WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Dec 19, 2023 – 08:06 AM EST

PDB ID	:	1IU3
Title	:	CRYSTAL STRUCTURE OF THE E.COLI SEQA PROTEIN COMPLEXED WITH HEMIMETHYLATED DNA
Authors	:	Fujikawa, N.; Kurumizaka, H.; Nureki, O.; Tanaka, Y.; Yamazoe, M.; Hiraga, S.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2002-02-26
Resolution	:	3.00 Å(reported)

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

|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
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

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 3.00 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.
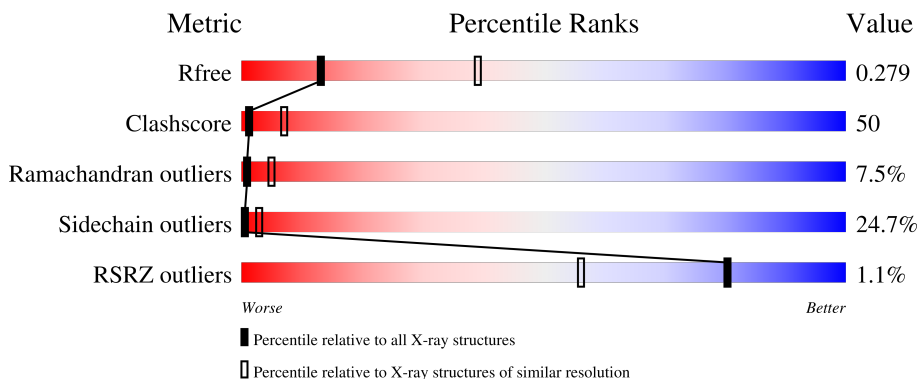
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	10	30% 30% 20% 20%
1	D	10	10% 60% 20% 10%
2	B	10	20% 70% 10%
2	E	10	10% / 60% 40%
3	C	116	% / 21% 47% 20% 11% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	116	

%

27%   47%   21%   5% •

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 2755 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 DNA chain called 5'-D(*AP*AP*GP*GP*AP*TP*CP*CP*AP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	10	Total	C	N	O	P	0	0	0
			204	98	43	54	9			
1	D	10	Total	C	N	O	P	0	0	0
			204	98	43	54	9			

- Molecule 2 is a DNA chain called 5'-D(*TP*TP*GP*GP*AP*TP*CP*CP*TP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	10	Total	C	N	O	P	0	0	0
			200	98	31	62	9			
2	E	10	Total	C	N	O	P	0	0	0
			200	98	31	62	9			

- Molecule 3 is a protein called SeqA protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	115	Total	C	N	O	S	0	0	0
			908	572	157	172	7			
3	F	115	Total	C	N	O	S	0	0	0
			908	572	157	172	7			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1	GLY	-	cloning artifact	UNP P36658
C	2	PRO	-	cloning artifact	UNP P36658
C	3	LEU	-	cloning artifact	UNP P36658
C	4	GLY	-	cloning artifact	UNP P36658
C	5	SER	-	cloning artifact	UNP P36658
F	1	GLY	-	cloning artifact	UNP P36658
F	2	PRO	-	cloning artifact	UNP P36658
F	3	LEU	-	cloning artifact	UNP P36658

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	4	GLY	-	cloning artifact	UNP P36658
F	5	SER	-	cloning artifact	UNP P36658

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	9	Total 9	O 9	0	0
4	B	9	Total 9	O 9	0	0
4	D	12	Total 12	O 12	0	0
4	E	13	Total 13	O 13	0	0
4	C	59	Total 59	O 59	0	0
4	F	29	Total 29	O 29	0	0

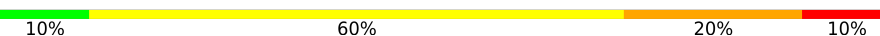# 3 Residue-property plots ⓘ

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.
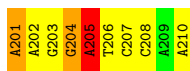
• Molecule 1: 5'-D(*AP*AP*GP*GP*AP*TP*CP*CP*AP*A)-3'

Chain A:

• Molecule 1: 5'-D(*AP*AP*GP*GP*AP*TP*CP*CP*AP*A)-3'

Chain D:

• Molecule 2: 5'-D(*TP*TP*GP*GP*AP*TP*CP*CP*TP*T)-3'

Chain B:

• Molecule 2: 5'-D(*TP*TP*GP*GP*AP*TP*CP*CP*TP*T)-3'

Chain E:

• Molecule 3: SeqA protein

Chain C:

• Molecule 3: SeqA protein

Chain F:

%

27%   47%   21%   5%

GLY P2 L3 G4 S5 R8 E9 L10 L11 D14 E15 Y16 A17 E18 Q19 K20 R21 A22 V23 N24 R25 F26 M27 L28 S31 T32 L33 Y34 S35 L36 D37 A38 Q39 A40 T45 E46 S47 L48 R51 T52 R53 V54 Y55 E60 Q61 T62 L63 L64 K65 N66 G67 N68 Q69 T70

K71 P72 K73 H74 V75 P76 G77 T78 P79 Y80 W81 V82 I83 T84 N85 T86 N87 T88 G89 R90 K91 C92 S93 M94 I95 E96 H97 I98 M99 Q100 S101 M102 Q103 F104 P105 A106 E107 L108 I109 E110 K111 V112 C113 G114 T115 I116

