wwPDB X-ray Structure Validation Summary Report (i)

May 21, 2020 – 11:53 pm BST

PDB ID	:	4ZJD
Title	:	Small heat shock protein AgsA from Salmonella typhimurium: Truncations at N- and C- termini
Authors	:	Mani, N.; Suguna, K.
Deposited on	:	2015-04-29
Resolution	:	7.50 Å(reported)

This is a wwPDB X-ray 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/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

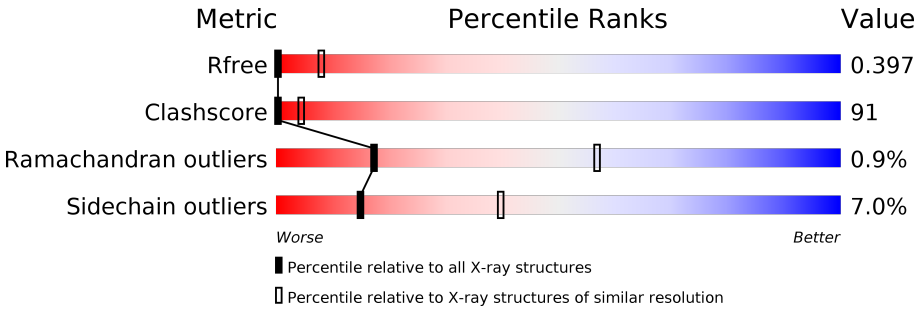
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 7.50 Å.

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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1004 (10.00-3.90)
Clashscore	141614	1069 (10.00-3.90)
Ramachandran outliers	138981	1002 (10.00-3.90)
Sidechain outliers	138945	1002 (10.00-3.86)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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	137	28% 34% 7% 32%
1	B	137	23% 43% • 32%
1	C	137	23% 39% 5% • 32%
1	D	137	20% 42% 7% 32%
1	E	137	23% 38% 5% • 32%
1	F	137	26% 38% • 32%

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 4294 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Aggregation suppressing protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	93	Total	C	N	O	0	0	0
			714	458	117	139			
1	B	93	Total	C	N	O	0	0	0
			717	459	117	141			
1	C	93	Total	C	N	O	0	0	0
			713	456	116	141			
1	D	93	Total	C	N	O	0	0	0
			716	458	117	141			
1	E	93	Total	C	N	O	0	0	0
			717	459	117	141			
1	F	93	Total	C	N	O	0	0	0
			717	459	117	141			

There are 6 discrepancies between the modelled and reference sequences:

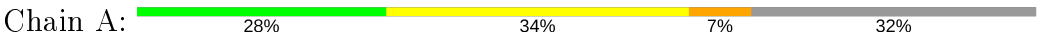
Chain	Residue	Modelled	Actual	Comment	Reference
A	11	MET	-	expression tag	UNP D1MC98
B	11	MET	-	expression tag	UNP D1MC98
C	11	MET	-	expression tag	UNP D1MC98
D	11	MET	-	expression tag	UNP D1MC98
E	11	MET	-	expression tag	UNP D1MC98
F	11	MET	-	expression tag	UNP D1MC98

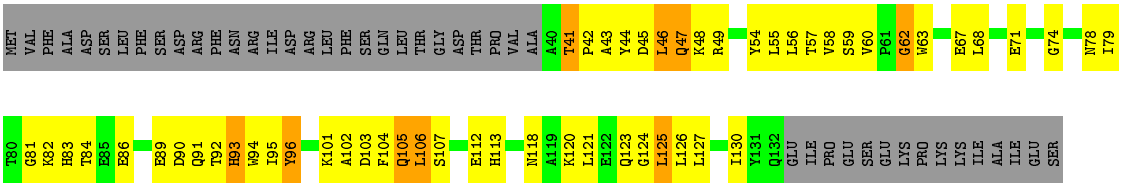
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-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 outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Aggregation suppressing protein

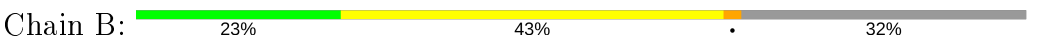
Chain A:

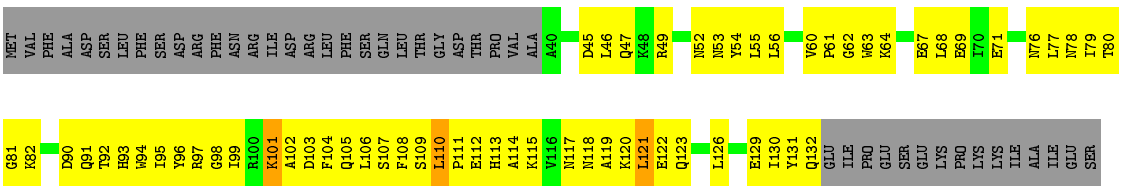




- Molecule 1: Aggregation suppressing protein

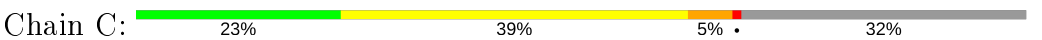
Chain B:

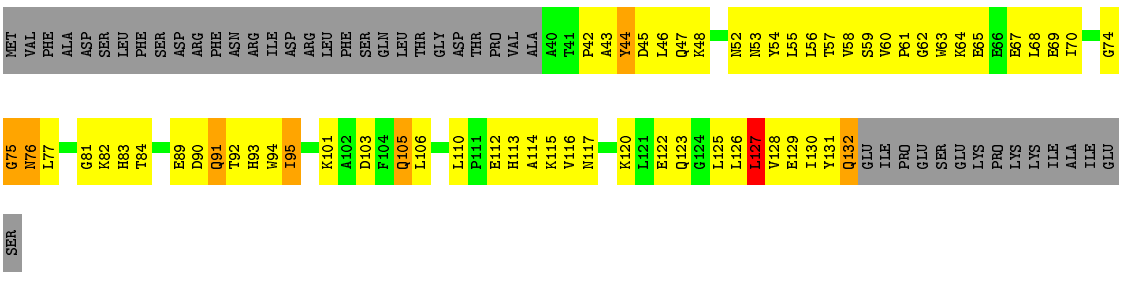




- Molecule 1: Aggregation suppressing protein

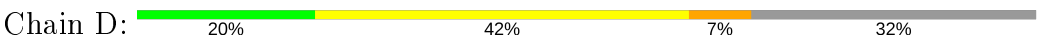
Chain C:

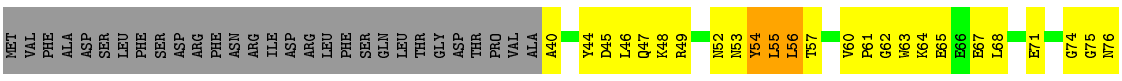


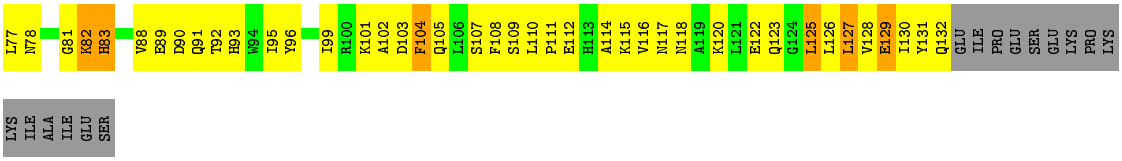


- Molecule 1: Aggregation suppressing protein

Chain D:

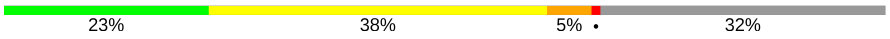


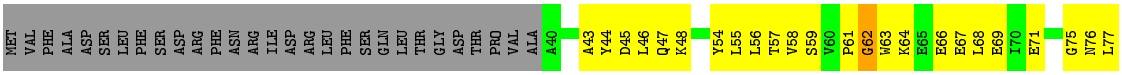


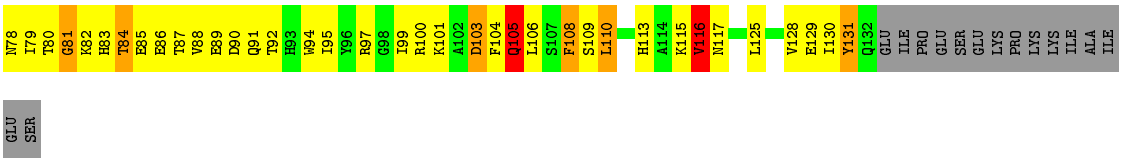


- Molecule 1: Aggregation suppressing protein

Chain E:

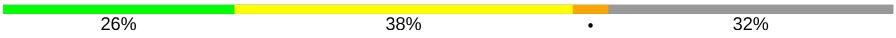


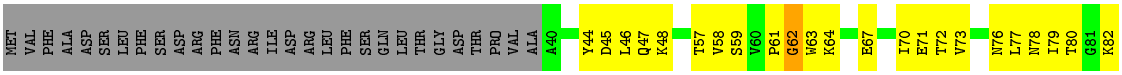




- Molecule 1: Aggregation suppressing protein

Chain F:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α, β, γ	89.78Å 89.78Å 707.48Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	68.14 – 7.50 78.61 – 7.50	Depositor EDS
% Data completeness (in resolution range)	99.4 (68.14-7.50) 100.0 (78.61-7.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.77 (at 7.42Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.351 , 0.396 0.380 , 0.397	Depositor DCC
R_{free} test set	74 reflections (4.55%)	wwPDB-VP
Wilson B-factor (Å^2)	574.3	Xtriage
Anisotropy	0.226	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 999.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.82	EDS
Total number of atoms	4294	wwPDB-VP
Average B, all atoms (Å^2)	333.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 12.64% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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 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	0.72	0/728	1.36	7/992 (0.7%)
1	B	0.73	0/731	1.35	2/996 (0.2%)
1	C	0.67	0/727	1.28	3/992 (0.3%)
1	D	0.96	3/730 (0.4%)	1.30	6/995 (0.6%)
1	E	0.74	0/731	1.37	9/996 (0.9%)
1	F	0.73	0/731	1.39	6/996 (0.6%)
All	All	0.76	3/4378 (0.1%)	1.34	33/5967 (0.6%)

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 outliers	#Planarity outliers
1	B	0	1
1	D	0	1
1	E	0	1
All	All	0	3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	54	TYR	CE1-CZ	-7.81	1.28	1.38
1	D	54	TYR	CD1-CE1	-7.55	1.28	1.39
1	D	129	GLU	CG-CD	6.17	1.61	1.51

The worst 5 of 33 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	46	LEU	CB-CA-C	11.71	132.45	110.20
1	B	121	LEU	CB-CG-CD2	-10.45	93.24	111.00
1	A	47	GLN	N-CA-CB	9.51	127.72	110.60
1	F	110	LEU	CB-CG-CD2	-8.85	95.96	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	103	ASP	CB-CG-OD1	-7.17	111.85	118.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	76	ASN	Mainchain
1	D	56	LEU	Mainchain
1	E	108	PHE	Mainchain

5.2 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 the 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 within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	714	0	675	242	0
1	B	717	0	676	245	0
1	C	713	0	668	95	0
1	D	716	0	676	141	0
1	E	717	0	679	191	0
1	F	717	0	679	168	0
All	All	4294	0	4053	760	0

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 91.

The worst 5 of 760 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:48:LYS:HB3	1:F:93:HIS:CG	1.12	1.64
1:E:44:TYR:CZ	1:F:96:TYR:CD2	1.75	1.63
1:A:47:GLN:CG	1:B:94:TRP:HA	1.26	1.61
1:A:47:GLN:HG2	1:B:94:TRP:CA	1.24	1.59
1:A:125:LEU:CD2	1:B:99:ILE:HG21	1.35	1.55

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 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 X-ray entries followed by that with respect to entries of similar resolution.

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	91/137 (66%)	86 (94%)	5 (6%)	0	100 100
1	B	91/137 (66%)	83 (91%)	8 (9%)	0	100 100
1	C	91/137 (66%)	80 (88%)	7 (8%)	4 (4%)	2 22
1	D	91/137 (66%)	80 (88%)	11 (12%)	0	100 100
1	E	91/137 (66%)	81 (89%)	9 (10%)	1 (1%)	14 52
1	F	91/137 (66%)	84 (92%)	7 (8%)	0	100 100
All	All	546/822 (66%)	494 (90%)	47 (9%)	5 (1%)	17 57

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	76	ASN
1	C	75	GLY
1	E	116	VAL
1	C	77	LEU
1	C	95	ILE

5.3.2 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 X-ray entries followed by that with respect to entries of similar resolution.

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	71/121 (59%)	68 (96%)	3 (4%)	30 54
1	B	72/121 (60%)	70 (97%)	2 (3%)	43 65
1	C	71/121 (59%)	64 (90%)	7 (10%)	8 26

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	D	72/121 (60%)	62 (86%)	10 (14%)	3 17
1	E	72/121 (60%)	69 (96%)	3 (4%)	30 54
1	F	72/121 (60%)	67 (93%)	5 (7%)	15 40
All	All	430/726 (59%)	400 (93%)	30 (7%)	15 40

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	65	GLU
1	D	88	VAL
1	F	97	ARG
1	D	83	HIS
1	D	89	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	123	GLN
1	F	113	HIS
1	E	113	HIS
1	D	47	GLN
1	E	105	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.