wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 03:17 PM BST

PDB ID	:	1HCT
Title	:	NMR STRUCTURE OF THE HUMAN SRC SH2 DOMAIN COMPLEX
Authors	:	Gampe Junior, R.T.; Xu, R.X.
Deposited on	:	1994-09-02

This is a wwPDB NMR Structure Validation Summary 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:

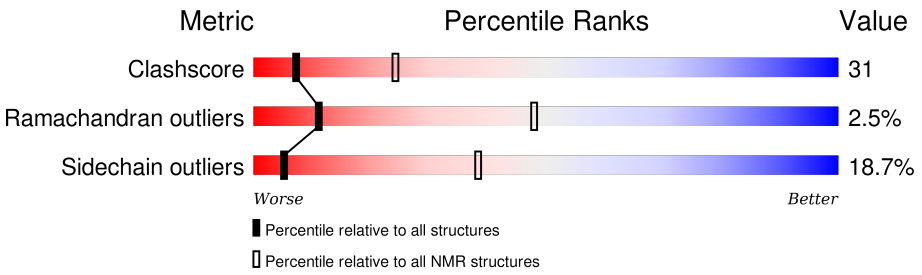
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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 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	6	17% 33% 50%
2	B	107	43% 43% 8% 6%

2 Ensemble composition and analysis (i)

This entry contains 23 models. Model 3 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:102-A:104, B:146-B:246 (104)	0.42	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 4 clusters and 6 single-model clusters were found.

Cluster number	Models
1	1, 3, 5, 9, 11, 13, 15, 17, 21, 22
2	10, 18, 19
3	12, 14
4	20, 23
Single-model clusters	2; 4; 6; 7; 8; 16

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1811 atoms, of which 891 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called ACETYL-PYEEIE-OH.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	6	95	32	40	5	17	1	0

- Molecule 2 is a protein called HUMAN SRC.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	107	1716	543	851	154	164	4	0

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: ACETYL-PYEEIE-OH

Chain A: 17% 33% 50%

?100 Y101 E102 E103 I104 E105

- Molecule 2: HUMAN SRC

Chain B: 43% 43% 8% 6%

M140 D141 S142 I143 Q144 A145 E146 E147 W148 Y149 F150 G151 K152 R155 R156 R160 L161 L162 L163 N164 R169 F172 L173 V174 R175 E176 S177 E178 T179 T180 K181 Y184 C185 L186 S187 D192 L197 N198 V199 Y202 K203 I204 R205 S209 G210 G211 F212 Y213 I214 R217

F220 L223 L226 V227 Y230 S231 K232 H233 A234 D235 G236 L237 R240 L241 T242 T243 V244 C245 P246

4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 3. Colouring as in section 4.1 above.

- Molecule 1: ACETYL-PYEEIE-OH

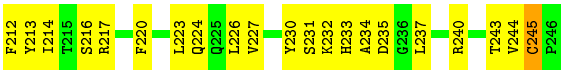
Chain A: 17% 33% 50%

?100 Y101 E102 E103 I104 E105

- Molecule 2: HUMAN SRC

Chain B: 41% 46% 7% 6%

M140 D141 S142 I143 Q144 A145 E146 E147 W148 Y149 F150 G151 K152 R155 R156 E157 S158 E159 R160 L161 L162 L163 N164 N167 P168 R169 F172 L173 V174 R175 E176 S177 E178 T179 T180 K181 L186 S187 V188 S189 D192 L197 N198 V199 Y202 K203 I204 R205 K206 S209 G210 G211


F212
Y213
I214
T215
S216
R217
F220
L223
Q224
Q225
L226
V227
Y230
S231
K232
H233
A234
D235
G236
L237
R240
T243
V244
G245
P246

5 Refinement protocol and experimental data overview (i)

Of the ? calculated structures, 23 were deposited, based on the following criterion: *?*.

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

Software name	Classification	Version
X-PLOR	refinement	

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 (i)

6.1 Standard geometry (i)

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

There are no covalent bond-length or bond-angle outliers.

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
2	B	0.0±0.0	8.0±0.2
All	All	0	183

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

5 of 8 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
2	B	160	ARG	Sidechain	23
2	B	169	ARG	Sidechain	23
2	B	217	ARG	Sidechain	23
2	B	156	ARG	Sidechain	23
2	B	175	ARG	Sidechain	23

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	26	23	23	2±2
2	B	821	807	807	51±9
All	All	19481	19090	19090	1182

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 31.

5 of 428 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:184:TYR:CE2	2:B:223:LEU:HD11	0.94	1.95	9	6
2:B:163:LEU:HD21	2:B:199:VAL:HG23	0.93	1.37	18	23
2:B:174:VAL:HG22	2:B:186:LEU:HD13	0.91	1.42	19	11
2:B:231:SER:OG	2:B:241:LEU:HD12	0.90	1.65	2	7
2:B:173:LEU:HD12	2:B:175:ARG:NH1	0.85	1.86	20	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	3/6 (50%)	3±0 (96±11%)	0±0 (3±9%)	0±0 (1±7%)	19	64
2	B	100/107 (93%)	85±3 (85±3%)	13±3 (13±3%)	3±1 (3±1%)	11	47
All	All	2369/2599 (91%)	2017 (85%)	292 (12%)	60 (3%)	11	48

5 of 13 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	210	GLY	20
2	B	177	SER	6
2	B	237	LEU	6
2	B	217	ARG	6
2	B	213	TYR	4

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	3/4 (75%)	3±0 (94±13%)	0±0 (6±13%)	29 75
2	B	89/94 (95%)	72±2 (81±3%)	17±2 (19±3%)	5 37
All	All	2116/2254 (94%)	1721 (81%)	395 (19%)	5 38

5 of 55 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)
2	B	180	THR	21
2	B	197	LEU	17
2	B	150	PHE	17
2	B	203	LYS	17
2	B	178	GLU	16

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

1 non-standard protein/DNA/RNA residue 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
1	PTR	A	101	1	13,16,17	1.10±0.01	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
1	PTR	A	101	1	19,22,24	1.05±0.02	0±0 (0±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
1	PTR	A	101	1	-	0±0,9,11,13	0±0,1,1,1

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.5 Carbohydrates (i)

There are no carbohydrates 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