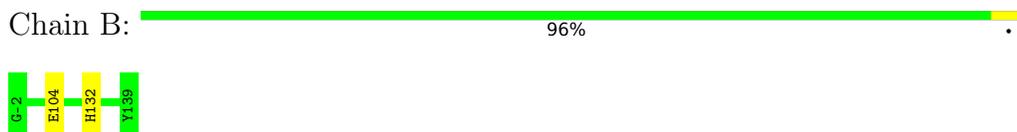
4.2.6 Score per residue for model 6

• Molecule 1: Plasminogen

Chain A:

85%   7%   8%

TYR VAL GLU PHE SER E-2 G169 S179 P218 T228 T229 R242 C243 ALA ALA

• Molecule 2: M protein

Chain B:

95%   5%

G-2 T86 E104 L133 Y139
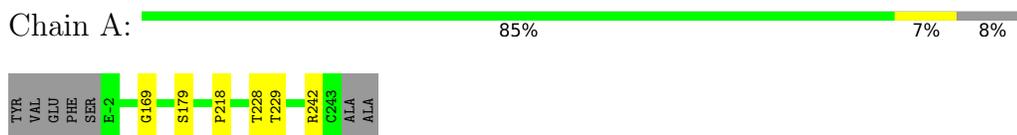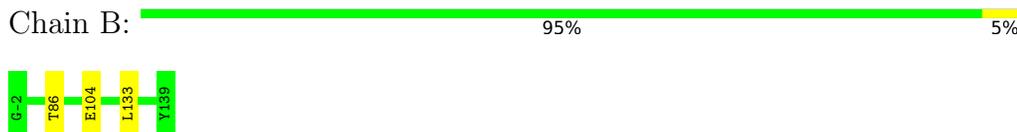
4.2.7 Score per residue for model 7

• Molecule 1: Plasminogen

Chain A:

80%   11%   8%

TYR VAL GLU PHE SER E-2 G169 S170 T181 M182 S183 S203 D219 T228 T229 E236 R242 C243 ALA ALA

• Molecule 2: M protein

Chain B:

95%   5%

G-2 D99 E104 R138 Y139
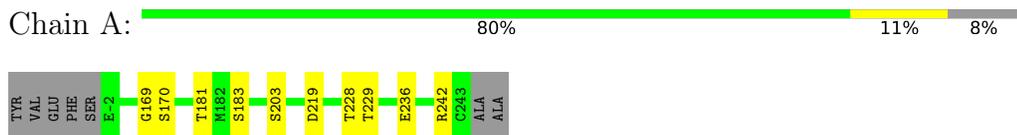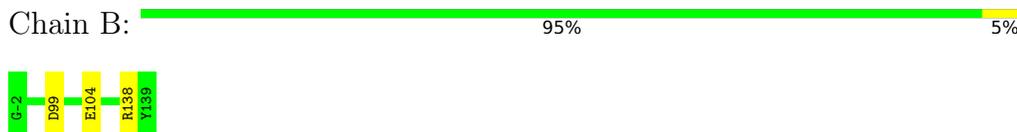
4.2.8 Score per residue for model 8

• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.9 Score per residue for model 9

• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.10 Score per residue for model 10

• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.11 Score per residue for model 11

• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.12 Score per residue for model 12

• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.13 Score per residue for model 13

• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.14　Score per residue for model 14

• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.15　Score per residue for model 15

• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.16　Score per residue for model 16

• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.17 Score per residue for model 17

• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.18 Score per residue for model 18
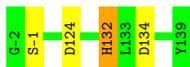
• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.19 Score per residue for model 19
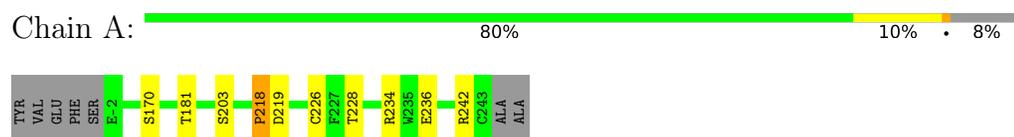
• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:

4.2.20 Score per residue for model 20

• Molecule 1: Plasminogen

Chain A:

• Molecule 2: M protein

Chain B:
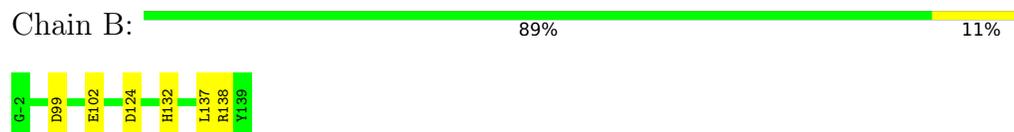
5 Refinement protocol and experimental data overview ⓘ

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

Of the 2000 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy.*

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

Software name	Classification	Version
X-PLOR NIH	structure calculation	
X-PLOR NIH	refinement	
HADDOCK	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	578
Number of shifts mapped to atoms	578
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	30%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality ⓘ

6.1 Standard geometry ⓘ

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in each chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	655	600	602	1±1
2	B	476	467	466	0±1
All	All	22620	21340	21360	27

