Full wwPDB NMR Structure Validation Report (i)

Feb 24, 2022 – 06:35 AM EST

PDB ID	:	1Z9E
Title	:	Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75
Authors	:	Cherepanov, P.; Sun, Z.-Y.J.; Rahman, S.; Maertens, G.; Wagner, G.; Engelman, A.
Deposited on	:	2005-04-01

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:

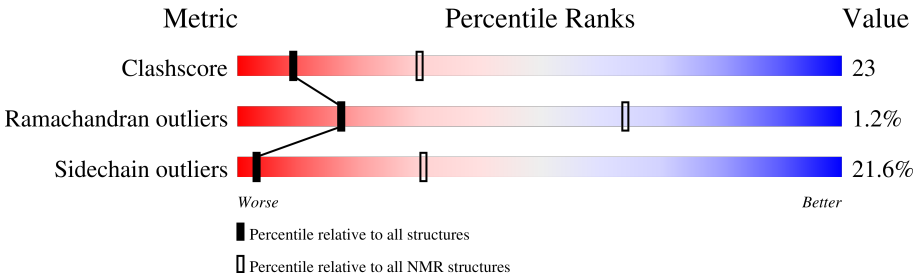
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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	127	33% 28% • • 35%

2 Ensemble composition and analysis (i)

This entry contains 15 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:348-A:428 (81)	0.20	1

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 1 clusters and 5 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1400 atoms, of which 720 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PC4 and SFRS1 interacting protein 2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	83	1400	427	720	122	124	7	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	345	GLY	-	cloning artifact	UNP O75475
A	346	SER	-	cloning artifact	UNP O75475

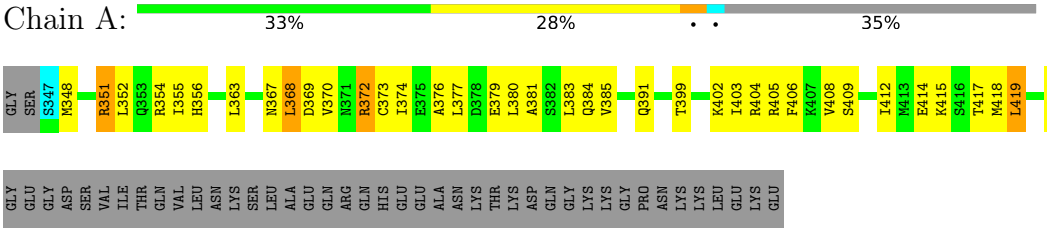
4 Residue-property plots

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: PC4 and SFRS1 interacting protein 2

Chain A:



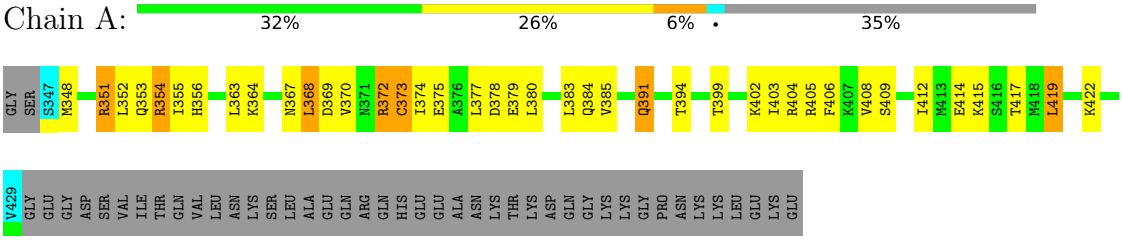
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

• Molecule 1: PC4 and SFRS1 interacting protein 2

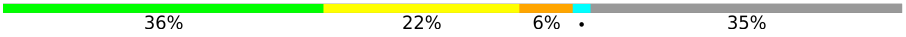
Chain A:

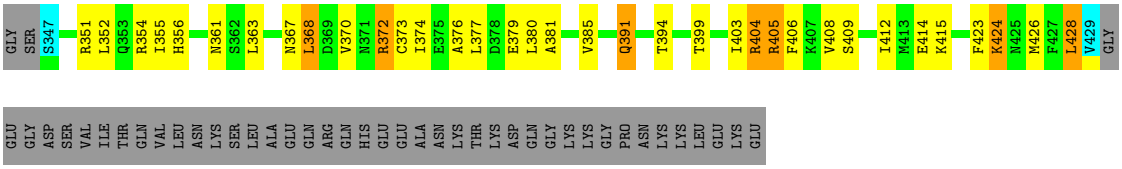


4.2.2 Score per residue for model 2

• Molecule 1: PC4 and SFRS1 interacting protein 2

Chain A:

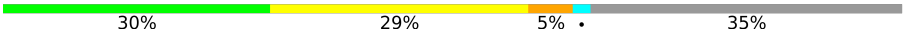


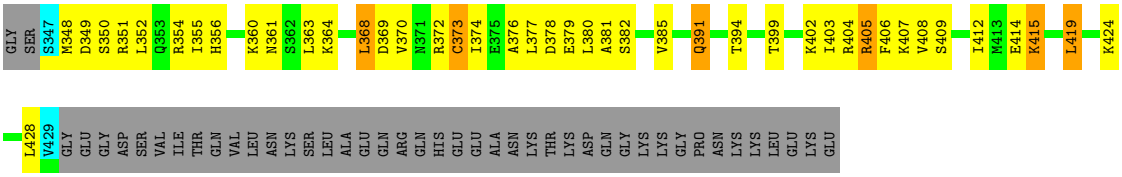


4.2.3 Score per residue for model 3

• Molecule 1: PC4 and SFRS1 interacting protein 2

Chain A:

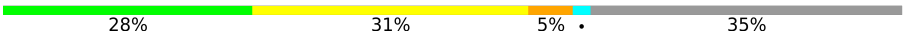


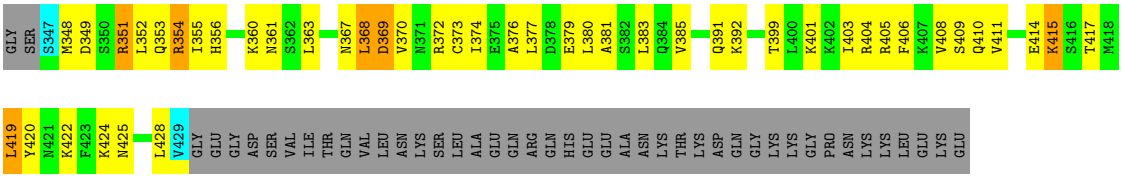


4.2.4 Score per residue for model 4

• Molecule 1: PC4 and SFRS1 interacting protein 2

Chain A:

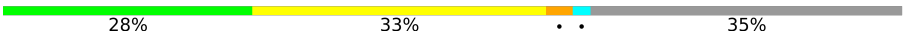


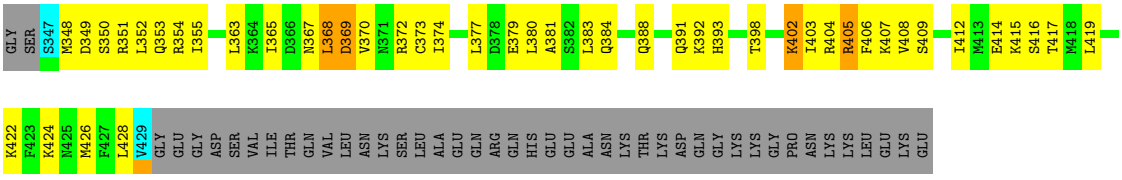


4.2.5 Score per residue for model 5

• Molecule 1: PC4 and SFRS1 interacting protein 2

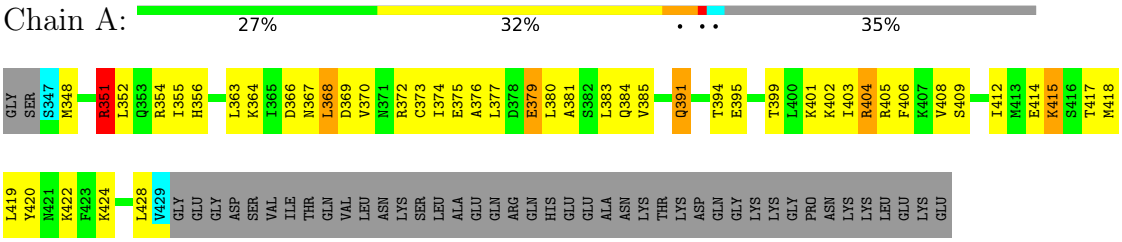
Chain A:





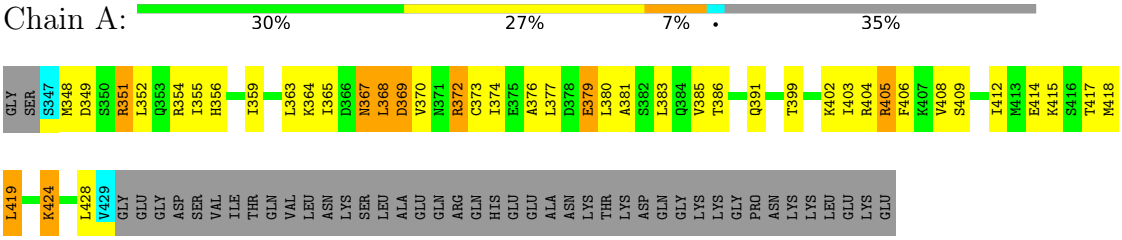
4.2.6 Score per residue for model 6

• Molecule 1: PC4 and SFRS1 interacting protein 2



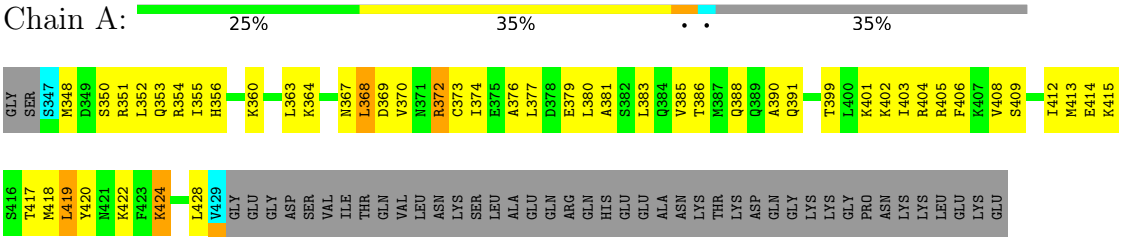
4.2.7 Score per residue for model 7

• Molecule 1: PC4 and SFRS1 interacting protein 2



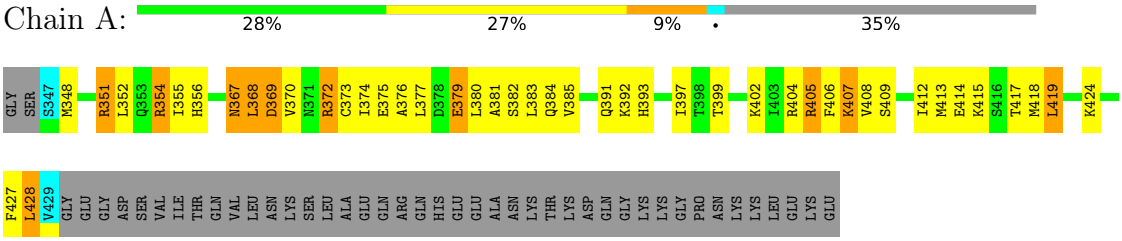
4.2.8 Score per residue for model 8

• Molecule 1: PC4 and SFRS1 interacting protein 2



4.2.9 Score per residue for model 9

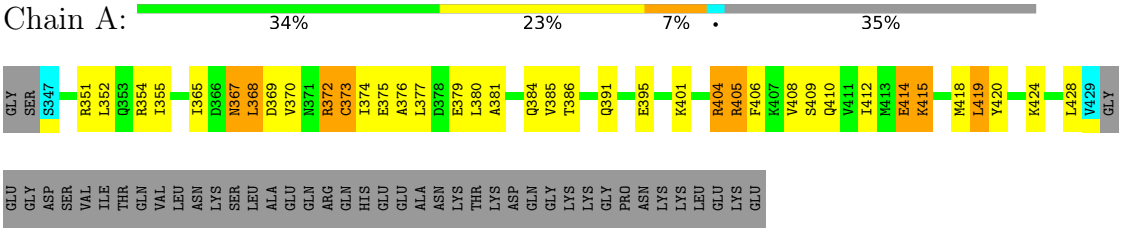
• Molecule 1: PC4 and SFRS1 interacting protein 2



4.2.10 Score per residue for model 10

• Molecule 1: PC4 and SFRS1 interacting protein 2

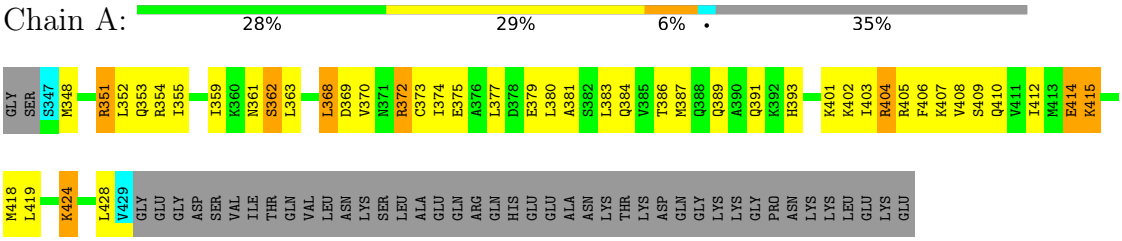
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: PC4 and SFRS1 interacting protein 2

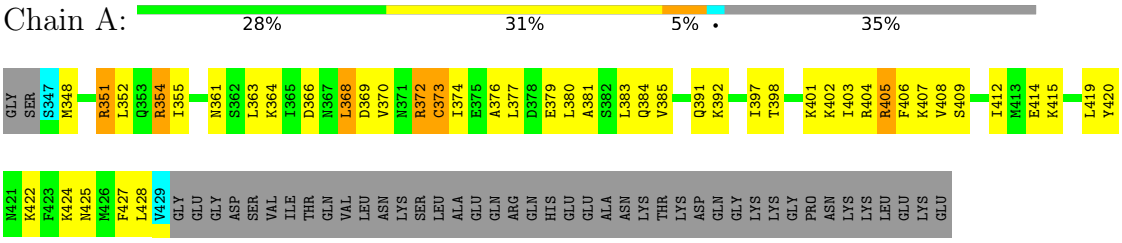
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: PC4 and SFRS1 interacting protein 2

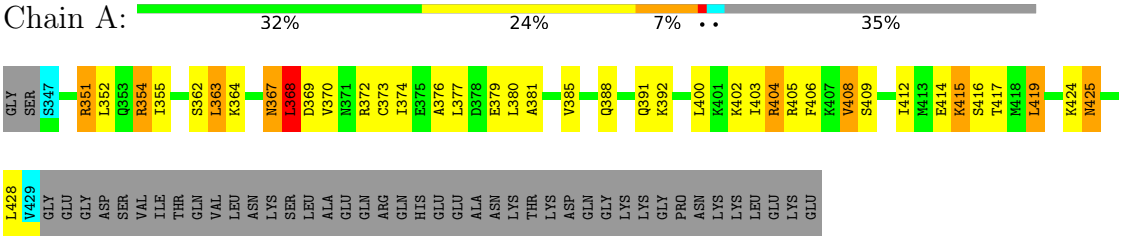
Chain A:



4.2.13 Score per residue for model 13

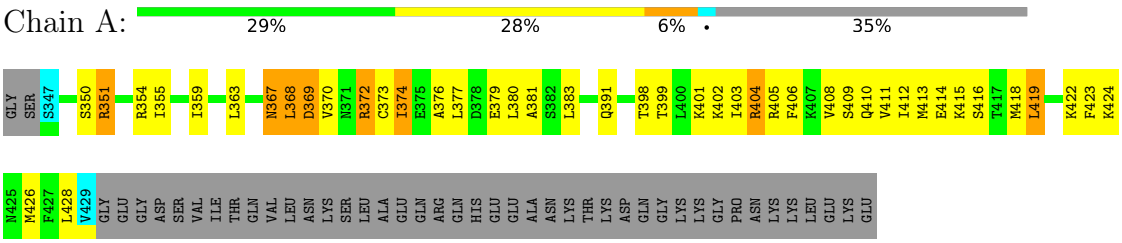
• Molecule 1: PC4 and SFRS1 interacting protein 2

Chain A:



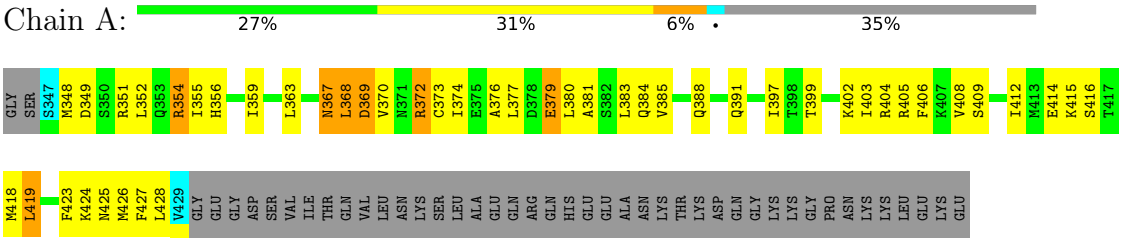
4.2.14 Score per residue for model 14

• Molecule 1: PC4 and SFRS1 interacting protein 2



4.2.15 Score per residue for model 15

• Molecule 1: PC4 and SFRS1 interacting protein 2



5 Refinement protocol and experimental data overview

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

Of the 20 calculated structures, 15 were deposited, based on the following criterion: *structures with the least restraint violations,structures with the lowest energy*.

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

Software name	Classification	Version
X-PLOR	structure solution	3.851
X-PLOR	refinement	3.851

No chemical shift data was provided.

6 Model quality

6.1 Standard geometry

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	5.0±0.0
All	All	0	75

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	351	ARG	Sidechain	15
1	A	354	ARG	Sidechain	15
1	A	372	ARG	Sidechain	15
1	A	404	ARG	Sidechain	15
1	A	405	ARG	Sidechain	15

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	667	707	707	31±4
All	All	10005	10605	10605	470

