wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 03:47 AM BST

PDB ID	:	2MUT
Title	:	Solution structure of the F231L mutant ERCC1-XPF dimerization region
Authors	:	Faridounnia, M.; Wienk, H.; Kovacic, L.; Folkers, G.E.; Jaspers, N.G.J.; Kaptein, R.; Hoeijmakers, J.H.J.; Boelens, R.
Deposited on	:	2014-09-17

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:

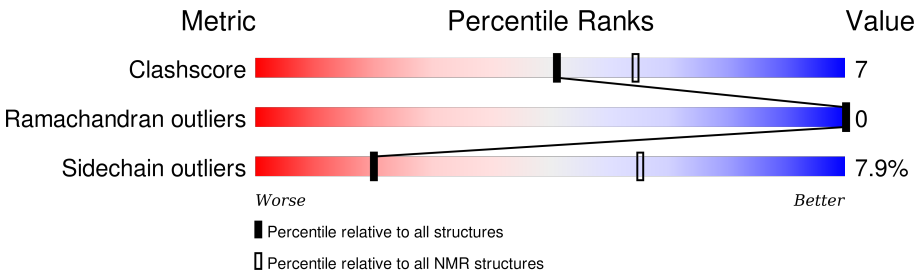
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 is 87%.

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	96	52% 17% 31%
2	B	84	57% 17% 26%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 14 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:230-A:295, B:833-B:894 (128)	0.21	14

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 1 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2846 atoms, of which 1431 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called DNA excision repair protein ERCC-1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	96	1553	481	784	145	139	4	0

There are 19 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	212	ARG	-	EXPRESSION TAG	UNP P07992
A	213	ILE	-	EXPRESSION TAG	UNP P07992
A	214	ARG	-	EXPRESSION TAG	UNP P07992
A	215	ARG	-	EXPRESSION TAG	UNP P07992
A	216	ARG	-	EXPRESSION TAG	UNP P07992
A	217	TYR	-	EXPRESSION TAG	UNP P07992
A	218	ASN	-	EXPRESSION TAG	UNP P07992
A	219	MET	-	EXPRESSION TAG	UNP P07992
A	231	LEU	PHE	ENGINEERED MUTATION	UNP P07992
A	298	GLY	-	EXPRESSION TAG	UNP P07992
A	299	GLY	-	EXPRESSION TAG	UNP P07992
A	300	LEU	-	EXPRESSION TAG	UNP P07992
A	301	GLU	-	EXPRESSION TAG	UNP P07992
A	302	HIS	-	EXPRESSION TAG	UNP P07992
A	303	HIS	-	EXPRESSION TAG	UNP P07992
A	304	HIS	-	EXPRESSION TAG	UNP P07992
A	305	HIS	-	EXPRESSION TAG	UNP P07992
A	306	HIS	-	EXPRESSION TAG	UNP P07992
A	307	HIS	-	EXPRESSION TAG	UNP P07992

- Molecule 2 is a protein called DNA repair endonuclease XPF.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	84	1293	405	647	111	126	4	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	822	MET	-	INITIATING METHIONINE	UNP Q92889

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: DNA excision repair protein ERCC-1

Chain A: 52% 17% 31%

R212 I213 R214 R215 R216 Y217 N218 M219 A220 D221 L222 L223 M224 E225 K226 L227 E228 Q229 D230 R234 V235 T236 E237 C238 L239 T240 T241 L263 I264 D270 L273 C274 R284 L285 V288 L289 P292 V296 P297 G298 G299 L300 E301 H302 H303 H304 H305 H306 H307

- Molecule 2: DNA repair endonuclease XPF

Chain B: 57% 17% 26%

M822 D823 S824 E825 T826 L827 P828 E829 S830 E831 K832 P837 F840 K850 N851 L855 K860 N861 I862 A863 E864 N879 A880 A881 K884 Q885 S893 F894 A895 E896 V897 V898 S899 K900 G901 K902 G903 K904 K905

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

The representative model is number 14. Colouring as in section 4.1 above.

- Molecule 1: DNA excision repair protein ERCC-1

Chain A: 57% 9% • 31%

R212 I213 R214 R215 R216 Y217 N218 M219 A220 D221 L222 L223 M224 E225 K226 L227 E228 Q229 D230 R234 V235 L239 E261 Q262 L263 I264 R283 R284 L285 V288 L289 V296 P297 G298 G299 L300 E301 H302 H303 H304 H305 H306 H307

- Molecule 2: DNA repair endonuclease XPF

Chain B: 58% 14% • 26%

M822 D823 S824 E825 T826 L827 P828 E829 S830 E831 K832 P837 F840 N848 A849 K850 N851 N861 I862 A863 E864 A881 Q885 L886 I890 S893 F894 A895 E896 V897 V898 S899 K900 G901 K902 G903 K904 K905

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *restrained molecular dynamics*.

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
CYANA	structure solution	
CNS	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	2mut_cs.str
Number of chemical shift lists	1
Total number of shifts	2040
Number of shifts mapped to atoms	2040
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	87%

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.

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
All	All	0	2

There are no bond-length outliers.

There are no bond-angle outliers.

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	283	ARG	Sidechain	1
1	A	284	ARG	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	509	530	530	9±2
2	B	481	476	476	7±2
All	All	19800	20120	20120	284

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:235:VAL:O	1:A:239:LEU:HG	0.62	1.94	1	20
2:B:838:GLN:O	2:B:842:LEU:HG	0.62	1.95	2	3
1:A:249:ASP:HA	1:A:277:LEU:HD21	0.62	1.72	3	1
1:A:256:THR:HG21	1:A:274:CYS:SG	0.57	2.39	1	9
1:A:261:GLU:HG3	2:B:890:ILE:O	0.57	1.99	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	66/96 (69%)	63±1 (95±2%)	3±1 (5±2%)	0±0 (0±0%)	100	100
2	B	62/84 (74%)	61±1 (99±1%)	1±1 (1±1%)	0±0 (0±0%)	100	100
All	All	2560/3600 (71%)	2478 (97%)	82 (3%)	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	59/86 (69%)	56±1 (95±2%)	3±1 (5±2%)	33	78
2	B	52/71 (73%)	46±2 (89±3%)	6±2 (11±3%)	12	56
All	All	2220/3140 (71%)	2045 (92%)	175 (8%)	20	65

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	862	ILE	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	264	ILE	20
2	B	851	ASN	16
2	B	879	ASN	14
2	B	840	PHE	13

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)

The completeness of assignment taking into account all chemical shift lists is 87% for the well-defined parts and 81% for the entire structure.

7.1 Chemical shift list 1

File name: 2mut_cs.str

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	2040
Number of shifts mapped to atoms	2040
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	174	-0.61 ± 0.09	Should be applied
$^{13}C_{\beta}$	163	0.18 ± 0.06	None needed (< 0.5 ppm)
$^{13}C'$	148	-0.21 ± 0.09	None needed (< 0.5 ppm)
^{15}N	161	0.23 ± 0.16	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 87%, i.e. 1357 atoms were assigned a chemical shift out of a possible 1557. 25 out of 27 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	601/628 (96%)	248/250 (99%)	234/256 (91%)	119/122 (98%)
Sidechain	686/831 (83%)	428/485 (88%)	248/310 (80%)	10/36 (28%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	70/98 (71%)	41/54 (76%)	29/40 (72%)	0/4 (0%)
Overall	1357/1557 (87%)	717/789 (91%)	511/606 (84%)	129/162 (80%)

7.1.4 Statistically unusual chemical shifts (i)

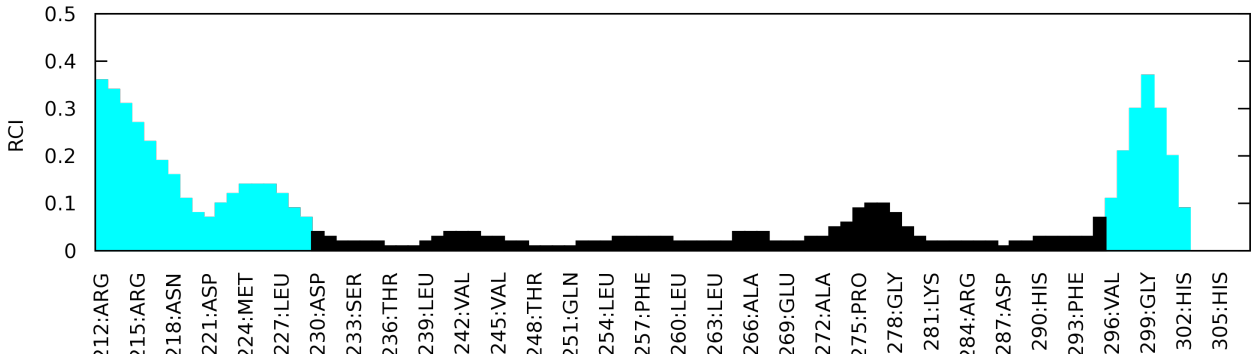
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	213	ILE	CG2	27.27	24.63 – 10.43	6.9
2	B	845	PRO	HD2	1.67	5.45 – 1.85	-5.5
2	B	884	LYS	CG	30.79	30.67 – 19.17	5.1

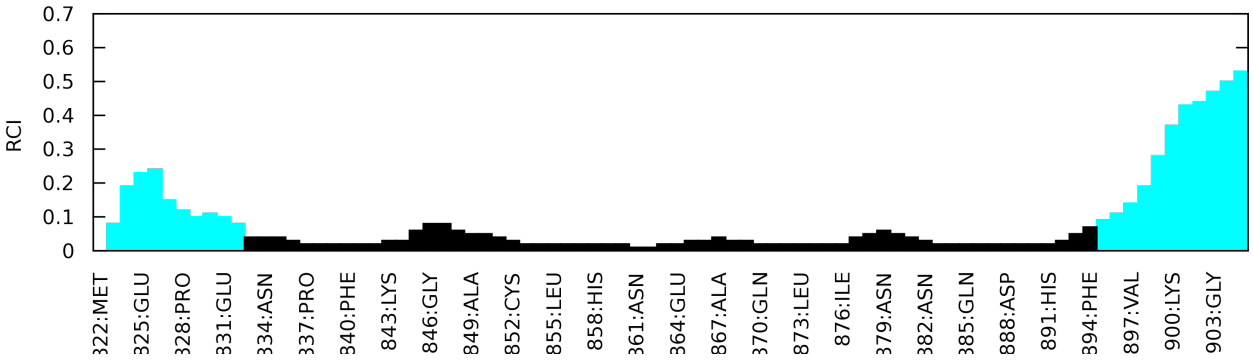
7.1.5 Random Coil Index (RCI) plots (i)

The images below report *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A:



Random coil index (RCI) for chain B:


RCI
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
322:MET
325:GLU
328:PRO
331:GLU
334:ASN
337:PRO
340:PHE
843:LYS
846:GLY
849:ALA
852:CYS
855:LEU
858:HIS
361:ASN
364:GLU
867:ALA
370:GLN
873:LEU
876:ILE
379:ASN
382:ASN
385:GLN
888:ASP
891:HIS
894:PHE
897:VAL
900:LYS
903:GLY