Full wwPDB NMR Structure Validation Report (i)

Oct 4, 2021 – 06:12 PM EDT

PDB ID	:	2KLU
Title	:	NMR structure of the transmembrane and cytoplasmic domains of human CD4
Authors	:	Wittlich, M.; Willbold, D.
Deposited on	:	2009-07-08

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:

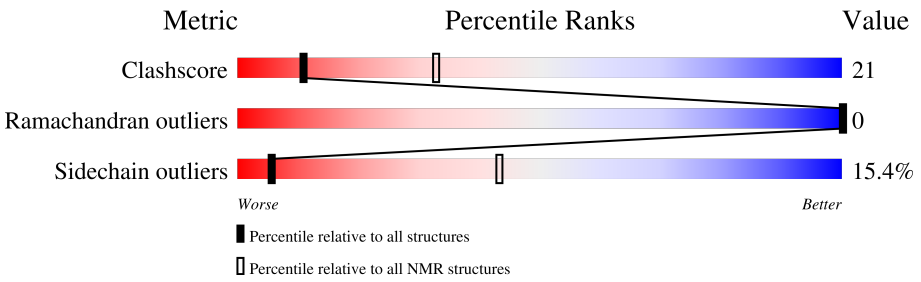
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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	70	17% 7% • 74%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 10 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:375-A:392 (18)	0.11	10

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 5 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1153 atoms, of which 601 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called T-cell surface glycoprotein CD4.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	70	1153	351	601	111	88	2	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	364	GLY	-	expression tag	UNP P01730
A	365	PRO	-	expression tag	UNP P01730
A	366	LEU	-	expression tag	UNP P01730
A	367	VAL	-	expression tag	UNP P01730
A	368	PRO	-	expression tag	UNP P01730
A	369	ARG	-	expression tag	UNP P01730
A	370	GLY	-	expression tag	UNP P01730
A	371	SER	-	expression tag	UNP P01730
A	394	SER	CYS	engineered mutation	UNP P01730
A	397	SER	CYS	engineered mutation	UNP P01730
A	420	SER	CYS	engineered mutation	UNP P01730
A	422	SER	CYS	engineered mutation	UNP P01730
A	430	HIS	CYS	engineered mutation	UNP P01730

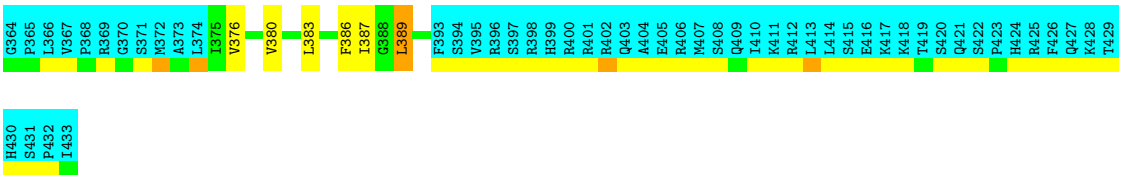
4 Residue-property plots (i)

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: T-cell surface glycoprotein CD4

Chain A: 17% 7% • 74%



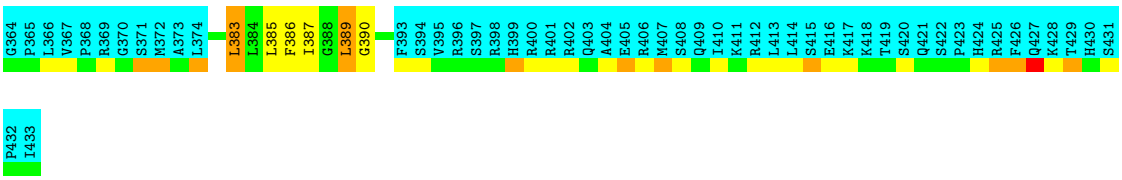
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: T-cell surface glycoprotein CD4

