wwPDB X-ray Structure Validation Summary Report (i)

Jun 12, 2024 – 11:24 PM EDT

PDB ID	:	1JOG
Title	:	Structure of HI0074 from Heamophilus Influenzae reveals the fold of a substrate binding domain of a nucleotidyltransferase
Authors	:	Lehmann, C.; Lim, K.; Herzberg, O.; Structure 2 Function Project (S2F)
Deposited on	:	2001-07-29
Resolution	:	2.40 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 (i)) were used in the production of this report:

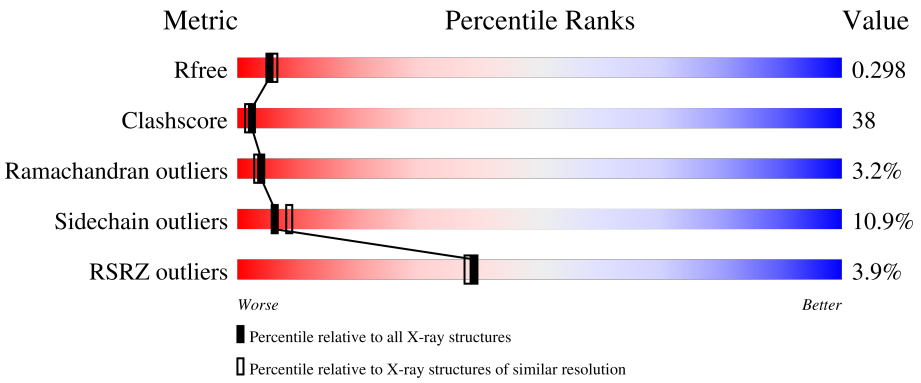
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	146	4%; 52%, 32%, 8%, •, 8%
1	B	146	48%, 36%, 6%, •, 8%
1	C	146	8%; 33%, 49%, 11%, 8%
1	D	146	2%; 45%, 40%, 8%, •, 7%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4575 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 HYPOTHETICAL PROTEIN HI0074.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	135	1106	706	183	211	6	0	0	0
1	B	135	1106	706	183	211	6	0	0	0
1	C	135	1106	706	183	211	6	0	0	0
1	D	136	1117	712	187	212	6	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP P43934
A	2	MSE	MET	MODIFIED RESIDUE	UNP P43934
A	33	MSE	MET	MODIFIED RESIDUE	UNP P43934
A	60	MSE	MET	MODIFIED RESIDUE	UNP P43934
A	61	MSE	MET	MODIFIED RESIDUE	UNP P43934
A	96	MSE	MET	MODIFIED RESIDUE	UNP P43934
A	105	MSE	MET	MODIFIED RESIDUE	UNP P43934
A	119	MSE	MET	MODIFIED RESIDUE	UNP P43934
B	1	MSE	MET	MODIFIED RESIDUE	UNP P43934
B	2	MSE	MET	MODIFIED RESIDUE	UNP P43934
B	33	MSE	MET	MODIFIED RESIDUE	UNP P43934
B	60	MSE	MET	MODIFIED RESIDUE	UNP P43934
B	61	MSE	MET	MODIFIED RESIDUE	UNP P43934
B	96	MSE	MET	MODIFIED RESIDUE	UNP P43934
B	105	MSE	MET	MODIFIED RESIDUE	UNP P43934
B	119	MSE	MET	MODIFIED RESIDUE	UNP P43934
C	1	MSE	MET	MODIFIED RESIDUE	UNP P43934
C	2	MSE	MET	MODIFIED RESIDUE	UNP P43934
C	33	MSE	MET	MODIFIED RESIDUE	UNP P43934
C	60	MSE	MET	MODIFIED RESIDUE	UNP P43934
C	61	MSE	MET	MODIFIED RESIDUE	UNP P43934

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	96	MSE	MET	MODIFIED RESIDUE	UNP P43934
C	105	MSE	MET	MODIFIED RESIDUE	UNP P43934
C	119	MSE	MET	MODIFIED RESIDUE	UNP P43934
D	1	MSE	MET	MODIFIED RESIDUE	UNP P43934
D	2	MSE	MET	MODIFIED RESIDUE	UNP P43934
D	33	MSE	MET	MODIFIED RESIDUE	UNP P43934
D	60	MSE	MET	MODIFIED RESIDUE	UNP P43934
D	61	MSE	MET	MODIFIED RESIDUE	UNP P43934
D	96	MSE	MET	MODIFIED RESIDUE	UNP P43934
D	105	MSE	MET	MODIFIED RESIDUE	UNP P43934
D	119	MSE	MET	MODIFIED RESIDUE	UNP P43934

- Molecule 2 is water.

