Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 08:51 PM BST

PDB ID	:	2J2S
Title	:	SOLUTION STRUCTURE OF THE NONMETHYL-CPG-BINDING CXXC DOMAIN OF THE LEUKAEMIA-ASSOCIATED MLL HISTONE METHYLTRANSFERASE
Authors	:	Allen, M.D.; Grummitt, C.G.; Hilcenko, C.; Young-Min, S.; Tonkin, L.M.; Johnson, C.M.; Bycroft, M.; Warren, A.J.
Deposited on	:	2006-08-17

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:

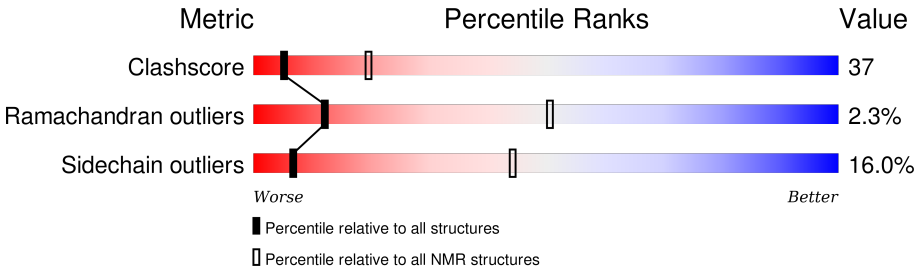
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	72	28% 26% 6% • 39%

2 Ensemble composition and analysis

This entry contains 20 models. Model 20 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:1154-A:1197 (44)	0.17	20

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 4 clusters and 2 single-model clusters were found.

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1136 atoms, of which 580 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ZINC FINGER PROTEIN HRX.

Mol	Chain	Residues	Atoms						Trace
1	A	72	Total	C	H	N	O	S	0
			1134	336	580	113	94	11	

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1143	GLY	GLU	ENGINEERED MUTATION	UNP Q03164
A	1144	GLY	PRO	ENGINEERED MUTATION	UNP Q03164
A	1145	SER	PRO	ENGINEERED MUTATION	UNP Q03164

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
2	A	2	Total	Zn
			2	2

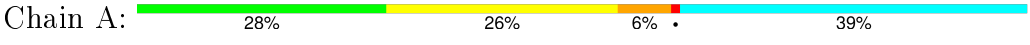
4 Residue-property plots

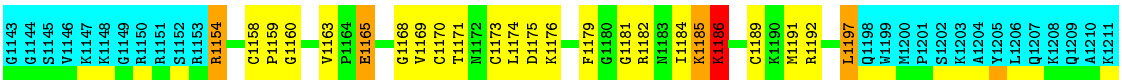
4.1 Average score per residue in the NMR ensemble

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: ZINC FINGER PROTEIN HRX

Chain A:





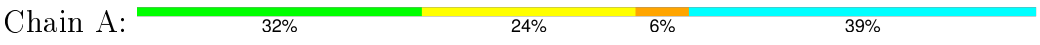
4.2 Scores per residue for each member of the ensemble

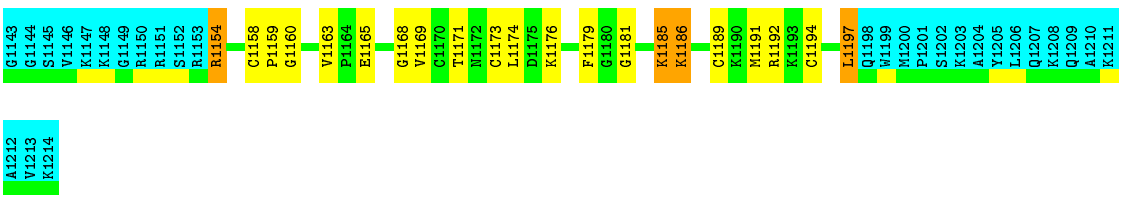
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: ZINC FINGER PROTEIN HRX

Chain A:

