wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 11:34 PM BST

PDB ID	:	2KHW
Title	:	Solution Structure of the human Polymerase iota UBM2-Ubiquitin Complex
Authors	:	Bomar, M.G.; D'Souza, S.; Bienko, M.; Dikic, I.; Walker, G.
Deposited on	:	2009-04-13

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:

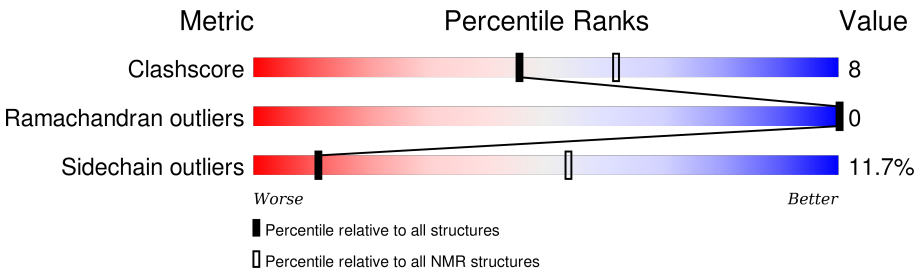
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	108	18% 6% 6% 70%
2	B	79	67% 24% • 6%

2 Ensemble composition and analysis (i)

This entry contains 25 models. Model 4 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:64-A:88, B:201-B:272 (97)	0.12	4

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 6 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1743 atoms, of which 885 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Immunoglobulin G-binding protein G, DNA polymerase iota.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	32	528	174	263	41	50	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	EXPRESSION TAG	UNP P19909
A	2	GLN	-	EXPRESSION TAG	UNP P19909
A	57	GLY	-	SEE REMARK 999	UNP Q9UNA4
A	58	SER	-	SEE REMARK 999	UNP Q9UNA4
A	59	ASP	-	SEE REMARK 999	UNP Q9UNA4
A	60	GLU	-	SEE REMARK 999	UNP Q9UNA4
A	101	LEU	-	EXPRESSION TAG	UNP Q9UNA4
A	102	GLU	-	EXPRESSION TAG	UNP Q9UNA4
A	103	HIS	-	EXPRESSION TAG	UNP Q9UNA4
A	104	HIS	-	EXPRESSION TAG	UNP Q9UNA4
A	105	HIS	-	EXPRESSION TAG	UNP Q9UNA4
A	106	HIS	-	EXPRESSION TAG	UNP Q9UNA4
A	107	HIS	-	EXPRESSION TAG	UNP Q9UNA4
A	108	HIS	-	EXPRESSION TAG	UNP Q9UNA4

- Molecule 2 is a protein called Ubiquitin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	74	1215	374	622	103	115	1	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	198	GLY	-	EXPRESSION TAG	UNP P62988
B	199	SER	-	EXPRESSION TAG	UNP P62988
B	200	HIS	-	EXPRESSION TAG	UNP P62988

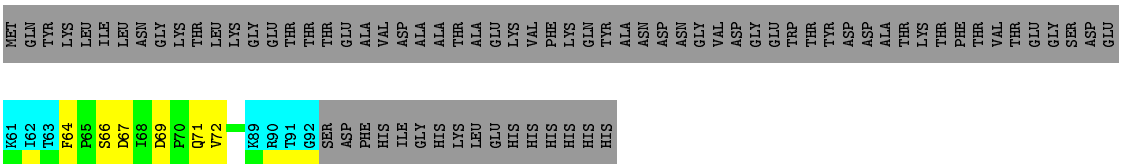
4 Residue-property plots (i)

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.

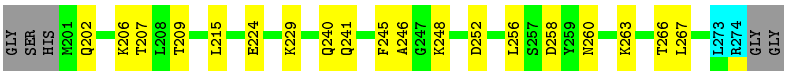
- Molecule 1: Immunoglobulin G-binding protein G, DNA polymerase iota

Chain A: 18% 6% 6% 70%



- Molecule 2: Ubiquitin

Chain B: 67% 24% • 6%

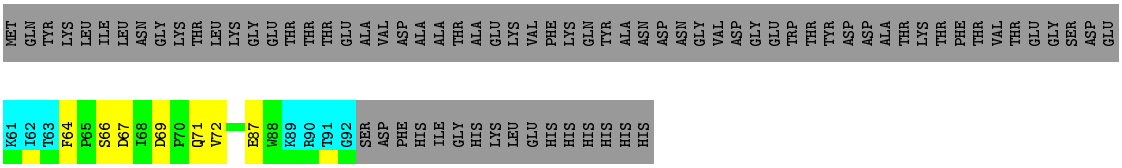


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 4. Colouring as in section 4.1 above.

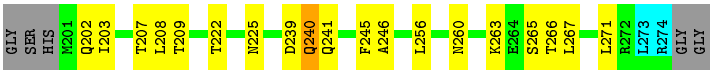
- Molecule 1: Immunoglobulin G-binding protein G, DNA polymerase iota

Chain A: 17% 6% 6% 70%



- Molecule 2: Ubiquitin

Chain B: 67% 23% • • 6%


GLY
SER
HIS
M201
Q202
T203
T207
L208
T209
T222
N225
D239
Q240
Q241
F245
A246
L256
N260
K263
E264
S265
T266
L267
L271
R272
L273
R274
GLY
GLY

5 Refinement protocol and experimental data overview (i)

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

Of the 40 calculated structures, 25 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 NIH	refinement	2.19
CYANA	structure solution	

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	210	196	196	4±1
2	B	574	598	596	11±3
All	All	19600	19850	19800	318

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 8.

5 of 92 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:67:ASP:O	2:B:209:THR:HG22	0.83	1.73	13	25
2:B:206:LYS:NZ	2:B:266:THR:HG21	0.76	1.95	19	3
2:B:266:THR:C	2:B:267:LEU:HD12	0.73	2.04	16	25
2:B:259:TYR:O	2:B:260:ASN:ND2	0.66	2.29	23	2
2:B:216:GLU:O	2:B:229:LYS:NZ	0.65	2.28	7	2

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	25/108 (23%)	24±0 (94±2%)	1±0 (6±2%)	0±0 (0±0%)	100	100
2	B	71/79 (90%)	66±1 (93±1%)	5±1 (7±1%)	0±0 (0±0%)	100	100
All	All	2400/4675 (51%)	2245 (94%)	155 (6%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	23/93 (25%)	22±1 (95±4%)	1±1 (5±4%)	37	80
2	B	66/70 (94%)	57±2 (86±3%)	9±2 (14±3%)	8	48
All	All	2225/4075 (55%)	1964 (88%)	261 (12%)	11	54

5 of 26 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)
2	B	263	LYS	25
2	B	240	GLN	24
2	B	202	GLN	23
2	B	241	GLN	21
2	B	252	ASP	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