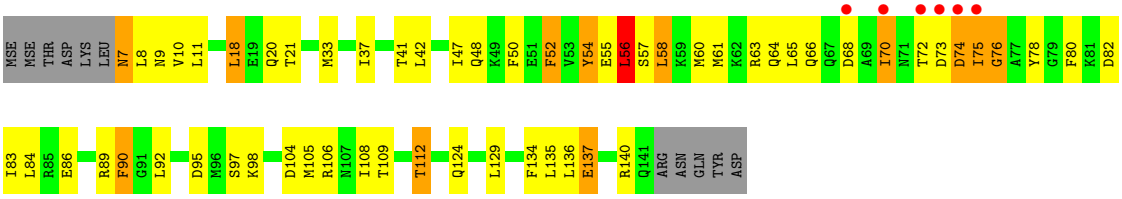
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	45	45	45	0	0
			Total	O		
2	B	40	40	40	0	0
			Total	O		
2	C	12	12	12	0	0
			Total	O		
2	D	43	43	43	0	0

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.

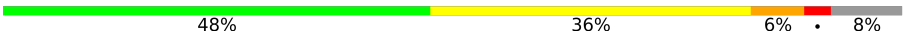
• Molecule 1: HYPOTHETICAL PROTEIN HI0074

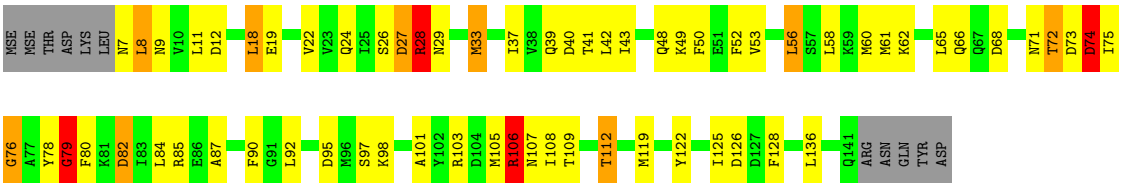
Chain A:



• Molecule 1: HYPOTHETICAL PROTEIN HI0074

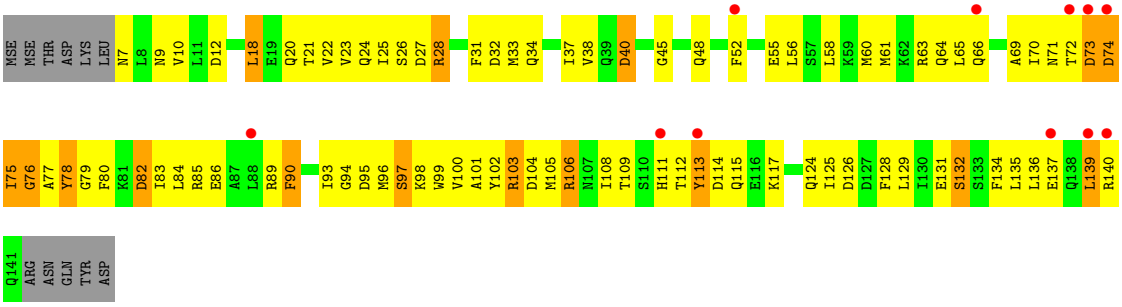
Chain B:





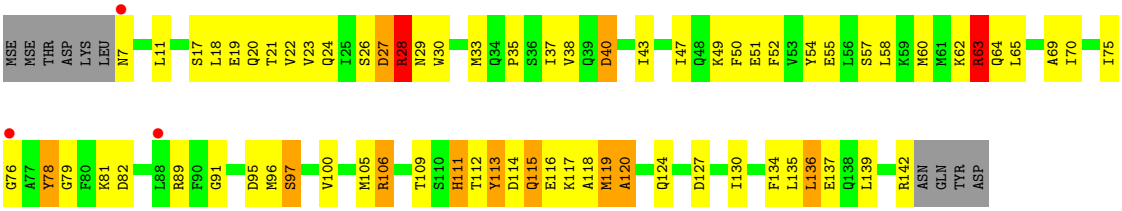
• Molecule 1: HYPOTHETICAL PROTEIN HI0074

Chain C:



• Molecule 1: HYPOTHETICAL PROTEIN HI0074

Chain D:


MSE
MSE
THR
ASP
LYS
LEU
N7
L11
S17
L18
E19
Q20
T21
V22
V23
Q24
I25
S26
D27
R28
N29
W30
M33
Q34
P35
S36
I37
V38
Q39
D40
I43
I47
Q48
K49
F50
E51
F52
V53
Y54
E55
L56
S57
L58
K59
M60
M61
K62
R63
Q64
L65
A69
I70
I75
G76
A77
Y78
G79
F80
K81
D82
L88
R89
F90
G91
D95
M96
S97
V100
M105
R106
T109
S110
H111
T112
Y113
D114
Q115
E116
K117
A118
M119
A120
Q124
D127
I130
F134
L135
L136
E137
Q138
L139
R142
ASN
GLN
TYR
ASP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	38.30Å 87.10Å 165.60Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.40 29.60 – 2.39	Depositor EDS
% Data completeness (in resolution range)	(Not available) (20.00-2.40) 95.9 (29.60-2.39)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.69 (at 2.39Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.211 , 0.282 0.232 , 0.298	Depositor DCC
R_{free} test set	1131 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	42.8	Xtriage
Anisotropy	0.672	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 62.0	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	4575	wwPDB-VP
Average B, all atoms (Å^2)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.83% 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	1.07	2/1119 (0.2%)	1.11	6/1499 (0.4%)
1	B	0.96	0/1119	1.05	2/1499 (0.1%)
1	C	0.87	0/1119	0.96	3/1499 (0.2%)
1	D	0.90	0/1130	1.01	1/1513 (0.1%)
All	All	0.95	2/4487 (0.0%)	1.03	12/6010 (0.2%)

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

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	52	PHE	CE2-CZ	5.77	1.48	1.37
1	A	54	TYR	CE1-CZ	-5.09	1.31	1.38

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	56	LEU	CA-CB-CG	7.04	131.50	115.30
1	A	104	ASP	CB-CG-OD1	6.66	124.30	118.30
1	B	79	GLY	N-CA-C	5.79	127.57	113.10
1	C	23	VAL	CB-CA-C	-5.60	100.75	111.40
1	A	58	LEU	CA-CB-CG	-5.43	102.82	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	122	TYR	Sidechain

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	1106	0	1085	79	0
1	B	1106	0	1085	65	0
1	C	1106	0	1085	139	0
1	D	1117	0	1098	91	0
2	A	45	0	0	0	0
2	B	40	0	0	0	0
2	C	12	0	0	2	0
2	D	43	0	0	4	0
All	All	4575	0	4353	334	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 38.

The worst 5 of 334 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:B:75:ILE:HG23	1:B:78:TYR:HB3	1.18	1.11
1:D:7:ASN:ND2	1:D:63:ARG:HH22	1.52	1.07
1:A:106:ARG:HG2	1:A:106:ARG:HH11	1.11	1.05
1:C:20:GLN:HB3	1:D:37:ILE:HD12	1.35	1.04
1:C:75:ILE:HG23	1:C:78:TYR:HD1	1.19	1.04

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	133/146 (91%)	123 (92%)	8 (6%)	2 (2%)	10	14
1	B	133/146 (91%)	121 (91%)	5 (4%)	7 (5%)	2	1
1	C	133/146 (91%)	112 (84%)	17 (13%)	4 (3%)	4	3
1	D	134/146 (92%)	121 (90%)	9 (7%)	4 (3%)	4	3
All	All	533/584 (91%)	477 (90%)	39 (7%)	17 (3%)	4	3

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	27	ASP
1	B	74	ASP
1	C	27	ASP
1	D	27	ASP
1	D	28	ARG

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	119/122 (98%)	108 (91%)	11 (9%)	9	13
1	B	119/122 (98%)	105 (88%)	14 (12%)	5	7
1	C	119/122 (98%)	107 (90%)	12 (10%)	7	11
1	D	120/122 (98%)	105 (88%)	15 (12%)	4	5
All	All	477/488 (98%)	425 (89%)	52 (11%)	6	8

5 of 52 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	73	ASP
1	C	136	LEU
1	D	115	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	78	TYR
1	C	103	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	7	ASN
1	C	66	GLN
1	D	115	GLN
1	D	7	ASN
1	D	64	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 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	129/146 (88%)	0.02	6 (4%) 31 30	23, 42, 82, 85	0
1	B	129/146 (88%)	-0.19	0 100 100	21, 50, 80, 83	0
1	C	129/146 (88%)	0.40	11 (8%) 10 10	38, 72, 83, 85	0
1	D	130/146 (89%)	0.10	3 (2%) 60 58	34, 59, 82, 84	0
All	All	517/584 (88%)	0.08	20 (3%) 39 38	21, 56, 82, 85	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	73	ASP	6.8
1	A	72	THR	5.2
1	C	72	THR	4.9
1	C	74	ASP	4.5
1	A	74	ASP	3.9

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.