Full wwPDB NMR Structure Validation Report (i)

Feb 26, 2022 – 03:03 PM EST

PDB ID	:	2CPE
Title	:	Solution structure of the RNA recognition motif of Ewing Sarcoma(EWS) protein
Authors	:	Nagata, T.; Muto, Y.; Inoue, M.; Kigawa, T.; Terada, T.; Shirouzu, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2005-05-19

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:

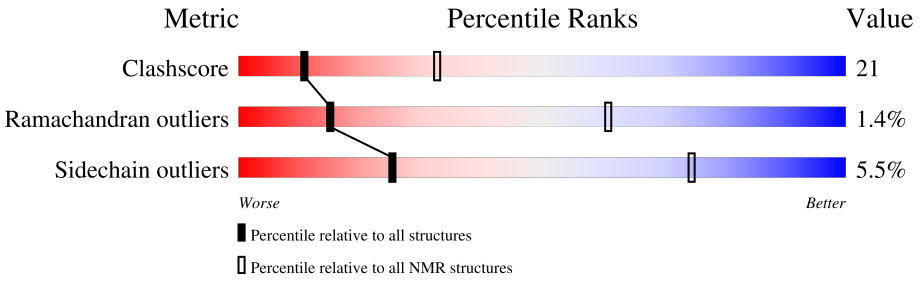
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
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	113	50% 24% • 26%

2 Ensemble composition and analysis (i)

This entry contains 20 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:361-A:444 (84)	0.10	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 4 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1660 atoms, of which 815 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called RNA-binding protein EWS.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	113	1660	523	815	142	176	4	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	346	GLY	-	cloning artifact	UNP Q01844
A	347	SER	-	cloning artifact	UNP Q01844
A	348	SER	-	cloning artifact	UNP Q01844
A	349	GLY	-	cloning artifact	UNP Q01844
A	350	SER	-	cloning artifact	UNP Q01844
A	351	SER	-	cloning artifact	UNP Q01844
A	352	GLY	-	cloning artifact	UNP Q01844
A	454	GLY	-	cloning artifact	UNP Q01844
A	455	PRO	-	cloning artifact	UNP Q01844
A	456	SER	-	cloning artifact	UNP Q01844
A	457	SER	-	cloning artifact	UNP Q01844
A	458	GLY	-	cloning artifact	UNP Q01844

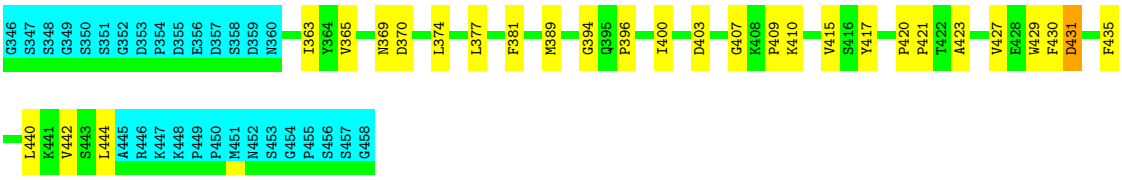
4 Residue-property plots (i)

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: RNA-binding protein EWS

Chain A: 50% 24% • 26%



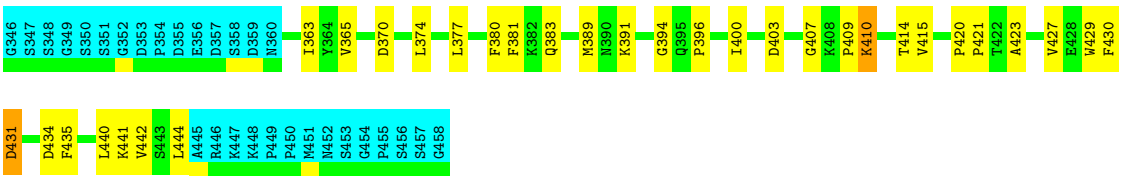
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: RNA-binding protein EWS

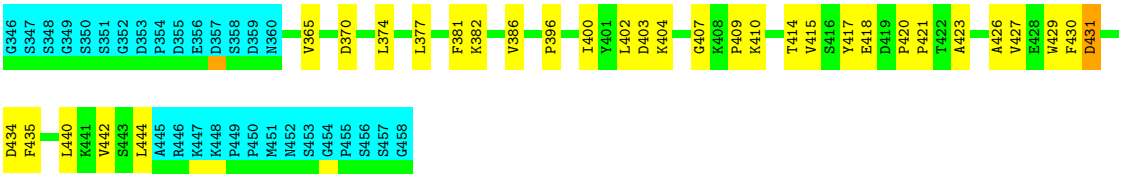
Chain A: 46% 27% • 26%



4.2.2 Score per residue for model 2

• Molecule 1: RNA-binding protein EWS

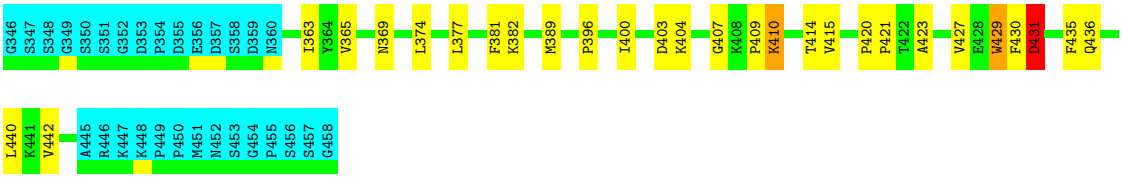
Chain A: 46% 27% • 26%



4.2.3 Score per residue for model 3

- Molecule 1: RNA-binding protein EWS

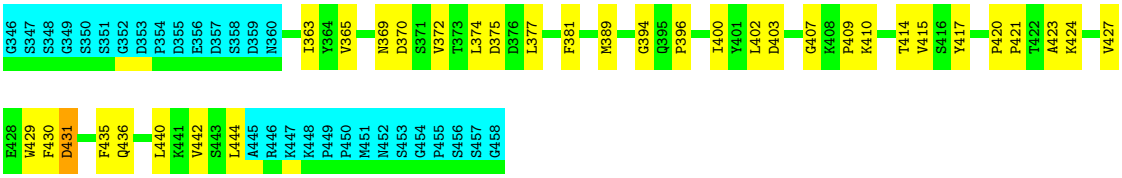
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: RNA-binding protein EWS

Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: RNA-binding protein EWS

Chain A:

4.2.6 Score per residue for model 6

- Molecule 1: RNA-binding protein EWS

Chain A: 48% 26% • 26%

G346 S347 S348 G349 S350 S351 G352 D353 P354 D355 E356 D357 S358 D359 N360 S361 A362 I363 Y364 V365 D370 L374 D375 D376 L377 F381 K382 M389 N390 K391 G394 Q395 P396 I400 D403 G407 K408 P409 K410 V415 S416 Y417 P420 P421 T422 A423 V427 E428 W429

F430 D431 F435 L440 K441 V442 A445 R446 K447 K448 P449 P450 M451 N452 S453 G454 P455 S456 S457 G458

4.2.7 Score per residue for model 7

- Molecule 1: RNA-binding protein EWS

Chain A: 49% 25% • 26%

G346 S347 S348 G349 S350 S351 G352 D353 P354 D355 E356 D357 S358 D359 N360 V365 D370 L374 L377 F381 K382 M389 R392 P396 I400 D403 G407 K408 P409 K410 V415 S416 Y417 E418 D419 P420 P421 T422 A423 V427 E428 W429 F430 D431 D434 F435

L440 K441 V442 A445 R446 K447 K448 P449 P450 M451 N452 S453 G454 P455 S456 S457 G458

4.2.8 Score per residue for model 8

- Molecule 1: RNA-binding protein EWS

Chain A: 47% 27% • 26%

G346 S347 S348 G349 S350 S351 G352 D353 P354 D355 E356 D357 S358 D359 N360 I363 Y364 V365 N369 D370 S371 V372 T373 L374 D375 D376 L377 F381 N390 T393 G394 Q395 P396 I400 D403 K404 G407 K408 P409 K410 V415 S416 Y417 P420 P421 T422 A423 V427 E428

W429 F430 D431 G432 K433 D434 V442 A445 R446 K447 K448 P449 P450 M451 N452 S453 G454 P455 S456 S457 G458

4.2.9 Score per residue for model 9

- Molecule 1: RNA-binding protein EWS

Chain A: 46% 28% 26%

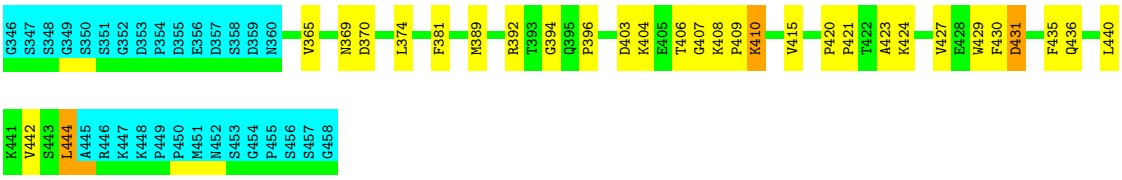
G346 S347 S348 G349 S350 S351 G352 D353 P354 D355 E356 D357 S358 D359 N360 I363 Y364 V365 D370 L374 D375 D376 L377 F380 F381 M389 N390 T393 G394 Q395 P396 I400 D403 K404 G407 K408 P409 K410 V415 S416 Y417 P420 P421 T422 A423 A426 V427 E428

W429 F430 D431 F435 L440 K441 V442 A445 R446 K447 K448 P449 P450 M451 N452 S453 G454 P455 S456 S457 G458

4.2.10 Score per residue for model 10

- Molecule 1: RNA-binding protein EWS

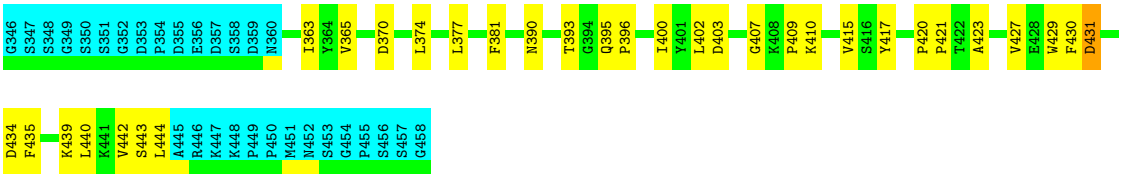
Chain A: 48% 24% • 26%



4.2.11 Score per residue for model 11

- Molecule 1: RNA-binding protein EWS

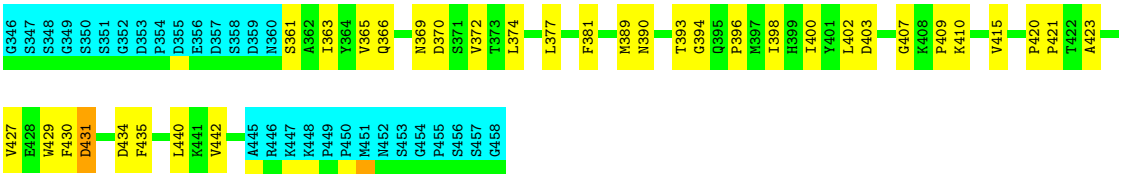
Chain A: 46% 27% • 26%



4.2.12 Score per residue for model 12

- Molecule 1: RNA-binding protein EWS

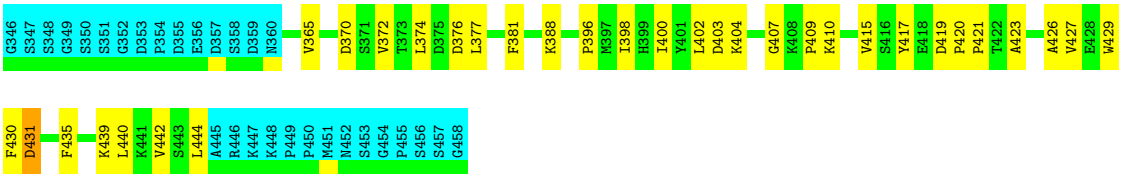
Chain A: 44% 29% • 26%



4.2.13 Score per residue for model 13

- Molecule 1: RNA-binding protein EWS

Chain A: 45% 28% • 26%



4.2.14 Score per residue for model 14 (medoid)

- Molecule 1: RNA-binding protein EWS

Chain A: 47% 26% •• 26%

G346 S347 S348 G349 S350 S351 G352 D353 P354 D355 E356 D357 S358 D359 N360 S361 A362 I363 Y364 V365 N369 D370 L374 F380 F381 K382 Q383 N390 T393 P396 D403 G407 K408 P409 K410 V415 S416 Y417 P420 P421 T422 A423 V427 E428 W429 F430 D431 F435

Q436 L440 K441 V442 S443 L444 A445 R446 K447 K448 P449 P450 M451 N452 S453 G454 P455 S456 S457 G458

4.2.15 Score per residue for model 15

- Molecule 1: RNA-binding protein EWS

Chain A: 49% 23% • 26%

G346 S347 S348 G349 S350 S351 G352 D353 P354 D355 E356 D357 S358 D359 N360 V365 N369 L374 L377 F381 M389 N390 K391 G394 Q395 P396 D403 K404 E405 T406 G407 K408 P409 K410 T414 V415 S416 Y417 P420 P421 T422 A423 A426 V427 E428 W429 F430 D431

D434 F435 L440 K441 V442 A445 R446 K447 K448 P449 P450 M451 N452 S453 G454 P455 S456 S457 G458

4.2.16 Score per residue for model 16

- Molecule 1: RNA-binding protein EWS

Chain A: 48% 26% • 26%

G346 S347 S348 G349 S350 S351 G352 D353 P354 D355 E356 D357 S358 D359 N360 I363 Y364 V365 N369 D370 L374 L377 F381 M389 G394 Q395 P396 I400 D403 G407 K408 P409 K410 T414 V415 D419 P420 P421 T422 A423 V427 E428 W429 F430 D431 G432 K433

D434 F435 Q436 L440 K441 V442 A445 R446 K447 K448 P449 P450 M451 N452 S453 G454 P455 S456 S457 G458

4.2.17 Score per residue for model 17

- Molecule 1: RNA-binding protein EWS

Chain A: 49% 24% • 26%

G346 S347 S348 G349 S350 S351 G352 D353 P354 D355 E356 D357 S358 D359 N360 V365 N369 D370 L374 D375 D376 F381 K382 M389 P396 D403 K404 G407 K408 P409 K410 V415 S416 Y417 E418 D419 P420 P421 T422 A423 K424 V427 E428 W429 F430 D431 F435 L440

K441 V442 S443 L444 A445 R446 K447 K448 P449 P450 M451 N452 S453 G454 P455 S456 S457 G458

4.2.18 Score per residue for model 18

- Molecule 1: RNA-binding protein EWS

Chain A: 48% 26% • 26%

G346 S347 S348 G349 S350 S351 G352 D353 P354 D355 E356 D357 S358 D359 N360 V365 D370 L374 F380 F381 C384 M389 P396 D403 K404 G407 K408 P409 K410 V415 S416 Y417 E418 D419 P420 P421 T422 A423 K424 A425 A426 V427 E428 W429 F430 D431 F435 K439

L440 K441 V442 S443 L444 A445 R446 K447 K448 P449 P450 M451 N452 S453 G454 P455 S456 S457 G458

4.2.19 Score per residue for model 19

- Molecule 1: RNA-binding protein EWS

Chain A: 46% 27% • 26%

G346 S347 S348 G349 S350 S351 G352 D353 P354 D355 E356 D357 S358 D359 N360 I363 Y364 V365 N369 D370 L374 L377 F381 K382 M389 G394 Q395 P396 I400 D403 G407 K408 P409 K410 V415 S416 Y417 E418 D419 P420 P421 T422 A423 K424 A425 A426 V427 E428 W429

F430 D431 F435 L440 K441 V442 S443 L444 A445 R446 K447 K448 P449 P450 M451 N452 S453 G454 P455 S456 S457 G458

4.2.20 Score per residue for model 20

- Molecule 1: RNA-binding protein EWS

Chain A: 48% 25% •• 26%

G346 S347 S348 G349 S350 S351 G352 D353 P354 D355 E356 D357 S358 D359 N360 I363 Y364 V365 D370 L374 F381 M389 M390 T393 G394 Q396 P396 H399 D403 G407 K408 P409 K410 A413 T414 V415 S416 Y417 P420 P421 T422 A423 V427 E428 W429 F430 D431