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
2:B:113:HIS:HB2	2:B:115:GLU:OE1	0.62	1.94	8	1
1:A:181:THR:HG23	1:A:226:CYS:SG	0.53	2.44	4	3
1:A:181:THR:HG21	1:A:236:GLU:HB2	0.52	1.80	7	1
1:A:172:GLU:OE1	1:A:220:ARG:HD2	0.49	2.07	2	1
2:B:104:GLU:OE2	2:B:106:LEU:HB2	0.49	2.07	13	1
1:A:181:THR:HG21	1:A:236:GLU:HB3	0.47	1.87	2	2
2:B:97:LEU:O	2:B:101:VAL:HG23	0.47	2.09	18	3
1:A:196:HIS:CE1	1:A:234:ARG:HA	0.46	2.45	1	1
1:A:182:MET:HG3	1:A:236:GLU:OE2	0.43	2.13	3	1
1:A:181:THR:HB	1:A:236:GLU:OE1	0.43	2.12	11	1
1:A:174:TYR:CZ	1:A:176:GLY:HA3	0.43	2.48	4	1
1:A:230:ASP:OD2	1:A:232:ASN:HB2	0.43	2.13	14	1
1:A:187:CYS:HA	1:A:227:PHE:O	0.42	2.14	12	3

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:233:LYS:HE3	1:A:235:TRP:O	0.42	2.14	13	2
1:A:181:THR:HB	1:A:236:GLU:OE2	0.42	2.14	17	4
1:A:236:GLU:HG3	1:A:237:TYR:N	0.40	2.30	19	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	78/87 (90%)	72±1 (93±2%)	6±1 (7±2%)	0±1 (0±1%)	38	78
2	B	55/57 (96%)	51±1 (92±2%)	3±1 (6±2%)	1±1 (1±1%)	15	61
All	All	2660/2880 (92%)	2460 (92%)	179 (7%)	21 (1%)	24	71

All 7 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	132	HIS	10
2	B	133	LEU	3
1	A	218	PRO	3
1	A	169	GLY	2
2	B	128	ASP	1
1	A	231	PRO	1
2	B	135	GLY	1

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	73/78 (94%)	67±2 (91±3%)	6±2 (9±3%)	14	61

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	51/51 (100%)	48±1 (94±2%)	3±1 (6±2%)	22	71
All	All	2480/2580 (96%)	2292 (92%)	188 (8%)	17	65

All 36 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	242	ARG	20
1	A	229	THR	18
1	A	203	SER	16
1	A	219	ASP	13
1	A	228	THR	12
2	B	104	GLU	11
2	B	86	THR	11
1	A	183	SER	10
2	B	99	ASP	9
1	A	170	SER	8
1	A	236	GLU	6
2	B	-1	SER	6
1	A	220	ARG	5
2	B	132	HIS	5
2	B	134	ASP	5
1	A	191	ASP	3
1	A	172	GLU	3
2	B	138	ARG	3
2	B	102	GLU	2
2	B	124	ASP	2
1	A	182	MET	2
1	A	243	CYS	2
1	A	234	ARG	2
2	B	108	ASN	2
1	A	181	THR	1
2	B	137	LEU	1
2	B	95	LYS	1
1	A	221	ASP	1
1	A	173	ASN	1
2	B	93	GLU	1
1	A	218	PRO	1
2	B	128	ASP	1
2	B	118	GLU	1
2	B	105	ARG	1
2	B	110	ARG	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	179	SER	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 ⓘ

The completeness of assignment taking into account all chemical shift lists is 30% for the well-defined parts and 30% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *KTI55DL_K2_CS.str*

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	578
Number of shifts mapped to atoms	578
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 ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	56	-0.04 ± 0.17	None needed (< 0.5 ppm)
$^{13}C_\beta$	54	-0.05 ± 0.09	None needed (< 0.5 ppm)
$^{13}C'$	56	-0.20 ± 0.20	None needed (< 0.5 ppm)
^{15}N	54	-1.81 ± 0.36	Should be applied

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 30%, i.e. 539 atoms were assigned a chemical shift out of a possible 1775. 0 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	274/671 (41%)	108/267 (40%)	112/274 (41%)	54/130 (42%)
Sidechain	265/965 (27%)	159/575 (28%)	106/334 (32%)	0/56 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/139 (0%)	0/72 (0%)	0/55 (0%)	0/12 (0%)
Overall	539/1775 (30%)	267/914 (29%)	218/663 (33%)	54/198 (27%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 30%, i.e. 539 atoms were assigned a chemical shift out of a possible 1775. 0 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	274/671 (41%)	108/267 (40%)	112/274 (41%)	54/130 (42%)
Sidechain	265/965 (27%)	159/575 (28%)	106/334 (32%)	0/56 (0%)
Aromatic	0/139 (0%)	0/72 (0%)	0/55 (0%)	0/12 (0%)
Overall	539/1775 (30%)	267/914 (29%)	218/663 (33%)	54/198 (27%)

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.

Random coil index (RCI) for chain B: