W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 30, 2021 – 04:36 PM EST

PDB ID	:	3HUU
Title	:	Crystal structure of transcription regulator like protein from Staphylococcus haemolyticus
Authors	:	Agarwal, R.; Burley, S.K.; Swaminathan, S.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2009-06-15
Resolution	:	1.95 Å(reported)

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

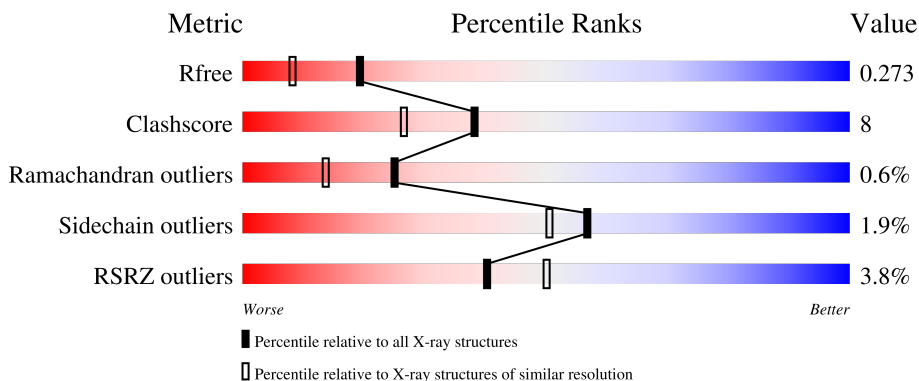|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.16
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.16

1 Overall quality at a glance ⓘ

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

The reported resolution of this entry is 1.95 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	Percentile Ranks	Value
Rfree		0.273
Clashscore		8
Ramachandran outliers		0.6%
Sidechain outliers		1.9%
RSRZ outliers		3.8%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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	305	3% 69% 17% • 13%
1	B	305	3% 70% 14% • 15%
1	C	305	5% 70% 17% 12%
1	D	305	3% 73% 13% • 13%

2   Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 8497 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 Transcription regulator like protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	265	Total 2100	C 1325	N 354	O 412	S 9	0	0	0
1	B	260	Total 2059	C 1302	N 342	O 406	S 9	0	0	0
1	C	267	Total 2115	C 1333	N 358	O 415	S 9	0	0	0
1	D	265	Total 2095	C 1321	N 353	O 412	S 9	0	0	0

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	expression tag	UNP Q4L6K9
A	0	SER	-	expression tag	UNP Q4L6K9
A	1	LEU	-	expression tag	UNP Q4L6K9
A	296	GLU	-	expression tag	UNP Q4L6K9
A	297	GLY	-	expression tag	UNP Q4L6K9
A	298	HIS	-	expression tag	UNP Q4L6K9
A	299	HIS	-	expression tag	UNP Q4L6K9
A	300	HIS	-	expression tag	UNP Q4L6K9
A	301	HIS	-	expression tag	UNP Q4L6K9
A	302	HIS	-	expression tag	UNP Q4L6K9
A	303	HIS	-	expression tag	UNP Q4L6K9
B	-1	MET	-	expression tag	UNP Q4L6K9
B	0	SER	-	expression tag	UNP Q4L6K9
B	1	LEU	-	expression tag	UNP Q4L6K9
B	296	GLU	-	expression tag	UNP Q4L6K9
B	297	GLY	-	expression tag	UNP Q4L6K9
B	298	HIS	-	expression tag	UNP Q4L6K9
B	299	HIS	-	expression tag	UNP Q4L6K9
B	300	HIS	-	expression tag	UNP Q4L6K9
B	301	HIS	-	expression tag	UNP Q4L6K9
B	302	HIS	-	expression tag	UNP Q4L6K9

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	303	HIS	-	expression tag	UNP Q4L6K9
C	-1	MET	-	expression tag	UNP Q4L6K9
C	0	SER	-	expression tag	UNP Q4L6K9
C	1	LEU	-	expression tag	UNP Q4L6K9
C	296	GLU	-	expression tag	UNP Q4L6K9
C	297	GLY	-	expression tag	UNP Q4L6K9
C	298	HIS	-	expression tag	UNP Q4L6K9
C	299	HIS	-	expression tag	UNP Q4L6K9
C	300	HIS	-	expression tag	UNP Q4L6K9
C	301	HIS	-	expression tag	UNP Q4L6K9
C	302	HIS	-	expression tag	UNP Q4L6K9
C	303	HIS	-	expression tag	UNP Q4L6K9
D	-1	MET	-	expression tag	UNP Q4L6K9
D	0	SER	-	expression tag	UNP Q4L6K9
D	1	LEU	-	expression tag	UNP Q4L6K9
D	296	GLU	-	expression tag	UNP Q4L6K9
D	297	GLY	-	expression tag	UNP Q4L6K9
D	298	HIS	-	expression tag	UNP Q4L6K9
D	299	HIS	-	expression tag	UNP Q4L6K9
D	300	HIS	-	expression tag	UNP Q4L6K9
D	301	HIS	-	expression tag	UNP Q4L6K9
D	302	HIS	-	expression tag	UNP Q4L6K9
D	303	HIS	-	expression tag	UNP Q4L6K9

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	38	Total 38	O 38	0	0
2	B	30	Total 30	O 30	0	0
2	C	26	Total 26	O 26	0	0
2	D	34	Total 34	O 34	0	0

3 Residue-property plots ⓘ

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.
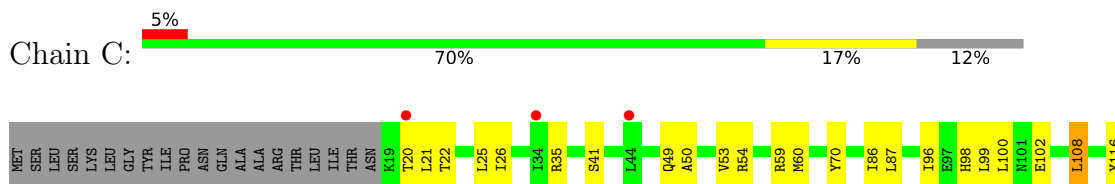
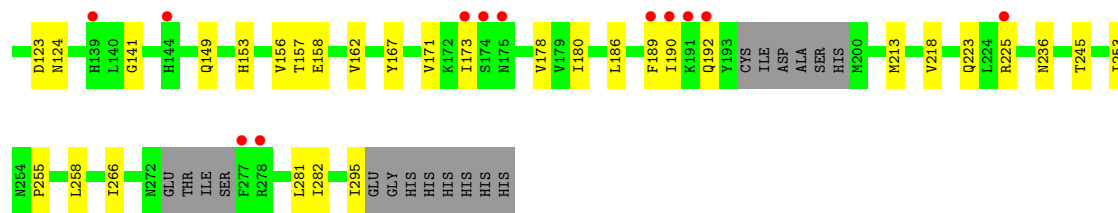- Molecule 1: Transcription regulator like protein
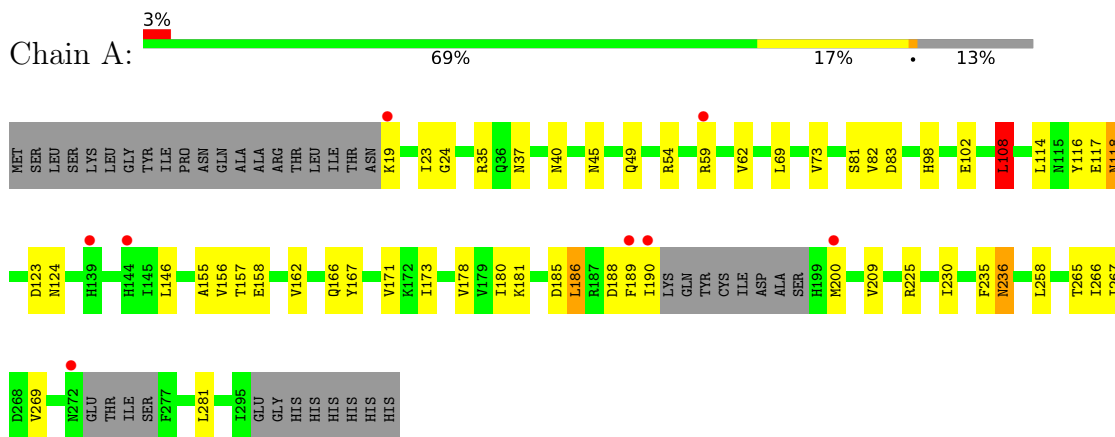
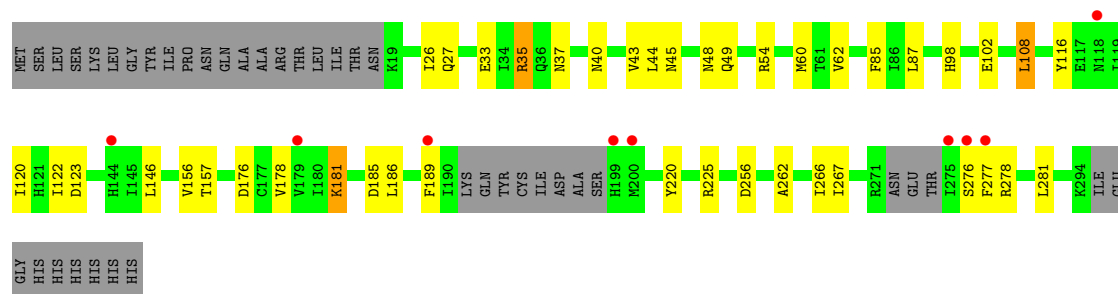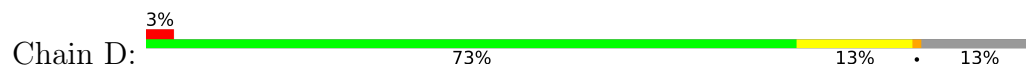- Molecule 1: Transcription regulator like protein

- Molecule 1: Transcription regulator like protein

• Molecule 1: Transcription regulator like protein

Chain D:
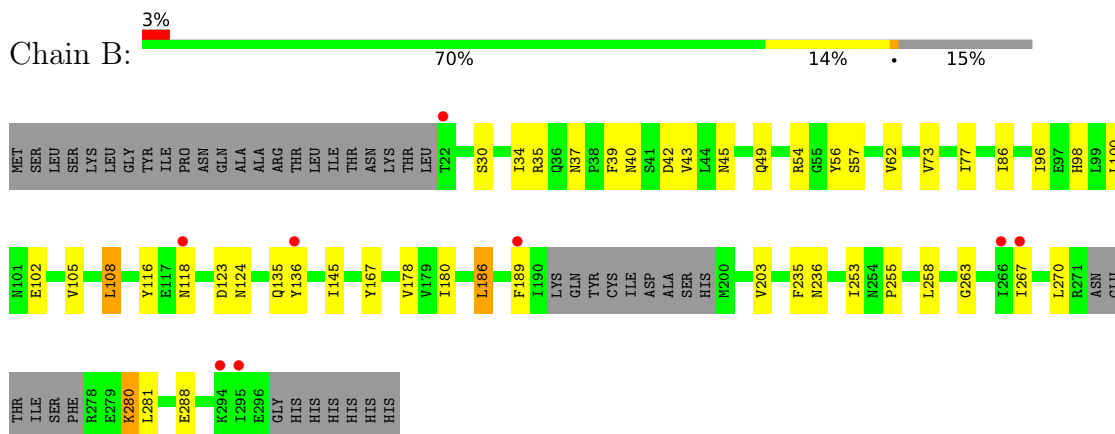
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	51.87Å 68.08Å 97.83Å 82.42° 75.95° 79.39°	Depositor
Resolution (Å)	43.76 – 1.95 43.76 – 1.95	Depositor EDS
% Data completeness (in resolution range)	84.2 (43.76-1.95) 84.3 (43.76-1.95)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.68 (at 1.95Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.237 , 0.274 0.238 , 0.273	Depositor DCC
R_{free} test set	2336 reflections (2.82%)	wwPDB-VP
Wilson B-factor (Å^2)	23.2	Xtriage
Anisotropy	0.399	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 42.4	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	8497	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 57.68 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.2846e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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	0.34	0/2133	0.61	1/2892 (0.0%)
1	B	0.34	0/2092	0.58	0/2840
1	C	0.38	1/2148 (0.0%)	0.61	1/2912 (0.0%)
1	D	0.35	0/2128	0.63	1/2886 (0.0%)
All	All	0.35	1/8501 (0.0%)	0.61	3/11530 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	180	ILE	C-N	-5.77	1.20	1.34

