wwPDB NMR Structure Validation Summary Report (i)

May 28, 2020 – 09:08 pm BST

PDB ID	:	1YXE
Title	:	Structure and inter-domain interactions of domain II from the blood stage malarial protein, apical membrane antigen 1
Authors	:	Feng, Z.P.; Keizer, D.W.; Stevenson, R.A.; Yao, S.; Babon, J.J.; Murphy, V.J.; Anders, R.F.; Norton, R.S.
Deposited on	:	2005-02-21

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 following versions of software and data (see references (1)) were used in the production of this report:

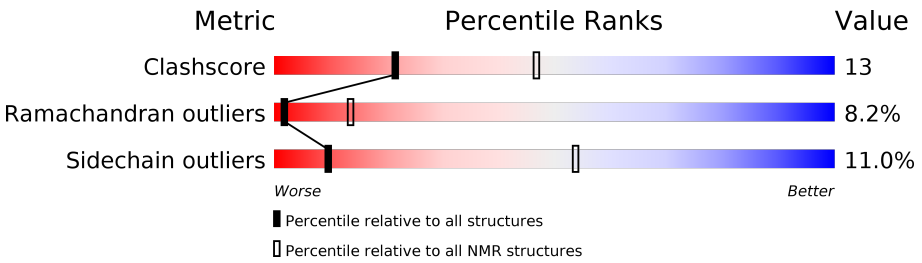
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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	140	25% 14% 61%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 11 is the overall representative, medoid model (most similar to other models). The authors have identified model 12 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:320-A:344, A:398-A:427 (55)	1.42	11

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 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2186 atoms, of which 1060 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called apical membrane antigen 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	140	2186	711	1060	197	212	6	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-12	MET	-	EXPRESSION TAG	UNP P22621
A	-11	ARG	-	EXPRESSION TAG	UNP P22621
A	-10	GLY	-	EXPRESSION TAG	UNP P22621
A	-9	SER	-	EXPRESSION TAG	UNP P22621
A	-8	HIS	-	EXPRESSION TAG	UNP P22621
A	-7	HIS	-	EXPRESSION TAG	UNP P22621
A	-6	HIS	-	EXPRESSION TAG	UNP P22621
A	-5	HIS	-	EXPRESSION TAG	UNP P22621
A	-4	HIS	-	EXPRESSION TAG	UNP P22621
A	-3	HIS	-	EXPRESSION TAG	UNP P22621
A	-2	GLY	-	EXPRESSION TAG	UNP P22621
A	-1	SER	-	EXPRESSION TAG	UNP P22621

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: apical membrane antigen 1

Chain A: 25% 14% 61%

M-12 R-11 G-10 S-9 H-8 H-7 H-6 H-5 H-4 H-3 G-2 S-1 N309 A310 K311 F312 G313 L314 W315 V316 D317 G318 N319 D322 I323 V326 F329 P330 A331 I332 L340 V341 F342 E343 L344 S345 A346 S347 D348 Q349 P350 K351 Q352 Y353 E354 Q355 H356 L357 T358 D359 Y360 E361 K362 I363 K364 E365

G366 F367 K368 N369 K370 N371 A372 S373 N374 I375 K376 S377 A378 F379 L380 P381 T382 G383 A384 F385 K386 A387 D388 R389 Y390 K391 S392 H393 G394 K395 G396 Y397 N398 W399 Y402 Q407 I411 F412 K415 P416 T417 C418 L419 I420 I426 A427 T428 T429 A430 L431 S432 H433 P434 I435 E436

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

The representative model is number 11. Colouring as in section 4.1 above.

- Molecule 1: apical membrane antigen 1

Chain A: 26% 10% • 61%

M-12 R-11 G-10 S-9 H-8 H-7 H-6 H-5 H-4 H-3 G-2 S-1 N309 A310 K311 F312 G313 L314 W315 V316 D317 G318 N319 D322 I323 V326 I332 D333 L334 F335 E336 C337 V341 S345 A346 S347 D348 Q349 P350 K351 Q352 Y353 E354 Q355 H356 L357 T358 D359 Y360 E361 K362 I363 K364 E365 G366

F367 K368 N369 K370 N371 A372 S373 N374 I375 K376 S377 A378 F379 L380 P381 T382 G383 A384 F385 K386 A387 D388 R389 Y390 K391 S392 H393 G394 K395 G396 Y397 N398 W399 Y402 Q407 K408 C409 E410 I411 F412 M413 P416 L419 I426 A427 T428 T429 A430 L431 S432 H433 P434 I435 E436

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing*.

Of the 100 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
X-PLOR	structure solution	NIH
X-PLOR	refinement	NIH

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	447	417	417	11±4
All	All	8940	8340	8340	225

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 13.

5 of 151 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:411:ILE:HD13	1:A:412:PHE:N	0.78	1.94	9	5
1:A:400:GLY:HA3	1:A:411:ILE:HD13	0.78	1.56	2	1
1:A:344:LEU:HD11	1:A:399:TRP:CZ3	0.75	2.16	6	1
1:A:326:VAL:HG12	1:A:412:PHE:CD1	0.73	2.18	2	3
1:A:323:ILE:HG21	1:A:420:ILE:HG21	0.72	1.60	5	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	55/140 (39%)	35±3 (64±6%)	16±3 (28±5%)	5±2 (8±3%)	2 13
All	All	1100/2800 (39%)	700 (64%)	310 (28%)	90 (8%)	2 13

5 of 26 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	407	GLN	14
1	A	412	PHE	10
1	A	322	ASP	6
1	A	416	PRO	6
1	A	331	ALA	6

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	52/122 (43%)	46±1 (89±3%)	6±1 (11±3%)	10 54
All	All	1040/2440 (43%)	926 (89%)	114 (11%)	10 54

5 of 38 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	329	PHE	10
1	A	411	ILE	9
1	A	415	LYS	6
1	A	334	LEU	6
1	A	419	LEU	5

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