wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 05:13 AM BST

PDB ID	:	2ROQ
Title	:	Solution Structure of the thiolation-thioesterase di-domain of enterobactin synthetase component F
Authors	:	Frueh, D.P.; Arthanari, H.; Koglin, A.; Vosburg, D.A.; Bennett, A.E.; Walsh, C.T.; Wagner, G.
Deposited on	:	2008-04-05

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:

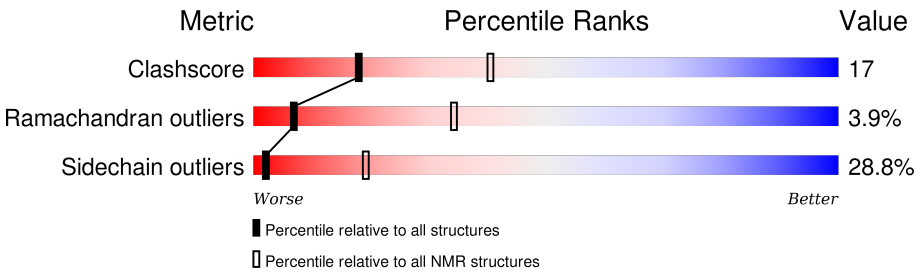
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	343	44% 38% • 14%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:11-A:86, A:97-A:208, A:216-A:219, A:224-A:287, A:301-A:338 (294)	1.12	8

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 8 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 5227 atoms, of which 2585 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Enterobactin synthetase component F.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	343	5227	1670	2585	465	497	10	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	EXPRESSION TAG	UNP P11454
A	48	ALA	SER	ENGINEERED	UNP P11454
A	336	LEU	-	EXPRESSION TAG	UNP P11454
A	337	GLU	-	EXPRESSION TAG	UNP P11454
A	338	HIS	-	EXPRESSION TAG	UNP P11454
A	339	HIS	-	EXPRESSION TAG	UNP P11454
A	340	HIS	-	EXPRESSION TAG	UNP P11454
A	341	HIS	-	EXPRESSION TAG	UNP P11454
A	342	HIS	-	EXPRESSION TAG	UNP P11454
A	343	HIS	-	EXPRESSION TAG	UNP P11454

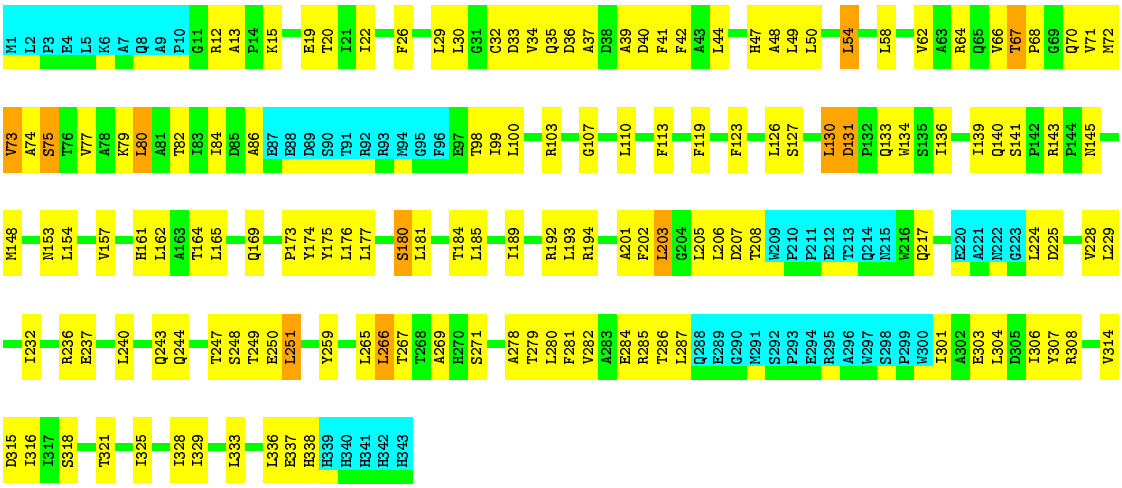
4 Residue-property plots (i)

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: Enterobactin synthetase component F

Chain A: 44% 38% • 14%

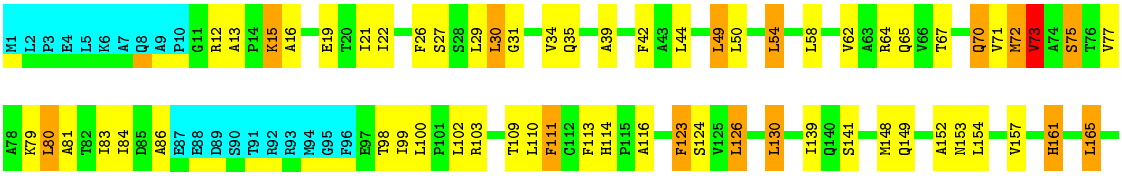


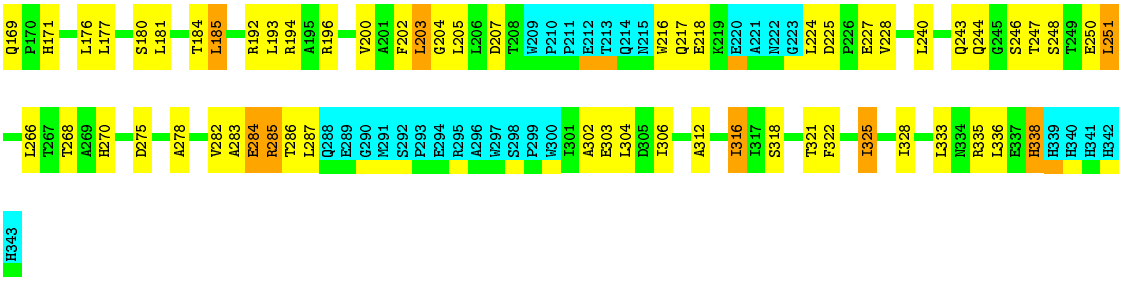
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 8. Colouring as in section 4.1 above.

- Molecule 1: Enterobactin synthetase component F

Chain A: 50% 29% 6% 14%




Q169
P170
H171
L176
L177
S180
L181
T184
L185
R192
L193
R194
A195
R196
V200
A201
F202
L203
G204
L205
L206
D207
T208
W209
P210
P211
E212
T213
Q214
N215
W216
Q217
E218
K219
E220
A221
N222
G223
L224
D225
P226
E227
V228
L240
Q243
Q244
G245
S246
T247
S248
T249
E250
L251
L266
T267
T268
A269
H270
D275
A278
V282
A283
E284
R285
T286
L287
Q288
E289
G290
M291
S292
P293
E294
R295
A296
W297
S298
P299
W300
I301
A302
E303
L304
D305
I306
A312
I316
I317
S318
T321
F322
I325
I328
L333
N334
R335
L336
E337
H338
H339
H340
H341
H342
H343

5 Refinement protocol and experimental data overview (i)

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
CYANA	structure solution	2.1
CNS	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)

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	2237	2228	2228	78±9
All	All	44740	44560	44560	1556

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 17.

5 of 791 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:62:VAL:HG21	1:A:84:ILE:HD12	0.99	1.29	11	1
1:A:99:ILE:HD12	1:A:139:ILE:HD12	0.99	1.30	8	4
1:A:72:MET:O	1:A:73:VAL:HG22	0.97	1.58	8	3
1:A:71:VAL:HG13	1:A:80:LEU:HD11	0.96	1.33	8	3
1:A:154:LEU:HD22	1:A:185:LEU:HD21	0.95	1.33	12	10

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	294/343 (86%)	241±5 (82±2%)	41±6 (14±2%)	11±3 (4±1%)	7 34
All	All	5880/6860 (86%)	4825 (82%)	826 (14%)	229 (4%)	7 34

5 of 55 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	19
1	A	75	SER	18
1	A	107	GLY	11
1	A	145	ASN	11
1	A	73	VAL	10

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	231/273 (85%)	164±5 (71±2%)	67±5 (29±2%)	2 19
All	All	4620/5460 (85%)	3289 (71%)	1331 (29%)	2 19

5 of 187 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	251	LEU	20
1	A	285	ARG	19
1	A	67	THR	19
1	A	318	SER	19
1	A	266	LEU	18

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)

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