wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 06:12 PM BST

PDB ID	:	1X9A
Title	:	Solution NMR Structure of Protein Tm0979 from Thermotoga maritima. Ontario Center for Structural Proteomics Target TM0979_1_87; Northeast Structural Genomics Consortium Target VT98.
Authors	:	Gaspar, J.A.; Liu, C.; Vassall, K.A.; Stathopulos, P.B.; Meglei, G.; Stephen, R.; Pineda-Lucena, A.; Wu, B.; Yee, A.; Arrowsmith, C.H.; Meiering, E.M.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2004-08-20

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:

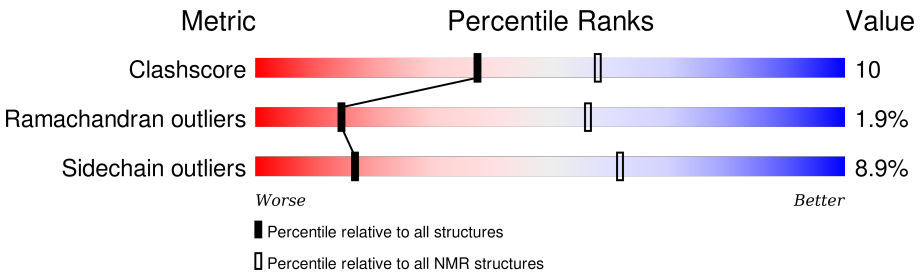
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 is 76%.

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	107	62% 15% • • 19%
1	B	107	55% 17% • 7% 19%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 3 is the overall representative, medoid model (most similar to other models). The authors have identified model 9 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:2-A:87, B:201-B:280 (166)	0.57	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 3 clusters. No single-model clusters were found.

Cluster number	Models
1	2, 4, 7, 9
2	3, 6, 8
3	1, 5, 10

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2812 atoms, of which 1416 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called hypothetical protein TM0979.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	87	1406	455	708	107	135	1	0
1	B	87	1406	455	708	107	135	1	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	EXPRESSION TAG	UNP Q9X074
A	-18	GLY	-	EXPRESSION TAG	UNP Q9X074
A	-17	SER	-	EXPRESSION TAG	UNP Q9X074
A	-16	SER	-	EXPRESSION TAG	UNP Q9X074
A	-15	HIS	-	EXPRESSION TAG	UNP Q9X074
A	-14	HIS	-	EXPRESSION TAG	UNP Q9X074
A	-13	HIS	-	EXPRESSION TAG	UNP Q9X074
A	-12	HIS	-	EXPRESSION TAG	UNP Q9X074
A	-11	HIS	-	EXPRESSION TAG	UNP Q9X074
A	-10	HIS	-	EXPRESSION TAG	UNP Q9X074
A	-9	SER	-	EXPRESSION TAG	UNP Q9X074
A	-8	SER	-	EXPRESSION TAG	UNP Q9X074
A	-7	GLY	-	EXPRESSION TAG	UNP Q9X074
A	-6	LEU	-	EXPRESSION TAG	UNP Q9X074
A	-5	VAL	-	EXPRESSION TAG	UNP Q9X074
A	-4	PRO	-	EXPRESSION TAG	UNP Q9X074
A	-3	ARG	-	EXPRESSION TAG	UNP Q9X074
A	-2	GLY	-	EXPRESSION TAG	UNP Q9X074
A	-1	SER	-	EXPRESSION TAG	UNP Q9X074
A	0	HIS	-	EXPRESSION TAG	UNP Q9X074
B	181	MET	-	EXPRESSION TAG	UNP Q9X074
B	182	GLY	-	EXPRESSION TAG	UNP Q9X074
B	183	SER	-	EXPRESSION TAG	UNP Q9X074
B	184	SER	-	EXPRESSION TAG	UNP Q9X074
B	185	HIS	-	EXPRESSION TAG	UNP Q9X074
B	186	HIS	-	EXPRESSION TAG	UNP Q9X074
B	187	HIS	-	EXPRESSION TAG	UNP Q9X074
B	188	HIS	-	EXPRESSION TAG	UNP Q9X074
B	189	HIS	-	EXPRESSION TAG	UNP Q9X074

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	190	HIS	-	EXPRESSION TAG	UNP Q9X074
B	191	SER	-	EXPRESSION TAG	UNP Q9X074
B	192	SER	-	EXPRESSION TAG	UNP Q9X074
B	193	GLY	-	EXPRESSION TAG	UNP Q9X074
B	194	LEU	-	EXPRESSION TAG	UNP Q9X074
B	195	VAL	-	EXPRESSION TAG	UNP Q9X074
B	196	PRO	-	EXPRESSION TAG	UNP Q9X074
B	197	ARG	-	EXPRESSION TAG	UNP Q9X074
B	198	GLY	-	EXPRESSION TAG	UNP Q9X074
B	199	SER	-	EXPRESSION TAG	UNP Q9X074
B	200	HIS	-	EXPRESSION TAG	UNP Q9X074