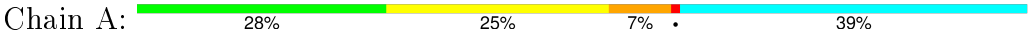


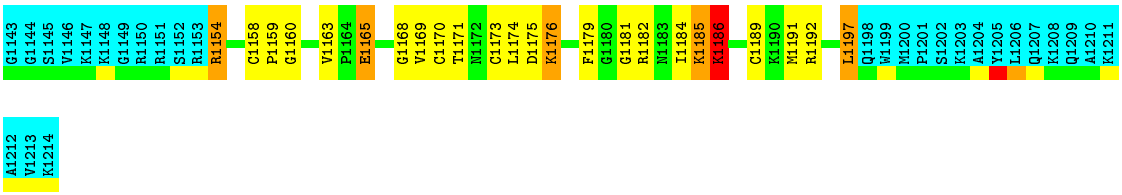


4.2.2 Score per residue for model 2

- Molecule 1: ZINC FINGER PROTEIN HRX

Chain A:

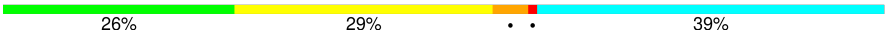


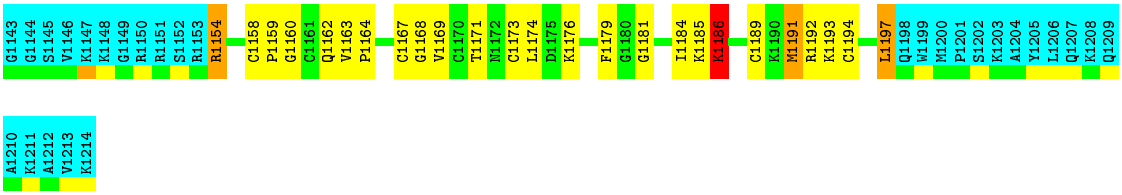


4.2.3 Score per residue for model 3

- Molecule 1: ZINC FINGER PROTEIN HRX

Chain A:

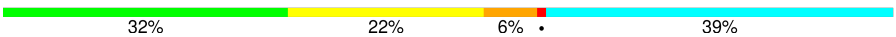


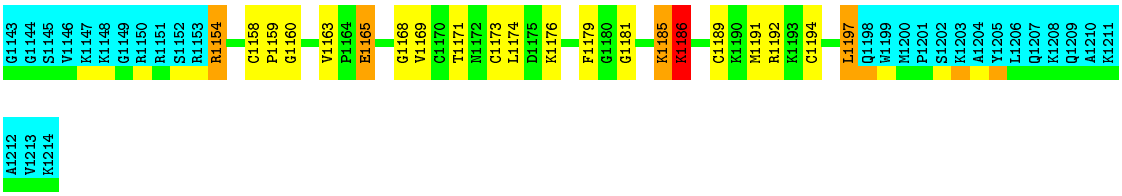


4.2.4 Score per residue for model 4

- Molecule 1: ZINC FINGER PROTEIN HRX

Chain A:

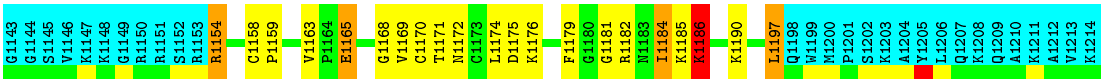




4.2.5 Score per residue for model 5

- Molecule 1: ZINC FINGER PROTEIN HRX

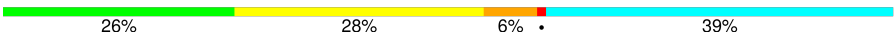
Chain A:

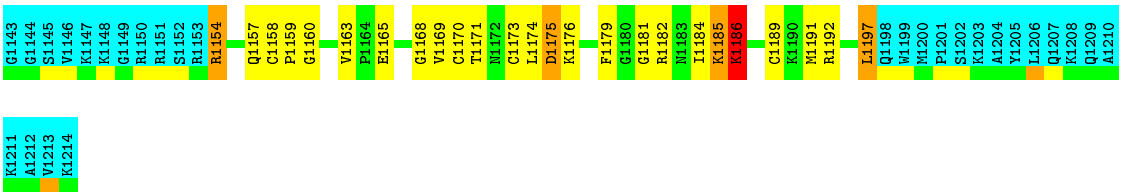


4.2.6 Score per residue for model 6

- Molecule 1: ZINC FINGER PROTEIN HRX

Chain A:

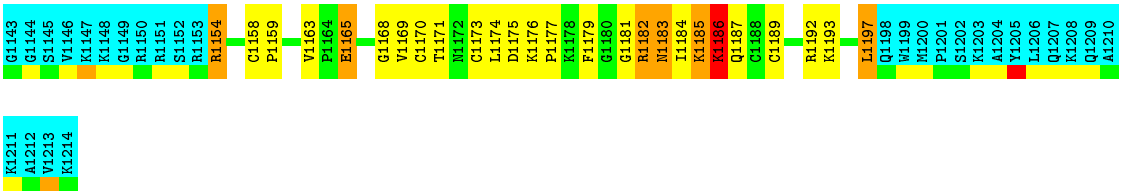




4.2.7 Score per residue for model 7

- Molecule 1: ZINC FINGER PROTEIN HRX

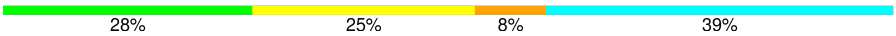
Chain A:

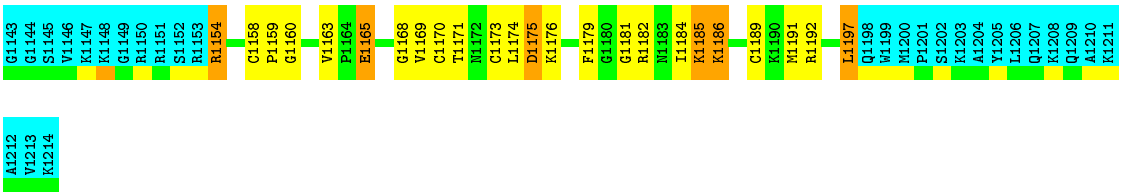


4.2.8 Score per residue for model 8

- Molecule 1: ZINC FINGER PROTEIN HRX

Chain A:

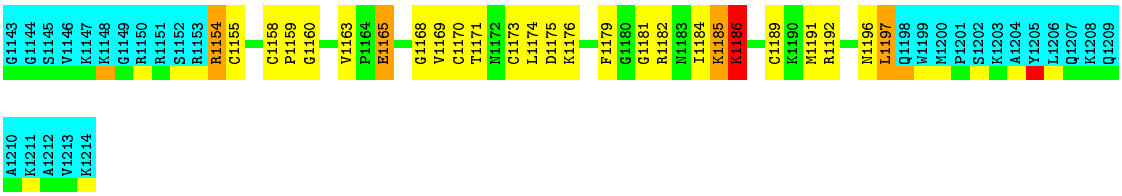




4.2.9 Score per residue for model 9

- Molecule 1: ZINC FINGER PROTEIN HRX

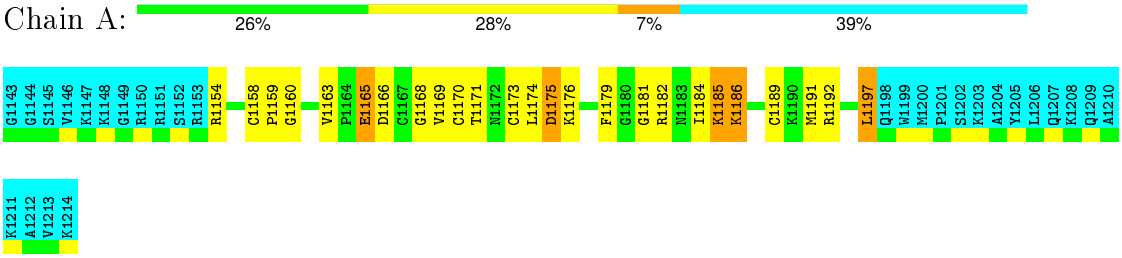
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: ZINC FINGER PROTEIN HRX

