Full wwPDB NMR Structure Validation Report (i)

Mar 6, 2022 – 07:45 PM EST

PDB ID	:	2KUO
Title	:	Structure and identification of ADP-ribose recognition motifs of APLF and role in the DNA damage response
Authors	:	Li, G.Y.; McCulloch, R.D.; Fenton, A.; Cheung, M.; Meng, L.; Ikura, M.; Koch, C.A.
Deposited on	:	2010-02-23

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

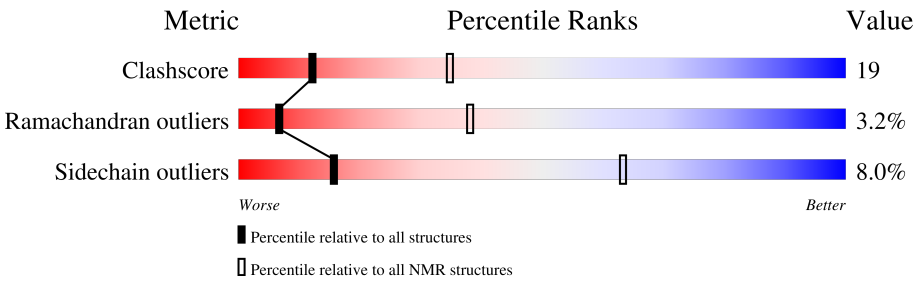
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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 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	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	91	47% 21% 32%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 15 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:376-A:404 (29)	0.22	8
2	A:416-A:448 (33)	0.91	15

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 4 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1392 atoms, of which 671 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Aprataxin and PNK-like factor.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	91	1390	439	671	136	139	5	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP Q8IW19
A	-1	SER	-	expression tag	UNP Q8IW19

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

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	2	2	2

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: Aprataxin and PNK-like factor

Chain A: 47% 21% 32%

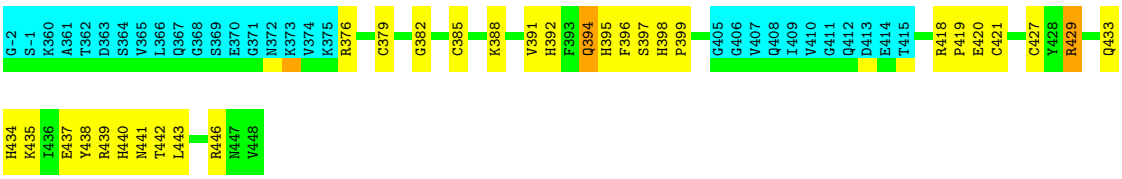
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Aprataxin and PNK-like factor

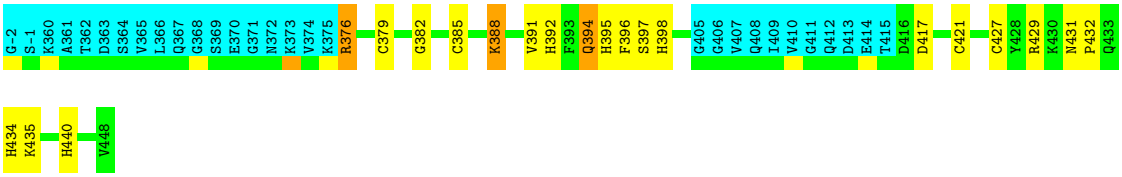
Chain A: 35% 31% • 32%



4.2.2 Score per residue for model 2

• Molecule 1: Aprataxin and PNK-like factor

Chain A: 45% 20% • 32%



4.2.3 Score per residue for model 3

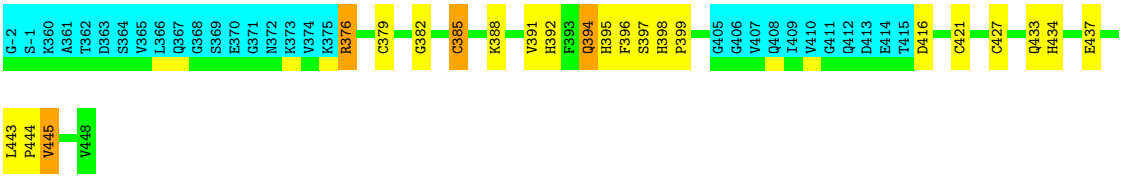
- Molecule 1: Aprataxin and PNK-like factor

