wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 04:00 PM BST

PDB ID	:	1OPI
Title	:	SOLUTION STRUCTURE OF THE THIRD RNA RECOGNITION MOTIF (RRM) OF U2AF65 IN COMPLEX WITH AN N-TERMINAL SF1 PEPTIDE
Authors	:	Selenko, P.; Gregorovic, G.; Sprangers, R.; Stier, G.; Rhani, Z.; Kramer, A.; Sattler, M.
Deposited on	:	2003-03-05

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:

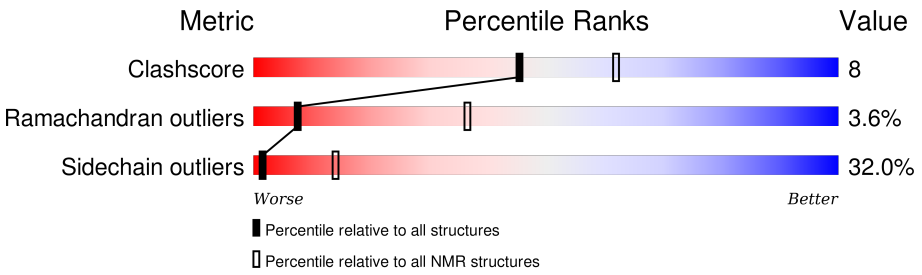
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	104	48% 45% 7%
2	B	13	15% 85%

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 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:374-A:424, A:428-A:473, B:22-B:23 (99)	0.38	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 2 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1889 atoms, of which 933 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called SPLICING FACTOR U2AF 65 KDA SUBUNIT.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	104	1647	529	810	137	163	8	0

- Molecule 2 is a protein called SPLICING FACTOR SF1.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
2	B	13	242	71	123	28	20	0

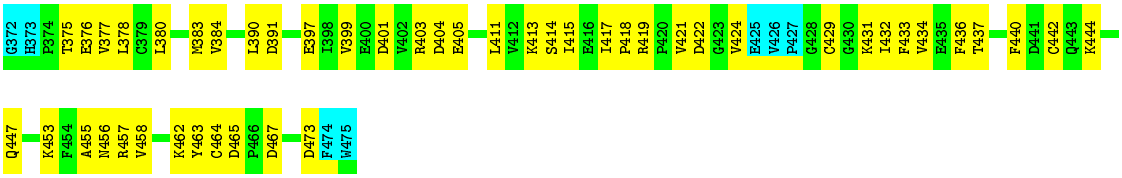
4 Residue-property plots (i)

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.

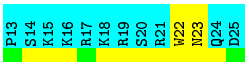
- Molecule 1: SPLICING FACTOR U2AF 65 KDA SUBUNIT

Chain A: 48% 45% 7%



- Molecule 2: SPLICING FACTOR SF1

Chain B: 15% 85%

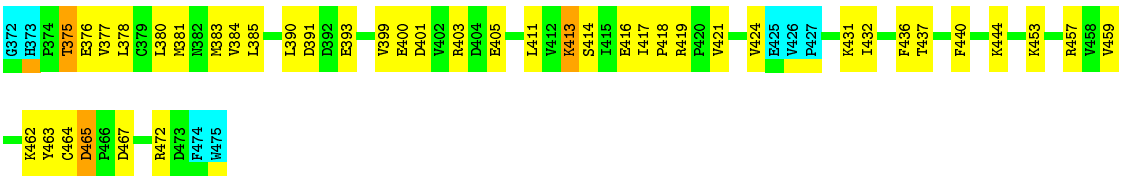


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 3. Colouring as in section 4.1 above.

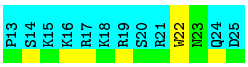
- Molecule 1: SPLICING FACTOR U2AF 65 KDA SUBUNIT

Chain A: 54% 37% • 7%



- Molecule 2: SPLICING FACTOR SF1

Chain B: 8% 8% 85%


P13
S14
K15
K16
R17
K18
R19
S20
R21
W22
N23
Q24
D25

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *RESTRAINED MOLECULAR DYNAMICS USING CNS AND ARIA FOR AMBIGUOUS DISTANCE RESTRAINTS.*

Of the 100 calculated structures, 10 were deposited, based on the following criterion: *STRUCTURES WITH THE LEAST RESTRAINT VIOLATIONS, STRUCTURES WITH THE LOWEST ENERGY.*

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

Software name	Classification	Version
CNS	refinement	1.1
ARIA	refinement	1.2

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	774	758	757	13±3
2	B	22	16	16	0±0
All	All	7960	7740	7730	127

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 56 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:417:ILE:HG23	1:A:432:ILE:HG21	1.01	1.31	9	7
1:A:417:ILE:HG23	1:A:432:ILE:CG2	0.69	2.17	3	7
1:A:421:VAL:O	1:A:424:VAL:HG23	0.65	1.92	6	6
1:A:418:PRO:HD2	1:A:432:ILE:HG22	0.65	1.68	8	6
1:A:399:VAL:HG22	1:A:417:ILE:HD12	0.64	1.70	6	5

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	97/104 (93%)	83±3 (86±4%)	11±2 (11±3%)	3±1 (3±1%)	8 39
2	B	2/13 (15%)	1±1 (25±25%)	1±1 (55±42%)	0±0 (20±24%)	0 2
All	All	990/1170 (85%)	835 (84%)	119 (12%)	36 (4%)	7 36

5 of 13 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	376	GLU	9
1	A	455	ALA	5
2	B	22	TRP	4
1	A	383	MET	3
1	A	467	ASP	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	89/95 (94%)	60±3 (68±3%)	29±3 (32±3%)	1 14
2	B	2/13 (15%)	2±1 (75±25%)	1±1 (25±25%)	3 26
All	All	910/1080 (84%)	619 (68%)	291 (32%)	1 14

5 of 66 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	375	THR	10
1	A	440	PHE	10
1	A	419	ARG	10
1	A	465	ASP	10
1	A	436	PHE	10

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