Full wwPDB NMR Structure Validation Report (i)

Jun 16, 2024 – 09:48 AM EDT

PDB ID	:	2J4M
BMRB ID	:	7285
Title	:	Double dockerin from Piromyces equi Cel45A
Authors	:	Nagy, T.; Tunnicliffe, R.B.; Higgins, L.D.; Walters, C.; Gilbert, H.J.; Williamson, M.P.
Deposited on	:	2006-09-01

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

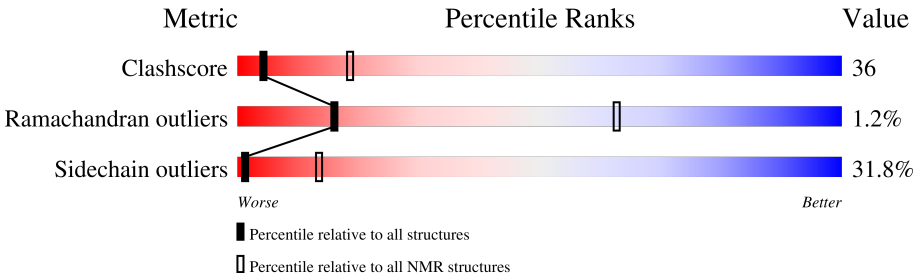
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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	100	41% 30% 9% 20%

2 Ensemble composition and analysis (i)

This entry contains 6 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:2-A:45 (44)	0.41	1
2	A:54-A:89 (36)	0.46	3

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 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 3, 4, 5
Single-model clusters	6

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1392 atoms, of which 633 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ENDOGLUCANASE 45A.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	100	1392	451	633	134	163	11	0

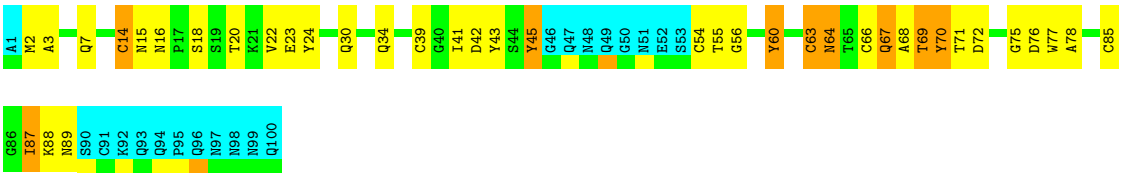
4 Residue-property plots (i)

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: ENDOGLUCANASE 45A

Chain A: 41% 30% 9% 20%



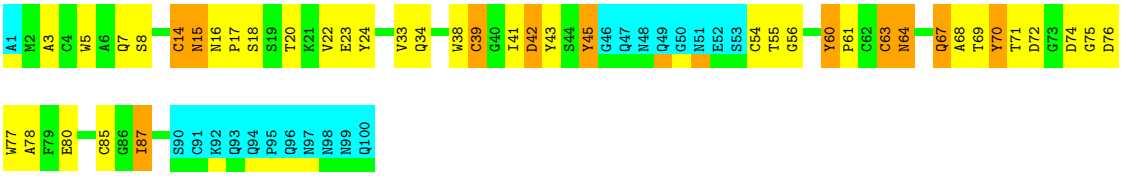
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

