wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 11:10 PM BST

PDB ID	:	2KHO
Title	:	NMR-RDC / XRAY structure of E. coli HSP70 (DNAK) chaperone (1-605) complexed with ADP and substrate
Authors	:	Zuiderweg, E.R.P.; Bertelsen, E.B.
Deposited on	:	2009-04-10

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:

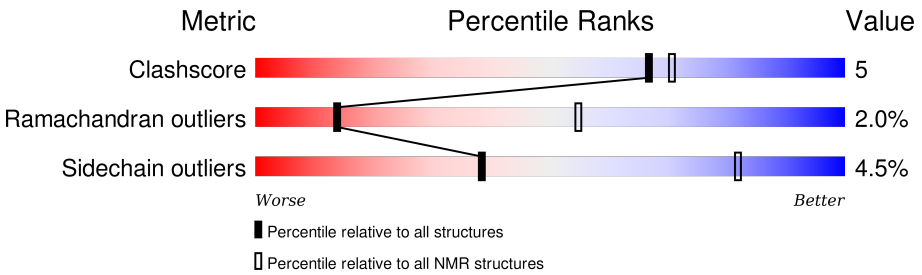
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	605	80% 16% • • •

2 Ensemble composition and analysis (i)

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 5078 atoms, of which 577 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Heat shock protein 70.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	600	5078	2830	577	741	915	15	0

There is a discrepancy between the modelled and reference sequences:

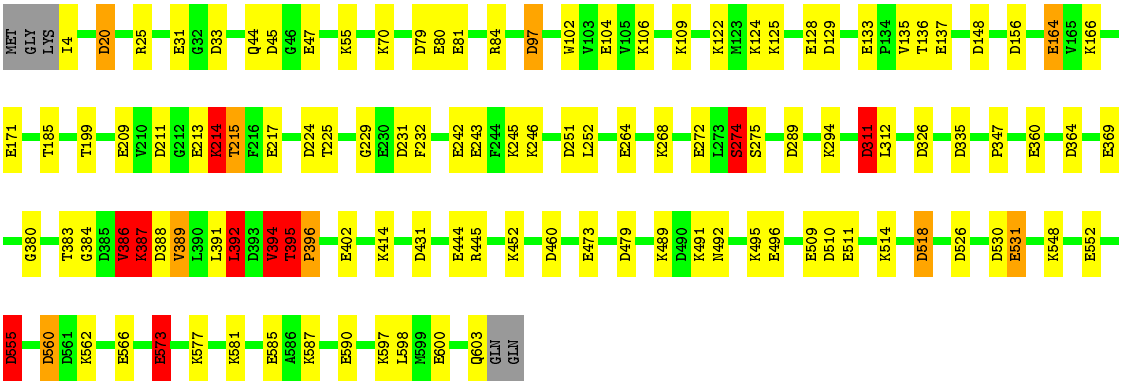
Chain	Residue	Modelled	Actual	Comment	Reference
A	530	ASP	GLU	SEE REMARK 999	UNP P0A6Y8

4 Residue-property plots (i)

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: Heat shock protein 70

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *RDC OPTIMIZATION*.

Of the 1 calculated structures, 1 were deposited, based on the following criterion: *TARGET FUNCTION*.

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

Software name	Classification	Version
REDCAT	refinement	
REDCAT	structure solution	
own software FORTRAN 77	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)

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	0.99	1/4552 (0.0%)	1.23	44/6142 (0.7%)
All	All	0.99	1/4552 (0.0%)	1.23	44/6142 (0.7%)

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	3	7
All	All	3	7

All bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	274	SER	C-O	8.77	1.40	1.23

5 of 44 angle outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	394	VAL	C-N-CA	9.35	145.07	121.70
1	A	392	LEU	CB-CA-C	8.69	126.71	110.20
1	A	389	VAL	N-CA-CB	8.36	129.90	111.50
1	A	394	VAL	CA-C-N	8.22	135.29	117.20
1	A	274	SER	O-C-N	-8.13	109.69	122.70

All chiral outliers are listed below.

Mol	Chain	Res	Type	Atoms
1	A	387	LYS	CA
1	A	389	VAL	CA
1	A	395	THR	CA

5 of 7 planar outliers are listed below.

Mol	Chain	Res	Type	Group
1	A	214	LYS	Peptide
1	A	386	VAL	Peptide
1	A	396	PRO	Peptide
1	A	274	SER	Mainchain
1	A	383	THR	Peptide

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	4501	577	4494	44
All	All	4501	577	4494	44

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 5.

5 of 44 clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)
1:A:102:TRP:CZ2	1:A:102:TRP:CZ3	1.00	2.43
1:A:70:LYS:NZ	1:A:171:GLU:OE2	0.67	2.26
1:A:514:LYS:NZ	1:A:518:ASP:OD2	0.67	2.28
1:A:231:ASP:HB2	1:A:312:LEU:HD11	0.66	1.67
1:A:55:LYS:NZ	1:A:264:GLU:OE2	0.66	2.28

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	598/605 (99%)	566 (95%)	20 (3%)	12 (2%)	14 55

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	598/605 (99%)	566 (95%)	20 (3%)	12 (2%)	14 55

5 of 12 Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
1	A	384	GLY
1	A	389	VAL
1	A	392	LEU
1	A	229	GLY
1	A	386	VAL

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	492/496 (99%)	470 (96%)	22 (4%)	38 82
All	All	492/496 (99%)	470 (96%)	22 (4%)	38 82

5 of 22 residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
1	A	526	ASP
1	A	394	VAL
1	A	573	GLU
1	A	79	ASP
1	A	136	THR

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