Chain A: 44% 23% • 32%

4.2.4 Score per residue for model 4

- Molecule 1: Aprataxin and PNK-like factor

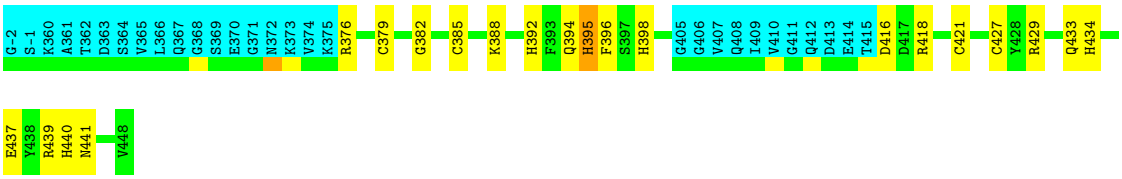
Chain A: 44% 20% • 32%



4.2.5 Score per residue for model 5

- Molecule 1: Aprataxin and PNK-like factor

Chain A: 45% 22% • 32%



4.2.6 Score per residue for model 6

- Molecule 1: Aprataxin and PNK-like factor

Chain A: 40% 29% 32%

G-2 S-1 K360 A361 T362 D363 S364 V365 L366 Q367 G368 S369 E370 G371 N372 K373 V374 K375 R376 C379 G382 C385 K388 V391 H392 F393 Q394 H395 F396 S397 H398 P399 D403 Y404 G405 G406 V407 Q408 I409 V410 G411 Q412 D413 E414 T415 D416 D417 R418 C421 C427 H434

K435 R439 H440 R446 N447 V448

4.2.7 Score per residue for model 7

- Molecule 1: Aprataxin and PNK-like factor

Chain A: 37% 27% • 32%

G-2 S-1 K360 A361 T362 D363 S364 V365 L366 Q367 G368 S369 E370 G371 N372 K373 V374 K375 R376 C379 G382 C385 Y386 R387 K388 H392 F393 Q394 H395 F396 S397 H398 P399 G405 G406 V407 Q408 I409 V410 G411 Q412 D413 E414 T415 R418 P419 E420 C421 P422 Y423 C427 Y428 R429

P432 Q433 H434 K435 I436 E437 Y438 R439 H440 N441 T442 L443 V448

4.2.8 Score per residue for model 8

- Molecule 1: Aprataxin and PNK-like factor

Chain A: 37% 27% • 32%

G-2 S-1 K360 A361 T362 D363 S364 V365 L366 Q367 G368 S369 E370 G371 N372 K373 V374 K375 R376 C379 M380 Y381 G382 C385 K388 V391 H392 F393 Q394 H395 F396 S397 H398 D403 Y404 G405 G406 V407 Q408 I409 V410 G411 Q412 D413 E414 T415 D416 D417 R418 C421 C427 N431

P432 Q433 H434 K435 I436 E437 Y438 R439 H440 L443 P444 V445 R446 N447 V448

4.2.9 Score per residue for model 9

- Molecule 1: Aprataxin and PNK-like factor

Chain A: 49% 18% • 32%

G-2 S-1 K360 A361 T362 D363 S364 V365 L366 Q367 G368 S369 E370 G371 N372 K373 V374 K375 R376 C379 G382 C385 K388 H392 H398 G405 G406 V407 Q408 I409 V410 G411 Q412 D413 E414 T415 D416 D417 C421 C427 Q433 H434 E437 P444 V445 R446 N447 V448

4.2.10 Score per residue for model 10

- Molecule 1: Aprataxin and PNK-like factor

Chain A: 45% 21% • 32%

G-2 S-1 K360 A361 T362 D363 S364 V365 L366 Q367 G368 S369 E370 G371 N372 K373 V374 K375 R376 C379 M380 Y381 G382 A383 N384 C385 K388 V391 H392 F393 Q394 H395 F396 S397 H398 P399 G400 G405 G406 V407 Q408 I409 V410 G411 Q412 D413 E414 T415 C421 C427 P432 Q433 H434

K435 I436 H440 V448

4.2.11 Score per residue for model 11

- Molecule 1: Aprataxin and PNK-like factor

Chain A: 49% 19% 32%

G-2 S-1 K360 A361 T362 D363 S364 V365 L366 Q367 G368 S369 E370 G371 N372 K373 V374 K375 R376 C379 G382 C385 K388 V391 H392 H395 F396 S397 H398 G405 G406 V407 Q408 I409 V410 G411 Q412 D413 E414 T415 C421 P422 Y423 C427 Y428 R429 H434 K435 H440

V448

4.2.12 Score per residue for model 12

- Molecule 1: Aprataxin and PNK-like factor

Chain A: 42% 24% • 32%

G-2 S-1 K360 A361 T362 D363 S364 V365 L366 Q367 G368 S369 E370 G371 N372 K373 V374 K375 R376 C379 G382 C385 K388 V391 H392 F393 Q394 H395 F396 P399 G405 G406 V407 Q408 I409 V410 G411 Q412 D413 E414 T415 D416 D417 C421 C427 Q433 H434 K435 I436 E437

H440 L443 P444 V445 R446 N447 V448

4.2.13 Score per residue for model 13

- Molecule 1: Aprataxin and PNK-like factor

Chain A: 40% 26% • 32%

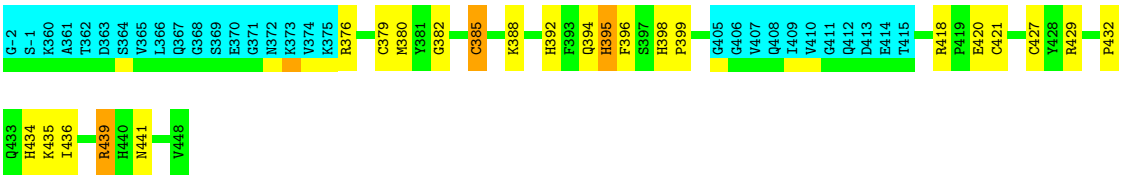
G-2 S-1 K360 A361 T362 D363 S364 V365 L366 Q367 G368 S369 E370 G371 N372 K373 V374 K375 R376 T377 S378 C379 G382 C385 K388 V391 H392 F393 Q394 H395 F396 S397 H398 P399 G405 G406 V407 Q408 I409 V410 G411 Q412 D413 E414 T415 D416 D417 R418 C421 C427 P432 Q433

