WORLDWIDE PDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

Apr 20, 2024 – 08:30 PM EDT

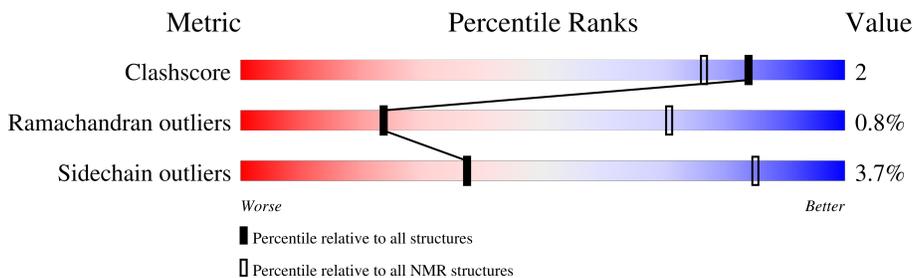|		
PDB ID	:	2KY6
Title	:	Structure of ARC92VBD/MED25ACID
Authors	:	Milbradt, A.G.; Sun, Z.J.; Selenko, P.; Takeuchi, K.; Naar, A.M.; Wagner, G.
Deposited on	:	2010-05-14

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment was not calculated.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Percentile Ranks	Value
Clashscore		2
Ramachandran outliers		0.8%
Sidechain outliers		3.7%

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	166	70% 7% 23%

2   Ensemble composition and analysis ⓘ

This entry contains 25 models. Model 14 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	A:398-A:410, A:423-A:501, A:508-A:542 (127)	0.65	14

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

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

The models can be grouped into 5 clusters and 7 single-model clusters were found.

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

3 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 2693 atoms, of which 1359 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Mediator of RNA polymerase II transcription subunit 25.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	166	2693	855	1359	235	236	8	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	388	GLY	-	expression tag	UNP Q71SY5
A	389	GLU	-	expression tag	UNP Q71SY5
A	390	PHE	-	expression tag	UNP Q71SY5

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.

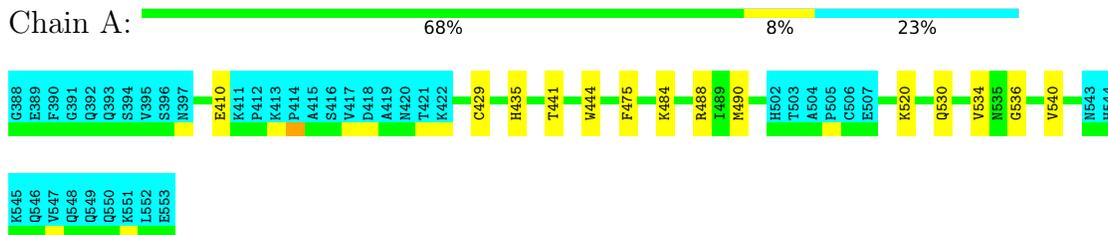• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

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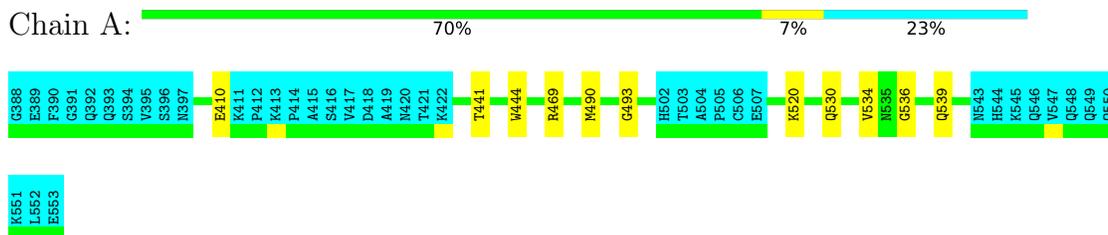
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A:

4.2.2 Score per residue for model 2

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A:

