



wwPDB NMR Structure Validation Summary Report ⓘ

Apr 26, 2016 – 04:02 PM BST

PDB ID : 1OQY
Title : Structure of the DNA repair protein hHR23a
Authors : Walters, K.J.; Lech, P.J.; Goh, A.M.; Wang, Q.; Howley, P.M.
Deposited on : 2003-03-11

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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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

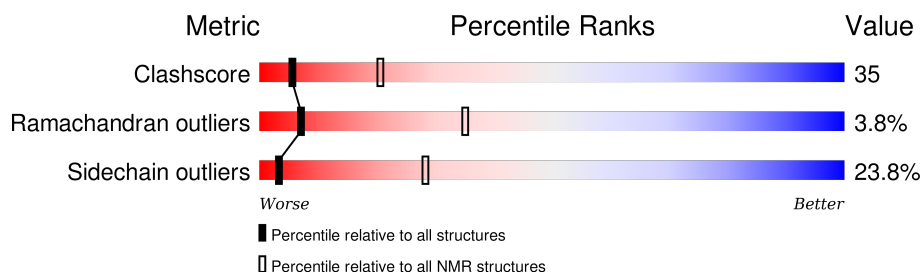
1 Overall quality at a glance

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	368	

2 Ensemble composition and analysis ⓘ

This entry contains 12 models. Model 9 is the overall representative, medoid model (most similar to other models). The authors have identified model 7 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:2-A:77 (76)	0.29	9
2	A:162-A:198 (37)	0.24	12
3	A:231-A:286 (56)	0.35	2
4	A:317-A:358 (42)	0.26	9

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	1, 2, 3, 6, 7, 8, 10, 12
2	4, 5, 9, 11

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 5484 atoms, of which 2706 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called UV excision repair protein RAD23 homolog A.

Mol	Chain	Residues	Atoms						Trace
1	A	363	Total	C	H	N	O	S	0
			5484	1725	2706	469	574	10	

There are 6 discrepancies between the modelled and reference sequences:

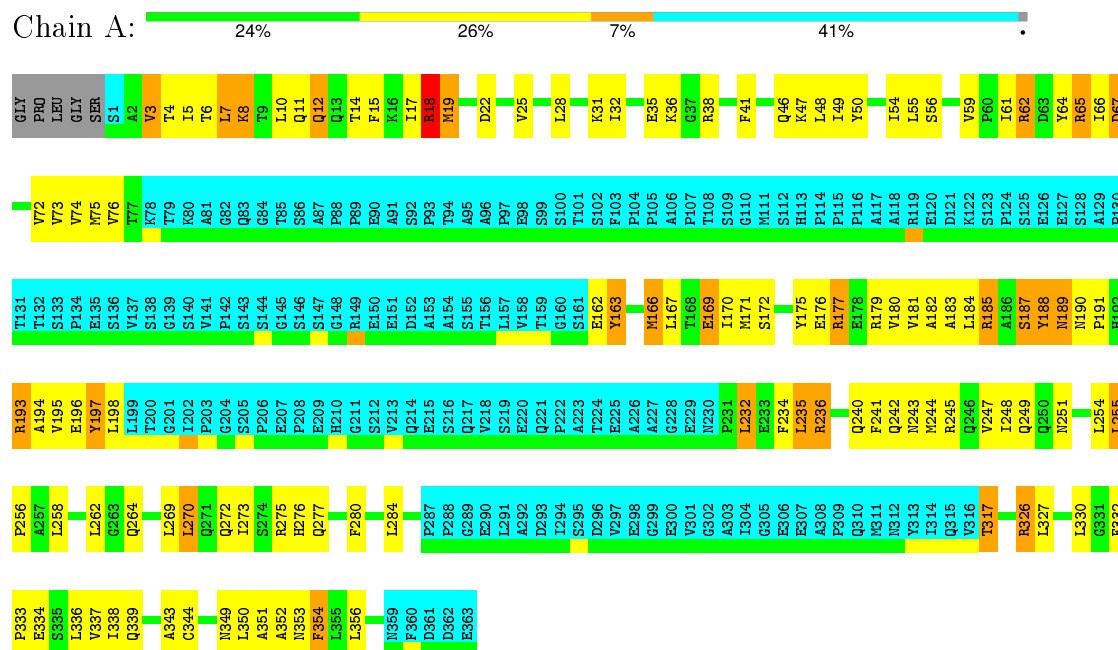
Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	CLONING ARTIFACT	UNP P54725
A	-3	PRO	-	CLONING ARTIFACT	UNP P54725
A	-2	LEU	-	CLONING ARTIFACT	UNP P54725
A	-1	GLY	-	CLONING ARTIFACT	UNP P54725
A	0	SER	-	CLONING ARTIFACT	UNP P54725
A	1	SER	MET	see remark 999	UNP P54725

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.

