wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 07:02 AM GMT

PDB ID	:	2Z9O
Title	:	Crystal structure of the dimeric form of RepE in complex with the repE operator DNA
Authors	:	Nakamura, A.; Wada, C.; Miki, K.
Deposited on	:	2007-09-21
Resolution	:	3.14 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
http://wwpdb.org/validation/2016/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:

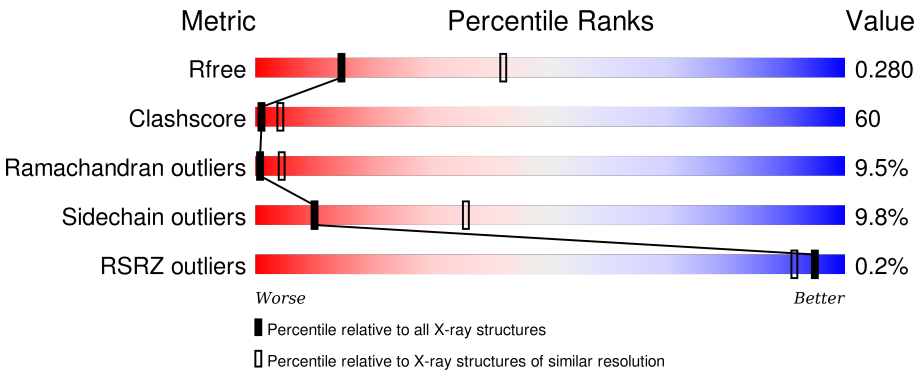
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 3.14 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}	91344	1095 (3.18-3.10)
Clashscore	102246	1202 (3.18-3.10)
Ramachandran outliers	100387	1162 (3.18-3.10)
Sidechain outliers	100360	1162 (3.18-3.10)
RSRZ outliers	91569	1097 (3.18-3.10)

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. 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	C	33	9% 88% •
2	D	33	9% 91%
3	A	266	20% 52% 11% • 17%
3	B	266	22% 51% 9% 18%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4926 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 DNA (33-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	33	669	323	118	196	32	0	0	0

- Molecule 2 is a DNA chain called DNA (33-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	33	678	326	124	196	32	0	0	0

- Molecule 3 is a protein called Replication initiation protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	222	1796	1151	312	325	8	0	0	0
			Total	C	N	O	S			
3	B	218	1783	1142	310	324	7	0	0	0

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-14	MET	-	EXPRESSION TAG	UNP P03856
A	-13	ARG	-	EXPRESSION TAG	UNP P03856
A	-12	GLY	-	EXPRESSION TAG	UNP P03856
A	-11	SER	-	EXPRESSION TAG	UNP P03856
A	-10	HIS	-	EXPRESSION TAG	UNP P03856
A	-9	HIS	-	EXPRESSION TAG	UNP P03856
A	-8	HIS	-	EXPRESSION TAG	UNP P03856
A	-7	HIS	-	EXPRESSION TAG	UNP P03856
A	-6	HIS	-	EXPRESSION TAG	UNP P03856
A	-5	HIS	-	EXPRESSION TAG	UNP P03856
A	-4	GLY	-	EXPRESSION TAG	UNP P03856
A	-3	SER	-	EXPRESSION TAG	UNP P03856

Continued on next page...

Continued from previous page...

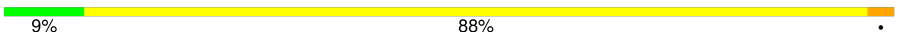
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	ILE	-	EXPRESSION TAG	UNP P03856
A	-1	GLU	-	EXPRESSION TAG	UNP P03856
A	0	GLY	-	EXPRESSION TAG	UNP P03856
A	1	ARG	-	EXPRESSION TAG	UNP P03856
B	-14	MET	-	EXPRESSION TAG	UNP P03856
B	-13	ARG	-	EXPRESSION TAG	UNP P03856
B	-12	GLY	-	EXPRESSION TAG	UNP P03856
B	-11	SER	-	EXPRESSION TAG	UNP P03856
B	-10	HIS	-	EXPRESSION TAG	UNP P03856
B	-9	HIS	-	EXPRESSION TAG	UNP P03856
B	-8	HIS	-	EXPRESSION TAG	UNP P03856
B	-7	HIS	-	EXPRESSION TAG	UNP P03856
B	-6	HIS	-	EXPRESSION TAG	UNP P03856
B	-5	HIS	-	EXPRESSION TAG	UNP P03856
B	-4	GLY	-	EXPRESSION TAG	UNP P03856
B	-3	SER	-	EXPRESSION TAG	UNP P03856
B	-2	ILE	-	EXPRESSION TAG	UNP P03856
B	-1	GLU	-	EXPRESSION TAG	UNP P03856
B	0	GLY	-	EXPRESSION TAG	UNP P03856
B	1	ARG	-	EXPRESSION TAG	UNP P03856

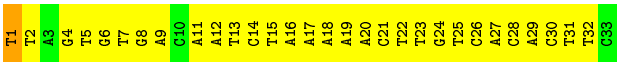
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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 (33-MER)

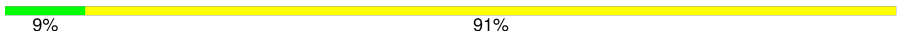
Chain C:

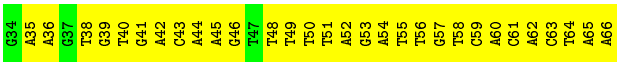




- Molecule 2: DNA (33-MER)

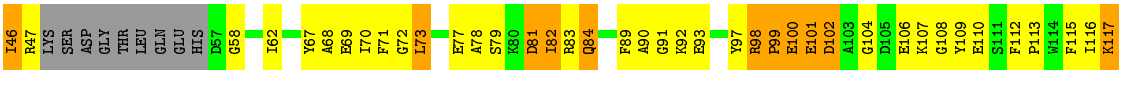
Chain D:





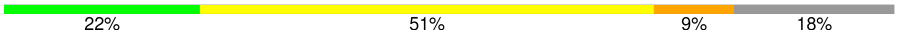
- Molecule 3: Replication initiation protein

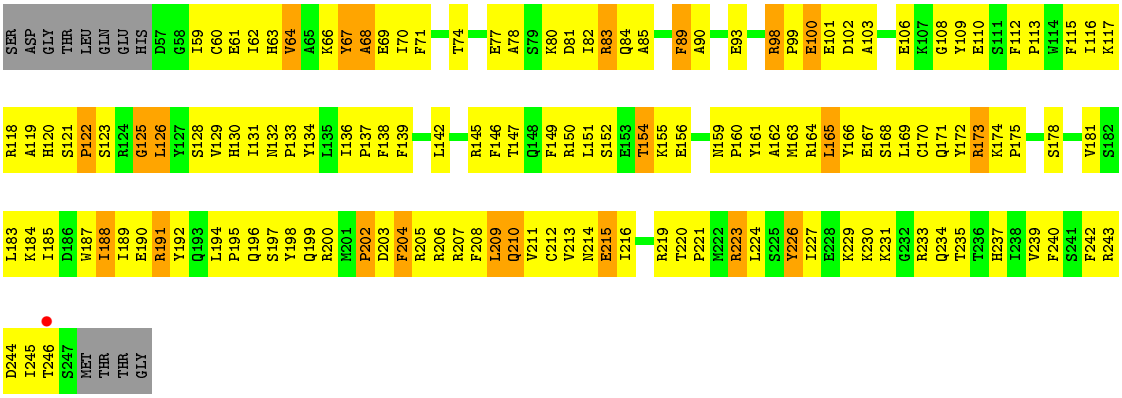
Chain A:



- Molecule 3: Replication initiation protein

Chain B:




SER ASP GLY THR LEU GLN GLU HIS D57 G58 I59 C60 E61 I62 H63 V64 A65 K66 Y67 A68 E69 I70 F71 T74 E77 A78 S79 K80 D81 I82 R83 Q84 A85 F89 A90 E93 R98 P99 E100 E101 D102 A103 E106 K107 G108 Y109 E110 S111 F112 P113 W114 F115 I116 K117
R118 A119 H120 S121 P122 S123 R124 G125 L126 Y127 S128 V129 L194 P195 Q196 S197 Y198 Q199 R200 L135 I136 P137 F138 F139 L142 R145 F146 T147 Q148 F149 R150 L151 S152 E153 T154 K155 E156 N159 P160 Y161 A162 M163 R164 L165 Y166 E167 S168 L169 C170 Q171 Y172 R173 K174 P175 S178 V181 S182
L183 K184 I185 D186 W187 I188 I189 E190 R191 Y192 Q193 M201 P202 D203 F204 R205 R206 R207 F208 L209 Q210 V211 C212 V213 N214 E215 I216 R219 T220 P221 M222 R223 L224 S225 Y226 I227 E228 K229 K230 K231 G232 R233 Q234 T235 T236 H237 I238 V239 F240 S241 F242 R243
D244 I245 T246 S247 MET THR THR GLY

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	60.73Å 99.32Å 95.00Å 90.00° 108.55° 90.00°	Depositor
Resolution (Å)	45.03 – 3.14 45.03 – 3.14	Depositor EDS
% Data completeness (in resolution range)	87.8 (45.03-3.14) 87.8 (45.03-3.14)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	3.01 (at 3.12Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.266 , 0.313 0.247 , 0.280	Depositor DCC
R_{free} test set	850 reflections (5.15%)	DCC
Wilson B-factor (Å^2)	51.2	Xtriage
Anisotropy	1.080	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , -6.0	EDS
Estimated twinning fraction	0.088 for h,-k,-h-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.25$	Xtriage
Outliers	0 of 16519 reflections	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	4926	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.43% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	C	0.46	0/749	0.72	0/1153
2	D	0.43	0/761	0.69	0/1174
3	A	0.47	0/1839	0.71	0/2480
3	B	0.47	0/1826	0.68	0/2462
All	All	0.46	0/5175	0.70	0/7269

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	C	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	1	DT	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	C	669	0	376	70	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	678	0	376	67	0
3	A	1796	0	1742	226	0
3	B	1783	0	1735	230	0
All	All	4926	0	4229	552	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 60.

The worst 5 of 552 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:B:200:ARG:HG2	3:B:202:PRO:HD2	1.15	1.10
2:D:39:DG:H2"	2:D:40:DT:H5"	1.25	1.10
2:D:41:DG:H2"	2:D:42:DA:H5'	1.38	1.04
1:C:6:DG:C2'	1:C:7:DT:H5"	1.86	1.04
2:D:60:DA:H2"	2:D:61:DC:H5'	1.40	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
3	A	218/266 (82%)	158 (72%)	40 (18%)	20 (9%)	1 4
3	B	214/266 (80%)	158 (74%)	35 (16%)	21 (10%)	1 4
All	All	432/532 (81%)	316 (73%)	75 (17%)	41 (10%)	1 4

5 of 41 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	91	GLY

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	A	99	PRO
3	B	68	ALA
3	B	147	THR
3	B	209	LEU

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	A	188/238 (79%)	162 (86%)	26 (14%)	4 19
3	B	189/238 (79%)	178 (94%)	11 (6%)	25 61
All	All	377/476 (79%)	340 (90%)	37 (10%)	10 36

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	165	LEU
3	A	205	ARG
3	B	215	GLU
3	A	176	ASP
3	A	189	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	196	GLN
3	A	237	HIS
3	B	148	GLN
3	A	171	GLN
3	B	130	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 carbohydrates 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	C	33/33 (100%)	-0.88	0	100	100	22, 65, 95, 99	0
2	D	33/33 (100%)	-0.85	0	100	100	39, 59, 103, 109	0
3	A	222/266 (83%)	-0.29	0	100	100	3, 33, 66, 79	0
3	B	218/266 (81%)	-0.22	1 (0%)	91	84	5, 41, 69, 80	0
All	All	506/598 (84%)	-0.33	1 (0%)	95	91	3, 41, 75, 109	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	B	246	THR	2.1

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 carbohydrates 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.