• Molecule 1: ENDOGLUCANASE 45A

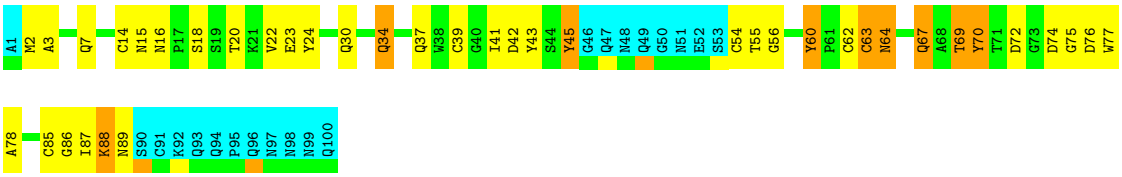
Chain A: 38% 31% 11% 20%



4.2.2 Score per residue for model 2

• Molecule 1: ENDOGLUCANASE 45A

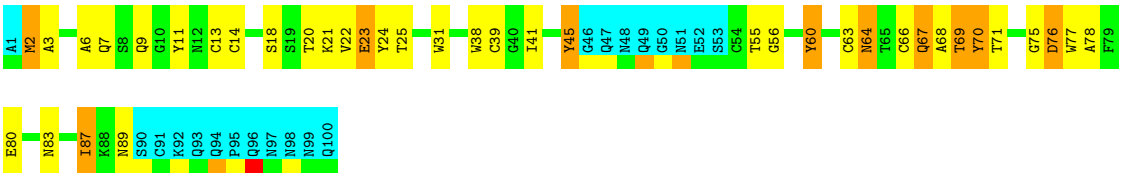
Chain A: 40% 31% 9% 20%



4.2.3 Score per residue for model 3

- Molecule 1: ENDOGLUCANASE 45A

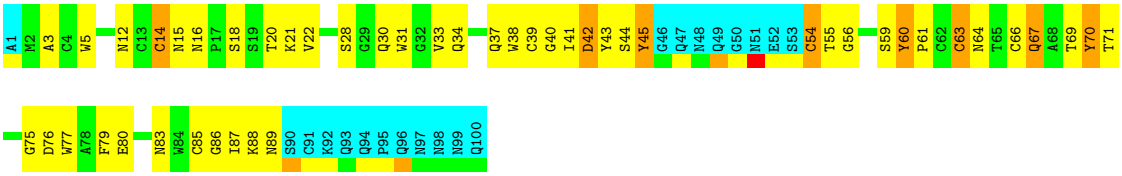
Chain A: 41% 29% 10% 20%



4.2.4 Score per residue for model 4

- Molecule 1: ENDOGLUCANASE 45A

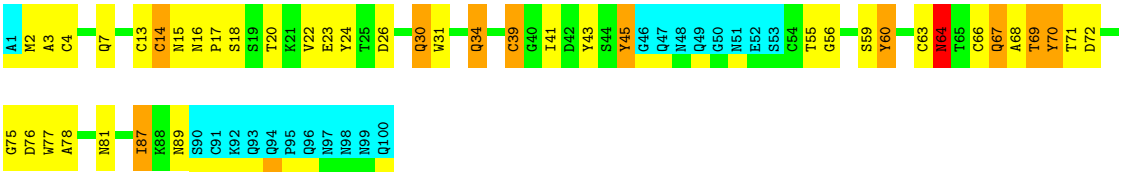
Chain A: 32% 40% 8% 20%



4.2.5 Score per residue for model 5

- Molecule 1: ENDOGLUCANASE 45A

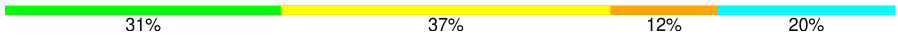
Chain A: 38% 31% 10% • 20%

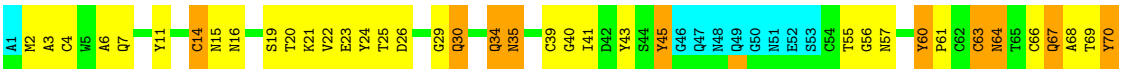


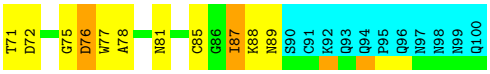
4.2.6 Score per residue for model 6

- Molecule 1: ENDOGLUCANASE 45A

Chain A:


31%
37%
12%
20%


A1 M2 A3 C4 W5 A6 Q7 Y11 C14 N15 N16 S19 T20 K21 Y22 E23 Y24 T25 D26 G29 Q30 Q34 N35 C39 G40 I41 D42 Y43 S44 Y45 G46 Q47 N48 Q49 G50 N51 E52 S53 C54 T55 G56 N57 Y60 P61 C62 C63 N64 T65 C66 Q67 A68 T69 Y70


T71 D72 G75 D76 W77 A78 N81 C85 G86 I87 K88 N89 S90 C91 K92 Q93 Q94 P95 Q96 N97 N98 N99 Q100

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *CNS*.