4   Data and refinement statistics   ⓘ

Property	Value	Source
Space group	P 6 2 2	Depositor
Cell constants a, b, c, α, β, γ	152.59Å 152.59Å 119.36Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	40.00 – 3.00 41.33 – 3.00	Depositor EDS
% Data completeness (in resolution range)	(Not available) (40.00-3.00) 97.8 (41.33-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.89 (at 3.01Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.238 , 0.286 0.237 , 0.279	Depositor DCC
R_{free} test set	816 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å^2)	85.1	Xtriage
Anisotropy	0.055	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 67.3	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	0.92	EDS
Total number of atoms	2755	wwPDB-VP
Average B, all atoms (Å^2)	79.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.32% 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	2.05	9/230 (3.9%)	1.96	11/353 (3.1%)
1	D	1.69	5/230 (2.2%)	1.60	4/353 (1.1%)
2	B	1.94	5/222 (2.3%)	1.77	7/341 (2.1%)
2	E	1.46	0/222	1.46	2/341 (0.6%)
3	C	2.02	22/925 (2.4%)	1.83	30/1248 (2.4%)
3	F	1.29	4/925 (0.4%)	1.39	8/1248 (0.6%)
All	All	1.73	45/2754 (1.6%)	1.66	62/3884 (1.6%)

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	A	0	2
1	D	0	2
2	B	0	3
2	E	0	3
3	C	0	2
All	All	0	12

The worst 5 of 45 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	75	VAL	CA-CB	11.86	1.79	1.54
3	C	16	TYR	CE2-CZ	-9.64	1.26	1.38
3	F	113	CYS	CB-SG	9.17	1.97	1.82
3	C	23	VAL	CB-CG2	-8.37	1.35	1.52
3	C	92	CYS	CB-SG	-7.78	1.69	1.82

The worst 5 of 62 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	203	DG	O5'-P-OP2	-13.65	93.42	105.70
1	D	202	DA	O5'-P-OP2	-12.14	94.78	105.70
3	C	75	VAL	C-N-CD	-10.59	97.30	120.60
3	C	25	ARG	NE-CZ-NH1	-10.29	115.16	120.30
1	A	203	DG	O5'-P-OP1	9.52	122.12	110.70

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	205	DA	Sidechain
1	A	210	DA	Sidechain
2	B	211	DT	Sidechain
2	B	217	DC	Sidechain
2	B	220	DT	Sidechain

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	204	0	113	7	0
1	D	204	0	113	10	4
2	B	200	0	117	17	0
2	E	200	0	117	14	0
3	C	908	0	909	98	0
3	F	908	0	909	108	0
4	A	9	0	0	4	0
4	B	9	0	0	4	0
4	C	59	0	0	21	0
4	D	12	0	0	4	0
4	E	13	0	0	7	0
4	F	29	0	0	12	0
All	All	2755	0	2278	243	4

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 50.

The worst 5 of 243 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:C:75:VAL:CB	3:C:75:VAL:CA	1.79	1.55
3:F:27:MET:SD	3:F:27:MET:CE	2.03	1.46
3:C:76:PRO:CG	3:C:76:PRO:CB	1.80	1.45
3:C:113:CYS:HB2	4:C:174:HOH:O	1.46	1.14
3:F:77:GLY:HA2	4:F:126:HOH:O	1.47	1.12

All (4) 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:D:201:DA:N3	1:D:201:DA:N3[10_665]	1.03	1.17
1:D:201:DA:N3	1:D:201:DA:C4[10_665]	2.05	0.15
1:D:201:DA:C2	1:D:201:DA:N3[10_665]	2.08	0.12
1:D:201:DA:C1'	1:D:201:DA:N9[10_665]	2.16	0.04

5.3   Torsion angles   ⓘ

5.3.1   Protein backbone   ⓘ

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	C	113/116 (97%)	89 (79%)	17 (15%)	7 (6%)	1	8
3	F	113/116 (97%)	89 (79%)	14 (12%)	10 (9%)	1	3
All	All	226/232 (97%)	178 (79%)	31 (14%)	17 (8%)	1	5

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	F	74	HIS
3	F	78	THR
3	C	67	GLY
3	F	65	LYS
3	C	13	SER

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	
3	C	99/99 (100%)	77 (78%)	22 (22%)	1	4
3	F	99/99 (100%)	72 (73%)	27 (27%)	0	2
All	All	198/198 (100%)	149 (75%)	49 (25%)	0	3

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	F	39	GLN
3	F	72	PRO
3	F	45	THR
3	F	66	ASN
3	F	75	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
3	F	68	ASN
3	F	69	GLN
3	F	100	GLN
3	C	74	HIS
3	C	24	ASN

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	10/10 (100%)	-0.55	0	100	100	59, 86, 100, 102	0
1	D	10/10 (100%)	-0.37	0	100	100	73, 96, 105, 130	0
2	B	10/10 (100%)	-0.21	0	100	100	62, 80, 120, 131	0
2	E	10/10 (100%)	0.06	1 (10%)	7	2	80, 89, 140, 157	0
3	C	115/116 (99%)	-0.39	1 (0%)	84	63	33, 55, 79, 93	0
3	F	115/116 (99%)	-0.18	1 (0%)	84	63	70, 91, 106, 116	0
All	All	270/272 (99%)	-0.28	3 (1%)	80	56	33, 78, 105, 157	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	220	DT	4.1
3	C	74	HIS	2.5
3	F	74	HIS	2.5

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 (i)

There are no such residues in this entry.