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: hypothetical protein TM0979

Chain A: 62% 15% • • 19%

MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS M1 A2 L5 H12 E15 K16 L17 I31 Q32 W37 E40 E41 L42 K52 L69 I70 T71 Y72 S73 E74 F75 I76 D77 L78 L79 E80 K84 G87

- Molecule 1: hypothetical protein TM0979

Chain B: 55% 17% • 7% 19%

MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS M201 A202 L203 Y204 L205 H212 E215 K216 L217 I231 Q232 W237 L242 Y249 K252 R258 L269 I270 E274 F275 I276 D277 L278 L279 E280 G281 E282 E283 K284 F285

I286 G287

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: hypothetical protein TM0979

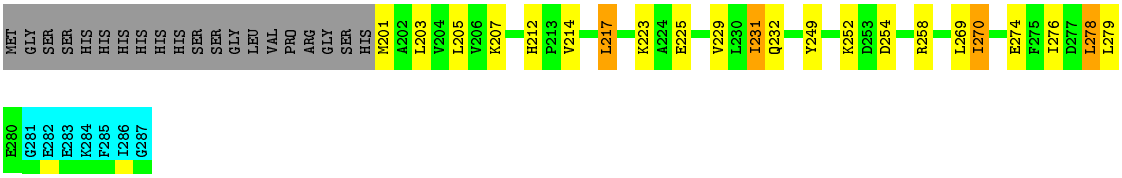
Chain A: 54% 21% 6% • 19%

MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS M1 A2 L5 D11 H12 P13 V14 E15 K16 L17 R20 E40 E41 L42 E43 I51 K52 L56 E62 S65 L69 I70 E74 F75 I76 D77 L78 L79 E80

K84 F85 I86 G87

- Molecule 1: hypothetical protein TM0979

Chain B: 54% 17% • 7% 19%


MET
GLY
SER
SER
HIS
HIS
HIS
HIS
HIS
HIS
SER
SER
GLY
LEU
VAL
PRO
ARG
GLY
SER
HIS
M201
A202
L203
V204
L205
V206
K207
H212
P213
V214
I217
K223
A224
E225
V229
I230
I231
Q232
Y249
K252
D253
D254
R258
L269
I270
E274
F275
I276
D277
I278
L279
E280
G281
E282
E283
K284
F285
I286
G287

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *distance geometry, simulated annealing, molecular dynamics.*

Of the 20 calculated structures, 10 were deposited, based on the following criterion: *structures with the least restraint violations.*

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

Software name	Classification	Version
DYANA	structure solution	
CNS	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	BMRB entry 6324
Number of chemical shift lists	1
Total number of shifts	1924
Number of shifts mapped to atoms	1924
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	76%

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	690	699	698	17±2
1	B	643	657	655	16±3
All	All	13330	13560	13530	263

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 10.

5 of 147 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:16:LYS:HE3	1:A:42:LEU:HB3	0.89	1.42	9	2
1:A:76:ILE:HG23	1:B:205:LEU:HD13	0.80	1.48	4	8
1:A:79:LEU:HB3	1:B:203:LEU:HD11	0.80	1.54	9	3
1:A:79:LEU:HD21	1:B:279:LEU:HD13	0.80	1.52	8	3
1:B:212:HIS:HB3	1:B:214:VAL:HG12	0.71	1.61	8	3

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	85/107 (79%)	76±2 (89±2%)	7±2 (9±2%)	2±1 (2±1%)	13	52
1	B	79/107 (74%)	71±2 (90±2%)	7±2 (8±2%)	1±1 (2±1%)	18	63
All	All	1640/2140 (77%)	1469 (90%)	140 (9%)	31 (2%)	14	56

5 of 9 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	84	LYS	7
1	B	242	LEU	6
1	A	82	GLU	4
1	B	224	ALA	4
1	A	2	ALA	3

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	74/92 (80%)	67±1 (90±1%)	7±1 (10±1%)	15	60
1	B	70/92 (76%)	64±1 (92±1%)	6±1 (8±1%)	19	64
All	All	1440/1840 (78%)	1312 (91%)	128 (9%)	17	62

5 of 43 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	B	278	LEU	10
1	A	78	LEU	10
1	A	70	ILE	10
1	B	217	LEU	10
1	B	270	ILE	9

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)

The completeness of assignment taking into account all chemical shift lists is 76% for the well-defined parts and 76% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 6324

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	1924
Number of shifts mapped to atoms	1924
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	172	0.54 ± 0.06	Should be applied
$^{13}C_{\beta}$	162	0.58 ± 0.13	Should be applied
$^{13}C'$	0	—	—
^{15}N	168	0.83 ± 0.38	Should be applied

