W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 11:10 pm BST

PDB ID	:	2L4K
Title	:	Water refined solution structure of the human Grb7-SH2 domain in complex with the 10 amino acid peptide pY1139
Authors	:	Pias, S.C.; Ivancic, M.; Brescia, P.J.; Johnson, D.L.; Smith, D.E.; Daly, R.J.; Lyons, B.A.
Deposited on	:	2010-10-07

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

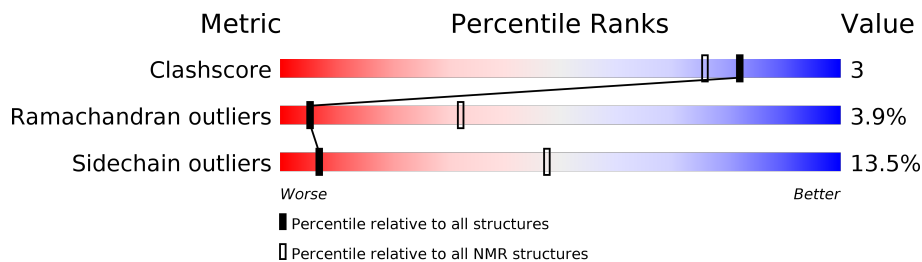|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
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 ⓘ

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	Percentile Ranks	Value
Clashscore		3
Ramachandran outliers		3.9%
Sidechain outliers		13.5%

Worse   *Better*

▌ Percentile relative to all structures
▯ Percentile relative to all NMR structures

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	120	66% 11% 5% · 18%
2	B	10	30% 20% 20% 30%

2 Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *best procheck statistics*.

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:19-A:117, B:1137-B:1138, B:1140-B:1144 (106)	0.72	2

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 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 5, 7, 8
2	3, 6, 9, 10
Single-model clusters	4

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2091 atoms, of which 1041 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Growth factor receptor-bound protein 7.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	120	1931	605	969	180	172	5	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION TAG	UNP Q14451
A	2	SER	-	EXPRESSION TAG	UNP Q14451

- Molecule 2 is a protein called Receptor tyrosine-protein kinase erbB-2.

| Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	10	160	52	72	13	22	1	0

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: Growth factor receptor-bound protein 7

Chain A:

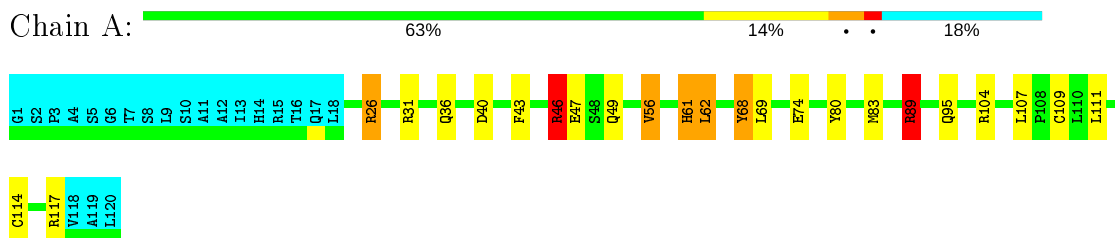- Molecule 2: Receptor tyrosine-protein kinase erbB-2

Chain B:
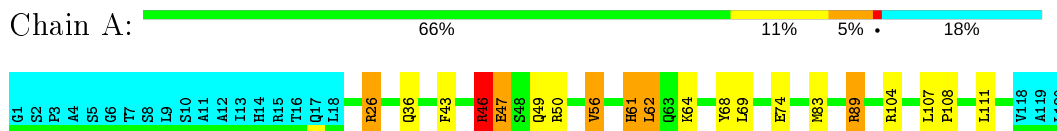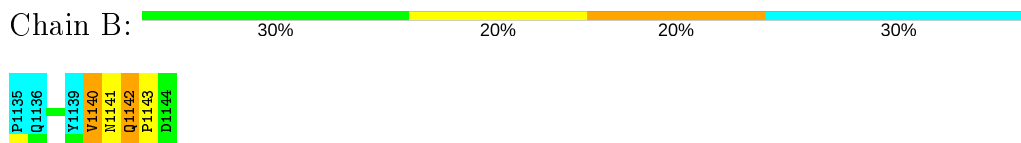
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Growth factor receptor-bound protein 7

Chain A:
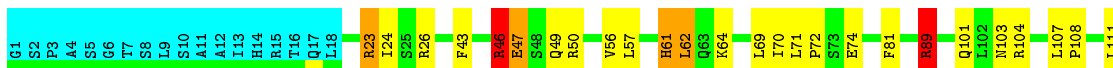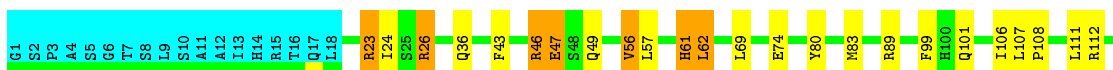

- Molecule 2: Receptor tyrosine-protein kinase erbB-2

Chain B:

P1135 Q1136 Y1139 V1140 N1141 Q1142 P1143 D1144

4.2.2 Score per residue for model 2 (medoid)

• Molecule 1: Growth factor receptor-bound protein 7

Chain A:

60% 18% •• 18%

G1 S2 P3 A4 S5 G6 T7 S8 L9 S10 A11 A12 I13 H14 R15 T16 Q17 L18 R23 I24 S25 R26 F43 R46 E47 S48 Q49 R50 V56 L57 H61 L62 Q63 K64 L69 I70 L71 P72 S73 E74 F81 R89 Q101 L102 N103 R104 L107 P108 L111

C114 V118 A119 L120

• Molecule 2: Receptor tyrosine-protein kinase erbB-2

Chain B:

40% 10% 10% 10% 30%

P1135 Q1136 Y1139 V1140 N1141 Q1142 P1143 D1144

4.2.3 Score per residue for model 3

• Molecule 1: Growth factor receptor-bound protein 7

Chain A:

68% 13% •• 18%

G1 S2 P3 A4 S5 G6 T7 S8 L9 S10 A11 A12 I13 H14 R15 T16 Q17 L18 R26 Q36 R46 E47 V56 H61 L62 Q63 K64 L69 I70 L71 E74 R89 Q101 R104 P108 L111 R112 V118 A119 L120

• Molecule 2: Receptor tyrosine-protein kinase erbB-2

Chain B:

30% 20% 20% 30%

P1135 Q1136 Y1139 V1140 N1141 Q1142 P1143 D1144

4.2.4 Score per residue for model 4

• Molecule 1: Growth factor receptor-bound protein 7

Chain A:

62% 15% 6% 18%

G1 S2 P3 A4 S5 G6 T7 S8 L9 S10 A11 A12 I13 H14 R15 T16 Q17 L18 R23 I24 S25 R26 Q36 F43 R46 E47 S48 Q49 V56 L57 H61 L62 L69 E74 Y80 M83 R89 F99 H100 Q101 I106 L107 P108 L111 R112

R117 V118 A119 L120

• Molecule 2: Receptor tyrosine-protein kinase erbB-2

Chain B: 40%   10%   20%   30%

P1135 Q1136 Y1139 V1140 N1141 Q1142 P1143 D1144
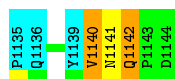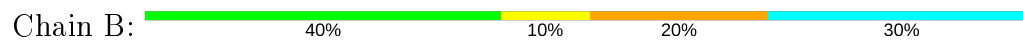
4.2.5 Score per residue for model 5

• Molecule 1: Growth factor receptor-bound protein 7

Chain A: 66%   13%   •   18%

G1 S2 P3 A4 S5 G6 T7 S8 L9 S10 A11 A12 I13 H14 R15 T16 Q17 L18 R26 Q36 F43 R46 E47 S48 Q49 R50 V56 L57 L62 Q63 K64 P72 S73 E74 M83 Q87 T88 R89 R104 L107 L110 L111 V118 A119 L120

• Molecule 2: Receptor tyrosine-protein kinase erbB-2

Chain B: 40%   30%   30%

P1135 Q1136 Y1139 V1140 N1141 Q1142 P1143 D1144
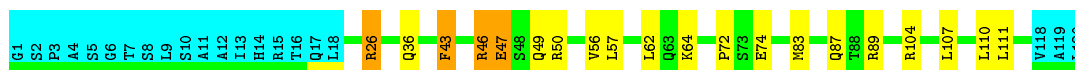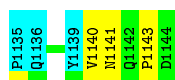
4.2.6 Score per residue for model 6

• Molecule 1: Growth factor receptor-bound protein 7

Chain A: 63%   14%   • •   18%

G1 S2 P3 A4 S5 G6 T7 S8 L9 S10 A11 A12 I13 H14 R15 T16 Q17 L18 R26 R31 Q36 F43 R46 E47 S48 Q49 R50 V56 H61 L62 L69 E74 S82 M83 Q87 T88 R89 V97 L102 I106 L107 P108 L111 V118

A119 L120

• Molecule 2: Receptor tyrosine-protein kinase erbB-2

Chain B: 20%   30%   20%   30%

P1135 Q1136 P1137 E1138 Y1139 V1140 N1141 Q1142 P1143 D1144
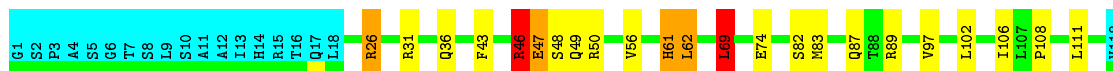
4.2.7 Score per residue for model 7

• Molecule 1: Growth factor receptor-bound protein 7

Chain A:

63%   14% · · 18%

G1 S2 P3 A4 S5 G6 T7 S8 L9 S10 A11 A12 I13 H14 R15 T16 Q17 L18 R26 D40 R46 E47 S48 Q49 R50 V56 L57 C60 H61 L62 Y68 L69 I70 L71 E74 Y80 M83 Q87 T88 R89 R104 L107 P108 L111 V118 A119

L120

• Molecule 2: Receptor tyrosine-protein kinase erbB-2

Chain B:

30%   20%   10%   10%   30%

P1135 Q1136 Y1139 V1140 N1141 Q1142 P1143 D1144
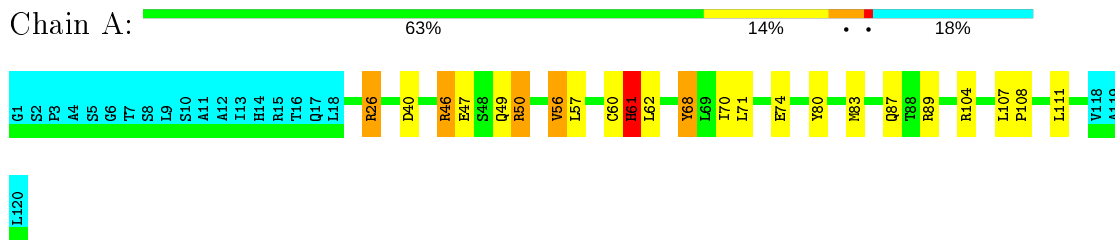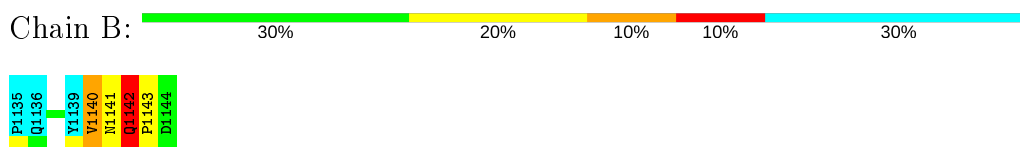
4.2.8 Score per residue for model 8

• Molecule 1: Growth factor receptor-bound protein 7

Chain A:

62%   15%   5% · 18%

G1 S2 P3 A4 S5 G6 T7 S8 L9 S10 A11 A12 I13 H14 R15 T16 Q17 L18 R26 R31 Q35 Q36 G41 L42 F43 R46 E47 S48 Q49 R50 N51 V56 H61 L62 Q63 K64 H67 Y68 L69 I70 L71 E74 Q87 T88 R89 R104 P108

L111 R112 H113 V118 A119 L120

• Molecule 2: Receptor tyrosine-protein kinase erbB-2

Chain B:

30%   30%   10%   30%

P1135 Q1136 Y1139 V1140 N1141 Q1142 P1143 D1144
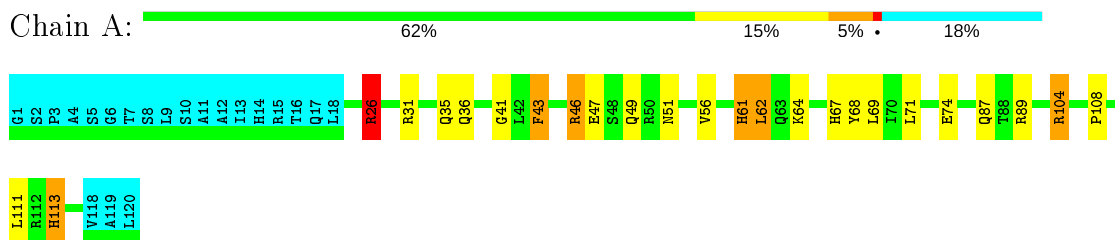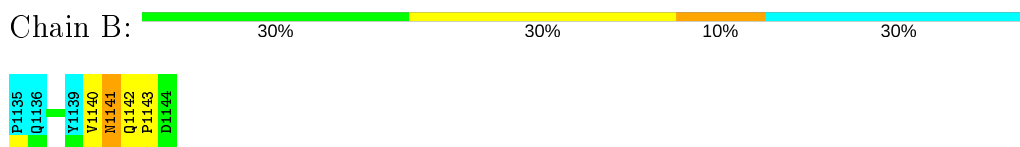
4.2.9 Score per residue for model 9

• Molecule 1: Growth factor receptor-bound protein 7

Chain A:

63%   12%   6% · 18%

G1 S2 P3 A4 S5 G6 T7 S8 L9 S10 A11 A12 I13 H14 R15 T16 Q17 L18 I24 S25 R26 Q35 Q36 D40 F43 R46 E47 S48 Q49 R50 G54 F55 V56 H61 L62 L69 E74 S82 M83 R89 I106 L107 P108 L111 C114

V118
A119
L120

• Molecule 2: Receptor tyrosine-protein kinase erbB-2

Chain B:
40%   10%   10%   10%   30%

P1135
Q1136
Y1139
V1140
N1141
Q1142
P1143
D1144
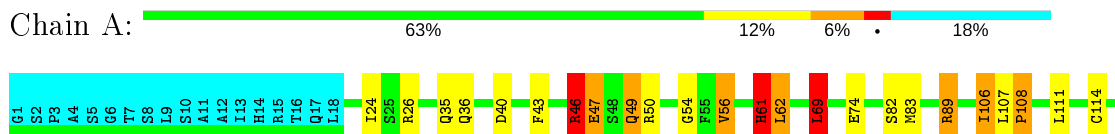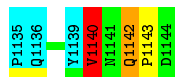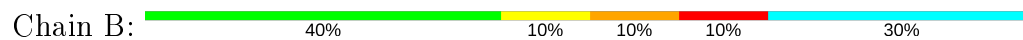
4.2.10   Score per residue for model 10

• Molecule 1: Growth factor receptor-bound protein 7

Chain A:
66%   13% • •   18%

G1
S2
P3
A4
S5
G6
T7
S8
L9
S10
A11
A12
I13
H14
R15
T16
Q17
L18
R26
R31
Q36
R46
E47
S48
Q49
G54
F55
V56
H61
L62
Q63
K64
L69
E74
R78
S82
R89
Q101
P108
L111
C114
V118
A119
L120

• Molecule 2: Receptor tyrosine-protein kinase erbB-2

Chain B:
50%   10%   10%   30%

P1135
Q1136
Y1139
V1140
N1141
Q1142
P1143
D1144
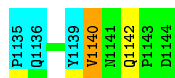
5   Refinement protocol and experimental data overview (i)

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

Of the 50 calculated structures, 10 were deposited, based on the following criterion: *structures with the lowest energy and the fewest restraint violations.*

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

Software name	Classification	Version
AMBER	refinement	9
CNS	structure calculation	1.1

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

COVALENT-GEOMETRY INFOmissingINFO

5.1   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	819	819	816	4±2
2	B	56	47	47	1±1
All	All	8750	8660	8629	44