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 23.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:380:LEU:HD23	1:A:419:LEU:HD22	0.99	1.30	7	12
1:A:352:LEU:HD11	1:A:385:VAL:HG11	0.91	1.42	6	9
1:A:370:VAL:HG13	1:A:412:ILE:HD11	0.89	1.41	11	12
1:A:377:LEU:HD13	1:A:415:LYS:CB	0.85	2.01	10	15
1:A:380:LEU:CD2	1:A:419:LEU:HD22	0.84	2.02	10	5
1:A:355:ILE:HD13	1:A:380:LEU:CD1	0.84	2.03	14	13
1:A:355:ILE:HD13	1:A:380:LEU:HD13	0.82	1.51	14	13
1:A:370:VAL:HG12	1:A:374:ILE:CD1	0.80	2.07	2	14
1:A:370:VAL:HG12	1:A:374:ILE:HD11	0.79	1.54	2	11
1:A:377:LEU:HD13	1:A:415:LYS:HB2	0.78	1.56	9	15
1:A:380:LEU:HD23	1:A:419:LEU:CD2	0.77	2.09	6	7
1:A:377:LEU:HD13	1:A:415:LYS:HB3	0.74	1.57	3	11
1:A:380:LEU:HD23	1:A:419:LEU:HD23	0.73	1.58	6	1
1:A:365:ILE:HA	1:A:368:LEU:HD13	0.71	1.62	7	3
1:A:398:THR:HG22	1:A:402:LYS:HE3	0.68	1.63	12	1
1:A:355:ILE:HD13	1:A:380:LEU:HD12	0.68	1.63	7	4
1:A:363:LEU:HD11	1:A:403:ILE:HG23	0.67	1.64	14	2
1:A:370:VAL:HG13	1:A:412:ILE:CD1	0.67	2.18	11	7
1:A:397:ILE:HD11	1:A:427:PHE:CE2	0.66	2.25	9	2
1:A:370:VAL:HG12	1:A:374:ILE:HD12	0.66	1.68	14	3
1:A:400:LEU:HA	1:A:403:ILE:HD12	0.66	1.67	13	1
1:A:351:ARG:HB3	1:A:383:LEU:HD11	0.66	1.68	7	5
1:A:352:LEU:HD21	1:A:385:VAL:HG21	0.65	1.65	10	5
1:A:363:LEU:CD1	1:A:403:ILE:HG23	0.64	2.22	14	1
1:A:348:MET:SD	1:A:383:LEU:HD22	0.64	2.32	11	4
1:A:356:HIS:CD2	1:A:399:THR:HG21	0.63	2.28	6	9
1:A:359:ILE:CG2	1:A:403:ILE:HD11	0.62	2.24	14	1
1:A:363:LEU:HD13	1:A:403:ILE:HD12	0.61	1.71	12	3
1:A:406:PHE:CZ	1:A:408:VAL:CG1	0.60	2.85	8	15
1:A:365:ILE:CA	1:A:368:LEU:HD13	0.59	2.26	7	3
1:A:355:ILE:HD11	1:A:379:GLU:HG3	0.59	1.74	6	13
1:A:391:GLN:O	1:A:394:THR:HG23	0.59	1.96	2	4
1:A:355:ILE:HG23	1:A:376:ALA:HB1	0.58	1.74	15	6
1:A:398:THR:HG22	1:A:402:LYS:HD3	0.57	1.76	5	1
1:A:352:LEU:HD11	1:A:385:VAL:CG1	0.56	2.26	7	7
1:A:385:VAL:CG1	1:A:390:ALA:HB2	0.56	2.31	8	1
1:A:406:PHE:CE2	1:A:412:ILE:HG21	0.55	2.37	11	2
1:A:356:HIS:NE2	1:A:399:THR:HG21	0.55	2.17	8	9
1:A:373:CYS:O	1:A:376:ALA:HB3	0.55	2.02	4	11
1:A:380:LEU:HG	1:A:419:LEU:HD22	0.55	1.79	15	1
1:A:410:GLN:HG3	1:A:411:VAL:HG23	0.55	1.76	14	2
1:A:359:ILE:HG21	1:A:403:ILE:HD11	0.55	1.79	14	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:380:LEU:CG	1:A:419:LEU:HD22	0.55	2.31	15	1
1:A:355:ILE:CD1	1:A:383:LEU:HD12	0.54	2.32	5	1
1:A:406:PHE:CD2	1:A:412:ILE:HG21	0.54	2.38	12	4
1:A:414:GLU:HG3	1:A:415:LYS:N	0.54	2.18	10	2
1:A:381:ALA:HA	1:A:419:LEU:HD21	0.53	1.79	15	5
1:A:363:LEU:HD11	1:A:403:ILE:HD13	0.53	1.81	5	7
1:A:355:ILE:HG22	1:A:359:ILE:HD12	0.53	1.80	15	3
1:A:370:VAL:CG1	1:A:412:ILE:HD11	0.52	2.27	11	2
1:A:373:CYS:O	1:A:377:LEU:HG	0.52	2.05	3	15
1:A:355:ILE:HD11	1:A:383:LEU:HD12	0.52	1.80	4	5
1:A:362:SER:O	1:A:368:LEU:HD12	0.52	2.05	11	1
1:A:370:VAL:CG1	1:A:374:ILE:HD11	0.51	2.31	2	4
