Full wwPDB NMR Structure Validation Report (i)

Apr 27, 2016 – 12:59 AM BST

PDB ID	:	2L4D
Title	:	cytochrome c domain of pp3183 protein from Pseudomonas putida
Authors	:	Banci, L.; Bertini, I.; Ciofi-Baffoni, S.; Kozyreva, T.; Mori, M.; Wang, S.
Deposited on	:	2010-10-04

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
http://wwpdb.org/validation/2016/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:

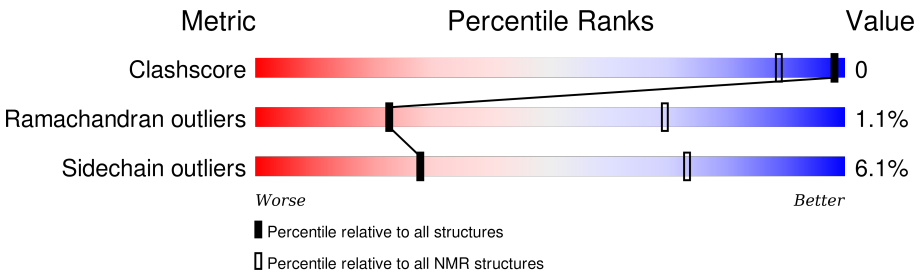
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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 36%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	110	75% · 21% ·

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 8 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:2-A:17, A:28-A:94 (83)	0.55	8

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 2 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1733 atoms, of which 858 are hydrogens and 0 are deuteriums.

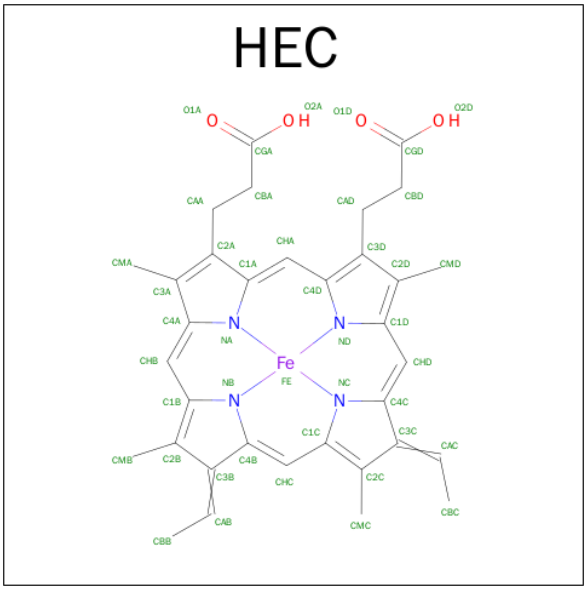
- Molecule 1 is a protein called SCO1/SenC family protein/cytochrome c.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	106	1658	518	826	150	158	6	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	EXPRESSION TAG	UNP Q88I19
A	-2	SER	-	EXPRESSION TAG	UNP Q88I19
A	-1	PHE	-	EXPRESSION TAG	UNP Q88I19
A	0	THR	-	EXPRESSION TAG	UNP Q88I19

- Molecule 2 is HEME C (three-letter code: HEC) (formula: $C_{34}H_{34}FeN_4O_4$).



Mol	Chain	Residues	Atoms					
			Total	C	Fe	H	N	O
2	A	1	75	34	1	32	4	4

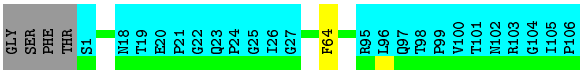
4 Residue-property plots (i)

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: SCO1/SenC family protein/cytochrome c

Chain A: 75% • 21% •



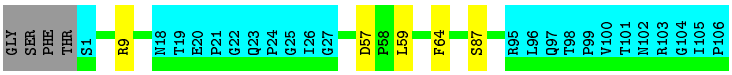
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: SCO1/SenC family protein/cytochrome c