Of the 80 calculated structures, 6 were deposited, based on the following criterion: *RANDOM FROM BEST 25*.

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

Software name	Classification	Version
CNS	refinement	1.1
FELIX	structure solution	

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	608	502	502	40±3
All	All	3648	3012	3012	238

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 36.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:22:VAL:HG22	1:A:41:ILE:HD13	0.79	1.55	5	3
1:A:3:ALA:HB3	1:A:45:TYR:C	0.73	2.03	5	5
1:A:14:CYS:N	1:A:39:CYS:SG	0.72	2.63	4	2
1:A:22:VAL:HG11	1:A:30:GLN:HG2	0.71	1.62	4	2
1:A:22:VAL:HG11	1:A:30:GLN:CG	0.71	2.15	5	2
1:A:71:THR:HG23	1:A:75:GLY:C	0.69	2.08	4	1
1:A:55:THR:HG22	1:A:77:TRP:CH2	0.69	2.23	5	3
1:A:22:VAL:HG22	1:A:41:ILE:CD1	0.68	2.16	5	3
1:A:67:GLN:O	1:A:78:ALA:HB2	0.67	1.89	3	2
1:A:68:ALA:HB2	1:A:87:ILE:HG12	0.67	1.65	5	4
1:A:3:ALA:HB3	1:A:45:TYR:O	0.66	1.90	5	5
1:A:56:GLY:O	1:A:60:TYR:O	0.64	2.16	5	6
1:A:78:ALA:HB2	1:A:87:ILE:CD1	0.63	2.24	6	2
1:A:71:THR:HG23	1:A:75:GLY:O	0.62	1.93	5	5

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:55:THR:HG21	1:A:76:ASP:O	0.62	1.95	4	3
1:A:25:THR:HG23	1:A:29:GLY:O	0.61	1.96	6	1
1:A:55:THR:HG22	1:A:77:TRP:CZ3	0.60	2.31	2	3
1:A:22:VAL:HG11	1:A:30:GLN:HB3	0.59	1.75	6	2
1:A:63:CYS:O	1:A:64:ASN:CB	0.58	2.52	2	5
1:A:41:ILE:CG2	1:A:43:TYR:CE1	0.57	2.87	1	5
1:A:20:THR:CG2	1:A:41:ILE:HD11	0.57	2.30	3	6
1:A:63:CYS:O	1:A:64:ASN:HB3	0.57	2.00	1	5
1:A:33:VAL:HG23	1:A:38:TRP:CZ3	0.56	2.35	1	2
1:A:45:TYR:C	1:A:45:TYR:CD1	0.56	2.79	4	3
1:A:70:TYR:C	1:A:70:TYR:CD1	0.54	2.80	6	2
1:A:3:ALA:N	1:A:45:TYR:O	0.54	2.40	6	5
1:A:45:TYR:O	1:A:45:TYR:CG	0.54	2.61	3	6
1:A:45:TYR:C	1:A:45:TYR:CD2	0.53	2.81	6	2
1:A:70:TYR:CE2	1:A:77:TRP:CD1	0.53	2.97	1	4
1:A:20:THR:HG21	1:A:41:ILE:HD11	0.53	1.81	4	3
1:A:6:ALA:HB1	1:A:11:TYR:HB2	0.52	1.81	6	2
1:A:60:TYR:CB	1:A:61:PRO:CD	0.51	2.89	4	3
1:A:60:TYR:CD1	1:A:60:TYR:N	0.51	2.79	3	3
1:A:5:TRP:CD1	1:A:42:ASP:OD1	0.51	2.64	1	1
1:A:15:ASN:O	1:A:34:GLN:NE2	0.51	2.43	4	4
1:A:70:TYR:CD2	1:A:70:TYR:C	0.49	2.86	4	4
1:A:78:ALA:HB2	1:A:87:ILE:HD11	0.49	1.83	6	2
1:A:67:GLN:O	1:A:67:GLN:NE2	0.48	2.46	2	4
1:A:22:VAL:HG12	1:A:24:TYR:O	0.48	2.08	5	1
1:A:13:CYS:CB	1:A:45:TYR:CE2	0.48	2.97	3	1
1:A:41:ILE:HG22	1:A:41:ILE:O	0.47	2.09	5	3
1:A:56:GLY:HA2	1:A:77:TRP:CH2	0.47	2.45	6	2
1:A:7:GLN:HA	1:A:11:TYR:O	0.47	2.09	6	1
1:A:3:ALA:CB	1:A:45:TYR:O	0.47	2.63	5	4
1:A:20:THR:HG22	1:A:41:ILE:HD11	0.47	1.86	5	2
1:A:13:CYS:HB3	1:A:45:TYR:CD2	0.47	2.45	3	1
1:A:70:TYR:CE1	1:A:77:TRP:CD1	0.47	3.02	6	2
1:A:68:ALA:CB	1:A:76:ASP:HB3	0.47	2.40	6	2
1:A:70:TYR:CZ	1:A:77:TRP:CB	0.46	2.98	1	4
1:A:14:CYS:N	1:A:39:CYS:HB2	0.46	2.26	5	2
1:A:56:GLY:CA	1:A:77:TRP:CZ3	0.46	2.98	6	2
1:A:70:TYR:O	1:A:77:TRP:N	0.46	2.49	6	6
1:A:55:THR:CG2	1:A:77:TRP:CZ2	0.46	2.99	5	2
1:A:42:ASP:OD2	1:A:42:ASP:N	0.46	2.47	4	1
1:A:66:CYS:SG	1:A:66:CYS:O	0.46	2.74	5	2

