wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 04:50 AM BST

PDB ID	:	2RNR
Title	:	Solution structure of the complex between TFIIE alpha C-terminal acidic domain and TFIIH p62 PH domain
Authors	:	Okuda, M.; Nishimura, Y.
Deposited on	:	2008-01-31

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:

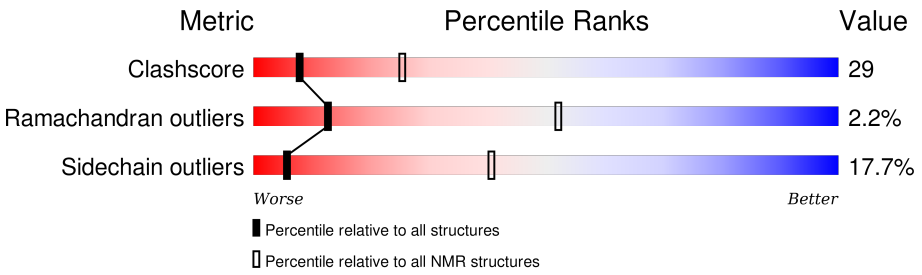
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	64	52% 31% • 11% •
2	B	110	30% 57% • 9% •

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 18 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:382-A:436, B:7-B:104 (153)	0.61	18

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 4 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2717 atoms, of which 1348 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Transcription initiation factor IIE subunit alpha.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	62	956	313	450	75	114	4	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	376	GLY	-	EXPRESSION TAG	UNP P29083
A	377	SER	-	EXPRESSION TAG	UNP P29083

- Molecule 2 is a protein called TFIIH basal transcription factor complex p62 subunit.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	108	1761	547	898	154	158	4	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	GLY	-	EXPRESSION TAG	UNP P32780
B	0	SER	-	EXPRESSION TAG	UNP P32780

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

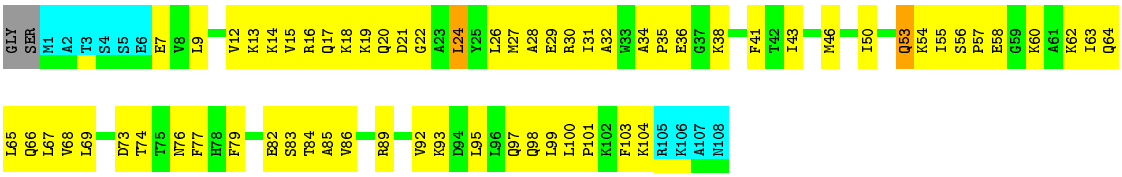
These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain 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: Transcription initiation factor IIE subunit alpha

Chain A: 52% 31% • 11% •

- Molecule 2: TFIIH basal transcription factor complex p62 subunit

Chain B: 30% 57% • 9% •



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

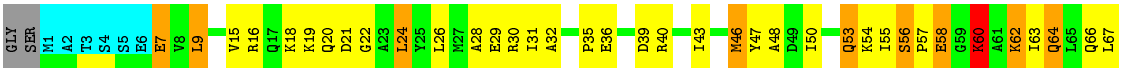
The representative model is number 18. Colouring as in section 4.1 above.

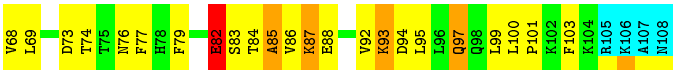
- Molecule 1: Transcription initiation factor IIE subunit alpha

Chain A: 61% 17% 5% • 11% •

- Molecule 2: TFIIH basal transcription factor complex p62 subunit

Chain B: 35% 41% 12% • 9% •




V68
L69
D73
T74
T75
N76
F77
H78
F79
E82
S83
T84
A85
V86
K87
E88
V92
K93
D94
L95
L96
Q97
Q98
L99
L100
P101
K102
F103
K104
R105
K106
A107
N108

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-distance geometry simulated annealing, 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 NIH	structure solution	
X-PLOR NIH	refinement	

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	442	402	402	25±4
2	B	788	824	824	63±7
All	All	24600	24520	24520	1429

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 29.

5 of 393 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:54:LYS:HB2	2:B:66:GLN:HB3	0.97	1.34	2	17
1:A:386:GLU:HB3	2:B:60:LYS:HD2	0.96	1.34	11	8
2:B:55:ILE:HG22	2:B:65:LEU:HD12	0.94	1.37	19	19
1:A:388:GLU:HG2	2:B:57:PRO:HG3	0.93	1.39	2	18
1:A:383:GLU:HB3	2:B:76:ASN:HD22	0.88	1.25	14	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	55/64 (86%)	49±1 (89±2%)	5±2 (9±3%)	1±1 (2±2%)	11	48
2	B	98/110 (89%)	87±2 (89±2%)	9±2 (9±2%)	2±1 (2±1%)	14	55
All	All	3060/3480 (88%)	2722 (89%)	271 (9%)	67 (2%)	13	52

5 of 20 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	82	GLU	10
1	A	386	GLU	8
1	A	392	ASP	6
2	B	36	GLU	5
2	B	83	SER	5

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	48/56 (86%)	40±2 (83±4%)	8±2 (17±4%)	6	42
2	B	85/94 (90%)	70±3 (82±3%)	15±3 (18±3%)	5	39
All	All	2660/3000 (89%)	2188 (82%)	472 (18%)	6	41

5 of 69 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	24	LEU	20
1	A	424	GLU	20
2	B	95	LEU	20
1	A	396	VAL	20
1	A	413	GLU	17

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