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 3.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:A:47:GLU:HA	1:A:56:VAL:HG13	0.67	1.65	9	3
1:A:70:ILE:HD12	2:B:1141:ASN:HB2	0.56	1.78	2	3
1:A:83:MET:HA	2:B:1140:VAL:HG12	0.52	1.80	6	2
1:A:61:HIS:CG	1:A:62:LEU:H	0.51	2.24	7	7
1:A:46:ARG:C	1:A:46:ARG:HD2	0.48	2.29	5	1
1:A:106:ILE:HD13	1:A:107:LEU:H	0.47	1.69	9	1
1:A:61:HIS:CG	1:A:62:LEU:N	0.47	2.82	2	8
1:A:41:GLY:HA2	1:A:113:HIS:CD2	0.47	2.44	8	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:A:69:LEU:C	1:A:69:LEU:HD13	0.46	2.29	3	5
2:B:1140:VAL:C	2:B:1142:GLN:N	0.44	2.71	6	5
1:A:106:ILE:HD13	1:A:107:LEU:N	0.43	2.28	9	1
1:A:54:GLY:HA3	1:A:69:LEU:HD22	0.43	1.91	10	2
1:A:81:PHE:H	1:A:89:ARG:HA	0.41	1.75	2	1
1:A:43:PHE:CG	1:A:57:LEU:HD21	0.41	2.50	5	1
1:A:54:GLY:CA	1:A:69:LEU:HD22	0.41	2.46	10	1
1:A:107:LEU:CB	1:A:108:PRO:HD3	0.41	2.46	9	1
1:A:26:ARG:HB2	1:A:67:HIS:CD2	0.40	2.51	8	1

5.2 Torsion angles (i)

5.2.1 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	99/120 (82%)	74±3 (75±3%)	22±3 (22±3%)	3±1 (3±1%)	7	38
2	B	6/10 (60%)	3±1 (57±13%)	2±0 (28±8%)	1±1 (15±14%)	0	4
All	All	1050/1300 (81%)	777 (74%)	232 (22%)	41 (4%)	5	32

All 13 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	89	ARG	10
1	A	108	PRO	8
1	A	83	MET	4
1	A	40	ASP	3
1	A	24	ILE	3
2	B	1140	VAL	2
2	B	1142	GLN	2
2	B	1141	ASN	2
2	B	1143	PRO	2
1	A	72	PRO	2
1	A	99	PHE	1
2	B	1138	GLU	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	97	VAL	1

5.2.2 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	91/106 (86%)	78±2 (86±2%)	13±2 (14±2%)	6	46
2	B	7/9 (78%)	7±0 (96±7%)	0±0 (4±7%)	33	81
All	All	980/1150 (85%)	848 (87%)	132 (13%)	7	47

All 37 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	111	LEU	10
1	A	26	ARG	10
1	A	46	ARG	9
1	A	56	VAL	9
1	A	36	GLN	8
1	A	62	LEU	8
1	A	61	HIS	7
1	A	43	PHE	7
1	A	64	LYS	5
1	A	49	GLN	5
1	A	114	CYS	4
1	A	71	LEU	4
1	A	87	GLN	4
1	A	101	GLN	4
1	A	57	LEU	3
1	A	107	LEU	3
1	A	106	ILE	3
1	A	80	TYR	3
1	A	69	LEU	3
1	A	68	TYR	2
1	A	23	ARG	2
1	A	47	GLU	2
1	A	104	ARG	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	35	GLN	2
1	A	109	CYS	1
1	A	95	GLN	1
1	A	103	ASN	1
2	B	1140	VAL	1
2	B	1141	ASN	1
1	A	51	ASN	1
2	B	1142	GLN	1
1	A	89	ARG	1
1	A	102	LEU	1
1	A	110	LEU	1
1	A	113	HIS	1
1	A	48	SER	1
1	A	50	ARG	1

5.2.3 RNA (i)

There are no RNA molecules in this entry.

MODRES-GEOMETRY INFOmissingINFO

5.3 Carbohydrates (i)

There are no carbohydrates in this entry.

5.4 Ligand geometry (i)

There are no ligands in this entry.

5.5 Other polymers (i)

There are no such molecules in this entry.

5.6 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation ⓘ

No chemical shift data were provided