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 76%, i.e. 1585 atoms were assigned a chemical shift out of a possible 2076. 30 out of 34 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	645/818 (79%)	323/326 (99%)	163/332 (49%)	159/160 (99%)
Sidechain	898/1091 (82%)	526/632 (83%)	368/426 (86%)	4/33 (12%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	42/167 (25%)	40/87 (46%)	0/74 (0%)	2/6 (33%)
Overall	1585/2076 (76%)	889/1045 (85%)	531/832 (64%)	165/199 (83%)

7.1.4 Statistically unusual chemical shifts (i)

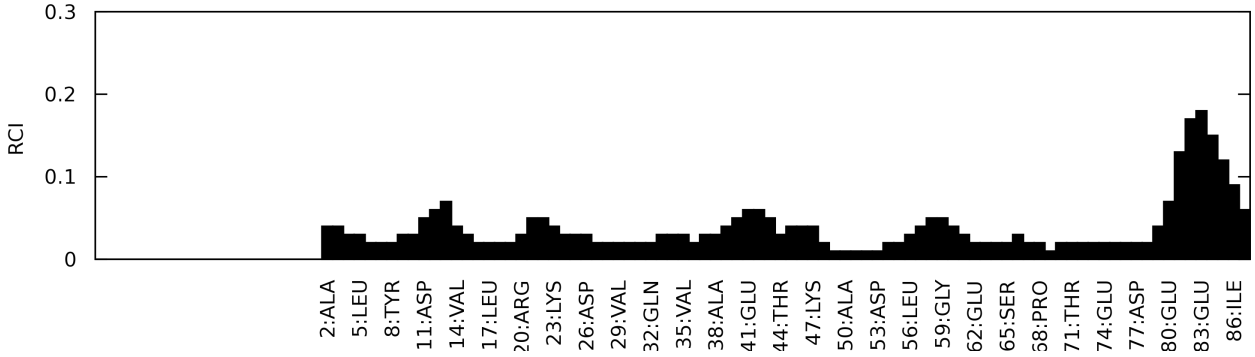
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	B	210	THR	HG21	-0.21	2.29 – -0.01	-5.9
1	B	210	THR	HG22	-0.21	2.29 – -0.01	-5.9
1	A	10	THR	HG21	-0.21	2.29 – -0.01	-5.9
1	B	210	THR	HG23	-0.21	2.29 – -0.01	-5.9
1	A	10	THR	HG22	-0.21	2.29 – -0.01	-5.9
1	A	10	THR	HG23	-0.21	2.29 – -0.01	-5.9

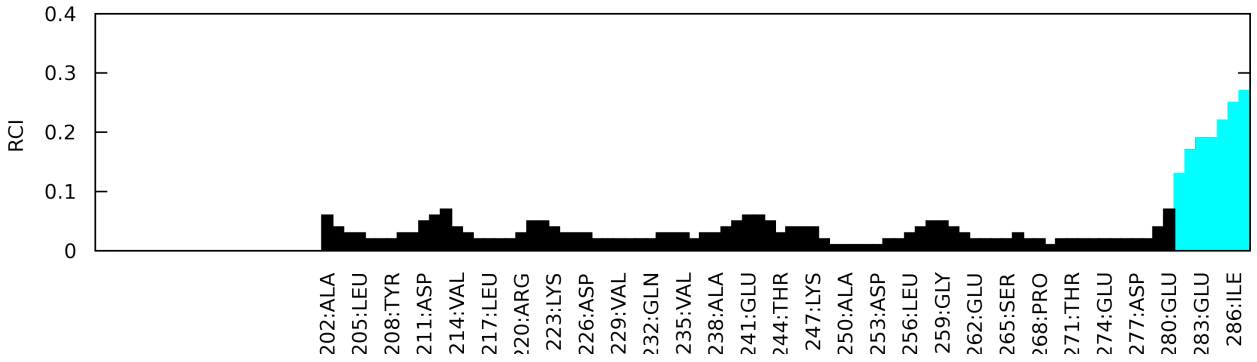
7.1.5 Random Coil Index (RCI) plots (i)

The images below report *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A:



Random coil index (RCI) for chain B:


RCI
0.4
0.3
0.2
0.1
0
202:ALA
205:LEU
208:TYR
211:ASP
214:VAL
217:LEU
220:ARG
223:LYS
226:ASP
229:VAL
232:GLN
235:VAL
238:ALA
241:GLU
244:THR
247:LYS
250:ALA
253:ASP
256:LEU
259:GLY
262:GLU
265:SER
268:PRO
271:THR
274:GLU
277:ASP
280:GLU
283:GLU
286:ILE