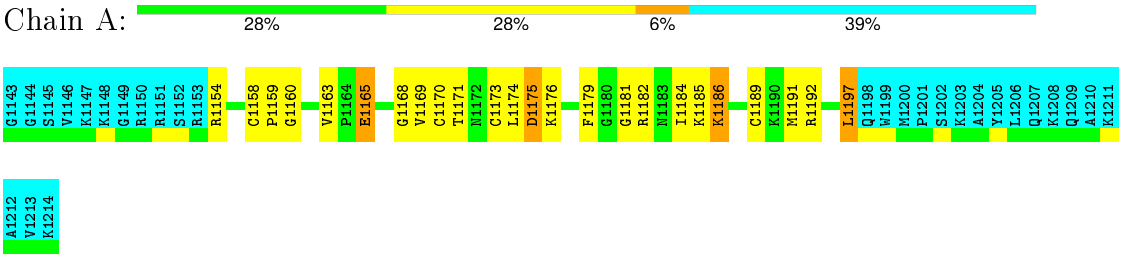
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: ZINC FINGER PROTEIN HRX

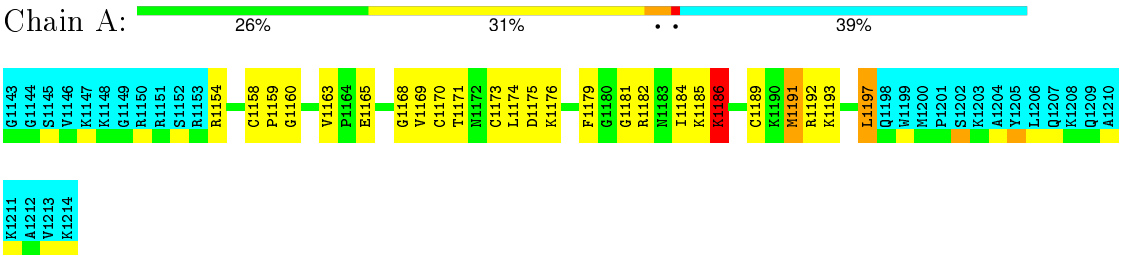
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: ZINC FINGER PROTEIN HRX

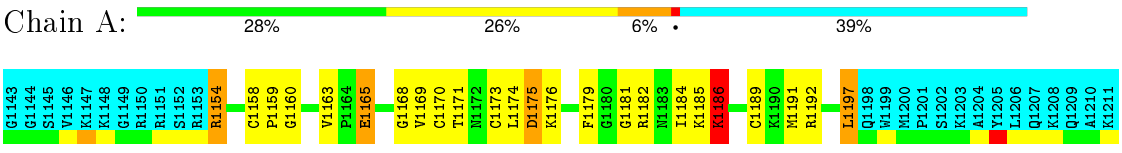
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: ZINC FINGER PROTEIN HRX

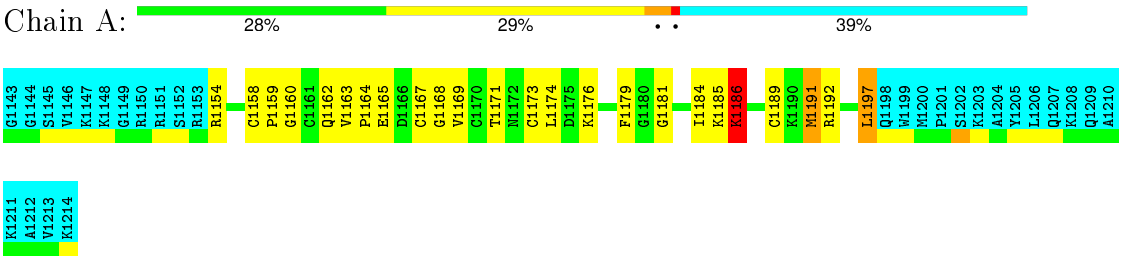
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: ZINC FINGER PROTEIN HRX

