wwPDB NMR Structure Validation Summary Report (i)

Dec 17, 2023 – 01:30 AM EST

PDB ID	:	2NXU
Title	:	Atomic structure of translation initiation factor aIF2 beta-subunit from Archaebacteria sulfolobus solfataricus: high resolution NMR in solution
Authors	:	Vasile, F.; Pechkova, E.; Stolboushkina, E.; Garber, M.; Nicolini, C.
Deposited on	:	2006-11-20

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
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

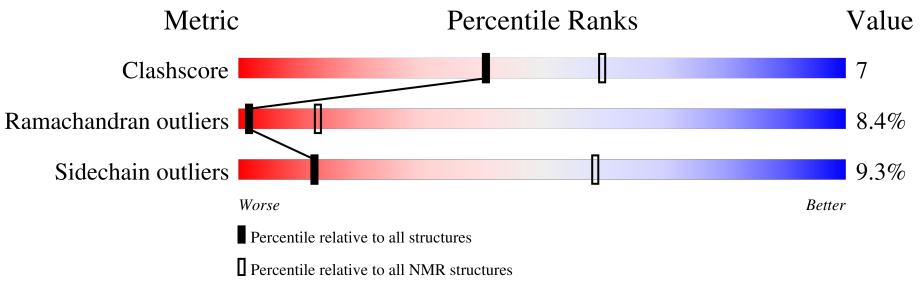
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	146	36% 16% • 38% 8%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 9 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:26-A:104 (79)	1.40	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 and 2 single-model clusters were found.

Cluster number	Models
1	3, 6, 7, 8, 9, 10
2	1, 2
Single-model clusters	4; 5

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2192 atoms, of which 1120 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Translation initiation factor 2 beta subunit.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	134	2192	682	1120	181	199	10	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	HIS	-	expression tag	UNP Q97W59
A	-4	HIS	-	expression tag	UNP Q97W59
A	-3	HIS	-	expression tag	UNP Q97W59
A	-2	HIS	-	expression tag	UNP Q97W59
A	-1	HIS	-	expression tag	UNP Q97W59
A	0	HIS	-	expression tag	UNP Q97W59
A	1	MET	-	expression tag	UNP Q97W59
A	2	GLY	-	expression tag	UNP Q97W59

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: Translation initiation factor 2 beta subunit

Chain A: 36% 16% • 38% 8%

HIS HIS HIS HIS HIS HIS MET GLY SER SER GLU LYS E7 Y8 V9 E10 M11 L12 D13 R14 L15 Y16 S17 K18 L19 P20 E21 K22 G23 R24 K25 Q29 S30 L31 P32 N33 I36 T43 I44 I45 R46 Y51 R54 I55 R56 R57 Y65 L70 E81 L82 Q85

F88 S89 V92 I93 L96 M97 E98 R99 A103 Y104 V105 E106 C107 S108 T109 C110 K111 S112 L113 D114 T115 I116 L117 K118 K119 E120 K121 K122 S123 W124 Y125 I126 V127 C128 L129 A130 C131 G132 A133 Q134 T135 P136 V137 K138 P139 L140

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: Translation initiation factor 2 beta subunit

Chain A: 38% 12% • 38% 8%

HIS HIS HIS HIS HIS HIS MET GLY SER SER GLU LYS E7 Y8 V9 E10 M11 L12 D13 R14 L15 Y16 S17 K18 L19 P20 E21 K22 G23 R24 K25 Q29 I36 G40 N41 T42 I43 I44 I45 Y51 I55 R56 D59 C62 Y65 D77 D78 K79 G80 E81 L82

Q85 F88 V92 I93 M97 A103 Y104 V105 E106 C107 S108 T109 C110 K111 S112 L113 D114 T115 I116 L117 K118 K119 E120 K121 K122 S123 W124 Y125 I126 V127 C128 L129 A130 C131 G132 A133 Q134 T135 P136 V137 K138 P139 L140

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *the structures were obtained by torsion angle dynamic calculation and energy minimization.*.

Of the 20 calculated structures, 10 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
DYANA	structure solution	1.5
GROMACS	refinement	3.3.1

No chemical shift data was provided.

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	0.69±0.01	0±0/640 (0.0± 0.0%)	1.13±0.04	3±2/860 (0.4± 0.2%)
All	All	0.69	0/6400 (0.0%)	1.13	32/8600 (0.4%)

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	1.8±0.9
All	All	0	18

There are no bond-length outliers.

5 of 11 unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	54	ARG	NE-CZ-NH1	8.18	124.39	120.30	8	5
1	A	51	TYR	CB-CG-CD2	-7.70	116.38	121.00	8	5
1	A	56	ARG	NE-CZ-NH1	6.85	123.72	120.30	10	4
1	A	57	ARG	NE-CZ-NH1	5.96	123.28	120.30	10	6
1	A	46	ARG	NE-CZ-NH1	5.91	123.26	120.30	2	4

There are no chirality outliers.

5 of 7 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	104	TYR	Sidechain	5
1	A	51	TYR	Sidechain	5
1	A	65	TYR	Sidechain	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	A	103	ALA	Mainchain	2
1	A	100	PHE	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	632	652	652	10±3
All	All	6320	6520	6520	96

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:36:ILE:HG12	1:A:43:THR:HB	0.74	1.60	7	1
1:A:55:ILE:HB	1:A:93:ILE:HG23	0.73	1.59	10	4
1:A:70:LEU:HG	1:A:86:GLY:HA3	0.69	1.65	1	1
1:A:75:ASN:HB2	1:A:84:ILE:HA	0.66	1.66	8	1
1:A:55:ILE:HD13	1:A:93:ILE:HA	0.66	1.68	5	3

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	79/146 (54%)	55±4 (70±4%)	17±4 (22±5%)	7±2 (8±2%)	2 13
All	All	790/1460 (54%)	550 (70%)	174 (22%)	66 (8%)	2 13

5 of 28 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	31	LEU	7
1	A	30	SER	5
1	A	40	GLY	5
1	A	33	ASN	4
1	A	92	VAL	4

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	70/132 (53%)	64±3 (91±4%)	6±3 (9±4%)	12	59
All	All	700/1320 (53%)	635 (91%)	65 (9%)	12	59

5 of 29 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	56	ARG	6
1	A	88	PHE	6
1	A	51	TYR	5
1	A	55	ILE	4
1	A	65	TYR	4

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 monosaccharides 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