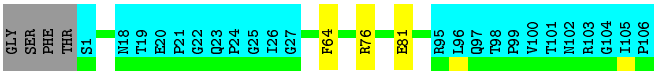
Chain A: 71% 5% 21% •



4.2.2 Score per residue for model 2

- Molecule 1: SCO1/SenC family protein/cytochrome c

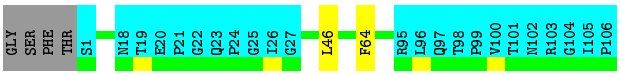
Chain A: 73% • 21% •



4.2.3 Score per residue for model 3

- Molecule 1: SCO1/SenC family protein/cytochrome c

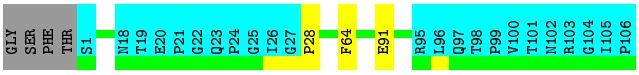
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: SCO1/SenC family protein/cytochrome c

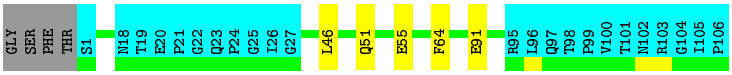
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: SCO1/SenC family protein/cytochrome c

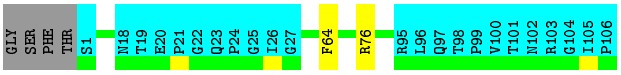
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: SCO1/SenC family protein/cytochrome c

Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: SCO1/SenC family protein/cytochrome c

Chain A:

4.2.8 Score per residue for model 8 (medoid)

- Molecule 1: SCO1/SenC family protein/cytochrome c

Chain A: 70% 5% 21% •

GLY SER PHE THR S1 N18 T19 E20 P21 G22 Q23 P24 G25 I26 G27 P28 D29 L63 F64 L70 R76 L85 R95 L96 Q97 T98 P99 V100 T101 N102 R103 G104 I105 P106

4.2.9 Score per residue for model 9

- Molecule 1: SCO1/SenC family protein/cytochrome c

Chain A: 69% 6% 21% •

GLY SER PHE THR S1 N18 T19 E20 P21 G22 Q23 P24 G25 I26 G27 P28 D29 L46 F64 M75 E81 L85 I86 R95 L96 Q97 T98 P99 V100 T101 N102 R103 G104 I105 P106

4.2.10 Score per residue for model 10

- Molecule 1: SCO1/SenC family protein/cytochrome c

Chain A: 74% • 21% •

GLY SER PHE THR S1 N18 T19 E20 P21 G22 Q23 P24 G25 I26 G27 L59 F64 R95 L96 Q97 T98 P99 V100 T101 N102 R103 G104 I105 P106

4.2.11 Score per residue for model 11

- Molecule 1: SCO1/SenC family protein/cytochrome c

Chain A: 65% 11% 21% •

GLY SER PHE THR S1 G2 E3 R9 C10 C13 V16 G17 N18 T19 E20 P21 G22 Q23 P24 G25 I26 G27 P28 D29 D38 L46 L59 L63 F64 R69 R95 L96 Q97 T98 P99 V100 T101 N102 R103 G104 I105 P106

4.2.12 Score per residue for model 12

- Molecule 1: SCO1/SenC family protein/cytochrome c

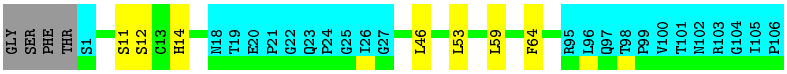
Chain A: 71% • • 21% •

GLY SER PHE THR S1 H14 N18 T19 E20 P21 G22 Q23 P24 G25 I26 G27 E55 K56 F64 L85 R95 L96 Q97 T98 P99 V100 T101 N102 R103 G104 I105 P106

4.2.13 Score per residue for model 13

- Molecule 1: SCO1/SenC family protein/cytochrome c

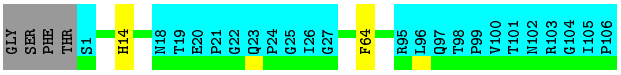
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: SCO1/SenC family protein/cytochrome c

Chain A:



4.2.15 Score per residue for model 15

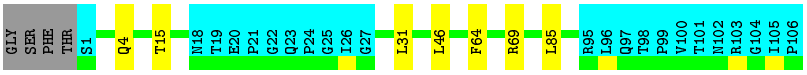
- Molecule 1: SCO1/SenC family protein/cytochrome c

Chain A:

4.2.16 Score per residue for model 16

- Molecule 1: SCO1/SenC family protein/cytochrome c

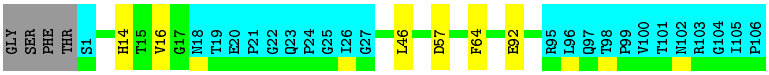
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: SCO1/SenC family protein/cytochrome c

Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: SCO1/SenC family protein/cytochrome c

Chain A:

4.2.19 Score per residue for model 19

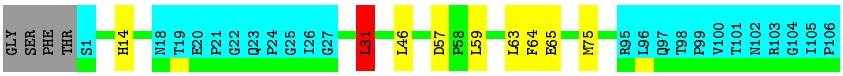
- Molecule 1: SCO1/SenC family protein/cytochrome c

Chain A:

4.2.20 Score per residue for model 20

- Molecule 1: SCO1/SenC family protein/cytochrome c

Chain A:



5 Refinement protocol and experimental data overview (i)

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

Of the 20 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	
AMBER	refinement	

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)	BMRB entry 17236
Number of chemical shift lists	1
Total number of shifts	507
Number of shifts mapped to atoms	507
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	36%

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

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: HEC

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	0.67±0.01	0±0/677 (0.0±0.0%)	1.06±0.03	0±0/917 (0.0±0.0%)
All	All	0.67	0/13540 (0.0%)	1.06	2/18340 (0.0%)

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	0.1±0.4
All	All	0	3

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	9	ARG	NE-CZ-NH2	-6.03	117.29	120.30	11	1
1	A	31	LEU	CB-CA-C	5.01	119.72	110.20	20	1

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	57	ASP	Peptide	2
1	A	56	LYS	Peptide	1

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	664	658	658	0±0
2	A	43	32	30	1±1
All	All	14140	13800	13760	13

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:A:107:HEC:HBC3	2:A:107:HEC:HMC1	0.68	1.65	20	3
2:A:107:HEC:HMC1	2:A:107:HEC:HBC3	0.60	1.70	16	2
2:A:107:HEC:HMB1	2:A:107:HEC:HBB3	0.56	1.76	11	1
2:A:107:HEC:CHA	2:A:107:HEC:CGD	0.44	2.95	1	1
1:A:85:LEU:HD21	2:A:107:HEC:HMB1	0.44	1.89	12	2
2:A:107:HEC:CBB	2:A:107:HEC:HMB1	0.43	2.44	11	1
2:A:107:HEC:HBB3	2:A:107:HEC:HMB1	0.41	1.92	16	1
1:A:31:LEU:HD22	1:A:31:LEU:C	0.41	2.36	20	1
1:A:10:CYS:SG	2:A:107:HEC:HMB1	0.41	2.56	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	83/110 (75%)	75±2 (91±2%)	7±2 (8±2%)	1±1 (1±1%)	23	69
All	All	1660/2200 (75%)	1507 (91%)	135 (8%)	18 (1%)	23	69

All 12 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	56	LYS	3
1	A	75	MET	3
1	A	59	LEU	2
1	A	28	PRO	2
1	A	71	ALA	1
1	A	32	GLY	1
1	A	2	GLY	1
1	A	55	GLU	1
1	A	76	ARG	1
1	A	57	ASP	1
1	A	12	SER	1
1	A	38	ASP	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	72/94 (77%)	68±2 (94±3%)	4±2 (6±3%)	28	73
All	All	1440/1880 (77%)	1352 (94%)	88 (6%)	28	73

All 34 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	64	PHE	20
1	A	46	LEU	10
1	A	14	HIS	6
1	A	59	LEU	4
1	A	63	LEU	4
1	A	29	ASP	4
1	A	3	GLU	3
1	A	16	VAL	3
1	A	91	GLU	2
1	A	55	GLU	2
1	A	76	ARG	2
1	A	31	LEU	2
1	A	85	LEU	2
1	A	86	ILE	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	69	ARG	2
1	A	81	GLU	2
1	A	87	SER	1
1	A	10	CYS	1
1	A	53	LEU	1
1	A	70	LEU	1
1	A	15	THR	1
1	A	35	ARG	1
1	A	36	GLN	1
1	A	51	GLN	1
1	A	34	THR	1
1	A	65	GLU	1
1	A	44	ARG	1
1	A	57	ASP	1
1	A	13	CYS	1
1	A	92	GLU	1
1	A	11	SER	1
1	A	48	VAL	1
1	A	4	GLN	1
1	A	9	ARG	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 carbohydrates in this entry.

6.6 Ligand geometry (i)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard

deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	HEC	A	107	1	24,50,50	0.80±0.03	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	HEC	A	107	1	19,82,82	2.44±0.13	2±0 (10±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	HEC	A	107	1	-	0±0,6,54,54	0±0,0,8,8

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	A	107	HEC	CBB-CAB-C3B	7.62	110.67	127.34	9	20
2	A	107	HEC	CBC-CAC-C3C	7.47	111.02	127.34	11	20

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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)

The completeness of assignment taking into account all chemical shift lists is 36% for the well-defined parts and 38% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 17236

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

7.1.1 Bookkeeping (i)

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	507
Number of shifts mapped to atoms	507
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	93	2.01 ± 0.44	Should be applied
$^{13}C_{\beta}$	78	3.66 ± 0.25	Should be applied
$^{13}C'$	89	1.80 ± 0.26	Should be applied
^{15}N	92	0.64 ± 0.24	Should be applied

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 36%, i.e. 378 atoms were assigned a chemical shift out of a possible 1047. 0 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	298/407 (73%)	80/162 (49%)	145/166 (87%)	73/79 (92%)
Sidechain	80/575 (14%)	11/337 (3%)	69/208 (33%)	0/30 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/65 (0%)	0/34 (0%)	0/28 (0%)	0/3 (0%)
Overall	378/1047 (36%)	91/533 (17%)	214/402 (53%)	73/112 (65%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 38%, i.e. 499 atoms were assigned a chemical shift out of a possible 1315. 1 out of 19 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	377/514 (73%)	106/204 (52%)	180/212 (85%)	91/98 (93%)
Sidechain	122/736 (17%)	24/433 (6%)	98/263 (37%)	0/40 (0%)
Aromatic	0/65 (0%)	0/34 (0%)	0/28 (0%)	0/3 (0%)
Overall	499/1315 (38%)	130/671 (19%)	278/503 (55%)	91/141 (65%)

7.1.4 Statistically unusual chemical shifts (i)

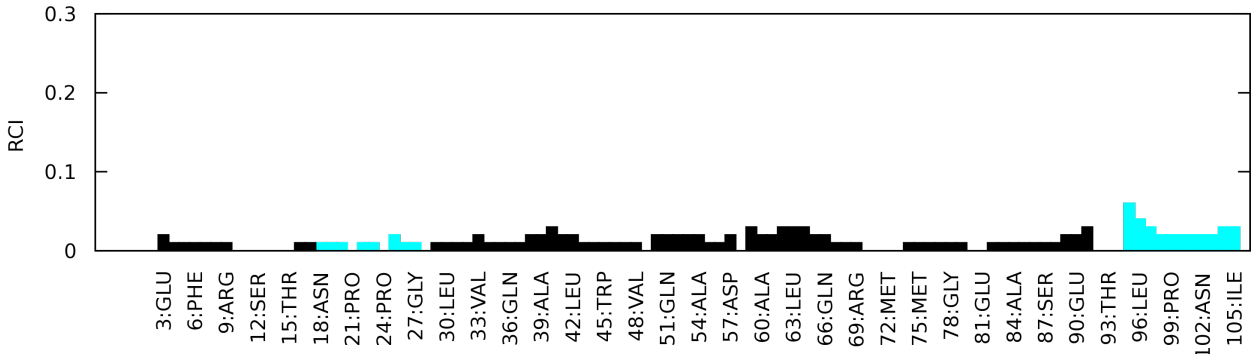
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	14	HIS	CA	78.40	68.24 – 44.74	9.3
1	A	72	MET	CB	13.70	44.20 – 21.80	-8.6
1	A	29	ASP	HA	6.17	6.15 – 3.05	5.1

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *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:


RCI
0.3
0.2
0.1
0
3:GLU
6:PHE
9:ARG
12:SER
15:THR
18:ASN
21:PRO
24:PRO
27:GLY
30:LEU
33:VAL
36:GLN
39:ALA
42:LEU
45:TRP
48:VAL
51:GLN
54:ALA
57:ASP
60:ALA
63:LEU
66:GLN
69:ARG
72:MET
75:MET
78:GLY
81:GLU
84:ALA
87:SER
90:GLU
93:THR
96:LEU
99:PRO
102:ASN
105:ILE