D434 F435 L440 K441 V442 A445 R446 K447 K448 P449 P450 M451 N452 S453 G454 P455 S456 S457 G458

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function,structures with the lowest energy,structures with the least restraint violations.*

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

Software name	Classification	Version
CYANA	structure solution	2.0.17
CYANA	refinement	2.0.17

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	650	645	645	27±2
All	All	13000	12900	12900	535

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:365:VAL:HG21	1:A:381:PHE:CZ	0.95	1.97	2	20
1:A:381:PHE:CD2	1:A:415:VAL:HG21	0.88	2.04	10	20
1:A:374:LEU:HD11	1:A:396:PRO:HG2	0.77	1.55	7	20
1:A:424:LYS:HG2	1:A:444:LEU:HD11	0.76	1.56	4	5
1:A:430:PHE:HB2	1:A:442:VAL:HG21	0.75	1.57	10	20
1:A:423:ALA:O	1:A:427:VAL:HG23	0.72	1.84	16	18
1:A:377:LEU:HD12	1:A:400:ILE:HD11	0.69	1.64	3	13
1:A:363:ILE:HD12	1:A:442:VAL:CG1	0.67	2.19	16	13
1:A:363:ILE:HD12	1:A:442:VAL:HG11	0.62	1.70	19	8
1:A:381:PHE:CD2	1:A:415:VAL:CG2	0.61	2.83	5	17
1:A:435:PHE:HB2	1:A:440:LEU:HD11	0.61	1.72	19	17
1:A:381:PHE:CE2	1:A:415:VAL:HG21	0.60	2.30	16	17
1:A:420:PRO:N	1:A:421:PRO:CD	0.60	2.64	13	20
1:A:381:PHE:CG	1:A:415:VAL:HG21	0.59	2.33	5	15

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:365:VAL:HG21	1:A:381:PHE:HZ	0.58	1.58	12	15
1:A:381:PHE:CE2	1:A:415:VAL:CG2	0.57	2.88	16	10
1:A:384:CYS:SG	1:A:429:TRP:CE3	0.57	2.97	18	1
1:A:429:TRP:CD1	1:A:429:TRP:C	0.56	2.79	3	20
1:A:417:TYR:CE2	1:A:426:ALA:CB	0.56	2.89	2	4
1:A:417:TYR:CZ	1:A:426:ALA:CB	0.52	2.93	13	3
1:A:427:VAL:HG21	1:A:444:LEU:HG	0.51	1.81	13	2
1:A:363:ILE:HD12	1:A:442:VAL:HG13	0.51	1.82	4	10
1:A:420:PRO:N	1:A:421:PRO:HD2	0.51	2.21	15	20
1:A:369:ASN:ND2	1:A:436:GLN:CD	0.51	2.65	14	2
1:A:389:MET:CE	1:A:394:GLY:O	0.50	2.60	16	11
1:A:389:MET:SD	1:A:396:PRO:N	0.50	2.85	10	6
1:A:380:PHE:CE1	1:A:383:GLN:NE2	0.50	2.79	1	2
1:A:427:VAL:O	1:A:431:ASP:CB	0.50	2.60	10	20
1:A:403:ASP:O	1:A:407:GLY:N	0.49	2.45	17	20
1:A:409:PRO:O	1:A:410:LYS:C	0.49	2.51	17	20
1:A:381:PHE:HA	1:A:430:PHE:CZ	0.49	2.43	15	20
1:A:369:ASN:OD1	1:A:436:GLN:NE2	0.49	2.45	3	2
1:A:420:PRO:CD	1:A:421:PRO:CD	0.48	2.92	5	20
1:A:369:ASN:ND2	1:A:369:ASN:H	0.48	2.06	15	1
1:A:365:VAL:HG21	1:A:381:PHE:CE1	0.47	2.43	5	1
1:A:369:ASN:HD21	1:A:436:GLN:NE2	0.47	2.08	10	1
1:A:417:TYR:CD2	1:A:423:ALA:HA	0.46	2.45	13	13
1:A:389:MET:SD	1:A:396:PRO:CA	0.45	3.04	10	8
1:A:399:HIS:O	1:A:413:ALA:HB1	0.45	2.11	20	1
1:A:390:ASN:ND2	1:A:393:THR:OG1	0.45	2.50	12	6
1:A:398:ILE:HG12	1:A:415:VAL:HG22	0.45	1.87	12	2
1:A:389:MET:HE1	1:A:394:GLY:O	0.45	2.12	1	1
1:A:369:ASN:ND2	1:A:436:GLN:NE2	0.45	2.65	14	1
1:A:386:VAL:CG1	1:A:418:GLU:OE2	0.44	2.66	2	1
1:A:434:ASP:CG	1:A:434:ASP:O	0.44	2.56	2	2
1:A:403:ASP:OD1	1:A:405:GLU:CG	0.44	2.66	15	1
1:A:380:PHE:CE2	1:A:430:PHE:CD1	0.44	3.05	18	2
1:A:443:SER:O	1:A:444:LEU:C	0.44	2.56	11	2
1:A:402:LEU:HD23	1:A:409:PRO:HA	0.44	1.90	13	5
1:A:417:TYR:OH	1:A:426:ALA:HB1	0.43	2.13	13	1
1:A:430:PHE:CB	1:A:442:VAL:HG21	0.43	2.36	10	1
1:A:406:THR:OG1	1:A:408:LYS:CG	0.43	2.66	10	1
1:A:369:ASN:OD1	1:A:372:VAL:N	0.43	2.52	4	3
1:A:434:ASP:O	1:A:434:ASP:CG	0.43	2.57	12	4
1:A:420:PRO:CD	1:A:421:PRO:HD3	0.42	2.44	13	8

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:374:LEU:O	1:A:377:LEU:N	0.42	2.51	5	3
1:A:435:PHE:CB	1:A:440:LEU:HD21	0.42	2.44	18	2
1:A:440:LEU:HD23	1:A:440:LEU:N	0.41	2.30	17	1
1:A:361:SER:O	1:A:417:TYR:N	0.41	2.53	14	2
1:A:420:PRO:HD2	1:A:421:PRO:CD	0.41	2.46	5	15
1:A:372:VAL:HG23	1:A:435:PHE:CE2	0.41	2.50	13	2
1:A:435:PHE:CD2	1:A:440:LEU:HD11	0.41	2.51	12	1
1:A:417:TYR:CE2	1:A:426:ALA:HB3	0.41	2.51	13	1
1:A:419:ASP:CB	1:A:421:PRO:HD2	0.41	2.46	7	6
1:A:435:PHE:HB3	1:A:440:LEU:HD21	0.41	1.93	14	1
1:A:392:ARG:NH1	1:A:392:ARG:CB	0.40	2.84	7	1
1:A:390:ASN:CB	1:A:393:THR:OG1	0.40	2.69	11	1
1:A:377:LEU:HD12	1:A:400:ILE:CD1	0.40	2.43	2	1
1:A:369:ASN:CG	1:A:436:GLN:CD	0.40	2.80	16	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	84/113 (74%)	71±2 (84±2%)	12±2 (14±2%)	1±1 (1±1%)	15	61
All	All	1680/2260 (74%)	1416 (84%)	240 (14%)	24 (1%)	15	61

All 4 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	431	ASP	19
1	A	444	LEU	2
1	A	433	LYS	2
1	A	388	LYS	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	71/94 (76%)	67±1 (95±2%)	4±1 (5±2%)	25	74
All	All	1420/1880 (76%)	1342 (95%)	78 (5%)	25	74

All 19 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	370	ASP	18
1	A	404	LYS	8
1	A	410	LYS	7
1	A	414	THR	6
1	A	382	LYS	6
1	A	431	ASP	6
1	A	376	ASP	5
1	A	439	LYS	4
1	A	391	LYS	3
1	A	369	ASN	3
1	A	434	ASP	2
1	A	441	LYS	2
1	A	395	GLN	2
1	A	429	TRP	1
1	A	375	ASP	1
1	A	408	LYS	1
1	A	392	ARG	1
1	A	361	SER	1
1	A	366	GLN	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