Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 04:42 PM BST

PDB ID	:	1QM3
Title	:	HUMAN PRION PROTEIN FRAGMENT 121-230
Authors	:	Zahn, R.; Liu, A.; Luhrs, T.; Wuthrich, K.
Deposited on	:	1999-09-20

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:

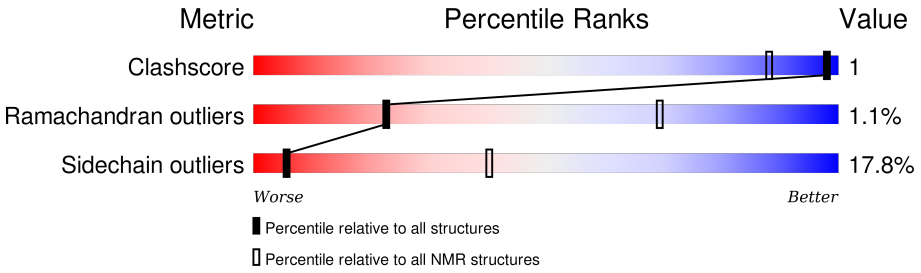
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 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	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	112	81% 12% 7%

2 Ensemble composition and analysis

This entry contains 20 models. Model 12 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:125-A:228 (104)	0.62	12

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 5 clusters and 7 single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1688 atoms, of which 811 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PRION PROTEIN.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	104	1688	544	811	153	171	9	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	119	GLY	-	CLONING ARTIFACT	UNP P04156
A	120	SER	-	CLONING ARTIFACT	UNP P04156

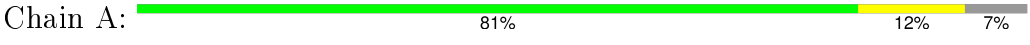
4 Residue-property plots (i)

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: PRION PROTEIN

Chain A:



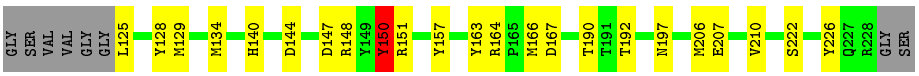
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: PRION PROTEIN

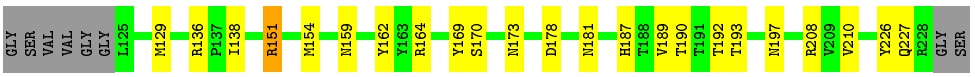
Chain A:



4.2.2 Score per residue for model 2

- Molecule 1: PRION PROTEIN

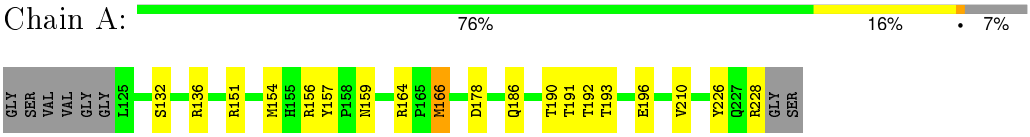
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: PRION PROTEIN

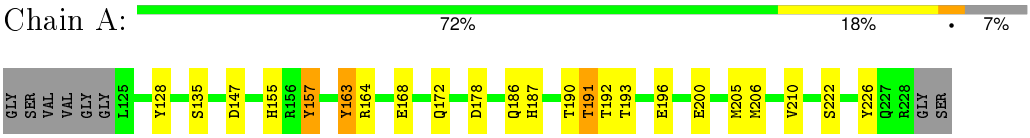
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: PRION PROTEIN

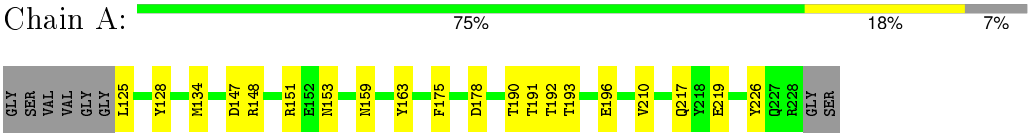
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: PRION PROTEIN

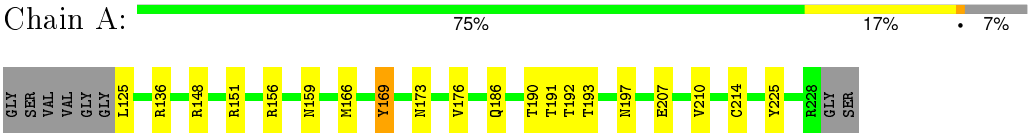
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: PRION PROTEIN