H434 K435 I436 E437 Y438 R439 H440 N441 V448

4.2.14 Score per residue for model 14

- Molecule 1: Aprataxin and PNK-like factor

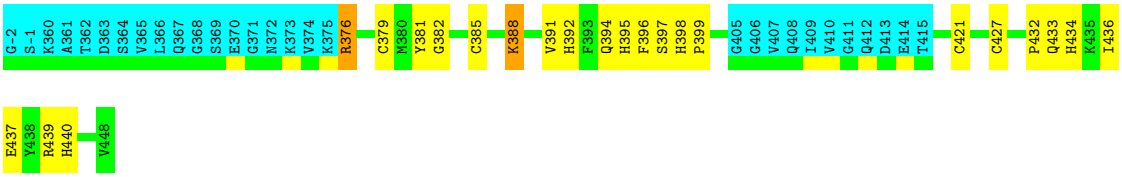
Chain A: 43% 22% • 32%



4.2.15 Score per residue for model 15 (medoid)

- Molecule 1: Aprataxin and PNK-like factor

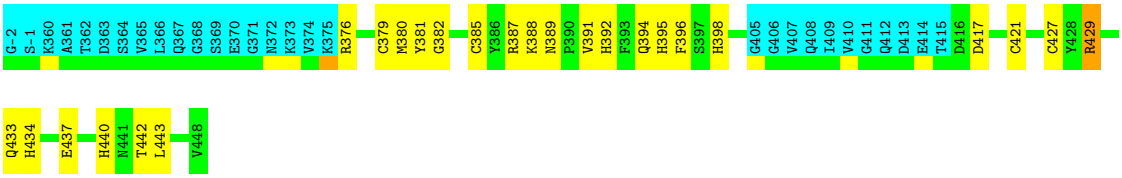
Chain A: 43% 23% • 32%



4.2.16 Score per residue for model 16

- Molecule 1: Aprataxin and PNK-like factor

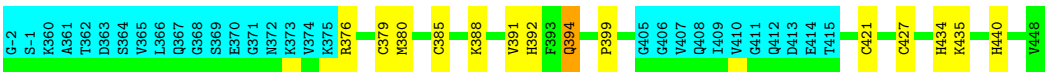
Chain A: 41% 26% • 32%



4.2.17 Score per residue for model 17

- Molecule 1: Aprataxin and PNK-like factor

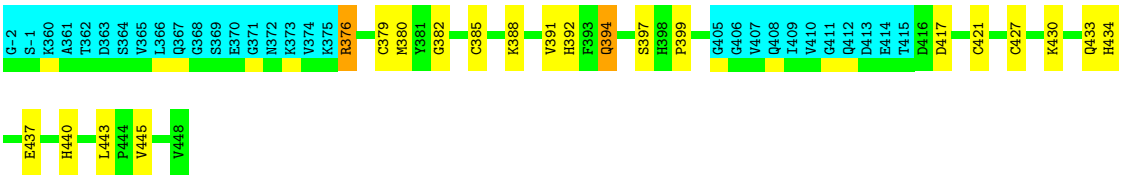
Chain A: 53% 14% • 32%



4.2.18 Score per residue for model 18

- Molecule 1: Aprataxin and PNK-like factor

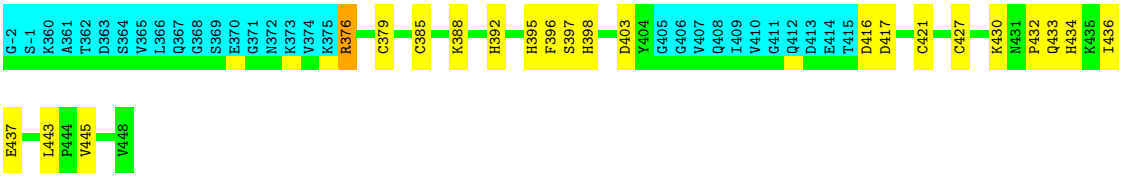
Chain A: 45% 21% • 32%



4.2.19 Score per residue for model 19

- Molecule 1: Aprataxin and PNK-like factor

Chain A: 44% 23% • 32%



4.2.20 Score per residue for model 20

- Molecule 1: Aprataxin and PNK-like factor

Chain A: 44% 22% • 32%

5 Refinement protocol and experimental data overview (i)

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

Of the 100 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
CYANA	structure solution	2.1
CNS	refinement	1.1

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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

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

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 Too-close contacts (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	518	470	472	19±4
2	A	2	0	0	2±1
All	All	10400	9400	9451	382

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 19.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:421:CYS:SG	1:A:427:CYS:SG	1.00	2.59	17	20
1:A:379:CYS:SG	1:A:385:CYS:SG	0.97	2.62	19	20
1:A:421:CYS:HG	2:A:450:ZN:ZN	0.96	0.71	3	13
1:A:379:CYS:HG	2:A:449:ZN:ZN	0.93	0.64	9	11
1:A:427:CYS:HG	2:A:450:ZN:ZN	0.89	0.65	15	13
1:A:385:CYS:HG	2:A:449:ZN:ZN	0.89	0.60	16	13
1:A:385:CYS:SG	1:A:392:HIS:NE2	0.79	2.56	9	19
1:A:427:CYS:SG	1:A:434:HIS:NE2	0.78	2.56	11	18
1:A:421:CYS:SG	1:A:440:HIS:NE2	0.75	2.59	8	15
1:A:379:CYS:SG	1:A:398:HIS:NE2	0.73	2.61	11	16
1:A:421:CYS:SG	1:A:434:HIS:NE2	0.72	2.62	5	13