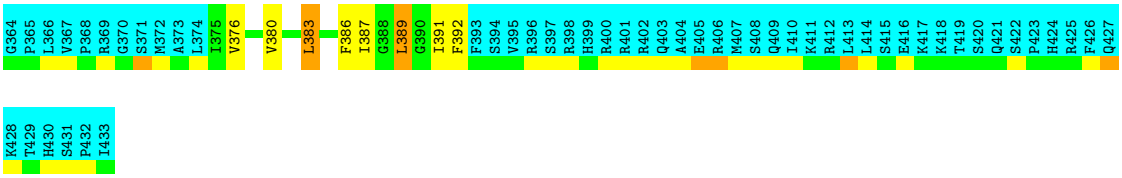
Chain A: 17% 6% • 74%



4.2.2 Score per residue for model 2

• Molecule 1: T-cell surface glycoprotein CD4

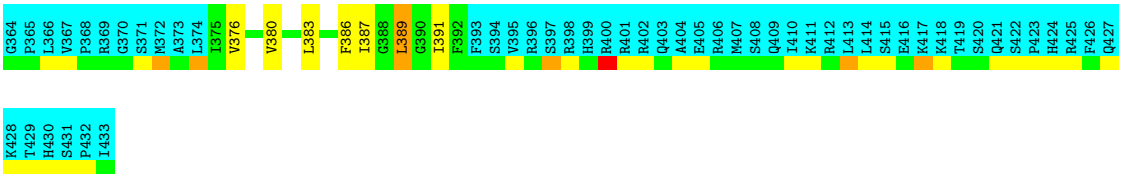
Chain A: 14% 9% • 74%



4.2.3 Score per residue for model 3

- Molecule 1: T-cell surface glycoprotein CD4

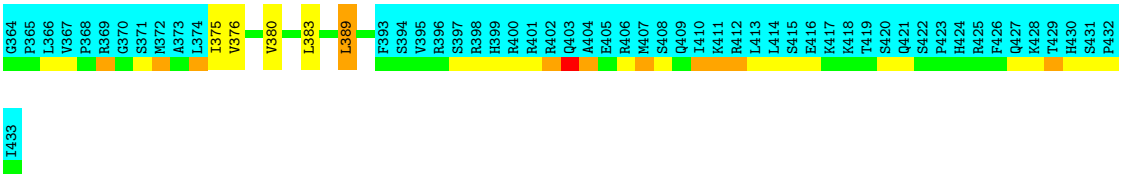
Chain A: 16% 9% • 74%



4.2.4 Score per residue for model 4

- Molecule 1: T-cell surface glycoprotein CD4

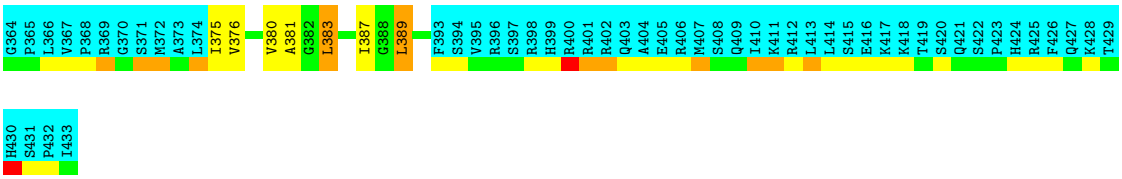
Chain A: 19% 6% • 74%



4.2.5 Score per residue for model 5

- Molecule 1: T-cell surface glycoprotein CD4

Chain A: 16% 7% • 74%



4.2.6 Score per residue for model 6

- Molecule 1: T-cell surface glycoprotein CD4

Chain A: 19% 6% • 74%

G364 P365 L366 V367 P368 R369 G370 S371 M372 A373 L374 V380 A381 G382 L383 F386 L389 F393 S394 V395 R396 S397 R398 H399 R400 R401 R402 Q403 A404 E405 R406 M407 S408 Q409 I410 K411 R412 L413 L414 S415 E416 K417 K418 T419 S420 Q421 S422 P423 H424 R425 F426 Q427 K428 T429 H430 S431

P432 I433

4.2.7 Score per residue for model 7

- Molecule 1: T-cell surface glycoprotein CD4

Chain A: 14% 10% • 74%

G364 P365 L366 V367 P368 R369 G370 S371 M372 A373 L374 I375 V376 L377 V380 A381 G382 L383 F386 L389 F392 F393 S394 V395 R396 S397 R398 H399 R400 R401 R402 Q403 A404 E405 R406 M407 S408 Q409 I410 K411 R412 L413 L414 S415 E416 K417 K418 T419 S420 Q421 S422 P423 H424 R425 F426 Q427

