



# Full wwPDB X-ray Structure Validation Report

Nov 12, 2014 – 10:47 AM EST

PDB ID : 4RKR  
Title : Crystal structure of LacI family transcriptional regulator from *Arthrobacter* sp. FB24, target EFI-560007, complex with lactose  
Authors : Patskovsky, Y.; Toro, R.; Bhosle, R.; Al Obaidi, N.; Chamala, S.; Attonito, J.D.; Scott Glenn, A.; Chowdhury, S.; Lafleur, J.; Siedel, R.D.; Hillerich, B.; Love, J.; Whalen, K.L.; Gerlt, J.A.; Burley, S.K.; Almo, S.C.; Enzyme Function Initiative  
Deposited on : 2014-10-13  
Resolution : 2.20 Å(reported)

This is a full wwPDB validation report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

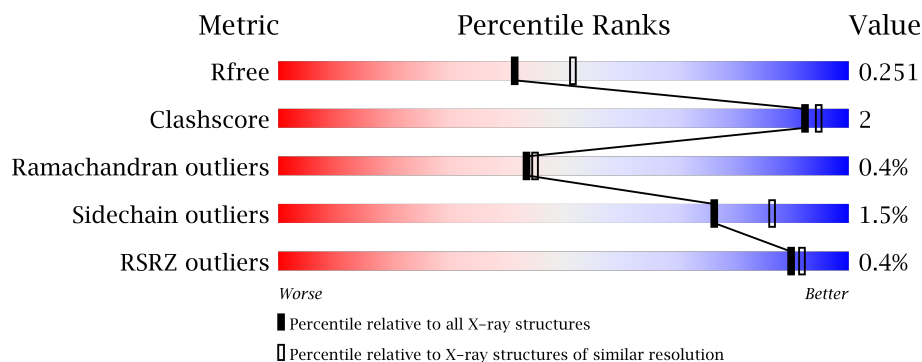
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.16 November 2013  
Xtriage (Phenix) : dev-1439  
EDS : stable24103  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.1.3  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable24103

# 1 Overall quality at a glance

The reported resolution of this entry is 2.20 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))
$R_{free}$	66092	2938 (2.20-2.20)
Clashscore	79885	3751 (2.20-2.20)
Ramachandran outliers	78287	3681 (2.20-2.20)
Sidechain outliers	78261	3682 (2.20-2.20)
RSRZ outliers	66119	2939 (2.20-2.20)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	294	
1	B	294	
1	C	294	
1	D	294	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 8508 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Transcriptional regulator, LacI family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	274	Total	C	N	O	S	0	0	0
			2009	1258	360	385	6			
1	B	267	Total	C	N	O	S	0	0	0
			1963	1229	351	377	6			
1	C	268	Total	C	N	O	S	0	1	0
			1980	1239	353	381	7			
1	D	273	Total	C	N	O	S	0	1	0
			2018	1261	358	393	6			

There are 44 discrepancies between the modelled and reference sequences:

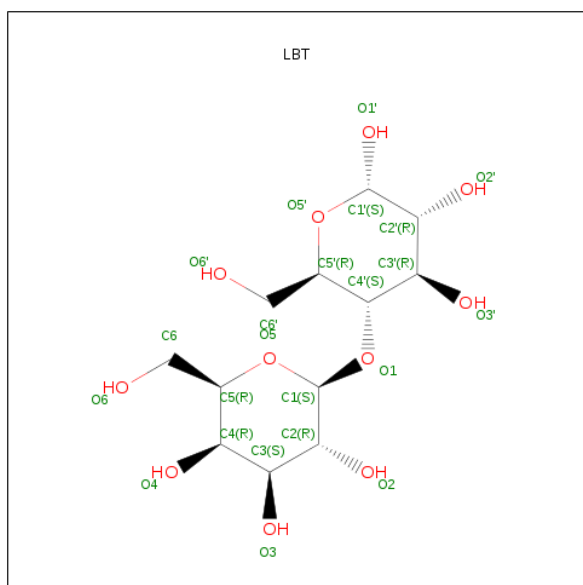
Chain	Residue	Modelled	Actual	Comment	Reference
A	56	MET	-	EXPRESSION TAG	UNP A0K1X3
A	57	SER	-	EXPRESSION TAG	UNP A0K1X3
A	58	LEU	-	EXPRESSION TAG	UNP A0K1X3
A	342	GLU	-	EXPRESSION TAG	UNP A0K1X3
A	343	GLY	-	EXPRESSION TAG	UNP A0K1X3
A	344	HIS	-	EXPRESSION TAG	UNP A0K1X3
A	345	HIS	-	EXPRESSION TAG	UNP A0K1X3
A	346	HIS	-	EXPRESSION TAG	UNP A0K1X3
A	347	HIS	-	EXPRESSION TAG	UNP A0K1X3
A	348	HIS	-	EXPRESSION TAG	UNP A0K1X3
A	349	HIS	-	EXPRESSION TAG	UNP A0K1X3
B	56	MET	-	EXPRESSION TAG	UNP A0K1X3
B	57	SER	-	EXPRESSION TAG	UNP A0K1X3
B	58	LEU	-	EXPRESSION TAG	UNP A0K1X3
B	342	GLU	-	EXPRESSION TAG	UNP A0K1X3
B	343	GLY	-	EXPRESSION TAG	UNP A0K1X3
B	344	HIS	-	EXPRESSION TAG	UNP A0K1X3
B	345	HIS	-	EXPRESSION TAG	UNP A0K1X3
B	346	HIS	-	EXPRESSION TAG	UNP A0K1X3
B	347	HIS	-	EXPRESSION TAG	UNP A0K1X3
B	348	HIS	-	EXPRESSION TAG	UNP A0K1X3

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	349	HIS	-	EXPRESSION TAG	UNP A0K1X3
C	56	MET	-	EXPRESSION TAG	UNP A0K1X3
C	57	SER	-	EXPRESSION TAG	UNP A0K1X3
C	58	LEU	-	EXPRESSION TAG	UNP A0K1X3
C	342	GLU	-	EXPRESSION TAG	UNP A0K1X3
C	343	GLY	-	EXPRESSION TAG	UNP A0K1X3
C	344	HIS	-	EXPRESSION TAG	UNP A0K1X3
C	345	HIS	-	EXPRESSION TAG	UNP A0K1X3
C	346	HIS	-	EXPRESSION TAG	UNP A0K1X3
C	347	HIS	-	EXPRESSION TAG	UNP A0K1X3
C	348	HIS	-	EXPRESSION TAG	UNP A0K1X3
C	349	HIS	-	EXPRESSION TAG	UNP A0K1X3
D	56	MET	-	EXPRESSION TAG	UNP A0K1X3
D	57	SER	-	EXPRESSION TAG	UNP A0K1X3
D	58	LEU	-	EXPRESSION TAG	UNP A0K1X3
D	342	GLU	-	EXPRESSION TAG	UNP A0K1X3
D	343	GLY	-	EXPRESSION TAG	UNP A0K1X3
D	344	HIS	-	EXPRESSION TAG	UNP A0K1X3
D	345	HIS	-	EXPRESSION TAG	UNP A0K1X3
D	346	HIS	-	EXPRESSION TAG	UNP A0K1X3
D	347	HIS	-	EXPRESSION TAG	UNP A0K1X3
D	348	HIS	-	EXPRESSION TAG	UNP A0K1X3
D	349	HIS	-	EXPRESSION TAG	UNP A0K1X3

- Molecule 2 is SUGAR (ALPHA-LACTOSE) (three-letter code: LBT) (formula:  $C_{12}H_{22}O_{11}$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			23	12	11		
2	B	1	Total	C	O	0	0
			23	12	11		
2	C	1	Total	C	O	0	0
			23	12	11		
2	D	1	Total	C	O	0	0
			23	12	11		

- Molecule 3 is water.

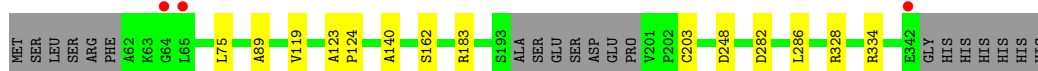
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	116	Total	O	0	0
			116	116		
3	B	103	Total	O	0	0
			103	103		
3	C	123	Total	O	0	0
			123	123		
3	D	104	Total	O	0	0
			104	104		

### 3 Residue-property plots

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.

