wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 08:45 PM BST

PDB ID	:	2HV1
Title	:	HADDOCK structure of ARNT PAS-B Homodimer
Authors	:	Card, P.B.; Gardner, K.H.
Deposited on	:	2006-07-27

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:

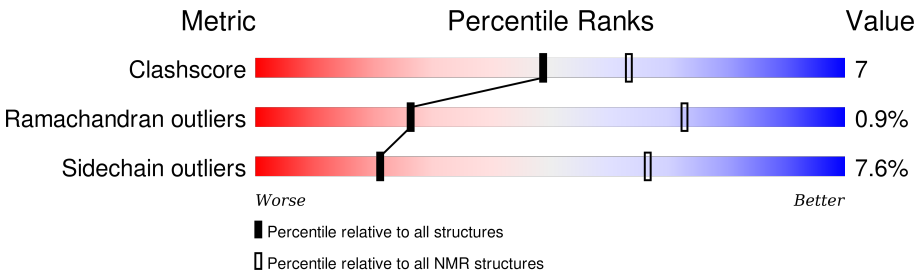
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	119	65% 24% •• 9%
1	B	119	65% 25% • 9%

2 Ensemble composition and analysis (i)

This entry contains 8 models. Model 6 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:8-A:114, B:7-B:114 (215)	0.29	6

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 2 clusters and 4 single-model clusters were found.

Cluster number	Models
1	1, 3
2	5, 6
Single-model clusters	2; 4; 7; 8

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 3556 atoms, of which 1754 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Aryl hydrocarbon receptor nuclear translocator.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	108	1778	575	877	155	165	6	0
1	B	108	1778	575	877	155	165	6	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP P27540
A	2	ALA	-	CLONING ARTIFACT	UNP P27540
A	3	MET	-	CLONING ARTIFACT	UNP P27540
A	4	ASP	-	CLONING ARTIFACT	UNP P27540
B	1	GLY	-	CLONING ARTIFACT	UNP P27540
B	2	ALA	-	CLONING ARTIFACT	UNP P27540
B	3	MET	-	CLONING ARTIFACT	UNP P27540
B	4	ASP	-	CLONING ARTIFACT	UNP P27540

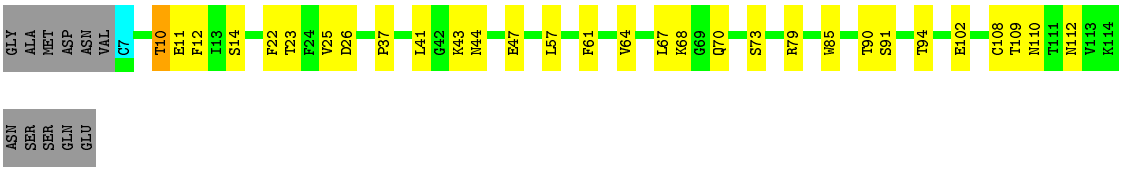
4 Residue-property plots (i)

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.

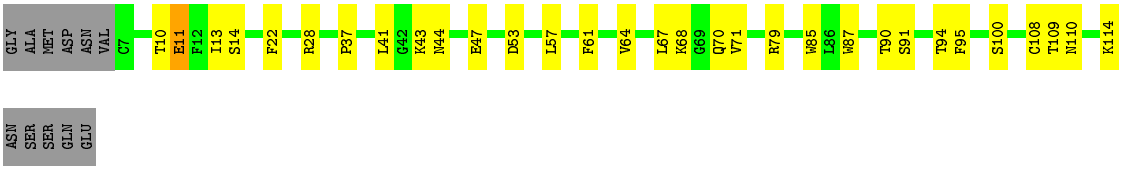
- Molecule 1: Aryl hydrocarbon receptor nuclear translocator

Chain A:



- Molecule 1: Aryl hydrocarbon receptor nuclear translocator

Chain B:

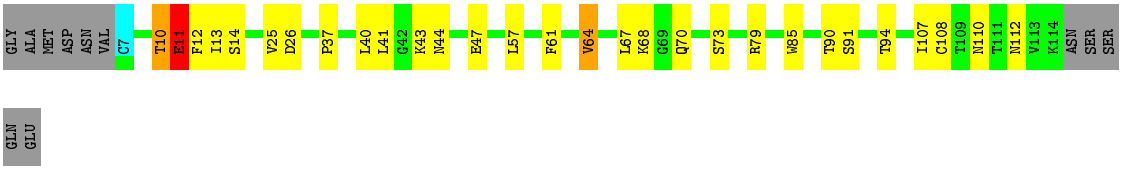


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 6. Colouring as in section 4.1 above.

- Molecule 1: Aryl hydrocarbon receptor nuclear translocator

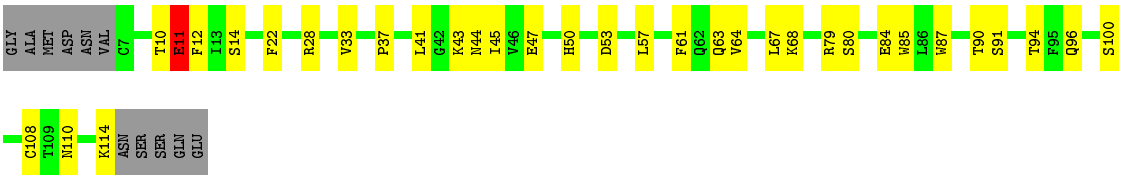
Chain A:



- Molecule 1: Aryl hydrocarbon receptor nuclear translocator

Chain B:

62%
28%
•
9%


GLY
ALA
MET
ASP
ASN
VAL
C7
T10
E11
F12
I13
S14
F22
R28
V33
P37
L41
G42
K43
N44
I45
V46
E47
H50
D53
L57
F61
Q62
Q63
V64
L67
K68
R79
S80
E84
W85
L86
W87
T90
S91
T94
F95
Q96
S100
C108
T109
N110
K114
ASN
SER
SER
GLN
GLU

5 Refinement protocol and experimental data overview (i)

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

Of the 50 calculated structures, 8 were deposited, based on the following criterion: *?*.

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

Software name	Classification	Version
HADDOCK	structure solution	1.2
CNS	structure solution	1.1
CNS	refinement	1.1

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)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	0.30±0.05	0±0/917 (0.0±0.0%)	0.46±0.25	0±1/1240 (0.0±0.1%)
1	B	0.31±0.05	0±0/923 (0.0±0.0%)	0.47±0.29	0±1/1248 (0.0±0.1%)
All	All	0.31	2/14720 (0.0%)	0.54	4/19904 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	0.1±0.3
1	B	0.0±0.0	0.1±0.3
All	All	0	2

All unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	B	11	GLU	CG-CD	10.13	1.67	1.51	6	1
1	A	11	GLU	CG-CD	9.97	1.67	1.51	6	1

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	B	11	GLU	CG-CD-OE2	-38.01	42.27	118.30	6	1
1	A	11	GLU	CG-CD-OE2	-30.65	57.00	118.30	6	1
1	A	11	GLU	CG-CD-OE1	-21.84	74.62	118.30	6	1
1	B	11	GLU	CG-CD-OE1	-16.41	85.48	118.30	6	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	11	GLU	Sidechain	1
1	B	11	GLU	Sidechain	1

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	895	872	869	13±2
1	B	901	877	873	14±2
All	All	14368	13992	13936	211

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 7.

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:12:PHE:HB2	1:A:26:ASP:HB3	0.77	1.55	7	8
1:B:8:GLN:HB3	1:B:28:ARG:HG2	0.75	1.56	8	2
1:B:44:ASN:HB2	1:B:47:GLU:HG3	0.72	1.58	1	7
1:A:87:TRP:HB2	1:A:114:LYS:HB2	0.71	1.60	1	2
1:A:14:SER:HB3	1:A:22:PHE:HE2	0.69	1.48	5	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	106/119 (89%)	101±1 (96±1%)	4±1 (4±1%)	1±0 (1±0%)	31 76

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	106/119 (89%)	101±2 (95±1%)	4±2 (4±1%)	1±0 (1±0%)	23	69
All	All	1696/1904 (89%)	1617 (95%)	64 (4%)	15 (1%)	26	73

All 3 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	B	68	LYS	8
1	A	68	LYS	6
1	B	28	ARG	1

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	102/112 (91%)	95±2 (93±2%)	7±2 (7±2%)	23	68
1	B	103/112 (92%)	95±2 (92±2%)	8±2 (8±2%)	20	65
All	All	1640/1792 (92%)	1516 (92%)	124 (8%)	21	66

5 of 39 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	94	THR	8
1	B	94	THR	8
1	A	57	LEU	8
1	B	10	THR	8
1	A	10	THR	8

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