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:379:CYS:SG	1:A:392:HIS:NE2	0.69	2.66	19	16
1:A:421:CYS:SG	1:A:434:HIS:CE1	0.68	2.86	7	15
1:A:385:CYS:SG	1:A:398:HIS:NE2	0.65	2.70	5	10
1:A:382:GLY:O	1:A:385:CYS:SG	0.62	2.57	10	17
1:A:418:ARG:HG3	1:A:439:ARG:HG2	0.60	1.71	1	1
1:A:429:ARG:HA	1:A:429:ARG:HE	0.60	1.55	20	6
1:A:420:GLU:HG3	1:A:441:ASN:ND2	0.59	2.12	7	1
1:A:439:ARG:NH2	1:A:441:ASN:HD21	0.56	1.98	7	2
1:A:433:GLN:O	1:A:437:GLU:HG3	0.56	2.01	5	13
1:A:379:CYS:HG	1:A:385:CYS:HG	0.55	1.42	20	1
1:A:377:THR:O	1:A:396:PHE:HA	0.54	2.03	13	1
1:A:421:CYS:SG	1:A:427:CYS:CB	0.53	2.97	9	12
1:A:391:VAL:O	1:A:394:GLN:HG3	0.53	2.03	13	11
1:A:418:ARG:O	1:A:439:ARG:HD3	0.52	2.03	7	3
1:A:421:CYS:HG	1:A:427:CYS:HG	0.52	1.47	15	1
1:A:376:ARG:HB2	1:A:397:SER:OG	0.51	2.05	15	5
1:A:439:ARG:NE	1:A:441:ASN:HD21	0.50	2.03	5	1
1:A:432:PRO:O	1:A:436:ILE:HG13	0.49	2.08	7	7
1:A:376:ARG:HB3	1:A:397:SER:N	0.49	2.23	18	1
1:A:391:VAL:HG12	1:A:395:HIS:NE2	0.49	2.23	4	3
1:A:395:HIS:HD1	1:A:396:PHE:HD2	0.49	1.50	2	5
1:A:395:HIS:ND1	1:A:396:PHE:HD1	0.49	2.05	19	5
1:A:421:CYS:HG	1:A:440:HIS:HE2	0.48	1.52	18	1
1:A:421:CYS:HG	1:A:434:HIS:CE1	0.48	2.27	7	2
1:A:427:CYS:SG	1:A:429:ARG:HG2	0.48	2.48	11	1
1:A:395:HIS:ND1	1:A:396:PHE:CD1	0.48	2.82	12	6
1:A:427:CYS:SG	1:A:440:HIS:NE2	0.48	2.87	15	1
1:A:388:LYS:N	1:A:388:LYS:HD3	0.47	2.24	15	3
1:A:420:GLU:HG2	1:A:441:ASN:ND2	0.47	2.24	14	1
1:A:391:VAL:O	1:A:395:HIS:CD2	0.47	2.67	20	6
1:A:387:ARG:NE	1:A:387:ARG:HA	0.46	2.26	7	1
1:A:433:GLN:O	1:A:437:GLU:HG2	0.46	2.11	13	1
1:A:395:HIS:ND1	1:A:396:PHE:CD2	0.46	2.83	13	2
1:A:431:ASN:O	1:A:434:HIS:HB3	0.46	2.11	3	1
1:A:444:PRO:O	1:A:445:VAL:HG12	0.46	2.11	9	2
1:A:421:CYS:SG	1:A:440:HIS:CD2	0.46	3.09	7	3
1:A:393:PHE:HA	1:A:396:PHE:O	0.45	2.11	13	1
1:A:379:CYS:SG	1:A:392:HIS:CE1	0.45	3.09	19	1
1:A:379:CYS:HB2	1:A:397:SER:O	0.45	2.10	6	1
1:A:430:LYS:NZ	1:A:430:LYS:HB2	0.45	2.26	18	1
1:A:395:HIS:HD1	1:A:396:PHE:HD1	0.45	1.54	14	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:392:HIS:CD2	1:A:398:HIS:CE1	0.44	3.05	13	4
1:A:421:CYS:SG	1:A:427:CYS:HB2	0.44	2.52	5	1
1:A:381:TYR:HB2	1:A:385:CYS:HB3	0.44	1.89	15	2
1:A:398:HIS:O	1:A:400:GLY:N	0.44	2.50	10	1
