wwPDB X-ray Structure Validation Summary Report (i)

Jun 18, 2024 – 12:36 AM EDT

PDB ID	:	5UI5
Title	:	Crystal structure of Aquifex aeolicus sigmaN bound to promoter DNA
Authors	:	Darst, S.A.; Campbell, E.A.; Rajashankar, K.
Deposited on	:	2017-01-12
Resolution	:	3.40 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

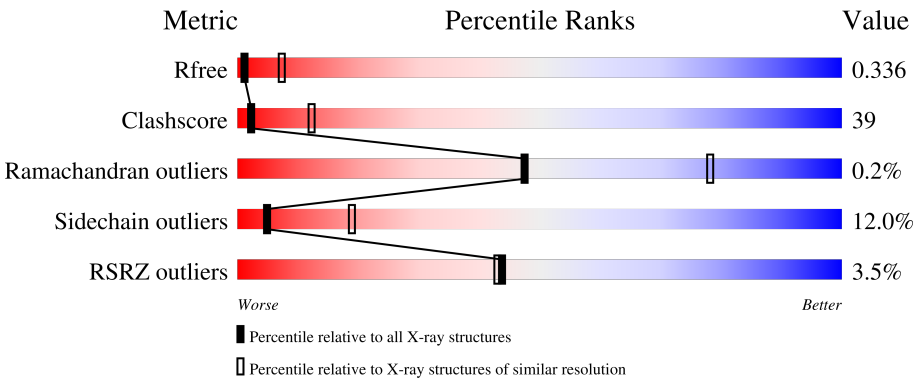
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 3.40 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	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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 of 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% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	31	19%, 45%, 55%
1	N	31	16%, 45%, 55%
2	B	31	13%, 42%, 42%, 13%, .
2	O	31	3%, 48%, 48%, .
3	I	342	%, 34%, 50%, 8%, 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	V	342	2% 32% 53% 9% 6%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7789 atoms, of which 1 is hydrogen 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 DNA chain called DNA (31-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	A	31	641	306	129	176	30	0	0	0
1	N	31	641	306	129	176	30	0	0	0

- Molecule 2 is a DNA chain called DNA (30-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	30	608	294	96	188	30	0	0	0
2	O	30	608	294	96	188	30	0	0	0

- Molecule 3 is a protein called RNA polymerase sigma factor RpoN.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C		N	O	S			
3	I	316	2605	1667		439	495	4	0	0	0
			Total	C	H	N	O	S			
3	V	323	2654	1695	1	446	508	4	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	57	GLY	-	expression tag	UNP O66858
I	58	PRO	-	expression tag	UNP O66858
I	59	HIS	-	expression tag	UNP O66858
I	60	MET	-	expression tag	UNP O66858
V	57	GLY	-	expression tag	UNP O66858
V	58	PRO	-	expression tag	UNP O66858
V	59	HIS	-	expression tag	UNP O66858
V	60	MET	-	expression tag	UNP O66858

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	5	Total O 5 5	0	0
4	B	4	Total O 4 4	0	0
4	I	7	Total O 7 7	0	0
4	N	6	Total O 6 6	0	0
4	O	1	Total O 1 1	0	0
4	V	9	Total O 9 9	0	0

3 Residue-property plots (i)

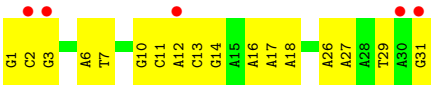
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). 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: DNA (31-MER)

Chain A:

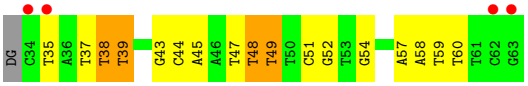
• Molecule 1: DNA (31-MER)

Chain N:



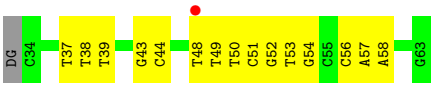
• Molecule 2: DNA (30-MER)

Chain B:



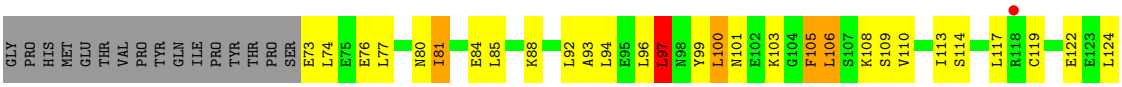
• Molecule 2: DNA (30-MER)

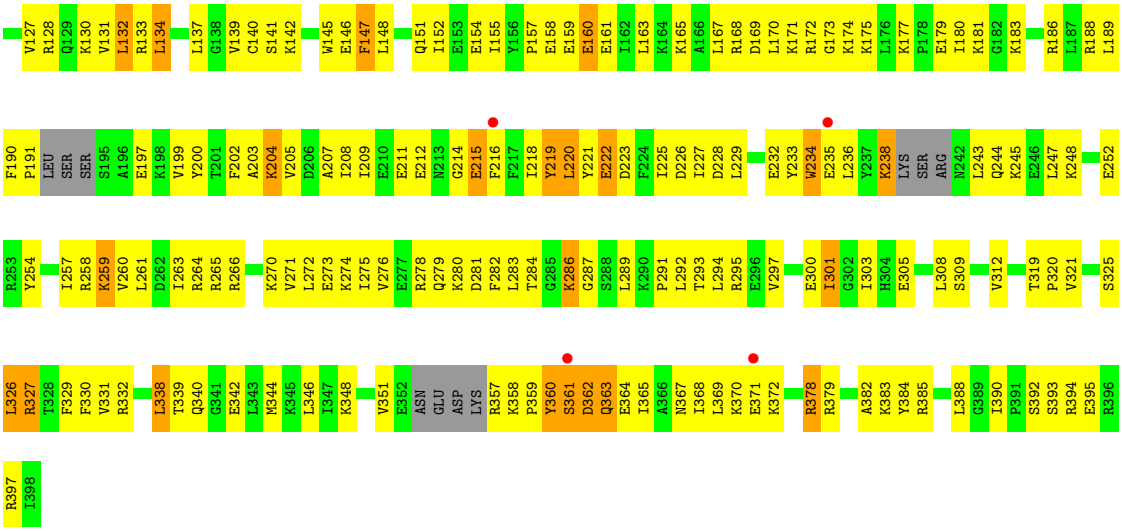
Chain O:



• Molecule 3: RNA polymerase sigma factor RpoN

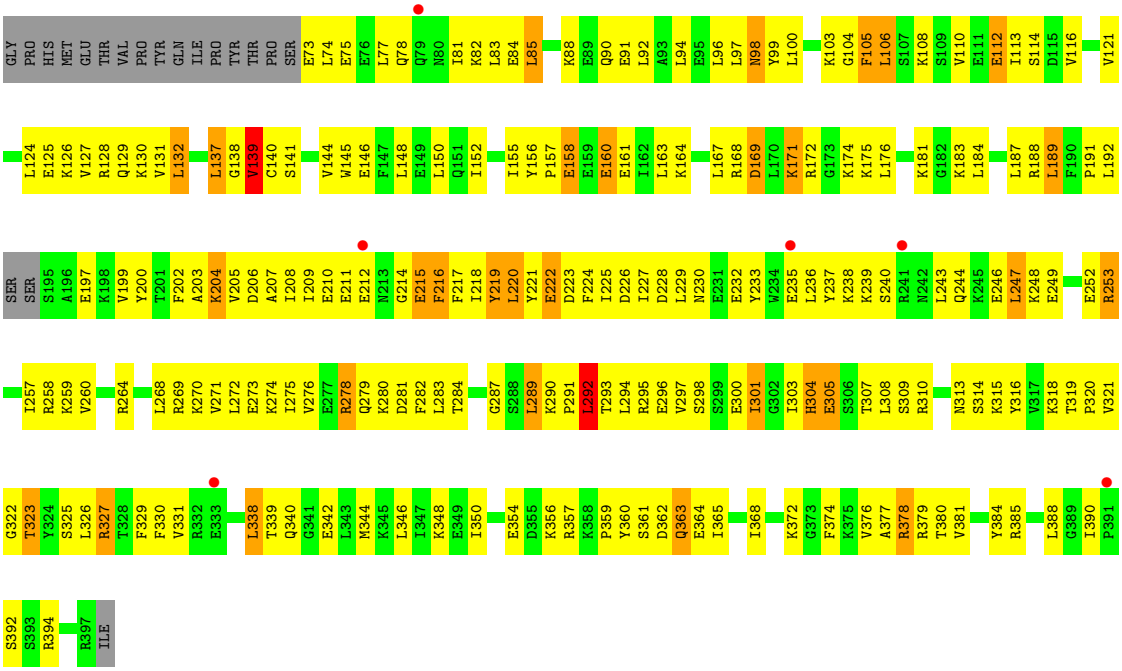
Chain I:





- Molecule 3: RNA polymerase sigma factor RpoN

Chain V:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α, β, γ	99.92Å 99.92Å 123.30Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	14.99 – 3.40 14.99 – 3.40	Depositor EDS
% Data completeness (in resolution range)	99.9 (14.99-3.40) 99.9 (14.99-3.40)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.26 (at 3.40Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, R_{free}	0.289 , 0.336 0.289 , 0.336	Depositor DCC
R_{free} test set	950 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	119.7	Xtriage
Anisotropy	0.273	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.25 , 53.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.017 for -h,-k,l 0.056 for h,-h-k,-l 0.027 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	7789	wwPDB-VP
Average B, all atoms (Å^2)	124.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.81% 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 (i)

5.1 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 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	0.53	0/723	1.01	0/1115
1	N	0.53	0/723	0.96	0/1115
2	B	0.60	0/677	1.43	7/1042 (0.7%)
2	O	0.58	0/677	1.18	0/1042
3	I	0.24	0/2638	0.46	0/3531
3	V	0.26	0/2688	0.49	0/3601
All	All	0.39	0/8126	0.80	7/11446 (0.1%)

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
3	I	0	2
3	V	0	1
All	All	0	3

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	48	DT	OP2-P-O3'	-12.29	78.17	105.20
2	B	39	DT	OP1-P-OP2	10.63	135.54	119.60
2	B	38	DT	OP2-P-O3'	-10.51	82.08	105.20
2	B	49	DT	OP1-P-OP2	9.57	133.96	119.60
2	B	38	DT	OP1-P-O3'	-9.36	84.61	105.20

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	I	284	THR	Peptide
3	I	97	LEU	Peptide
3	V	292	LEU	Peptide

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	641	0	349	23	1
1	N	641	0	349	18	1
2	B	608	0	345	19	0
2	O	608	0	345	20	0
3	I	2605	0	2674	239	0
3	V	2653	1	2714	284	0
4	A	5	0	0	0	0
4	B	4	0	0	0	0
4	I	7	0	0	3	0
4	N	6	0	0	1	0
4	O	1	0	0	0	0
4	V	9	0	0	3	0
All	All	7788	1	6776	565	2

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 39.

The worst 5 of 565 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)
3:V:281:ASP:HB2	3:V:287:GLY:HA3	1.31	1.12
3:V:210:GLU:HB2	3:V:211:GLU:HB2	1.37	1.06
3:V:100:LEU:HA	3:V:106:LEU:HA	1.39	1.05
2:B:48:DT:H4'	2:B:49:DT:OP1	1.60	1.02
3:I:281:ASP:HB2	3:I:287:GLY:HA3	1.41	1.00

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1:DG:O5'	1:A:31:DG:O3'[1_445]	2.11	0.09
1:N:1:DG:O5'	1:N:31:DG:O3'[1_565]	2.19	0.01

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	
3	I	308/342 (90%)	294 (96%)	14 (4%)	0	100	100
3	V	319/342 (93%)	305 (96%)	13 (4%)	1 (0%)	41	72
All	All	627/684 (92%)	599 (96%)	27 (4%)	1 (0%)	47	78

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	V	139	VAL

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	
3	I	286/317 (90%)	252 (88%)	34 (12%)	5	19
3	V	291/317 (92%)	256 (88%)	35 (12%)	5	19
All	All	577/634 (91%)	508 (88%)	69 (12%)	5	19

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	V	278	ARG
3	V	292	LEU
3	V	327	ARG
3	I	305	GLU
3	I	301	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
3	I	363	GLN
3	I	367	ASN
3	V	363	GLN
3	V	78	GLN
3	I	304	HIS

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 monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	31/31 (100%)	1.05	6 (19%)	1	1	101, 140, 188, 191	0
1	N	31/31 (100%)	0.79	5 (16%)	1	2	118, 156, 178, 182	0
2	B	30/31 (96%)	0.72	4 (13%)	3	4	112, 150, 184, 187	0
2	O	30/31 (96%)	0.62	1 (3%)	46	45	125, 148, 165, 175	0
3	I	316/342 (92%)	0.11	5 (1%)	72	70	85, 111, 146, 163	1 (0%)
3	V	323/342 (94%)	0.07	6 (1%)	66	65	85, 112, 145, 164	0
All	All	761/808 (94%)	0.20	27 (3%)	44	43	85, 116, 162, 191	1 (0%)

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	31	DG	4.8
1	A	1	DG	4.7
1	A	30	DA	4.4
3	I	118	ARG	3.5
2	B	63	DG	3.0

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

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

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.