K428 T429 H430 S431 P432 I433

4.2.8 Score per residue for model 8

- Molecule 1: T-cell surface glycoprotein CD4

Chain A: 14% 9% • 74%

G364 P365 L366 V367 P368 R369 G370 S371 M372 A373 L374 I375 V376 L377 V380 L383 F386 I387 G388 L389 F392 F393 S394 V395 R396 S397 R398 H399 R400 R401 R402 Q403 A404 E405 R406 M407 S408 Q409 I410 K411 R412 L413 L414 S415 E416 K417 K418 T419 S420 Q421 S422 P423 H424 R425 F426 Q427

K428 T429 H430 S431 P432 I433

4.2.9 Score per residue for model 9

- Molecule 1: T-cell surface glycoprotein CD4

Chain A: 13% 11% • 74%

G364 P365 L366 V367 P368 R369 G370 S371 M372 A373 L374 I375 V376 L377 V380 A381 G382 L383 F386 L389 G390 I391 F392 F393 S394 V395 R396 S397 R398 H399 R400 R401 R402 Q403 A404 E405 R406 M407 S408 Q409 I410 K411 R412 L413 L414 S415 E416 K417 K418 T419 S420 Q421 S422 P423 H424 R425 F426

Q427 K428 T429 H430 S431 P432 I433

4.2.10 Score per residue for model 10 (medoid)

- Molecule 1: T-cell surface glycoprotein CD4

Chain A: 14% 10% • 74%

G364 P365 L366 V367 P368 R369 G370 S371 M372 A373 L374 I375 V376 G379 L383 F386 I387 G388 L389 G390 I391 F392 F393 S394 V395 R396 S397 R398 H399 R400 R401 R402 Q403 A404 E405 R406 M407 S408 Q409 I410 K411 R412 L413 L414 S415 E416 K417 K418 T419 S420 Q421 S422 P423 H424 R425 F426 Q427

K428 T429 H430 S431 P432 I433

4.2.11 Score per residue for model 11

- Molecule 1: T-cell surface glycoprotein CD4

Chain A: 16% 9% • 74%

G364 P365 L366 V367 P368 R369 G370 S371 M372 A373 L374 I375 V376 G379 L383 I387 G388 L389 G390 I391 F392 F393 S394 V395 R396 S397 R398 H399 R400 R401 R402 Q403 A404 E405 R406 M407 S408 Q409 I410 K411 R412 L413 L414 S415 E416 K417 K418 T419 S420 Q421 S422 P423 H424 R425 F426 Q427 K428

T429 H430 S431 P432 I433

4.2.12 Score per residue for model 12

- Molecule 1: T-cell surface glycoprotein CD4

Chain A: 11% 11% • 74%

G364 P365 L366 V367 P368 R369 G370 S371 M372 A373 L374 I375 V376 G379 V380 A381 G382 L383 F386 I387 G388 L389 G390 I391 F392 F393 S394 V395 R396 S397 R398 H399 R400 R401 R402 Q403 A404 E405 R406 M407 S408 Q409 I410 K411 R412 L413 L414 S415 E416 K417 K418 T419 S420 Q421 S422 P423 H424 R425

F426 Q427 K428 T429 H430 S431 P432 I433

4.2.13 Score per residue for model 13

- Molecule 1: T-cell surface glycoprotein CD4

Chain A: 17% 7% • 74%

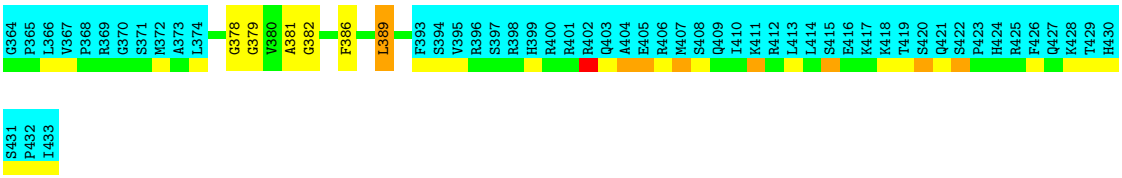
G364 P365 L366 V367 P368 R369 G370 S371 M372 A373 L374 I375 V376 L377 L383 L389 G390 F393 S394 V395 R396 S397 R398 H399 R400 R401 R402 Q403 A404 E405 R406 M407 S408 Q409 I410 K411 R412 L413 L414 S415 E416 K417 K418 T419 S420 Q421 S422 P423 H424 R425 F426 Q427 K428 T429 H430 S431 P432