1:A:352:LEU:HD12	1:A:393:HIS:CD2	0.51	2.40	9	3
1:A:363:LEU:CD1	1:A:403:ILE:HD12	0.50	2.36	4	3
1:A:363:LEU:HD11	1:A:403:ILE:HG12	0.50	1.83	13	1
1:A:408:VAL:CG1	1:A:409:SER:N	0.50	2.75	14	15
1:A:377:LEU:O	1:A:381:ALA:HB2	0.49	2.07	14	7
1:A:363:LEU:HD11	1:A:403:ILE:CG1	0.49	2.37	13	1
1:A:365:ILE:N	1:A:368:LEU:HD13	0.49	2.22	10	1
1:A:363:LEU:HD12	1:A:403:ILE:HG12	0.49	1.83	14	1
1:A:424:LYS:O	1:A:428:LEU:N	0.49	2.46	14	14
1:A:381:ALA:N	1:A:419:LEU:HD21	0.48	2.23	11	4
1:A:406:PHE:CZ	1:A:408:VAL:HG11	0.47	2.43	10	15
1:A:363:LEU:HD11	1:A:403:ILE:CG2	0.47	2.39	14	2
1:A:380:LEU:CD2	1:A:419:LEU:HD23	0.47	2.36	6	1
1:A:381:ALA:CA	1:A:419:LEU:HD21	0.47	2.40	15	2
1:A:355:ILE:HD13	1:A:380:LEU:HB2	0.47	1.87	15	1
1:A:363:LEU:CD1	1:A:403:ILE:HG12	0.46	2.40	13	2
1:A:352:LEU:CD1	1:A:385:VAL:HG11	0.46	2.30	6	3
1:A:352:LEU:CD1	1:A:393:HIS:CD2	0.46	2.98	11	2
1:A:401:LYS:CG	1:A:420:TYR:CE1	0.45	2.99	12	2
1:A:351:ARG:CD	1:A:383:LEU:HD21	0.45	2.41	4	1
1:A:367:ASN:O	1:A:369:ASP:N	0.45	2.50	7	9
1:A:401:LYS:CG	1:A:420:TYR:CZ	0.45	3.00	10	3
1:A:363:LEU:CD2	1:A:363:LEU:N	0.44	2.80	13	1
1:A:380:LEU:HG	1:A:419:LEU:HD23	0.43	1.90	4	1
1:A:423:PHE:O	1:A:426:MET:CG	0.43	2.66	2	3
1:A:376:ALA:HA	1:A:379:GLU:HG2	0.43	1.91	2	6
1:A:351:ARG:HG2	1:A:383:LEU:HD21	0.43	1.90	6	1
1:A:356:HIS:CE1	1:A:399:THR:HG1	0.43	2.31	6	2
1:A:407:LYS:HG2	1:A:413:MET:HE1	0.43	1.91	9	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:428:LEU:O	1:A:428:LEU:HD23	0.43	2.14	12	1
1:A:352:LEU:CD2	1:A:380:LEU:CD1	0.43	2.96	8	9
1:A:397:ILE:CD1	1:A:427:PHE:CE2	0.43	3.00	9	3
1:A:356:HIS:NE2	1:A:399:THR:CG2	0.42	2.82	6	6
1:A:351:ARG:HD2	1:A:383:LEU:HD21	0.42	1.90	4	1
1:A:397:ILE:HD12	1:A:427:PHE:CE2	0.42	2.49	12	1
1:A:363:LEU:CD1	1:A:403:ILE:CD1	0.42	2.98	11	5
1:A:401:LYS:HG3	1:A:420:TYR:CE1	0.42	2.50	12	1
1:A:363:LEU:N	1:A:363:LEU:HD23	0.41	2.30	13	1
1:A:351:ARG:HD3	1:A:383:LEU:HD21	0.41	1.92	12	1
1:A:363:LEU:HD22	1:A:406:PHE:CG	0.41	2.50	14	1
1:A:362:SER:O	1:A:368:LEU:HD13	0.41	2.14	13	1
1:A:370:VAL:HG22	1:A:412:ILE:HD11	0.41	1.92	13	1
1:A:352:LEU:HD23	1:A:380:LEU:HD12	0.41	1.92	6	2
1:A:399:THR:O	1:A:403:ILE:HG13	0.41	2.15	14	1
1:A:363:LEU:HD22	1:A:406:PHE:CE1	0.41	2.51	7	1
1:A:398:THR:O	1:A:401:LYS:CG	0.41	2.69	14	1
1:A:370:VAL:HG13	1:A:412:ILE:CG1	0.41	2.46	2	1
1:A:379:GLU:CA	1:A:379:GLU:OE1	0.41	2.69	6	3
1:A:363:LEU:CD1	1:A:403:ILE:HD13	0.40	2.46	7	1
1:A:406:PHE:CE2	1:A:408:VAL:HG12	0.40	2.51	13	1
1:A:355:ILE:CD1	1:A:383:LEU:CD1	0.40	2.99	8	1
1:A:363:LEU:HD12	1:A:403:ILE:CD1	0.40	2.46	7	1
1:A:373:CYS:O	1:A:376:ALA:N	0.40	2.54	8	1
1:A:348:MET:CE	1:A:383:LEU:CD2	0.40	2.99	15	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	81/127 (64%)	73±1 (90±2%)	7±1 (9±1%)	1±0 (1±0%)	17	64
All	All	1215/1905 (64%)	1096 (90%)	105 (9%)	14 (1%)	17	64