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	108	LEU	CA-CB-CG	5.83	128.72	115.30
1	D	108	LEU	CA-CB-CG	5.64	128.27	115.30
1	C	124	ASN	N-CA-C	-5.45	96.28	111.00

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts ⓘ

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	2100	0	2078	39	0
1	B	2059	0	2032	32	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	2115	0	2087	39	0
1	D	2095	0	2066	27	0
2	A	38	0	0	0	0
2	B	30	0	0	2	0
2	C	26	0	0	1	0
2	D	34	0	0	1	0
All	All	8497	0	8263	126	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 8.

The worst 5 of 126 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:A:62:VAL:HG11	1:B:49:GLN:HG3	1.39	1.05
1:C:86:ILE:HG12	1:C:108:LEU:HD11	1.55	0.88
1:A:59:ARG:HD3	1:B:57:SER:HA	1.56	0.87
1:A:49:GLN:HG2	1:B:62:VAL:HG21	1.57	0.87
1:B:86:ILE:HG12	1:B:108:LEU:HD21	1.65	0.79

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	259/305 (85%)	254 (98%)	4 (2%)	1 (0%)	34	22
1	B	254/305 (83%)	248 (98%)	5 (2%)	1 (0%)	34	22
1	C	261/305 (86%)	250 (96%)	9 (3%)	2 (1%)	19	9
1	D	259/305 (85%)	252 (97%)	5 (2%)	2 (1%)	19	9
All	All	1033/1220 (85%)	1004 (97%)	23 (2%)	6 (1%)	25	14

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	21	LEU
1	D	276	SER
1	A	236	ASN
1	C	236	ASN
1	D	278	ARG

5.3.2   Protein sidechains ⓘ

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	242/281 (86%)	239 (99%)	3 (1%)	71	68
1	B	238/281 (85%)	234 (98%)	4 (2%)	60	55
1	C	243/281 (86%)	239 (98%)	4 (2%)	62	58
1	D	241/281 (86%)	234 (97%)	7 (3%)	42	31
All	All	964/1124 (86%)	946 (98%)	18 (2%)	57	50

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	108	LEU
1	C	153	HIS
1	D	176	ASP
1	B	280	LYS
1	C	35	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 30 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	124	ASN
1	C	40	ASN
1	D	124	ASN
1	B	248	GLN
1	C	98	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates ⓘ

There are no monosaccharides 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 ⓘ

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	265/305 (86%)	0.39	8 (3%)	50	59	19, 28, 45, 56	0
1	B	260/305 (85%)	0.49	8 (3%)	49	58	18, 30, 47, 57	0
1	C	267/305 (87%)	0.59	15 (5%)	24	33	20, 31, 52, 66	0
1	D	265/305 (86%)	0.44	9 (3%)	45	55	18, 28, 47, 64	0
All	All	1057/1220 (86%)	0.48	40 (3%)	40	50	18, 29, 47, 66	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	277	PHE	16.8
1	D	277	PHE	11.4
1	D	275	ILE	7.1
1	A	272	ASN	5.9
1	D	189	PHE	5.0

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

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

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.