



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:06 AM GMT

PDB ID : 2PE5
Title : Crystal Structure of the Lac Repressor bound to ONPG in repressed state
Authors : Daber, R.; Stayrook, S.E.; Rosenberg, A.; Lewis, M.
Deposited on : 2007-04-02
Resolution : 3.50 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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

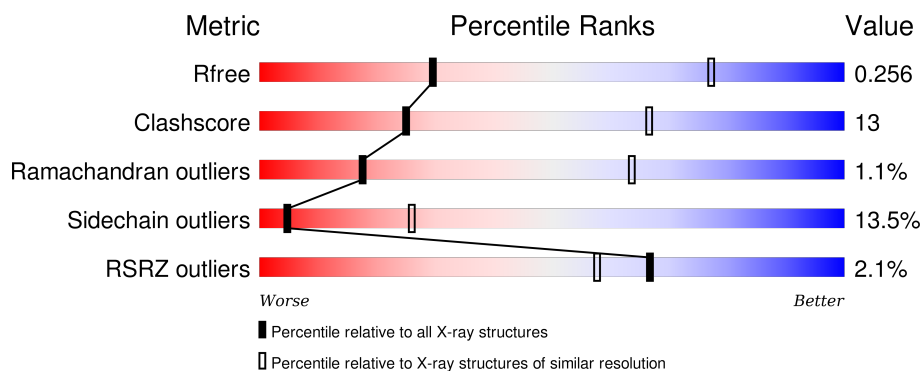
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.50 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.



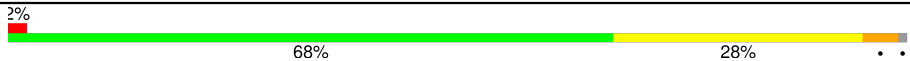
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1051 (3.60-3.40)
Clashscore	102246	1157 (3.60-3.40)
Ramachandran outliers	100387	1120 (3.60-3.40)
Sidechain outliers	100360	1121 (3.60-3.40)
RSRZ outliers	91569	1058 (3.60-3.40)

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 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 $\leq 5\%$. 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	D	20	<div> <div>30%</div> <div>50%</div> <div>5%</div> <div>15%</div> </div>
1	E	20	<div> <div>30%</div> <div>45%</div> <div>10%</div> <div>15%</div> </div>
1	F	20	<div> <div>40%</div> <div>40%</div> <div>35%</div> <div>10%</div> <div>15%</div> </div>
2	A	330	<div> <div>2%</div> <div>64%</div> <div>31%</div> <div>• •</div> </div>
2	B	330	<div> <div>%</div> <div>65%</div> <div>29%</div> <div>6%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	C	330	 A horizontal bar chart showing the quality of chain C. The bar is divided into four segments: a small red segment at the beginning labeled '2%', a large green segment labeled '68%', a large yellow segment labeled '28%', and a small grey segment at the end labeled with two dots '• •'.

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 8527 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 (5'-D(*DAP*DAP*DTP*DTP*DGP*DTP*DGP*DAP*DGP*DCP*DGP*DCP*DTP*DCP*DAP*DCP*DAP*DAP*DTP*DT)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	D	17	Total	C	N	O	P	0	0	0
			348	166	62	103	17			
1	E	17	Total	C	N	O	P	0	0	0
			349	166	65	101	17			
1	F	17	Total	C	N	O	P	0	0	0
			349	166	65	101	17			

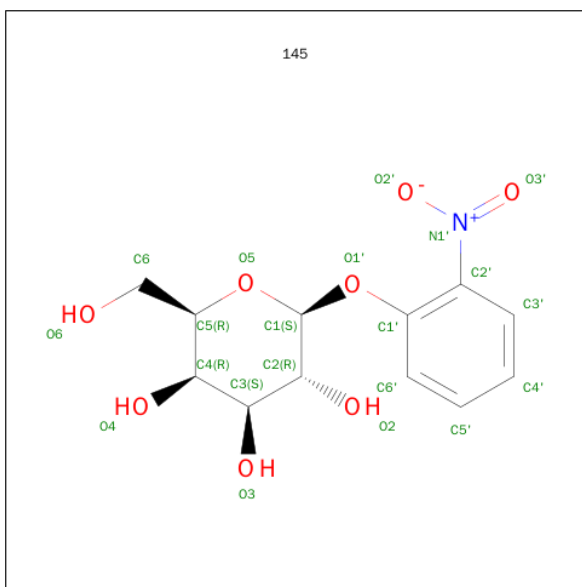
- Molecule 2 is a protein called Lactose operon repressor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	328	Total	C	N	O	S	0	0	0
			2468	1544	438	475	11			
2	B	330	Total	C	N	O	S	0	0	0
			2482	1553	440	478	11			
2	C	328	Total	C	N	O	S	0	0	0
			2468	1544	438	475	11			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	61	LEU	SER	ENGINEERED	UNP P03023
B	61	LEU	SER	ENGINEERED	UNP P03023
C	61	LEU	SER	ENGINEERED	UNP P03023

