Full wwPDB NMR Structure Validation Report (i)

May 31, 2020 – 11:26 am BST

PDB ID	:	6MPP
Title	:	MHC-I complex with NRAS Q16K mutation determined by NMR
Authors	:	Flores-Solis, D.; McShan, A.C.; Sgourakis, N.G.
Deposited on	:	2018-10-08

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 (1)) were used in the production of this report:

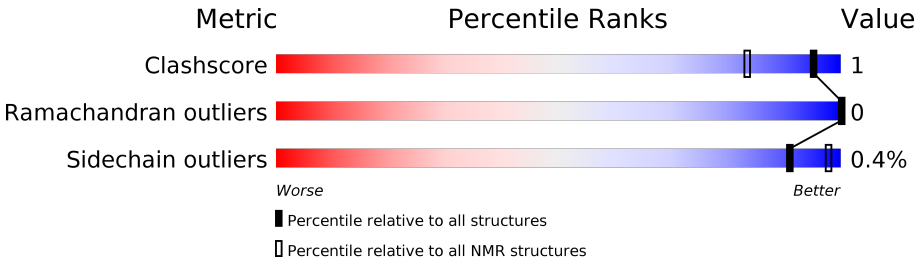
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 is 28%.

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	279	97%
2	B	10	50% 10% 40%
3	C	100	99%

2 Ensemble composition and analysis

This entry contains 10 models. Model 5 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:1-A:274, B:1-B:3, B:8-B:10, C:1-C:100 (380)	0.05	5

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. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

3 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 6110 atoms, of which 2966 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called HLA class I histocompatibility antigen, A-1 alpha chain.

Mol	Chain	Residues	Atoms						Trace
1	A	274	Total	C	H	N	O	S	0
			4313	1383	2086	408	426	10	

- Molecule 2 is a protein called NRAS Q16K peptide.

Mol	Chain	Residues	Atoms					Trace
2	B	10	Total	C	H	N	O	0
			157	50	77	11	19	

- Molecule 3 is a protein called Beta-2-microglobulin.

Mol	Chain	Residues	Atoms						Trace
3	C	100	Total	C	H	N	O	S	0
			1640	533	803	141	159	4	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	MET	-	initiating methionine	UNP P61769

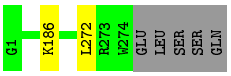
4 Residue-property plots

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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.

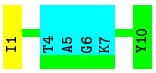
• Molecule 1: HLA class I histocompatibility antigen, A-1 alpha chain

Chain A:



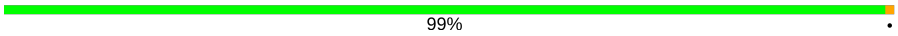
• Molecule 2: NRAS Q16K peptide

Chain B:



• Molecule 3: Beta-2-microglobulin

Chain C:



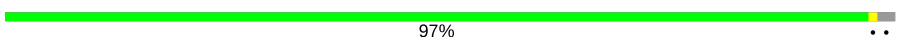
4.2 Scores per residue for each member of the ensemble

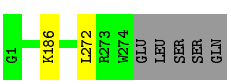
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: HLA class I histocompatibility antigen, A-1 alpha chain

Chain A:



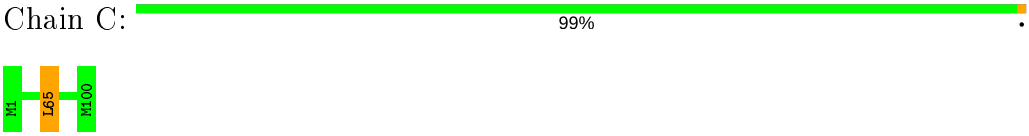


• Molecule 2: NRAS Q16K peptide

Chain B:

• Molecule 3: Beta-2-microglobulin

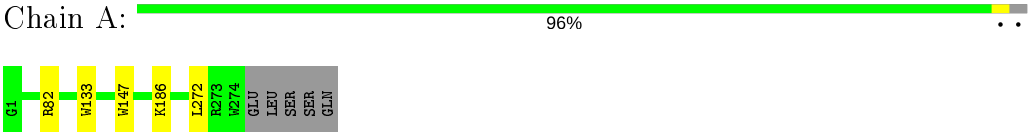
Chain C:



4.2.2 Score per residue for model 2

• Molecule 1: HLA class I histocompatibility antigen, A-1 alpha chain

Chain A:

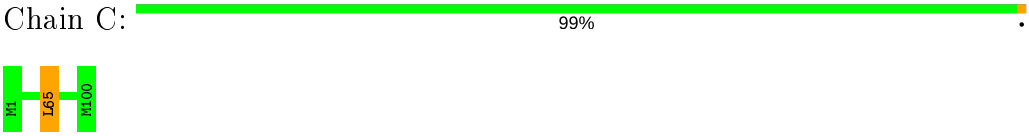


• Molecule 2: NRAS Q16K peptide

Chain B:

• Molecule 3: Beta-2-microglobulin

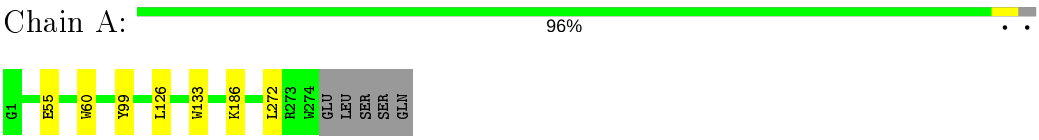
Chain C:



4.2.3 Score per residue for model 3

• Molecule 1: HLA class I histocompatibility antigen, A-1 alpha chain

Chain A:

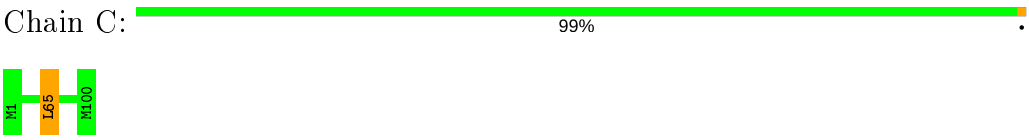


• Molecule 2: NRAS Q16K peptide

Chain B:

• Molecule 3: Beta-2-microglobulin

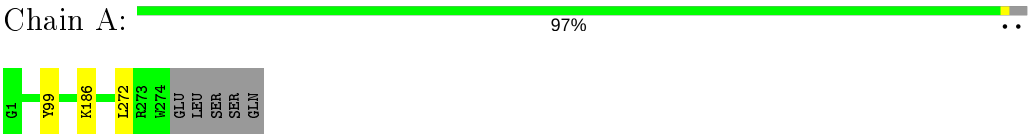
Chain C:



4.2.4 Score per residue for model 4

• Molecule 1: HLA class I histocompatibility antigen, A-1 alpha chain

Chain A:

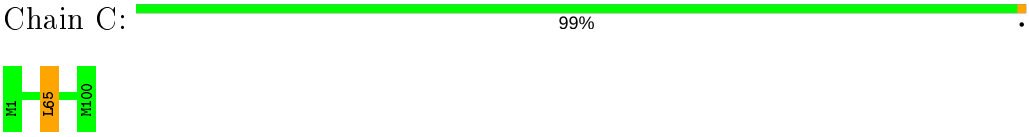


• Molecule 2: NRAS Q16K peptide

Chain B:

• Molecule 3: Beta-2-microglobulin

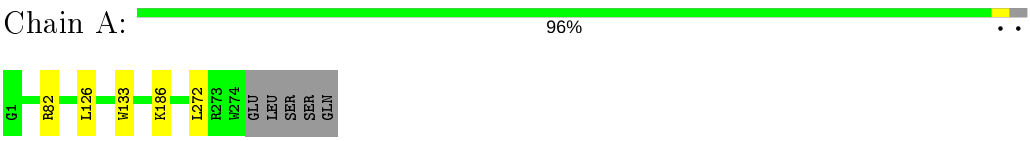
Chain C:



4.2.5 Score per residue for model 5 (medoid)

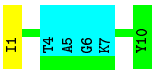
• Molecule 1: HLA class I histocompatibility antigen, A-1 alpha chain

Chain A:



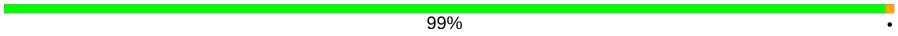
• Molecule 2: NRAS Q16K peptide

Chain B:



• Molecule 3: Beta-2-microglobulin

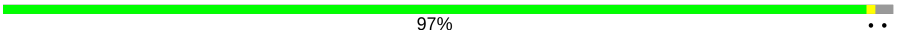
Chain C:

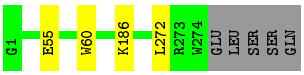


4.2.6 Score per residue for model 6

• Molecule 1: HLA class I histocompatibility antigen, A-1 alpha chain

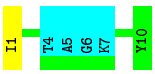
Chain A:





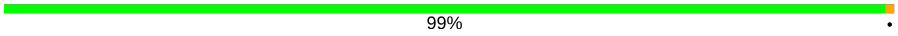
• Molecule 2: NRAS Q16K peptide

Chain B:



• Molecule 3: Beta-2-microglobulin

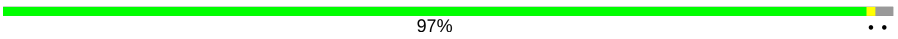
Chain C:

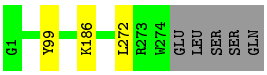


4.2.7 Score per residue for model 7

• Molecule 1: HLA class I histocompatibility antigen, A-1 alpha chain

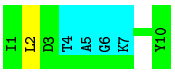
Chain A:





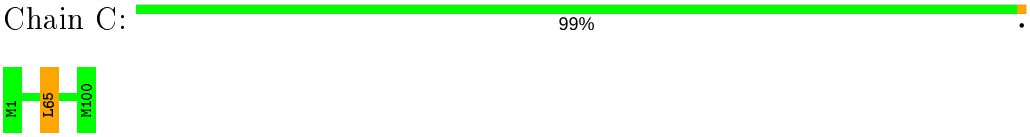
• Molecule 2: NRAS Q16K peptide

Chain B:



• Molecule 3: Beta-2-microglobulin

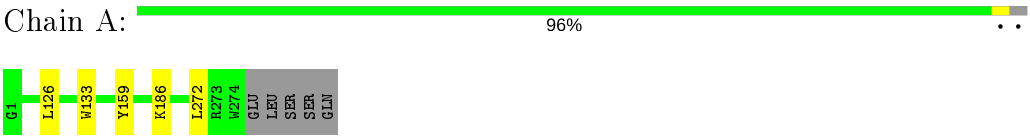
Chain C:



4.2.8 Score per residue for model 8

• Molecule 1: HLA class I histocompatibility antigen, A-1 alpha chain

Chain A:

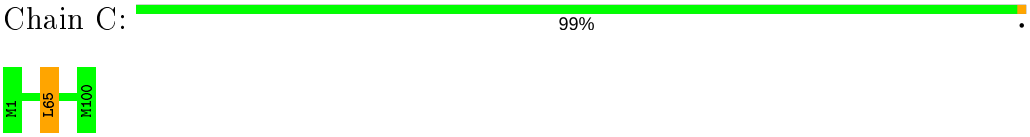


• Molecule 2: NRAS Q16K peptide

Chain B:

• Molecule 3: Beta-2-microglobulin

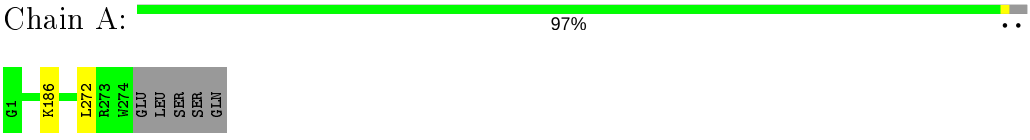
Chain C:



4.2.9 Score per residue for model 9

• Molecule 1: HLA class I histocompatibility antigen, A-1 alpha chain

Chain A:

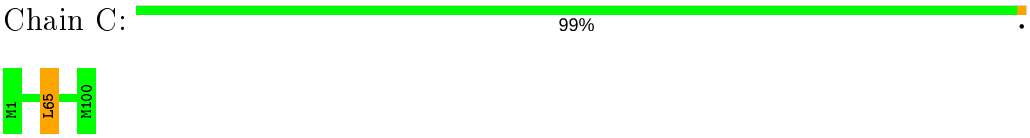


• Molecule 2: NRAS Q16K peptide

Chain B:

• Molecule 3: Beta-2-microglobulin

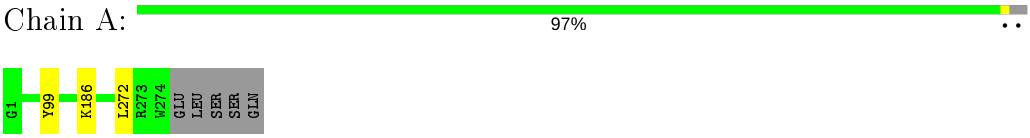
Chain C:



4.2.10 Score per residue for model 10

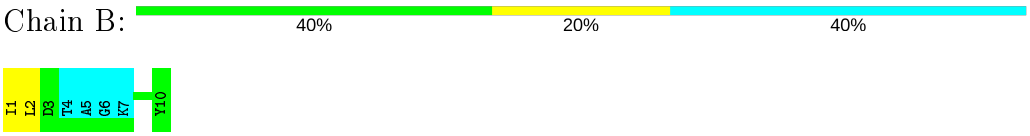
- Molecule 1: HLA class I histocompatibility antigen, A-1 alpha chain

Chain A:



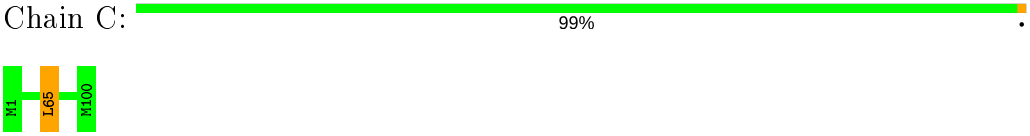
- Molecule 2: NRAS Q16K peptide

Chain B:



- Molecule 3: Beta-2-microglobulin

Chain C:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *distance geometry*.

Of the 50000 calculated structures, 10 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
CS-ROSETTA	refinement	
CS-ROSETTA	structure calculation	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1603
Number of shifts mapped to atoms	1603
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	28%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality

6.1 Standard geometry

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

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	2227	2086	2086	3±1
2	B	55	49	49	1±1
3	C	837	803	803	2±0
All	All	31190	29380	29380	62

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:1:ILE:HG23	2:B:1:ILE:O	0.70	1.85	9	1
2:B:1:ILE:HG13	2:B:1:ILE:O	0.62	1.92	10	1
2:B:1:ILE:O	2:B:1:ILE:HG22	0.57	1.99	6	1
1:A:159:TYR:OH	2:B:1:ILE:O	0.55	2.21	8	1
1:A:186:LYS:HD2	1:A:186:LYS:N	0.54	2.17	4	6
1:A:186:LYS:N	1:A:186:LYS:HD2	0.53	2.17	1	4
2:B:1:ILE:HG22	2:B:1:ILE:O	0.53	2.03	5	1
2:B:1:ILE:CG2	2:B:1:ILE:O	0.53	2.57	9	1
2:B:2:LEU:C	2:B:2:LEU:HD12	0.49	2.27	10	4
3:C:65:LEU:N	3:C:65:LEU:HD23	0.49	2.23	4	6
3:C:65:LEU:HD23	3:C:65:LEU:N	0.49	2.23	7	4
1:A:133:TRP:CZ2	1:A:147:TRP:CE3	0.47	3.03	2	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:272:LEU:HD12	1:A:272:LEU:N	0.47	2.25	4	3
1:A:272:LEU:N	1:A:272:LEU:HD12	0.47	2.25	7	7
2:B:2:LEU:O	2:B:2:LEU:HD12	0.47	2.10	10	1
1:A:55:GLU:OE2	1:A:60:TRP:CD1	0.46	2.69	6	2
3:C:65:LEU:N	3:C:65:LEU:CD2	0.45	2.79	1	4
1:A:99:TYR:C	1:A:99:TYR:CD1	0.45	2.90	10	3
3:C:65:LEU:CD2	3:C:65:LEU:N	0.45	2.79	5	6
1:A:99:TYR:CD1	1:A:99:TYR:C	0.45	2.90	4	1
1:A:126:LEU:HB2	1:A:133:TRP:CZ3	0.43	2.48	8	3
2:B:1:ILE:O	2:B:1:ILE:CG2	0.41	2.68	6	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	272/279 (97%)	263±0 (97±0%)	9±0 (3±0%)	0±0 (0±0%)	100	100
2	B	4/10 (40%)	4±0 (92±11%)	0±0 (8±11%)	0±0 (0±0%)	100	100
3	C	98/100 (98%)	94±0 (96±0%)	4±0 (4±0%)	0±0 (0±0%)	100	100
All	All	3740/3890 (96%)	3607 (96%)	133 (4%)	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	231/236 (98%)	231±0 (100±0%)	0±0 (0±0%)	93	98
2	B	6/8 (75%)	6±0 (100±0%)	0±0 (0±0%)	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
3	C	95/95 (100%)	94±0 (99±0%)	1±0 (1±0%)	74 96
All	All	3320/3390 (98%)	3308 (100%)	12 (0%)	91 98

All 2 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)
3	C	65	LEU	10
1	A	82	ARG	2

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 carbohydrates 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

The completeness of assignment taking into account all chemical shift lists is 28% for the well-defined parts and 28% for the entire structure.