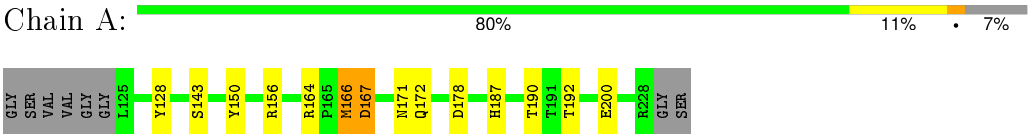
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: PRION PROTEIN

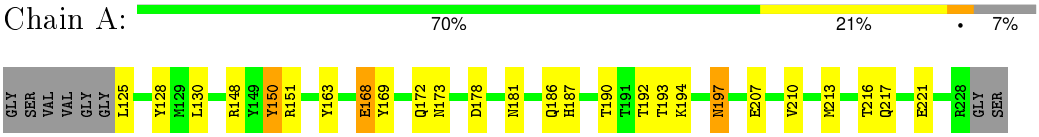
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: PRION PROTEIN

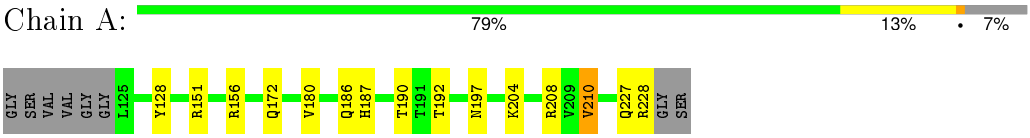
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: PRION PROTEIN

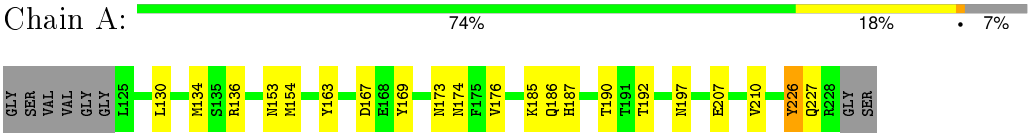
Chain A:



4.2.10 Score per residue for model 10

- Molecule 1: PRION PROTEIN

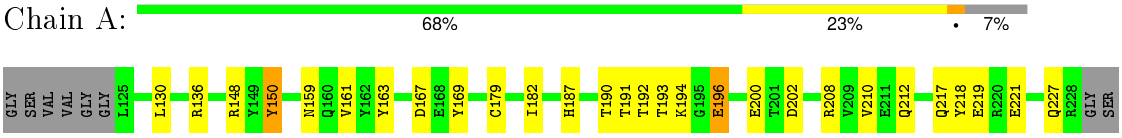
Chain A:



4.2.11 Score per residue for model 11

- Molecule 1: PRION PROTEIN

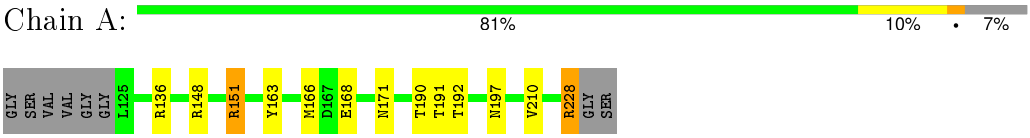
Chain A:



4.2.12 Score per residue for model 12 (medoid)

- Molecule 1: PRION PROTEIN

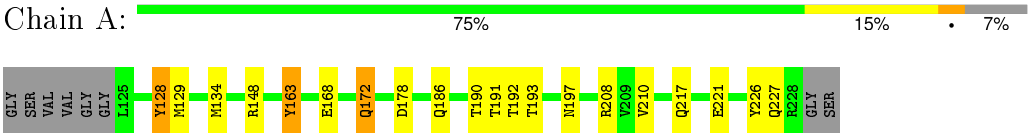
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: PRION PROTEIN

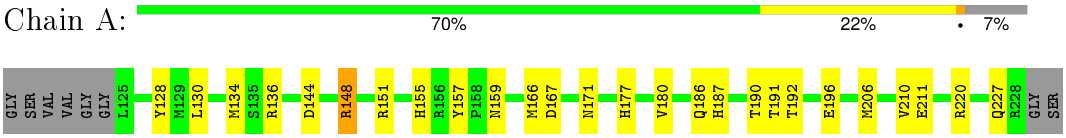
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: PRION PROTEIN

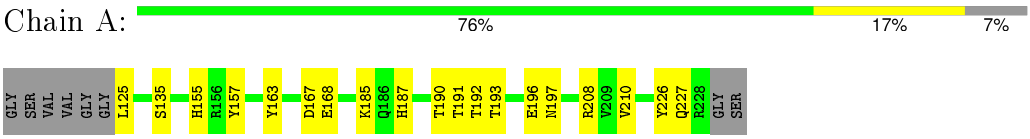
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: PRION PROTEIN