I433

4.2.14 Score per residue for model 14

- Molecule 1: T-cell surface glycoprotein CD4

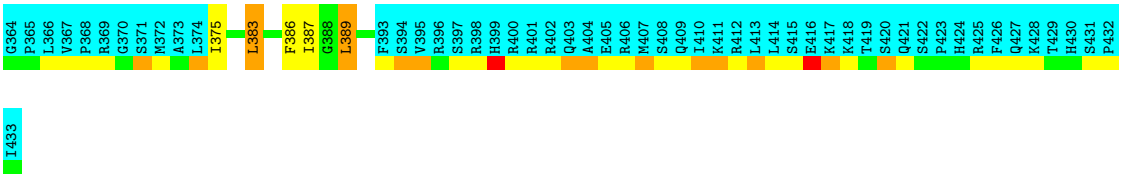
Chain A: 17% 7% • 74%



4.2.15 Score per residue for model 15

- Molecule 1: T-cell surface glycoprotein CD4

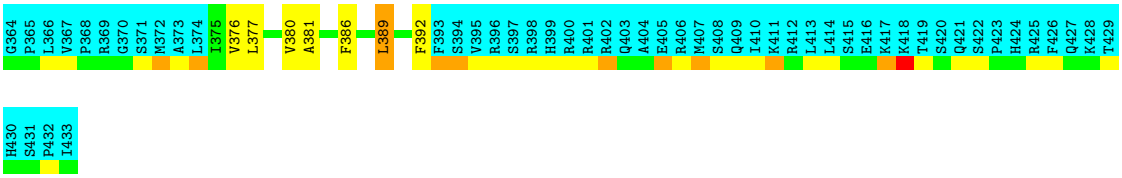
Chain A: 19% • • 74%



4.2.16 Score per residue for model 16

- Molecule 1: T-cell surface glycoprotein CD4

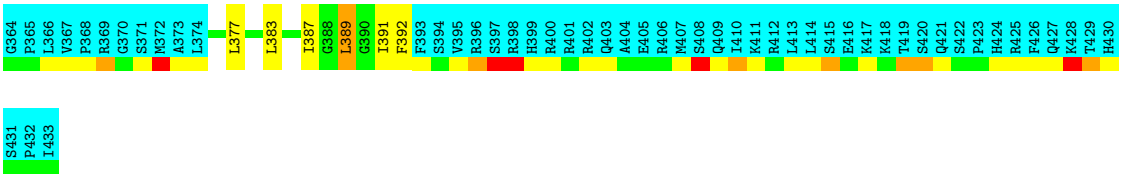
Chain A: 16% 9% • 74%



4.2.17 Score per residue for model 17

- Molecule 1: T-cell surface glycoprotein CD4

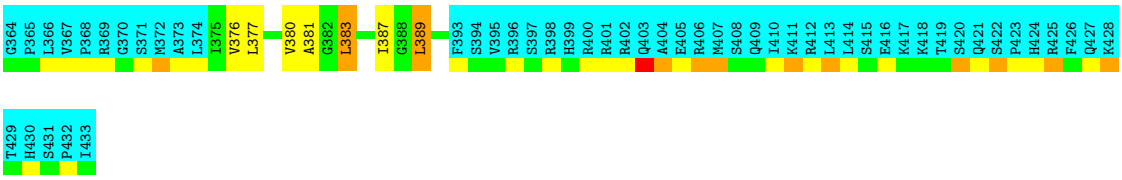
Chain A: 17% 7% • 74%



4.2.18 Score per residue for model 18

- Molecule 1: T-cell surface glycoprotein CD4

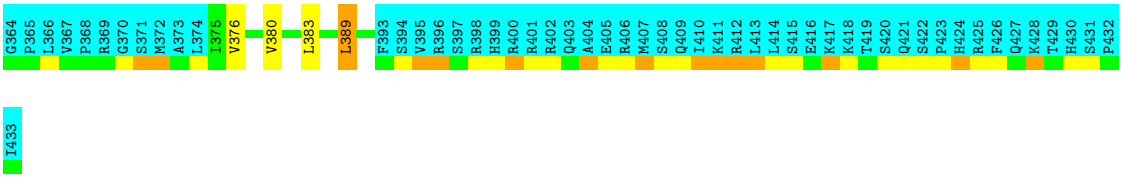
Chain A: 16% 7% • 74%



4.2.19 Score per residue for model 19

- Molecule 1: T-cell surface glycoprotein CD4

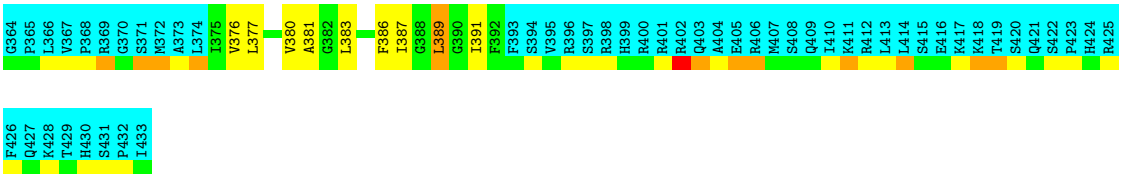
Chain A: 20% • • 74%



4.2.20 Score per residue for model 20

- Molecule 1: T-cell surface glycoprotein CD4

Chain A: 13% 11% • 74%



5 Refinement protocol and experimental data overview (i)

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

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	refinement	2.1

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.

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	125	144	144	6±1
All	All	2500	2880	2880	111

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 21.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:389:LEU:C	1:A:389:LEU:HD12	0.71	2.06	6	19