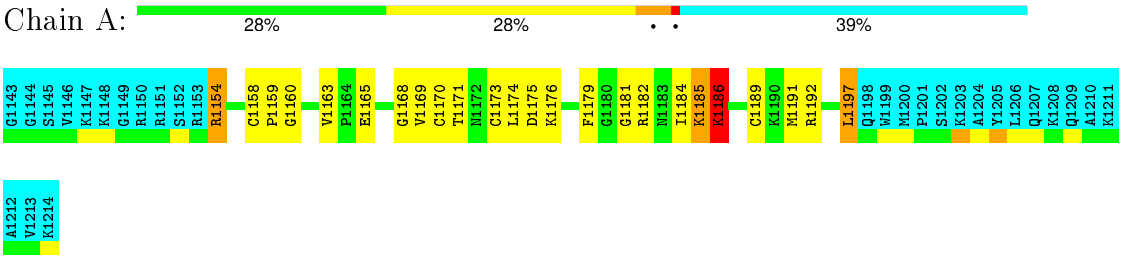
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: ZINC FINGER PROTEIN HRX

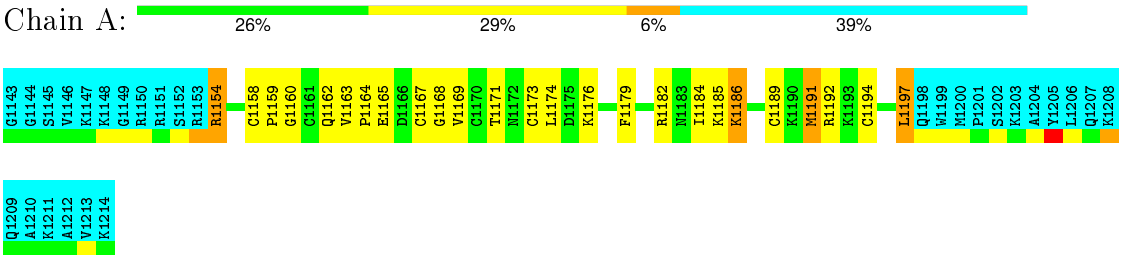
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: ZINC FINGER PROTEIN HRX

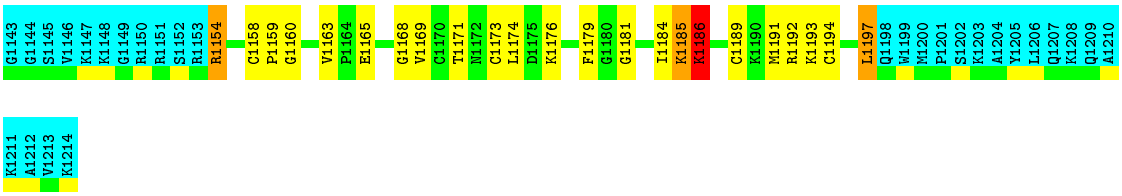
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: ZINC FINGER PROTEIN HRX

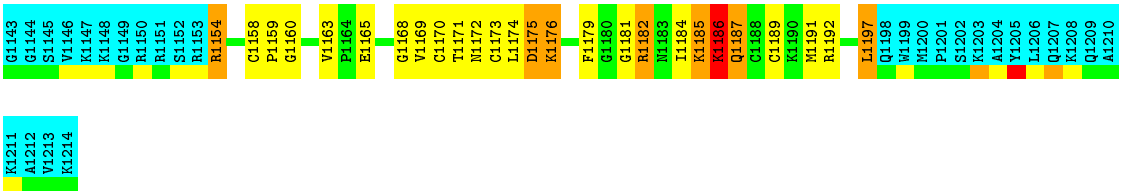
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: ZINC FINGER PROTEIN HRX