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:39:CYS:SG	1:A:40:GLY:N	0.45	2.89	4	2
1:A:22:VAL:CG1	1:A:24:TYR:O	0.45	2.64	3	5
1:A:11:TYR:CD2	1:A:31:TRP:CZ3	0.45	3.04	3	1
1:A:60:TYR:CD1	1:A:77:TRP:CZ3	0.45	3.05	2	1
1:A:54:CYS:SG	1:A:55:THR:N	0.45	2.89	4	1
1:A:70:TYR:CE2	1:A:77:TRP:CG	0.45	3.05	1	1
1:A:77:TRP:CE3	1:A:86:GLY:CA	0.45	2.99	2	1
1:A:63:CYS:O	1:A:64:ASN:OD1	0.45	2.35	5	1
1:A:22:VAL:HG13	1:A:31:TRP:O	0.45	2.11	5	2
1:A:5:TRP:CD1	1:A:42:ASP:CG	0.45	2.90	4	1
1:A:24:TYR:CD2	1:A:25:THR:N	0.44	2.85	3	1
1:A:76:ASP:O	1:A:77:TRP:CG	0.44	2.71	4	3
1:A:16:ASN:N	1:A:17:PRO:CD	0.44	2.81	1	2
1:A:15:ASN:O	1:A:16:ASN:HB2	0.44	2.13	5	1
1:A:55:THR:CG2	1:A:77:TRP:CH2	0.44	3.00	2	1
1:A:79:PHE:CE2	1:A:83:ASN:N	0.44	2.86	4	1
1:A:34:GLN:O	1:A:37:GLN:CG	0.44	2.65	4	1
1:A:56:GLY:HA2	1:A:77:TRP:CZ3	0.43	2.48	1	2
1:A:63:CYS:O	1:A:64:ASN:ND2	0.43	2.50	2	3
1:A:13:CYS:SG	1:A:45:TYR:CD2	0.43	3.11	3	1
1:A:33:VAL:HG23	1:A:38:TRP:CH2	0.43	2.48	1	2
1:A:33:VAL:CG2	1:A:38:TRP:CZ3	0.43	3.02	4	1
1:A:77:TRP:CE3	1:A:86:GLY:HA2	0.43	2.49	2	1
1:A:69:THR:CG2	1:A:78:ALA:HA	0.42	2.45	5	3
1:A:13:CYS:CB	1:A:45:TYR:CD2	0.42	3.02	3	1
1:A:5:TRP:NE1	1:A:42:ASP:OD2	0.42	2.51	4	1
1:A:60:TYR:CB	1:A:61:PRO:HD2	0.42	2.43	1	2
1:A:68:ALA:HB2	1:A:87:ILE:CG1	0.42	2.43	6	1
1:A:70:TYR:CZ	1:A:77:TRP:CG	0.42	3.08	1	3
1:A:16:ASN:OD1	1:A:34:GLN:NE2	0.42	2.53	2	2
1:A:55:THR:HG22	1:A:56:GLY:N	0.42	2.30	1	2
1:A:20:THR:OG1	1:A:34:GLN:NE2	0.42	2.53	6	2
1:A:70:TYR:CE1	1:A:77:TRP:CB	0.42	3.02	2	1
1:A:13:CYS:SG	1:A:45:TYR:CD1	0.42	3.13	5	1
1:A:15:ASN:OD1	1:A:16:ASN:ND2	0.42	2.53	1	1
1:A:75:GLY:O	1:A:77:TRP:CD1	0.42	2.73	2	1
1:A:70:TYR:CZ	1:A:77:TRP:HB2	0.42	2.50	3	2
1:A:35:ASN:O	1:A:35:ASN:CG	0.42	2.58	6	1
1:A:14:CYS:O	1:A:15:ASN:C	0.41	2.58	5	1
1:A:70:TYR:CE1	1:A:77:TRP:CG	0.41	3.08	6	1
1:A:63:CYS:SG	1:A:78:ALA:HB3	0.41	2.55	3	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:55:THR:HG22	1:A:86:GLY:HA3	0.41	1.91	4	1
1:A:22:VAL:HA	1:A:31:TRP:O	0.41	2.15	3	1
1:A:12:ASN:OD1	1:A:37:GLN:NE2	0.41	2.54	4	1
1:A:62:CYS:HB3	1:A:88:LYS:HE3	0.41	1.92	2	1
1:A:23:GLU:OE2	1:A:38:TRP:CZ3	0.41	2.74	3	1
1:A:67:GLN:O	1:A:78:ALA:CB	0.41	2.66	3	1
1:A:22:VAL:HG11	1:A:30:GLN:HG3	0.41	1.89	5	1
1:A:4:CYS:O	1:A:7:GLN:HG2	0.41	2.16	6	1
1:A:34:GLN:O	1:A:35:ASN:ND2	0.40	2.53	6	1
1:A:5:TRP:N	1:A:42:ASP:OD1	0.40	2.55	4	1
1:A:70:TYR:CD1	1:A:77:TRP:HB2	0.40	2.51	4	1
1:A:55:THR:CG2	1:A:56:GLY:N	0.40	2.85	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	80/100 (80%)	68±1 (84±1%)	12±1 (14±1%)	1±0 (1±0%)	16	63
All	All	480/600 (80%)	405 (84%)	69 (14%)	6 (1%)	16	63