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	368	LEU	14

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	77/116 (66%)	60±3 (78±3%)	17±3 (22±3%)	3 31
All	All	1155/1740 (66%)	906 (78%)	249 (22%)	3 31

All 49 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	391	GLN	15
1	A	414	GLU	15
1	A	368	LEU	12
1	A	369	ASP	11
1	A	402	LYS	11
1	A	372	ARG	10
1	A	419	LEU	10
1	A	367	ASN	9
1	A	384	GLN	8
1	A	417	THR	8
1	A	418	MET	8
1	A	364	LYS	7
1	A	422	LYS	7
1	A	405	ARG	7
1	A	348	MET	6
1	A	354	ARG	6
1	A	404	ARG	6
1	A	415	LYS	6
1	A	353	GLN	5
1	A	375	GLU	5
1	A	361	ASN	5
1	A	349	ASP	5
1	A	407	LYS	5
1	A	392	LYS	5
1	A	373	CYS	4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	424	LYS	4
1	A	350	SER	4
1	A	388	GLN	4
1	A	416	SER	4
1	A	379	GLU	4
1	A	386	THR	4
1	A	360	LYS	3
1	A	378	ASP	2
1	A	428	LEU	2
1	A	382	SER	2
1	A	351	ARG	2
1	A	366	ASP	2
1	A	395	GLU	2
1	A	413	MET	2
1	A	410	GLN	2
1	A	425	ASN	2
1	A	426	MET	1
1	A	362	SER	1
1	A	387	MET	1
1	A	389	GLN	1
1	A	401	LYS	1
1	A	363	LEU	1
1	A	408	VAL	1
1	A	374	ILE	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

No chemical shift data were provided