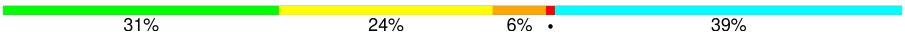
Chain A:

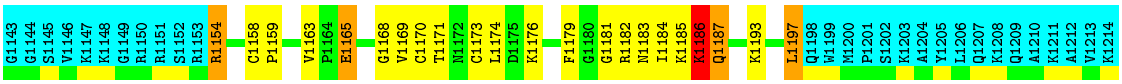


4.2.19 Score per residue for model 19

- Molecule 1: ZINC FINGER PROTEIN HRX

Chain A:

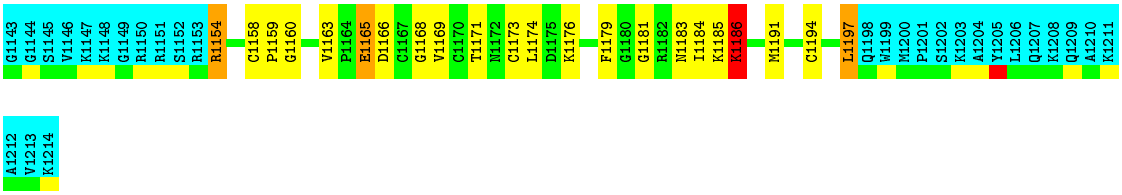




4.2.20 Score per residue for model 20 (medoid)

- Molecule 1: ZINC FINGER PROTEIN HRX

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *NO VIOLATIONS*.

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

Software name	Classification	Version
CNS	refinement	
ANSIG	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

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	331	334	330	24±2
All	All	6660	6680	6600	487