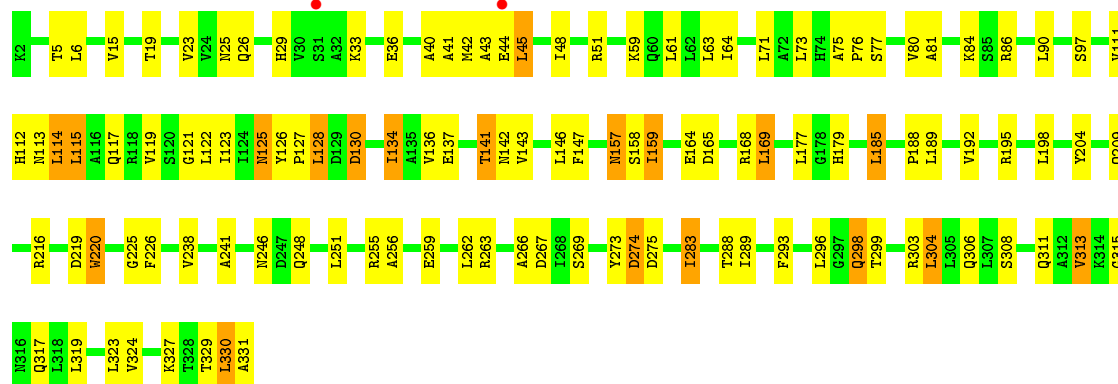
- Molecule 3 is 1-O-[O-NITROPHENYL]-BETA-D-GALACTOPYRANOSE (three-letter code: 145) (formula: C₁₂H₁₅NO₈).



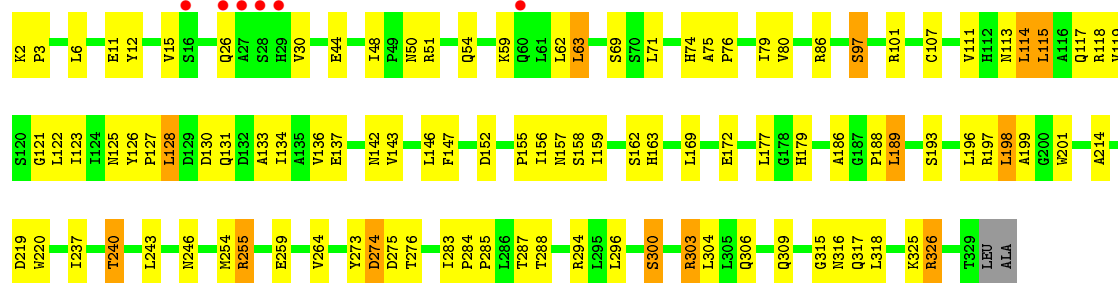
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			21	12	1	8		
3	B	1	Total	C	N	O	0	0
			21	12	1	8		
3	C	1	Total	C	N	O	0	0
			21	12	1	8		



• Molecule 2: Lactose operon repressor



• Molecule 2: Lactose operon repressor



4 Data and refinement statistics

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α , β , γ	253.08Å 253.08Å 203.25Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.98 – 3.50 19.98 – 3.50	Depositor EDS
% Data completeness (in resolution range)	99.9 (19.98-3.50) 99.9 (19.98-3.50)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	0.18	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.87 (at 3.52Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.233 , 0.287 0.217 , 0.256	Depositor DCC
R_{free} test set	1581 reflections (5.32%)	DCC
Wilson B-factor (Å ²)	63.9	Xtriage
Anisotropy	0.012	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 56.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	0 of 31298 reflections	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	8527	wwPDB-VP
Average B, all atoms (Å ²)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.25% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section:
145

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	D	0.88	0/389	1.84	14/598 (2.3%)
1	E	0.90	0/391	2.13	22/601 (3.7%)
1	F	0.87	0/391	2.00	22/601 (3.7%)
2	A	0.45	0/2502	0.64	0/3401
2	B	0.45	0/2516	0.65	0/3419
2	C	0.45	0/2502	0.62	0/3401
All	All	0.53	0/8691	0.97	58/12021 (0.5%)

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
2	B	1	1

There are no bond length outliers.

The worst 5 of 58 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	15	DC	O4'-C4'-C3'	-12.40	98.56	106.00
1	E	15	DC	O3'-P-O5'	-12.00	81.20	104.00
1	D	7	DT	O3'-P-O5'	-11.78	81.61	104.00
1	E	4	DT	P-O3'-C3'	10.98	132.88	119.70
1	F	4	DT	O3'-P-O5'	-10.35	84.34	104.00

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	44	GLU	CA

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	42	MET	Peptide

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	D	348	0	193	4	0
1	E	349	0	192	7	0
1	F	349	0	192	2	0
2	A	2468	0	2518	70	0
2	B	2482	0	2534	80	0
2	C	2468	0	2518	64	0
3	A	21	0	15	5	0
3	B	21	0	15	4	0
3	C	21	0	15	4	0
All	All	8527	0	8192	212	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 13.

The worst 5 of 212 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)
2:A:126:TYR:CD1	2:A:127:PRO:HD2	2.10	0.86
2:B:63:LEU:HD22	2:B:119:VAL:HG12	1.62	0.81
2:B:44:GLU:O	2:B:45:LEU:HG	1.81	0.80
2:C:126:TYR:CD2	2:C:127:PRO:HD2	2.17	0.79
2:C:71:LEU:HD21	2:C:80:VAL:HG21	1.66	0.77

There are no symmetry-related clashes.

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	
2	A	326/330 (99%)	286 (88%)	37 (11%)	3 (1%)	21	68
2	B	328/330 (99%)	294 (90%)	30 (9%)	4 (1%)	16	61
2	C	326/330 (99%)	285 (87%)	37 (11%)	4 (1%)	16	61
All	All	980/990 (99%)	865 (88%)	104 (11%)	11 (1%)	17	63

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	274	ASP
2	B	274	ASP
2	C	274	ASP
2	A	311	GLN
2	B	29	HIS

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	
2	A	268/269 (100%)	230 (86%)	38 (14%)	4	24
2	B	269/269 (100%)	229 (85%)	40 (15%)	4	22
2	C	268/269 (100%)	237 (88%)	31 (12%)	7	33
All	All	805/807 (100%)	696 (86%)	109 (14%)	5	26

5 of 109 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	117	GLN
2	B	177	LEU
2	C	198	LEU
2	B	125	ASN
2	B	141	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	125	ASN
2	B	227	GLN
2	C	306	GLN
2	B	153	GLN
2	B	157	ASN

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](#)

3 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
3	145	A	901	-	20,22,22	2.21	1 (5%)	26,31,31	2.11	4 (15%)
3	145	B	902	-	20,22,22	2.28	2 (10%)	26,31,31	2.40	9 (34%)
3	145	C	903	-	20,22,22	2.27	1 (5%)	26,31,31	2.19	3 (11%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	145	A	901	-	-	0/9/30/30	0/2/2/2
3	145	B	902	-	-	0/9/30/30	0/2/2/2
3	145	C	903	-	-	0/9/30/30	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	902	145	O1'-C1	2.75	1.45	1.41
3	A	901	145	O3'-N1'	9.60	1.41	1.22
3	B	902	145	O3'-N1'	9.63	1.41	1.22
3	C	903	145	O3'-N1'	9.84	1.42	1.22

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	903	145	O1'-C1'-C6'	-9.77	97.99	123.93
3	A	901	145	O1'-C1'-C6'	-7.37	104.37	123.93
3	B	902	145	O1'-C1'-C6'	-5.93	108.17	123.93
3	B	902	145	O5-C1-O1'	-4.62	96.05	108.39
3	A	901	145	O5-C1-O1'	-4.21	97.14	108.39

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 13 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	901	145	5	0
3	B	902	145	4	0
3	C	903	145	4	0

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	D	17/20 (85%)	0.24	0 100 100	61, 98, 117, 119	0
1	E	17/20 (85%)	0.22	0 100 100	57, 90, 133, 135	0
1	F	17/20 (85%)	1.76	8 (47%) 0 0	119, 165, 205, 209	0
2	A	328/330 (99%)	-0.40	6 (1%) 71 62	23, 40, 113, 126	0
2	B	330/330 (100%)	-0.53	2 (0%) 90 85	14, 30, 81, 97	0
2	C	328/330 (99%)	-0.39	6 (1%) 71 62	15, 30, 150, 155	0
All	All	1037/1050 (98%)	-0.38	22 (2%) 67 58	14, 35, 136, 209	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	2	DA	4.5
2	A	27	ALA	4.0
2	A	28	SER	3.5
1	F	17	DC	3.4
2	C	29	HIS	3.3

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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å ²)	Q<0.9
3	145	C	903	21/21	0.95	0.16	-0.15	10,15,23,23	0
3	145	B	902	21/21	0.96	0.15	-0.37	6,10,15,15	0
3	145	A	901	21/21	0.97	0.14	-0.60	15,17,25,25	0

6.5 Other polymers [i](#)

There are no such residues in this entry.