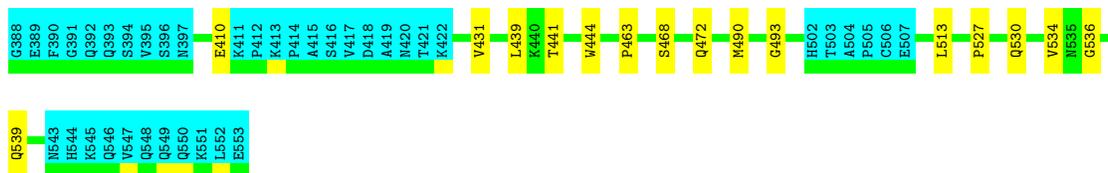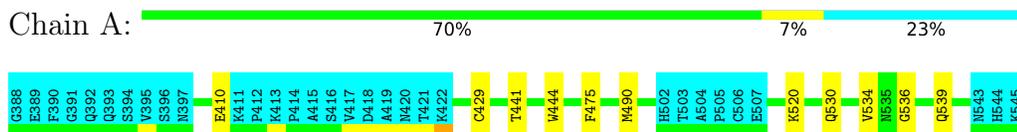### 4.2.3 Score per residue for model 3

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

4.2.4 Score per residue for model 4

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

4.2.5 Score per residue for model 5

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

4.2.6 Score per residue for model 6

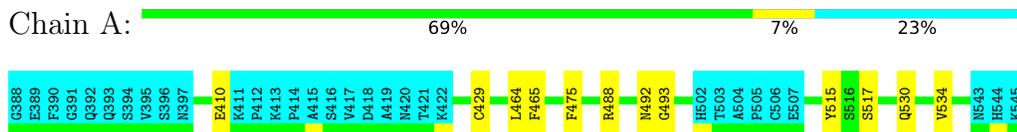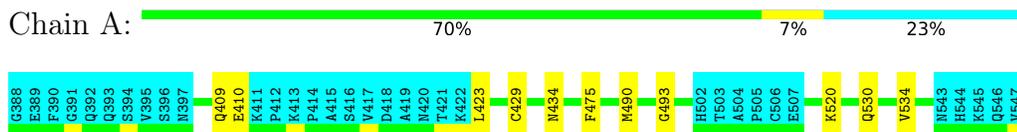• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A: 66%   10%   •   23%

4.2.7 Score per residue for model 7

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A: 67%   8%   •   23%

4.2.8 Score per residue for model 8

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A: 67%   8%   •   23%

4.2.9 Score per residue for model 9

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A: 68%   8%   •   23%
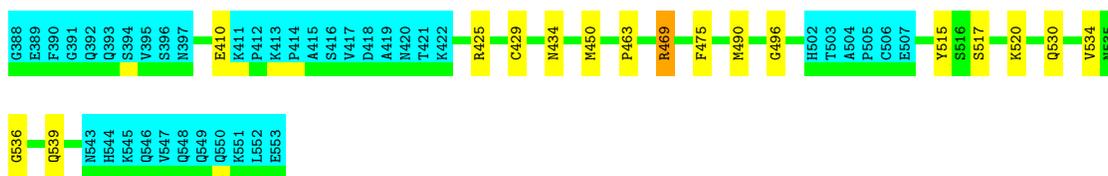
4.2.10 Score per residue for model 10

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

4.2.11 Score per residue for model 11

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

4.2.12 Score per residue for model 12

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

4.2.13 Score per residue for model 13

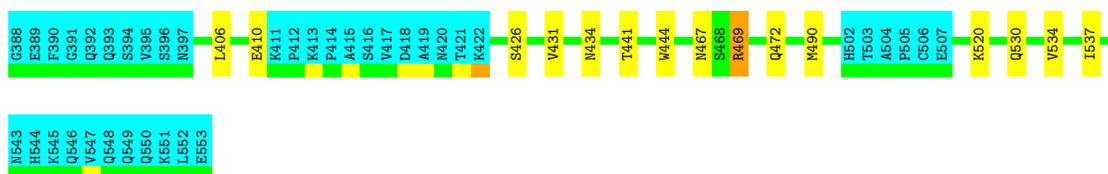• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

4.2.14 Score per residue for model 14 (medoid)

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A:

4.2.15 Score per residue for model 15

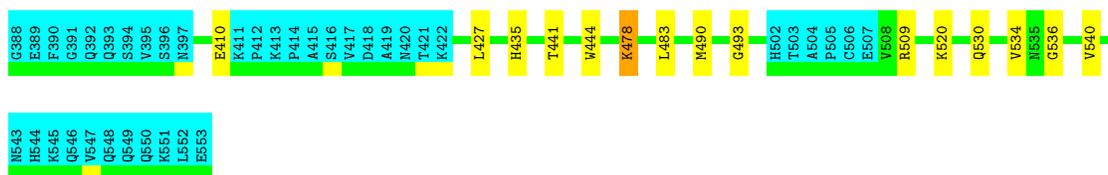• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A:

4.2.16 Score per residue for model 16

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A:

4.2.17 Score per residue for model 17

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A:

4.2.18 Score per residue for model 18

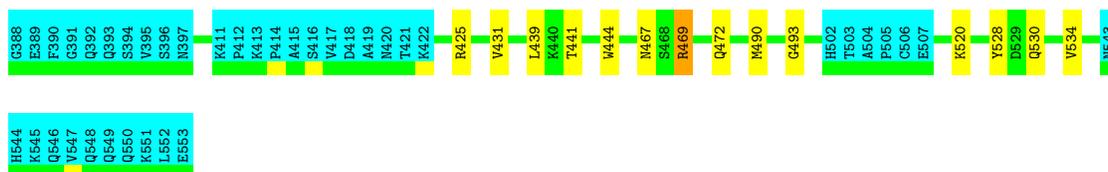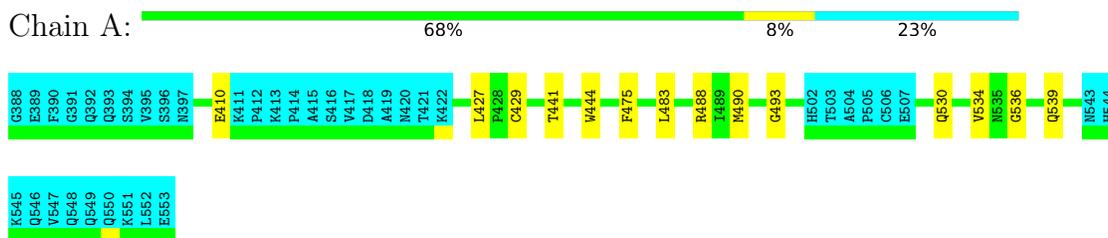• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

4.2.19 Score per residue for model 19

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

4.2.20 Score per residue for model 20

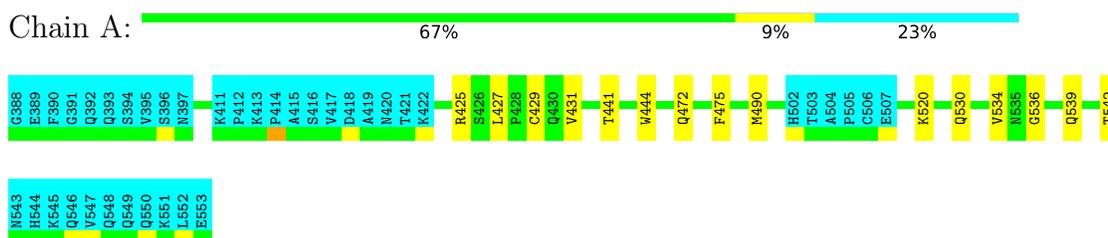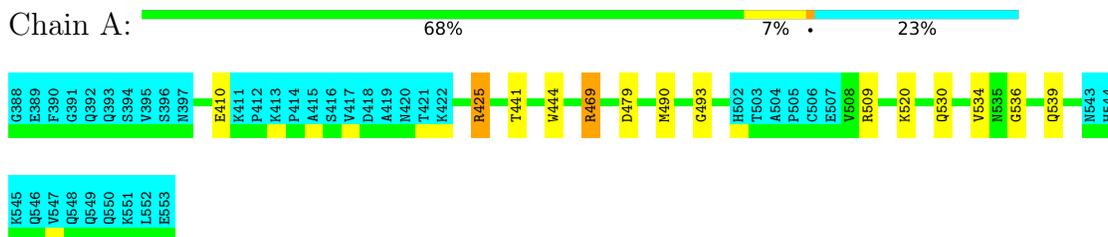• Molecule 1: Mediator of RNA polymerase II transcription subunit 25
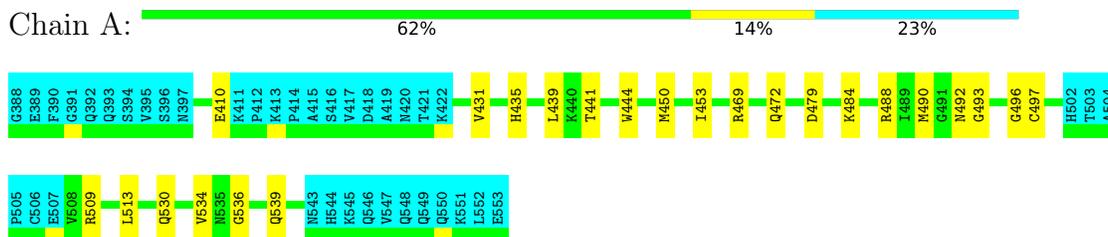
4.2.21 Score per residue for model 21