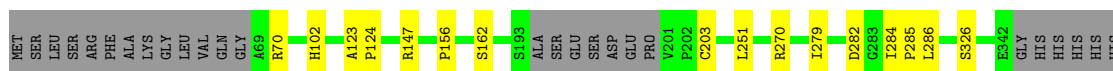
- Molecule 1: Transcriptional regulator, LacI family

Chain A: 



- Molecule 1: Transcriptional regulator, LacI family

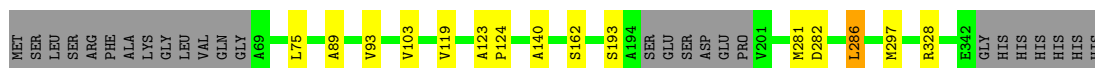
Chain B: 



HIS

- Molecule 1: Transcriptional regulator, LacI family

Chain C: 



- Molecule 1: Transcriptional regulator, LacI family

Chain D: 



HIS  
HIS  
HIS

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	36.94Å 124.91Å 117.04Å 90.00° 96.51° 90.00°	Depositor
Resolution (Å)	50.00 – 2.20 42.56 – 2.20	Depositor EDS
% Data completeness (in resolution range)	98.9 (50.00-2.20) 98.9 (42.56-2.20)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	0.14	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.71 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, $R_{free}$	0.171 , 0.252 0.179 , 0.251	Depositor DCC
$R_{free}$ test set	1588 reflections (3.10%)	DCC
Wilson B-factor (Å <sup>2</sup> )	33.6	Xtriage
Anisotropy	0.360	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 37.5	EDS
Estimated twinning fraction	0.032 for h,-k,-h-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	1 of 52882 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	8508	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.41% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality

### 5.1 Standard geometry

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

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	0.40	0/2041	0.64	0/2773
1	B	0.41	1/1995 (0.1%)	0.66	1/2712 (0.0%)
1	C	0.42	0/2012	0.64	0/2734
1	D	0.40	0/2055	0.64	0/2794
All	All	0.41	1/8103 (0.0%)	0.64	1/11013 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	285	PRO	N-CD	5.33	1.55	1.47

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	284	ILE	C-N-CD	5.21	139.35	128.40

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.



Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2009	0	2015	7	0
1	B	1963	0	1963	8	0
1	C	1980	0	1980	8	0
1	D	2018	0	2014	9	0
2	A	23	0	22	0	0
2	B	23	0	22	0	0
2	C	23	0	22	0	0
2	D	23	0	22	0	0
3	A	116	0	0	0	0
3	B	103	0	0	1	0
3	C	123	0	0	0	0
3	D	104	0	0	0	0
All	All	8508	0	8060	29	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 2.

All (29) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:C:286:LEU:HD12	1:D:286:LEU:HD13	1.63	0.81
1:A:286:LEU:HD13	1:B:286:LEU:CD1	2.28	0.62
1:D:341:THR:O	1:D:342:GLU:C	2.39	0.58
1:B:156:PRO:HB3	1:B:162:SER:HB2	1.86	0.58
1:A:286:LEU:HD13	1:B:286:LEU:HD13	1.88	0.56
1:C:123:ALA:HB3	1:C:124:PRO:HD3	1.91	0.51
1:D:123:ALA:HB3	1:D:124:PRO:HD3	1.91	0.51
1:D:122:LEU:O	1:D:126:VAL:HG22	2.11	0.50
1:D:186:ILE:O	1:D:221:TRP:HA	2.12	0.49
1:C:281[B]:MET:HG2	1:C:297:MET:HB2	1.93	0.49
1:B:270:ARG:NH2	3:B:537:HOH:O	2.46	0.49
1:C:162:SER:HB3	1:C:328:ARG:HG3	1.96	0.48
1:D:139:GLN:O	1:D:142:THR:HB	2.15	0.47
1:A:123:ALA:HB3	1:A:124:PRO:HD3	1.97	0.46
1:B:70:ARG:HD3	1:B:102:HIS:CE1	2.50	0.46
1:B:251:LEU:HD23	1:B:279:ILE:HB	1.98	0.46
1:D:183:ARG:HD2	1:D:248:ASP:O	2.17	0.45
1:C:75:LEU:HD21	1:C:89:ALA:HA	2.00	0.43
1:A:162:SER:HB3	1:A:328:ARG:HG3	2.01	0.43
1:C:93:VAL:HG22	1:C:103:VAL:HG11	2.01	0.43
1:A:119:VAL:HG21	1:A:140:ALA:HB1	2.00	0.43
1:C:119:VAL:HG21	1:C:140:ALA:HB1	2.01	0.42
1:C:281[B]:MET:HE3	1:C:281[B]:MET:HB3	1.74	0.42
1:B:123:ALA:HB3	1:B:124:PRO:HD3	2.02	0.42