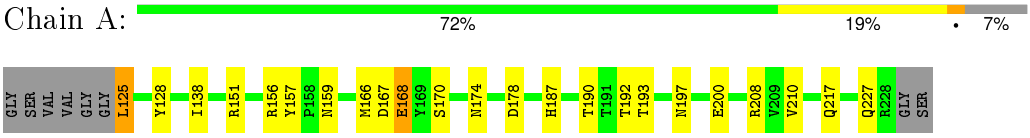
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: PRION PROTEIN

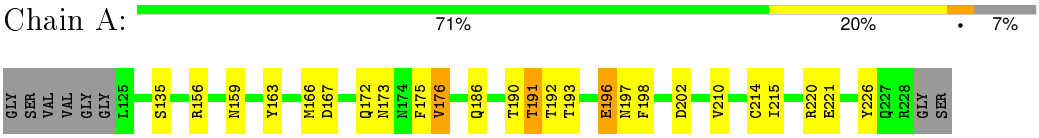
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: PRION PROTEIN

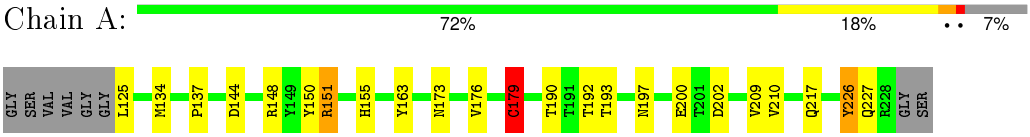
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: PRION PROTEIN

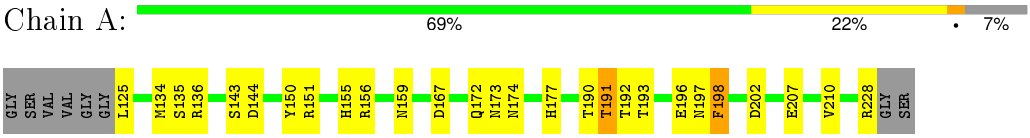
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: PRION PROTEIN

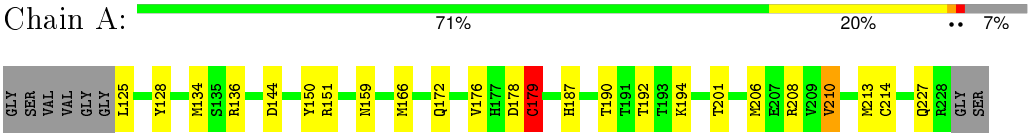
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: PRION PROTEIN

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *TORSION ANGLE DYNAMICS*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *LEAST RESTRAINT VIOLATION*.

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

Software name	Classification	Version
OPALP	refinement	
DYANA	structure solution	

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality

6.1 Standard geometry

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	0.68±0.01	0±0/897 (0.0±0.0%)	1.06±0.03	2±1/1210 (0.2±0.1%)
All	All	0.68	0/17940 (0.0%)	1.06	39/24200 (0.2%)

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	1.6±1.2
All	All	0	33

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	179	CYS	CA-CB-SG	7.57	127.62	114.00	20	2
1	A	208	ARG	NE-CZ-NH2	-7.47	116.56	120.30	20	1
1	A	156	ARG	NE-CZ-NH2	-7.07	116.77	120.30	16	1
1	A	210	VAL	CG1-CB-CG2	-7.06	99.61	110.90	4	18
1	A	150	TYR	CB-CG-CD2	-6.52	117.09	121.00	11	4
1	A	148	ARG	NE-CZ-NH2	-6.50	117.05	120.30	5	1
1	A	208	ARG	NE-CZ-NH1	6.37	123.48	120.30	20	1
1	A	220	ARG	NE-CZ-NH1	6.11	123.35	120.30	17	1
1	A	163	TYR	CB-CG-CD2	-5.88	117.47	121.00	4	1
1	A	169	TYR	CB-CG-CD2	-5.65	117.61	121.00	6	1
1	A	176	VAL	CG1-CB-CG2	-5.61	101.92	110.90	10	1
1	A	156	ARG	NE-CZ-NH1	5.56	123.08	120.30	7	3
1	A	220	ARG	NE-CZ-NH2	-5.51	117.54	120.30	14	1
1	A	164	ARG	NE-CZ-NH2	-5.37	117.61	120.30	4	1
1	A	136	ARG	NE-CZ-NH1	5.20	122.90	120.30	12	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	191	THR	CA-CB-CG2	5.01	119.42	112.40	4	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	151	ARG	Sidechain	4
1	A	157	TYR	Sidechain	4
1	A	156	ARG	Sidechain	4
1	A	163	TYR	Sidechain	3
1	A	148	ARG	Sidechain	3
1	A	208	ARG	Sidechain	3
1	A	228	ARG	Sidechain	2
1	A	226	TYR	Sidechain	2
1	A	225	TYR	Sidechain	1
1	A	153	ASN	Peptide	1
1	A	169	TYR	Sidechain	1
1	A	162	TYR	Sidechain	1
1	A	128	TYR	Sidechain	1
1	A	218	TYR	Sidechain	1
1	A	125	LEU	Peptide	1
1	A	136	ARG	Sidechain	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	877	811	813	1±1
All	All	17540	16220	16260	19