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 37.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:1168:GLY:HA3	1:A:1179:PHE:CE1	0.82	2.09	20	20
1:A:1169:VAL:HA	1:A:1174:LEU:HD21	0.77	1.56	8	20
1:A:1154:ARG:HA	1:A:1197:LEU:HD23	0.76	1.56	10	20
1:A:1168:GLY:CA	1:A:1179:PHE:CE1	0.74	2.71	10	20
1:A:1159:PRO:O	1:A:1163:VAL:HG23	0.71	1.85	8	20
1:A:1154:ARG:CA	1:A:1197:LEU:HD23	0.69	2.17	13	20
1:A:1169:VAL:O	1:A:1174:LEU:HD11	0.67	1.89	3	20
1:A:1163:VAL:HG21	1:A:1191:MET:HG3	0.67	1.67	14	2
1:A:1167:CYS:SG	1:A:1169:VAL:HG22	0.63	2.34	3	3
1:A:1163:VAL:HG11	1:A:1191:MET:CG	0.62	2.25	16	3
1:A:1163:VAL:HG12	1:A:1165:GLU:O	0.59	1.97	4	17
1:A:1182:ARG:NH1	1:A:1184:ILE:HD13	0.58	2.13	2	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:1168:GLY:CA	1:A:1179:PHE:CZ	0.56	2.89	14	20
1:A:1154:ARG:HG2	1:A:1197:LEU:CD2	0.56	2.31	10	3
1:A:1154:ARG:CG	1:A:1197:LEU:HD23	0.54	2.32	1	18
1:A:1185:LYS:O	1:A:1186:LYS:C	0.53	2.46	4	20
1:A:1158:CYS:HB2	1:A:1159:PRO:CD	0.53	2.34	1	20
1:A:1176:LYS:O	1:A:1181:GLY:N	0.53	2.41	20	19
1:A:1175:ASP:O	1:A:1182:ARG:O	0.53	2.27	18	9
1:A:1194:CYS:HB3	1:A:1197:LEU:HD21	0.53	1.80	16	6
1:A:1185:LYS:HB3	1:A:1187:GLN:NE2	0.50	2.21	19	1
1:A:1182:ARG:HB2	1:A:1184:ILE:HG23	0.50	1.84	7	3
1:A:1168:GLY:HA2	1:A:1179:PHE:CE1	0.48	2.42	2	16
1:A:1158:CYS:HB2	1:A:1159:PRO:HD2	0.48	1.86	3	20
1:A:1182:ARG:CB	1:A:1184:ILE:HG23	0.48	2.39	5	2
1:A:1168:GLY:HA2	1:A:1179:PHE:CZ	0.48	2.44	18	17
1:A:1160:GLY:HA2	1:A:1191:MET:O	0.48	2.08	16	17
1:A:1154:ARG:HG2	1:A:1197:LEU:HD23	0.47	1.86	10	3
1:A:1184:ILE:CG1	1:A:1185:LYS:N	0.46	2.78	11	9
1:A:1170:CYS:O	1:A:1174:LEU:HG	0.46	2.10	8	13
1:A:1154:ARG:HG3	1:A:1197:LEU:CD2	0.46	2.41	13	6
1:A:1173:CYS:O	1:A:1176:LYS:HB2	0.46	2.11	11	19
1:A:1189:CYS:HB3	1:A:1192:ARG:HB2	0.46	1.88	9	8
1:A:1154:ARG:CG	1:A:1197:LEU:CD2	0.45	2.94	5	11
1:A:1189:CYS:HB3	1:A:1192:ARG:CD	0.45	2.41	17	14
1:A:1162:GLN:O	1:A:1164:PRO:HD3	0.45	2.12	16	3
1:A:1172:ASN:OD1	1:A:1187:GLN:OE1	0.45	2.35	18	1
1:A:1168:GLY:CA	1:A:1179:PHE:CD1	0.45	3.00	20	3
1:A:1154:ARG:CB	1:A:1197:LEU:HD23	0.43	2.44	4	3
1:A:1184:ILE:HG13	1:A:1185:LYS:N	0.43	2.29	17	7
1:A:1172:ASN:ND2	1:A:1190:LYS:HG2	0.43	2.29	5	1
1:A:1174:LEU:HA	1:A:1179:PHE:HB3	0.42	1.91	14	18
1:A:1168:GLY:HA3	1:A:1179:PHE:CZ	0.42	2.48	14	1
1:A:1174:LEU:HA	1:A:1179:PHE:CB	0.42	2.45	14	1
1:A:1155:CYS:CB	1:A:1196:ASN:ND2	0.42	2.83	9	1
1:A:1166:ASP:HB3	1:A:1173:CYS:SG	0.41	2.55	10	2
1:A:1177:PRO:HD3	1:A:1183:ASN:HB2	0.41	1.91	7	1
1:A:1167:CYS:HB3	1:A:1191:MET:HG2	0.41	1.91	16	1
1:A:1154:ARG:HG3	1:A:1197:LEU:HD23	0.40	1.93	18	2
1:A:1168:GLY:HA2	1:A:1179:PHE:CE2	0.40	2.51	14	1
1:A:1182:ARG:HB2	1:A:1184:ILE:HD12	0.40	1.93	5	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	44/72 (61%)	40±1 (90±3%)	3±1 (8±3%)	1±0 (2±0%)	12 51
All	All	880/1440 (61%)	792 (90%)	68 (8%)	20 (2%)	12 51

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	1186	LYS	20

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	39/61 (64%)	33±2 (84±4%)	6±2 (16±4%)	7 44
All	All	780/1220 (64%)	655 (84%)	125 (16%)	7 44

All 15 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	1197	LEU	20
1	A	1171	THR	20
1	A	1154	ARG	16
1	A	1186	LYS	15
1	A	1165	GLU	13
1	A	1185	LYS	11
1	A	1175	ASP	9
1	A	1193	LYS	5
1	A	1191	MET	4
1	A	1183	ASN	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	1187	GLN	3
1	A	1176	LYS	2
1	A	1182	ARG	2
1	A	1157	GLN	1
1	A	1184	ILE	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

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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