wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 06:00 AM BST

PDB ID	:	2SRT
Title	:	CATALYTIC DOMAIN OF HUMAN STROMELYSIN-1 AT PH 5.5 AND 40OC COMPLEXED WITH INHIBITOR
Authors	:	Gooley, P.R.; O'Connell, J.F.
Deposited on	:	1995-03-22

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:

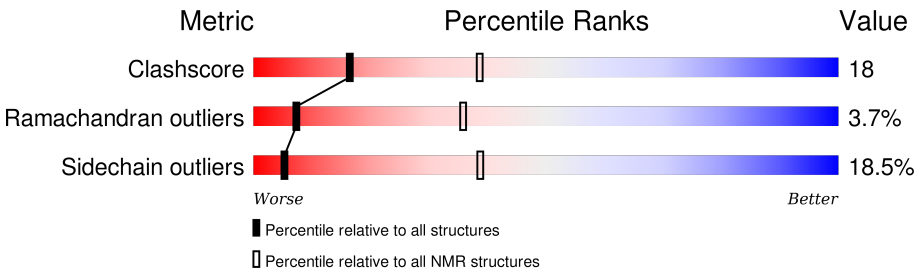
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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	173	61% 25% 7% 7%

2 Ensemble composition and analysis (i)

This entry contains 30 models. Model 10 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:84-A:157, A:161-A:170, A:174-A:250 (161)	0.34	10

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 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 4, 5, 6, 8, 9, 10, 11, 14, 16, 18, 19, 20, 25
2	3, 7, 12, 13, 15, 22, 24, 26
3	28, 29
4	17, 23
5	27, 30
Single-model clusters	21

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2746 atoms, of which 1333 are hydrogens and 0 are deuteriums.

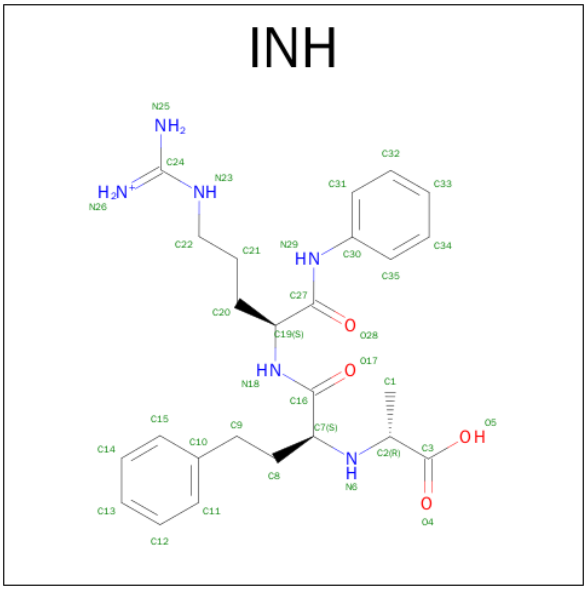
- Molecule 1 is a protein called STROMELYSIN-1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	173	2674	882	1298	228	264	2	0

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

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	2	2	2

- Molecule 3 is N-(R-CARBOXY-ETHYL)-ALPHA-(S)-(2-PHENYLETHYL)GLYCYL-L-ARGININE-N-PHENYLAMIDE (three-letter code: INH) (formula: $C_{25}H_{35}N_6O_4$).



Mol	Chain	Residues	Atoms				
			Total	C	H	N	O
3	A	1	70	25	35	6	4

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: STROMELYSIN-1

Chain A: 61% 25% 7% 7%

F83 F86 I89 P90 K91 T98 T105 P106 D107 S115 K119 A120 V123 W124 V127 R134 M143 V148 R149 E150 H151 G152 D153 F154 Y155 P156 F157 D158 G159 P160 G161 N162 V163 H166 A167 Y168 A169 P170 G171 P172 G173 I174 D177 D182 T187

G192 F196 L197 A200 H201 E202 H205 G208 L209 F210 H211 S212 A213 N214 L218 M219 Y220 P221 L222 L229 T230 R231 F232 R233 L234 I239 I242 D251 S252 P253 E254 T255

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

The representative model is number 10. Colouring as in section 4.1 above.

- Molecule 1: STROMELYSIN-1

Chain A: 58% 25% 8% • 7%

F83 F86 P87 G88 I89 P90 K91 T98 Y99 R100 I101 T105 P106 D107 K110 S115 K119 A120 V123 W124 V127 R134 M143 A147 V148 R149 E150 H151 G152 D153 F154 Y155 P156 F157 D158 G159 P160 G161 N162 V163 H166 A167 Y168 A169 P170 G171 P172 G173

I174 D177 D181 Q185 W186 T187 G192 F196 L197 A200 H201 E202 H205 L209 F210 H211 L218 D228 L229 T230 R231 F232 R233 L234 S235 Q236 I239 N240 G241 I242 L245 D251 S252 P253 E254 T255

5 Refinement protocol and experimental data overview (i)

Of the ? calculated structures, 30 were deposited, based on the following criterion: *?*.

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

Software name	Classification	Version
DIANA	refinement	
FANTOM	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: ZN, INH

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
1	A	0.0±0.0	1.9±0.9
All	All	0	57

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

5 of 7 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	149	ARG	Sidechain	15
1	A	93	ARG	Sidechain	9
1	A	134	ARG	Sidechain	8
1	A	233	ARG	Sidechain	8
1	A	231	ARG	Sidechain	7

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	1293	1231	1231	46±5
3	A	35	35	34	3±2
All	All	39900	37980	37950	1437

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 18.

5 of 300 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:209:LEU:HD21	1:A:242:ILE:HD12	0.93	1.34	17	8
3:A:256:INH:C16	3:A:256:INH:H36	0.88	1.97	25	4
3:A:256:INH:H36	3:A:256:INH:C16	0.87	1.97	20	3
1:A:200:ALA:HB3	1:A:218:LEU:HD11	0.87	1.44	21	29
3:A:256:INH:H36	3:A:256:INH:N18	0.85	1.86	25	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	161/173 (93%)	137±2 (85±1%)	18±3 (11±2%)	6±1 (4±1%)	7	36
All	All	4830/5190 (93%)	4106 (85%)	545 (11%)	179 (4%)	7	36

5 of 17 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	149	ARG	27
1	A	89	ILE	27
1	A	91	LYS	25
1	A	163	VAL	21
1	A	107	ASP	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	138/147 (94%)	113±3 (82±2%)	26±3 (18±2%)	5 39
All	All	4140/4410 (94%)	3375 (82%)	765 (18%)	5 39

5 of 73 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	162	ASN	30
1	A	218	LEU	30
1	A	155	TYR	30
1	A	229	LEU	30
1	A	98	THR	30

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 carbohydrates in this entry.

6.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 for Mogul analysis.

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
3	INH	A	256	2	33,36,36	1.04±0.00	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
3	INH	A	256	2	37,46,46	0.83±0.00	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
3	INH	A	256	2	-	0±0,32,36,36	0±0,2,2,2

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.

6.7 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