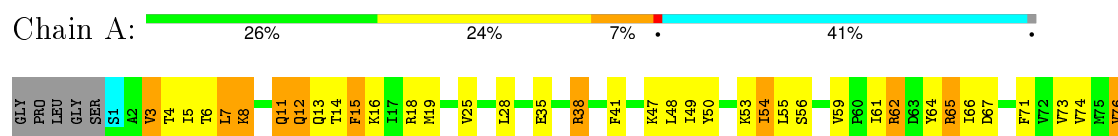
- Molecule 1: UV excision repair protein RAD23 homolog A



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

The representative model is number 9. Colouring as in section 4.1 above.

- Molecule 1: UV excision repair protein RAD23 homolog A



V337	L270	G204	V137	K77
I338	Q271	S205	S138	K78
Q339	Q272	P206	G139	I779
	I273	E207	S140	K80
A343	S274	P208	V141	A81
C344	R275	E209	P142	G82
E345		E210	S143	Q83
K346	E278	S211	S144	G84
	Q279	S212	G145	T85
K349	F280	Q213	S146	S86
L350		Q214	S147	A87
A351		E215	G148	P88
A352	L284	S216	R149	P89
K353	E285	Q217	E150	E90
F354	E286	V218	E151	A91
	P287	S219	D152	S92
K359	P288	E220	A153	P93
F360	G289	Q221	A154	T94
D361	E290	P222	S155	A95
D362	L291	A223	T156	A96
E363	D292	T224	L157	P97
	I293	E225	V158	E98
	I294	A226	T159	S99
	S295	A227	G160	S100
	D296	G228	S161	T101
	V297	E229	E162	S102
	E298	N230	Y163	F103
	G299	P231		P104
	E300		M166	P105
	V301	E232	L167	A106
	G302	F234	T168	P107
	A303	L235	E169	T108
	I304	R236	M171	S109
	G305			G110
	E306	Q240		M111
	E307	F241	R177	S112
	A308	Q242	E178	H113
	P309	N243	R179	P114
	Q310	R244	V180	P115
	K311	Q246	V181	P116
	K312	V247	A182	A117
	Y313	T248	A183	A118
	I314	Q249	L184	R119
	Q315	Q250	R185	E120
	V316	N251	A186	D121
	T317	P252	S187	K122
	F318	A253	N189	S123
	Q319	L254		P124
	E320	L255		S125
	K321	P256	H192	E126
		A257	R193	
		L258	A194	S127
	K326	L259		S128
	L327		Y197	A129
	K328	L262	L198	P130
	A329	G263	L199	T131
	L330	Q264	T200	T132
		S335	G201	S133
	P333	L269	I202	P134
	E334		E135	E135
	S335		P203	S136
	L336			

5 Refinement protocol and experimental data overview ⓘ

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

Of the 45 calculated structures, 12 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
X-PLOR	structure solution	3.851
X-PLOR	refinement	3.851

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.

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	12.0±0.0
All	All	0	144

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

5 of 12 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	236	ARG	Sidechain	12
1	A	245	ARG	Sidechain	12
1	A	18	ARG	Sidechain	12
1	A	179	ARG	Sidechain	12
1	A	38	ARG	Sidechain	12

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	1716	1740	1740	122±7
All	All	20592	20880	20880	1461

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 35.

5 of 589 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:247:VAL:HG22	1:A:254:LEU:HD13	1.11	1.12	11	4
1:A:269:LEU:HD23	1:A:273:ILE:HD11	1.06	1.27	9	4
1:A:327:LEU:HD13	1:A:337:VAL:HG13	1.05	1.18	12	1
1:A:327:LEU:HD23	1:A:337:VAL:HG13	0.99	1.34	3	6
1:A:49:ILE:CD1	1:A:54:ILE:HG22	0.93	1.94	11	10

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	211/368 (57%)	171±4 (81±2%)	32±3 (15±1%)	8±2 (4±1%)	7	35
All	All	2532/4416 (57%)	2053 (81%)	382 (15%)	97 (4%)	7	35

5 of 24 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	3	VAL	12
1	A	232	LEU	12
1	A	189	ASN	9
1	A	22	ASP	8
1	A	188	TYR	7

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	188/311 (60%)	143±4 (76±2%)	45±4 (24±2%)	3	28
All	All	2256/3732 (60%)	1718 (76%)	538 (24%)	3	28

5 of 128 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	242	GLN	12
1	A	344	CYS	12
1	A	354	PHE	12
1	A	166	MET	12
1	A	163	TYR	12

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

No chemical shift data were provided