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
1:A:176:VAL:HG22	1:A:214:CYS:HB3	0.55	1.78	6	3
1:A:176:VAL:HG21	1:A:215:ILE:HG13	0.51	1.82	17	1
1:A:191:THR:HG21	1:A:198:PHE:CE2	0.47	2.43	17	2
1:A:176:VAL:HG22	1:A:214:CYS:CB	0.47	2.39	17	1
1:A:206:MET:O	1:A:210:VAL:HG23	0.46	2.10	20	2
1:A:180:VAL:HA	1:A:210:VAL:HG12	0.45	1.89	14	2
1:A:176:VAL:HA	1:A:179:CYS:SG	0.44	2.53	20	2
1:A:157:TYR:CE2	1:A:206:MET:HG3	0.43	2.48	4	1
1:A:130:LEU:HD12	1:A:161:VAL:O	0.42	2.14	11	1
1:A:150:TYR:CD1	1:A:157:TYR:CD2	0.41	3.08	1	1
1:A:197:ASN:HD22	1:A:197:ASN:C	0.41	2.19	8	1
1:A:191:THR:HG23	1:A:196:GLU:O	0.40	2.16	17	1
1:A:137:PRO:HD2	1:A:209:VAL:HG13	0.40	1.93	18	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	102/112 (91%)	92±2 (90±2%)	9±2 (9±2%)	1±1 (1±1%)	23	69
All	All	2040/2240 (91%)	1835 (90%)	183 (9%)	22 (1%)	23	69

All 10 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	166	MET	5
1	A	172	GLN	3
1	A	196	GLU	3
1	A	168	GLU	3
1	A	167	ASP	3
1	A	157	TYR	1
1	A	170	SER	1
1	A	171	ASN	1
1	A	132	SER	1
1	A	189	VAL	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	97/101 (96%)	80±4 (82±4%)	17±4 (18±4%)	6 40
All	All	1940/2020 (96%)	1594 (82%)	346 (18%)	6 40

All 67 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	190	THR	20
1	A	192	THR	20
1	A	197	ASN	13
1	A	193	THR	13
1	A	151	ARG	12
1	A	187	HIS	11
1	A	191	THR	11
1	A	227	GLN	10
1	A	128	TYR	10
1	A	226	TYR	10
1	A	159	ASN	10
1	A	125	LEU	9
1	A	178	ASP	9
1	A	186	GLN	9
1	A	163	TYR	9
1	A	134	MET	8
1	A	173	ASN	7
1	A	167	ASP	7
1	A	136	ARG	7
1	A	148	ARG	6
1	A	196	GLU	6
1	A	150	TYR	6
1	A	172	GLN	6
1	A	217	GLN	6
1	A	166	MET	6
1	A	207	GLU	5
1	A	200	GLU	5
1	A	155	HIS	5
1	A	168	GLU	5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	144	ASP	5
1	A	135	SER	4
1	A	221	GLU	4
1	A	169	TYR	4
1	A	164	ARG	4
1	A	202	ASP	4
1	A	174	ASN	3
1	A	129	MET	3
1	A	208	ARG	3
1	A	147	ASP	3
1	A	194	LYS	3
1	A	179	CYS	3
1	A	130	LEU	3
1	A	228	ARG	3
1	A	154	MET	3
1	A	177	HIS	2
1	A	213	MET	2
1	A	185	LYS	2
1	A	219	GLU	2
1	A	143	SER	2
1	A	175	PHE	2
1	A	138	ILE	2
1	A	171	ASN	2
1	A	222	SER	2
1	A	181	ASN	2
1	A	198	PHE	1
1	A	201	THR	1
1	A	182	ILE	1
1	A	206	MET	1
1	A	170	SER	1
1	A	204	LYS	1
1	A	212	GLN	1
1	A	216	THR	1
1	A	153	ASN	1
1	A	211	GLU	1
1	A	205	MET	1
1	A	176	VAL	1
1	A	140	HIS	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 carbohydrates 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