1:A:383:LEU:O	1:A:387:ILE:HD12	0.66	1.91	17	5
1:A:376:VAL:O	1:A:380:VAL:HG23	0.65	1.91	18	7
1:A:389:LEU:HD12	1:A:389:LEU:O	0.61	1.95	16	11
1:A:383:LEU:C	1:A:387:ILE:HD12	0.61	2.16	1	2
1:A:387:ILE:O	1:A:391:ILE:HD12	0.57	1.98	20	7
1:A:383:LEU:CD1	1:A:387:ILE:HD11	0.55	2.32	2	4
1:A:383:LEU:CD1	1:A:387:ILE:CD1	0.52	2.88	18	4
1:A:383:LEU:HD13	1:A:387:ILE:CD1	0.52	2.35	5	4
1:A:389:LEU:C	1:A:389:LEU:CD1	0.49	2.80	7	13
1:A:386:PHE:HA	1:A:389:LEU:HD23	0.49	1.84	20	13
1:A:389:LEU:O	1:A:392:PHE:N	0.48	2.47	17	1
1:A:383:LEU:HD13	1:A:387:ILE:HD12	0.47	1.87	5	2
1:A:376:VAL:O	1:A:380:VAL:N	0.45	2.50	8	2

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:380:VAL:O	1:A:383:LEU:N	0.45	2.49	6	1
1:A:378:GLY:O	1:A:381:ALA:HB3	0.45	2.12	14	1
1:A:389:LEU:HD12	1:A:390:GLY:N	0.43	2.28	13	1
1:A:379:GLY:O	1:A:382:GLY:N	0.43	2.49	14	1
1:A:386:PHE:O	1:A:390:GLY:N	0.43	2.50	1	1
1:A:380:VAL:O	1:A:381:ALA:C	0.41	2.58	6	8
1:A:376:VAL:O	1:A:379:GLY:N	0.41	2.53	12	3
1:A:383:LEU:HD23	1:A:383:LEU:HA	0.40	1.78	19	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	18/70 (26%)	15±1 (85±8%)	3±1 (15±8%)	0±0 (0±0%)	100	100
All	All	360/1400 (26%)	307 (85%)	53 (15%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	12/60 (20%)	10±1 (85±5%)	2±1 (15±5%)	6	43
All	All	240/1200 (20%)	203 (85%)	37 (15%)	6	43

All 4 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	389	LEU	20
1	A	383	LEU	14
1	A	392	PHE	2
1	A	385	LEU	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)

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 (i)

No chemical shift data were provided