7.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *starfileV3.txt*

7.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	1603
Number of shifts mapped to atoms	1603
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	13

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	251	0.28 ± 0.12	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	228	-0.22 ± 0.19	None needed (< 0.5 ppm)
$^{13}C'$	241	-0.01 ± 0.12	None needed (< 0.5 ppm)
^{15}N	251	-0.18 ± 0.23	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 28%, i.e. 1362 atoms were assigned a chemical shift out of a possible 4820. 29 out of 44 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	969/1862 (52%)	244/741 (33%)	483/760 (64%)	242/361 (67%)
Sidechain	387/2463 (16%)	98/1455 (7%)	289/871 (33%)	0/137 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	6/495 (1%)	5/261 (2%)	0/210 (0%)	1/24 (4%)
Overall	1362/4820 (28%)	347/2457 (14%)	772/1841 (42%)	243/522 (47%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 28%, i.e. 1384 atoms were assigned a chemical shift out of a possible 4859. 29 out of 44 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	981/1882 (52%)	252/749 (34%)	483/768 (63%)	246/365 (67%)
Sidechain	397/2482 (16%)	108/1466 (7%)	289/878 (33%)	0/138 (0%)
Aromatic	6/495 (1%)	5/261 (2%)	0/210 (0%)	1/24 (4%)
Overall	1384/4859 (28%)	365/2476 (15%)	772/1856 (42%)	247/527 (47%)

7.1.4 Statistically unusual chemical shifts (i)

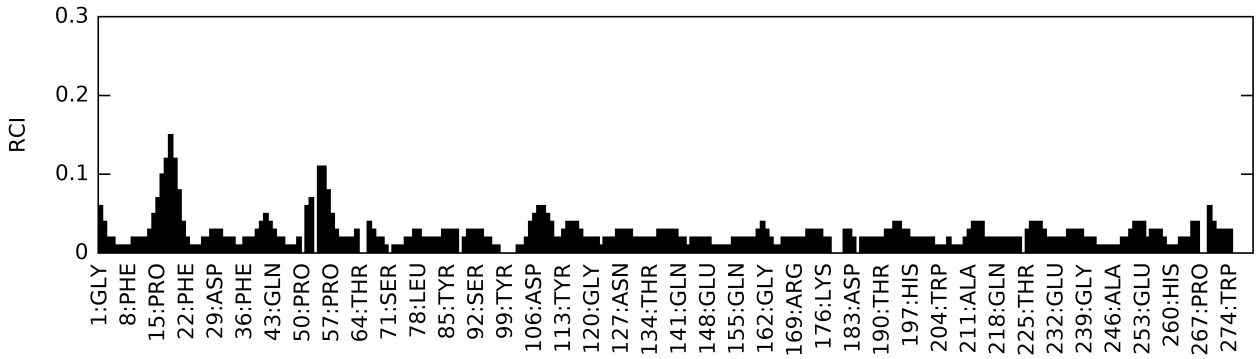
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	117	ALA	HB3	-0.86	2.61 – 0.11	-8.9
1	A	117	ALA	HB2	-0.86	2.61 – 0.11	-8.9
1	A	117	ALA	HB1	-0.86	2.61 – 0.11	-8.9
1	A	47	PRO	CA	52.62	71.13 – 55.53	-6.9
1	A	120	GLY	C	185.64	183.33 – 164.53	6.2
1	A	132	SER	C	185.27	183.48 – 165.88	6.0
1	A	112	GLY	C	184.36	183.33 – 164.53	5.5
1	A	100	GLY	C	184.09	183.33 – 164.53	5.4
1	A	211	ALA	HB3	0.04	2.61 – 0.11	-5.3
1	A	211	ALA	HB2	0.04	2.61 – 0.11	-5.3
1	A	211	ALA	HB1	0.04	2.61 – 0.11	-5.3
1	A	188	HIS	C	185.51	185.27 – 165.27	5.1
1	A	101	CYS	C	185.18	185.15 – 164.55	5.0

7.1.5 Random Coil Index (RCI) plots (i)

The images below report *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A:



Random coil index (RCI) for chain B: