wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 05:49 PM EDT

PDB ID	:	2RON
Title	:	The external thioesterase of the Surfactin-Synthetase
Authors	:	Koglin, A.; Lohr, F.; Bernhard, F.; Rogov, V.V.; Frueh, D.P.; Strieter, E.R.; Mofid, M.R.; Guentert, P.; Wagner, G.; Walsh, C.T.; Marahiel, M.A.; Doetsch, V.
Deposited on	:	2008-04-04

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
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

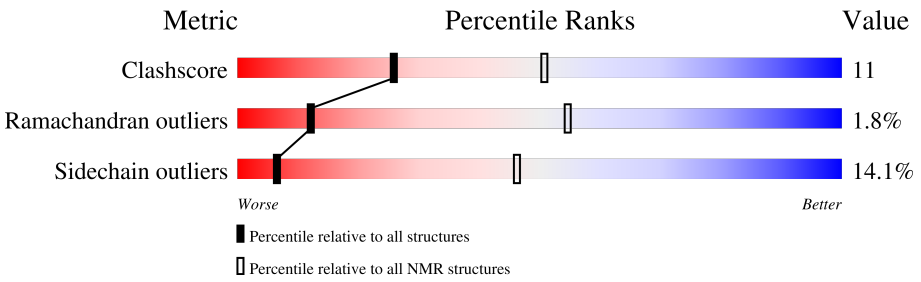
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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	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	242	71% 22% • 6%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 7 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:2-A:46, A:56-A:238 (228)	0.95	7

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 3 clusters and 5 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3850 atoms, of which 1902 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Surfactin synthetase thioesterase subunit.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	242	3850	1253	1902	329	356	10	0

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: Surfactin synthetase thioesterase subunit

Chain A: 71% 22% • 6%

M1 K6 L16 I17 P20 S26 F29 R30 P31 L32 L43 P47 P48 G49 H50 G51 T52 N53 Q54 T55 I58 E59 D60 L61 L64 R78 P79 L82 F83 G84 H85 G89 F93 L99 I104 F105 P106 Q107 A108 V109 I110 I111 I114

I119 A144 E145 L146 V147 V152 F156 L157 I178 Q179 S180 P181 V182 H183 V184 G187 L188 D189 D190 I194 R195 K201 K202 K205 T208 Q211 F212 D213 M217 L234 H237 P238 I239 I240 Q241 P242

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

The representative model is number 7. Colouring as in section 4.1 above.

- Molecule 1: Surfactin synthetase thioesterase subunit

Chain A: 66% 24% • 6%

M1 K6 S7 F8 Q15 L16 I17 R30 P31 L32 H33 L43 A44 P47 P48 G49 H50 G51 T52 N53 Q54 T55 S56 A57 I58 E59 D60 L61 L64 R78 P79 L82 F83 G84 H85 S86 M90 F93 R94 L95 E100 I104 F105 P106 Q107 A108 V109 I110

I111 I114 Q115 H118 I119 I136 I137 Q138 A144 E145 L146 V147 E148 N149 V152 F156 L157 F160 R161 L172 L175 S180 P181 V182 G187 L188 D189 D190 K191 K192 C193 I194 R195 K201 K205 D206 I207 T208 L234 N235 Q236 H237 P238 I239 I240 Q241

P242

5 Refinement protocol and experimental data overview (i)

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

Of the 150 calculated structures, 20 were deposited, based on the following criterion: *target function.*

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

Software name	Classification	Version
CYANA	structure solution	2.1
CYANA	refinement	2.1

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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 (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	1845	1801	1801	41±5
All	All	36900	36020	36020	822

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 11.

5 of 270 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:152:VAL:HG23	1:A:157:LEU:HD22	0.94	1.36	12	5
1:A:16:LEU:HD11	1:A:58:ILE:HD12	0.90	1.39	20	4
1:A:152:VAL:HG22	1:A:157:LEU:HD13	0.82	1.52	13	6
1:A:152:VAL:HG12	1:A:157:LEU:HD22	0.81	1.52	8	1
1:A:144:ALA:HB1	1:A:147:VAL:HG22	0.78	1.53	17	20

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	228/242 (94%)	178±4 (78±2%)	46±5 (20±2%)	4±1 (2±1%)	12 54
All	All	4560/4840 (94%)	3553 (78%)	926 (20%)	81 (2%)	12 54

5 of 23 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	180	SER	20
1	A	89	GLY	14
1	A	104	ILE	9
1	A	159	SER	5
1	A	141	GLY	5

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	198/210 (94%)	170±4 (86±2%)	28±4 (14±2%)	6 46
All	All	3960/4200 (94%)	3403 (86%)	557 (14%)	6 46

5 of 122 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	180	SER	20
1	A	194	ILE	20
1	A	201	LYS	20
1	A	93	PHE	19
1	A	111	ILE	19

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 monosaccharides 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