1:A:420:GLU:HG2	1:A:439:ARG:HE	0.43	1.72	1	1
1:A:439:ARG:NH1	1:A:441:ASN:HD21	0.43	2.12	13	1
1:A:385:CYS:SG	1:A:398:HIS:CD2	0.43	3.11	5	2
1:A:418:ARG:HB2	1:A:419:PRO:HD2	0.43	1.88	7	1
1:A:429:ARG:HA	1:A:429:ARG:NE	0.43	2.28	20	1
1:A:383:ALA:HB2	1:A:399:PRO:HG2	0.43	1.88	10	1
1:A:421:CYS:SG	1:A:434:HIS:HE1	0.43	2.34	11	1
1:A:418:ARG:NH1	1:A:435:LYS:O	0.43	2.52	14	3
1:A:395:HIS:ND1	1:A:396:PHE:HD2	0.43	2.12	2	1
1:A:431:ASN:OD1	1:A:432:PRO:HD2	0.42	2.13	2	2
1:A:446:ARG:O	1:A:448:VAL:N	0.42	2.52	6	1
1:A:379:CYS:HA	1:A:396:PHE:CD2	0.42	2.49	12	1
1:A:392:HIS:HA	1:A:395:HIS:CE1	0.42	2.50	2	2
1:A:418:ARG:HD2	1:A:437:GLU:O	0.42	2.14	5	1
1:A:423:TYR:HB2	1:A:427:CYS:HB2	0.42	1.92	11	2
1:A:387:ARG:HH21	1:A:389:ASN:ND2	0.42	2.13	16	1
1:A:388:LYS:HD3	1:A:388:LYS:H	0.41	1.75	2	1
1:A:419:PRO:O	1:A:438:TYR:HA	0.41	2.16	1	1
1:A:382:GLY:O	1:A:398:HIS:CD2	0.41	2.74	7	1
1:A:418:ARG:HB2	1:A:437:GLU:O	0.41	2.16	8	1
1:A:418:ARG:HG3	1:A:439:ARG:CG	0.41	2.46	13	1
1:A:447:ASN:HD22	1:A:447:ASN:N	0.40	2.14	12	1
1:A:430:LYS:HB2	1:A:430:LYS:NZ	0.40	2.31	19	1
1:A:385:CYS:SG	1:A:392:HIS:CD2	0.40	3.14	10	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	61/91 (67%)	54±2 (89±3%)	5±2 (8±3%)	2±1 (3±2%)	7 38

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1220/1820 (67%)	1085 (89%)	96 (8%)	39 (3%)	7 38

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	399	PRO	12
1	A	416	ASP	8
1	A	417	ASP	8
1	A	445	VAL	7
1	A	447	ASN	3
1	A	429	ARG	1

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	58/80 (72%)	53±1 (92±2%)	5±1 (8±2%)	16 63
All	All	1160/1600 (72%)	1067 (92%)	93 (8%)	16 63

All 14 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	388	LYS	20
1	A	376	ARG	19
1	A	394	GLN	17
1	A	443	LEU	9
1	A	435	LYS	7
1	A	429	ARG	3
1	A	442	THR	3
1	A	446	ARG	3
1	A	395	HIS	3
1	A	403	ASP	3
1	A	385	CYS	2
1	A	439	ARG	2
1	A	420	GLU	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	447	ASN	1

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates (i)

There are no monosaccharides in this entry.

6.6 Ligand geometry (i)

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

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

No chemical shift data were provided