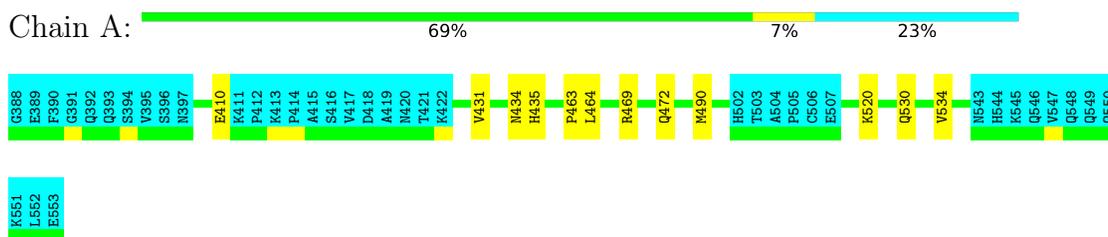• Molecule 1: Mediator of RNA polymerase II transcription subunit 25
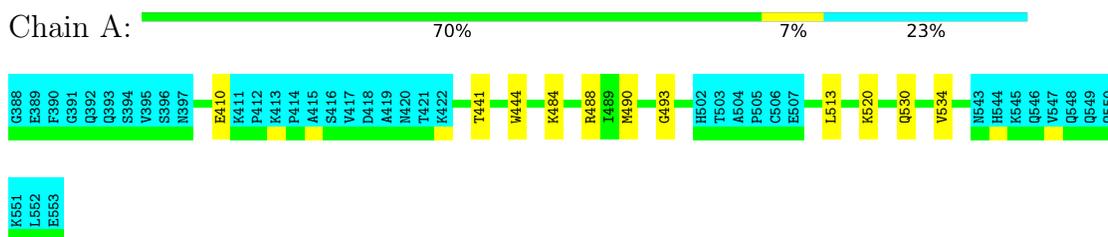
4.2.22 Score per residue for model 22

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A:
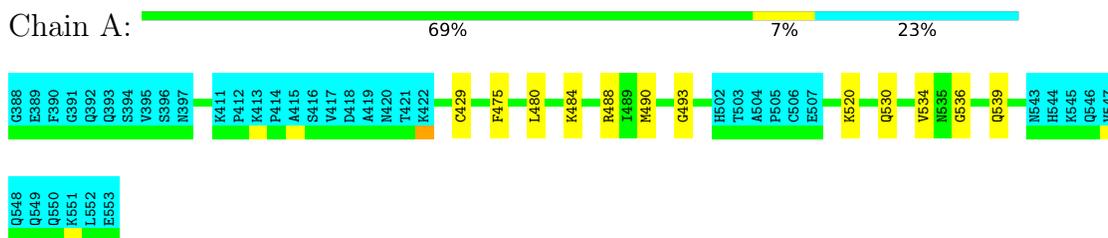
4.2.23 Score per residue for model 23

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25
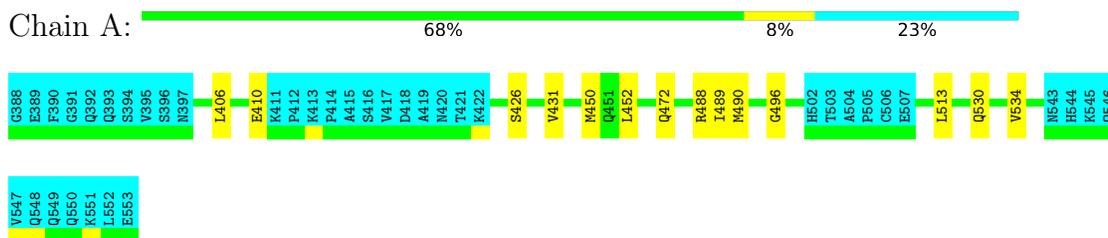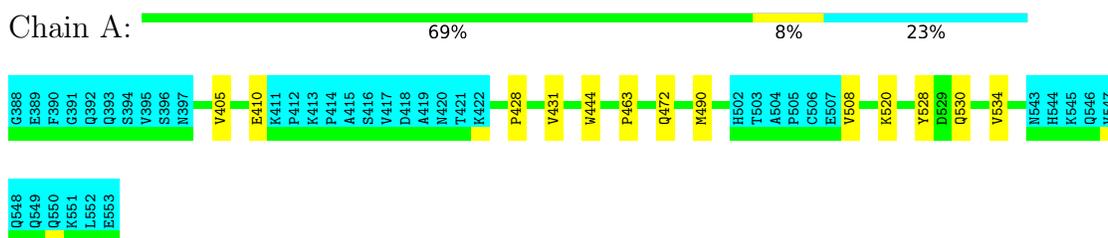
Chain A:
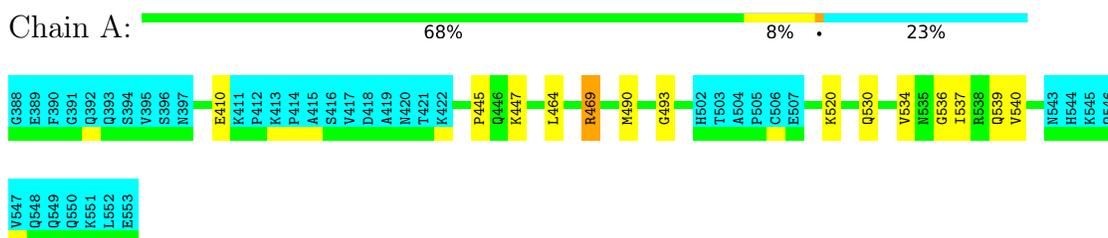
4.2.24 Score per residue for model 24

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A:
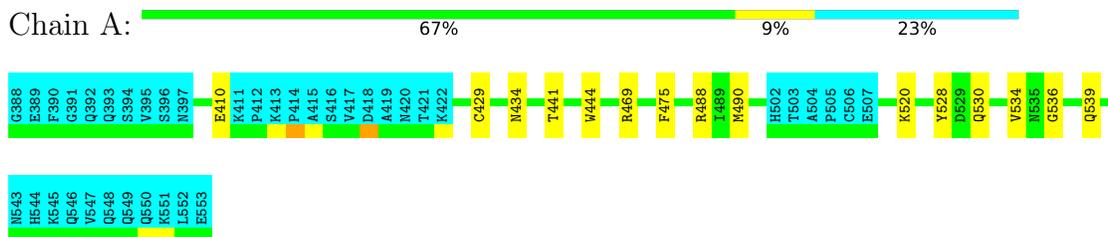
4.2.25 Score per residue for model 25

• Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A:
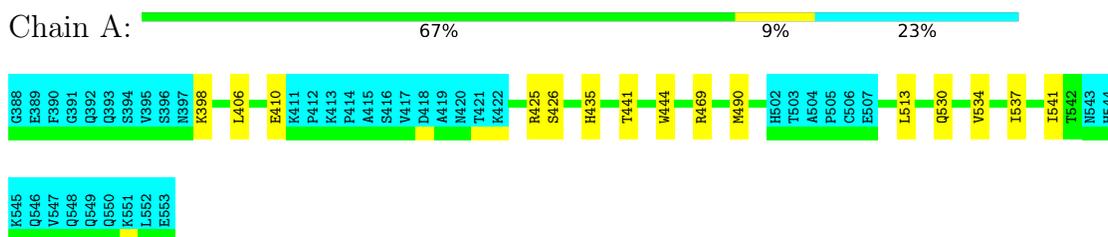
5 Refinement protocol and experimental data overview ⓘ

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

Of the 50 calculated structures, 25 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
CYANA	structure solution	3.0
CNS	refinement	1.1

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	0.2±0.5
All	All	0	4

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	469	ARG	Sidechain	3
1	A	425	ARG	Sidechain	1

6.2 Too-close contacts (i)

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

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	1033	1067	1067	5±2
All	All	25825	26675	26675	122

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 2.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:453:ILE:HD13	1:A:497:CYS:SG	0.80	2.17	20	1
1:A:536:GLY:O	1:A:539:GLN:HG2	0.58	1.98	6	13
1:A:441:THR:HA	1:A:444:TRP:CD1	0.58	2.34	3	16
1:A:429:CYS:HB3	1:A:475:PHE:CD1	0.57	2.35	24	4
1:A:469:ARG:NE	1:A:469:ARG:HA	0.56	2.16	7	5
1:A:530:GLN:O	1:A:534:VAL:HG23	0.56	2.01	9	25
1:A:453:ILE:HD11	1:A:497:CYS:SG	0.55	2.41	13	1
1:A:429:CYS:HB3	1:A:475:PHE:CE1	0.52	2.40	3	8
1:A:431:VAL:HA	1:A:472:GLN:O	0.52	2.05	7	10
1:A:537:ILE:HA	1:A:540:VAL:HG12	0.51	1.83	23	1
1:A:484:LYS:O	1:A:488:ARG:HG2	0.50	2.07	1	1
1:A:488:ARG:NE	1:A:488:ARG:HA	0.49	2.23	18	2
1:A:478:LYS:HE2	1:A:478:LYS:HA	0.48	1.85	8	1
1:A:425:ARG:NE	1:A:425:ARG:HA	0.46	2.26	10	4
1:A:484:LYS:O	1:A:488:ARG:HG3	0.46	2.10	20	2
1:A:468:SER:OG	1:A:527:PRO:HA	0.46	2.11	2	1
1:A:427:LEU:HD11	1:A:483:LEU:HA	0.45	1.88	14	3
1:A:515:TYR:CE2	1:A:517:SER:HA	0.45	2.46	4	2
1:A:480:LEU:O	1:A:484:LYS:HG2	0.45	2.12	22	1
1:A:469:ARG:HA	1:A:469:ARG:HE	0.44	1.72	7	1
1:A:537:ILE:O	1:A:541:ILE:HG12	0.44	2.13	25	2
1:A:450:MET:SD	1:A:496:GLY:HA3	0.43	2.53	20	3
1:A:536:GLY:O	1:A:540:VAL:HG23	0.43	2.13	1	2
1:A:450:MET:HA	1:A:497:CYS:O	0.43	2.14	20	1
1:A:452:LEU:HG	1:A:489:ILE:HD13	0.43	1.91	18	1
1:A:439:LEU:HD21	1:A:444:TRP:CH2	0.42	2.50	20	3
1:A:406:LEU:O	1:A:426:SER:HA	0.42	2.14	7	3
1:A:464:LEU:HD12	1:A:465:PHE:N	0.41	2.29	4	1
1:A:405:VAL:HG12	1:A:428:PRO:CA	0.41	2.46	19	1
1:A:444:TRP:CZ2	1:A:508:VAL:HG11	0.41	2.51	19	1
1:A:447:LYS:HB2	1:A:447:LYS:NZ	0.41	2.31	23	1
1:A:409:GLN:HA	1:A:423:LEU:O	0.40	2.16	5	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	127/166 (77%)	121±2 (95±1%)	5±2 (4±1%)	1±1 (1±1%)	24	71
All	All	3175/4150 (77%)	3022 (95%)	128 (4%)	25 (1%)	24	71

All 6 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	493	GLY	14
1	A	463	PRO	4
1	A	509	ARG	3
1	A	462	GLY	2
1	A	467	ASN	1
1	A	501	PRO	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	115/149 (77%)	111±1 (96±1%)	4±1 (4±1%)	37	85
All	All	2875/3725 (77%)	2769 (96%)	106 (4%)	37	85

All 21 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	490	MET	24
1	A	410	GLU	22
1	A	520	LYS	17
1	A	469	ARG	8
1	A	513	LEU	5
1	A	488	ARG	4
1	A	528	TYR	4
1	A	435	HIS	3
1	A	492	ASN	2
1	A	434	ASN	2
1	A	398	LYS	2
1	A	425	ARG	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	479	ASP	2
1	A	464	LEU	2
1	A	463	PRO	1
1	A	537	ILE	1
1	A	478	LYS	1
1	A	467	ASN	1
1	A	477	ASN	1
1	A	424	THR	1
1	A	427	LEU	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)

No chemical shift data were provided