All 2 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	64	ASN	5
1	A	59	SER	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	64/81 (79%)	44±2 (68±3%)	20±2 (32±3%)	1 13
All	All	384/486 (79%)	262 (68%)	122 (32%)	1 13

All 41 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	14	CYS	6
1	A	45	TYR	6
1	A	60	TYR	6
1	A	67	GLN	6
1	A	69	THR	6
1	A	70	TYR	6
1	A	87	ILE	6
1	A	18	SER	5
1	A	23	GLU	5
1	A	7	GLN	4
1	A	34	GLN	4
1	A	39	CYS	4
1	A	63	CYS	4
1	A	72	ASP	4
1	A	76	ASP	4
1	A	85	CYS	4
1	A	2	MET	4
1	A	42	ASP	3
1	A	54	CYS	3
1	A	80	GLU	3
1	A	88	LYS	3
1	A	21	LYS	3
1	A	74	ASP	2
1	A	26	ASP	2
1	A	30	GLN	2
1	A	81	ASN	2
1	A	8	SER	1
1	A	15	ASN	1
1	A	37	GLN	1
1	A	9	GLN	1
1	A	83	ASN	1
1	A	16	ASN	1
1	A	28	SER	1
1	A	44	SER	1
1	A	4	CYS	1
1	A	59	SER	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	64	ASN	1
1	A	19	SER	1
1	A	35	ASN	1
1	A	57	ASN	1
1	A	66	CYS	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