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:75:LEU:HD21	1:A:89:ALA:HA	2.00	0.41
1:B:123:ALA:O	1:B:147:ARG:NH1	2.54	0.41
1:D:123:ALA:O	1:D:147:ARG:NH1	2.54	0.41
1:A:183:ARG:HD2	1:A:248:ASP:O	2.22	0.41
1:D:281:MET:O	1:D:282:ASP:CB	2.69	0.40

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	
1	A	270/294 (92%)	266 (98%)	3 (1%)	1 (0%)	43	45
1	B	263/294 (90%)	259 (98%)	3 (1%)	1 (0%)	43	45
1	C	265/294 (90%)	260 (98%)	4 (2%)	1 (0%)	43	45
1	D	272/294 (92%)	267 (98%)	4 (2%)	1 (0%)	43	45
All	All	1070/1176 (91%)	1052 (98%)	14 (1%)	4 (0%)	43	45

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	282	ASP
1	B	282	ASP
1	C	282	ASP
1	D	282	ASP

### 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	204/223 (92%)	202 (99%)	2 (1%)	85	93
1	B	200/223 (90%)	198 (99%)	2 (1%)	85	93
1	C	202/223 (91%)	200 (99%)	2 (1%)	85	93
1	D	208/223 (93%)	202 (97%)	6 (3%)	55	66
All	All	814/892 (91%)	802 (98%)	12 (2%)	76	86

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	203	CYS
1	A	334	ARG
1	B	203	CYS
1	B	326	SER
1	C	193	SER
1	C	286	LEU
1	D	95	LEU
1	D	195	SER
1	D	198	ASP
1	D	222	SER
1	D	243	GLU
1	D	328	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	293	GLN
1	B	164	ASN
1	B	333	HIS
1	C	125	GLN
1	D	125	GLN
1	D	293	GLN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

## 5.6 Ligand geometry

4 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	$\# Z  > 2$
2	LBT	A	401	-	24,24,24	0.34	0	35,35,35	0.84	0
2	LBT	B	401	-	24,24,24	0.38	0	35,35,35	0.85	0
2	LBT	C	401	-	24,24,24	0.42	0	35,35,35	0.77	0
2	LBT	D	401	-	24,24,24	0.41	0	35,35,35	0.91	1 (2%)

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
2	LBT	A	401	-	-	0/8/48/48	0/2/2/2
2	LBT	B	401	-	-	0/8/48/48	0/2/2/2
2	LBT	C	401	-	-	0/8/48/48	0/2/2/2
2	LBT	D	401	-	-	0/8/48/48	0/2/2/2

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	401	LBT	C1-O1-C4'	-2.22	112.36	118.00

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	274/294 (93%)	-0.35	3 (1%) 77 78	25, 43, 78, 113	0
1	B	267/294 (90%)	-0.42	0 100 100	25, 43, 71, 99	0
1	C	268/294 (91%)	-0.42	0 100 100	26, 41, 73, 89	0
1	D	273/294 (92%)	-0.46	1 (0%) 90 92	26, 42, 76, 127	0
All	All	1082/1176 (92%)	-0.41	4 (0%) 90 92	25, 42, 76, 127	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	64	GLY	4.3
1	A	65	LEU	3.3
1	D	200	PRO	2.6
1	A	342	GLU	2.3

### 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 carbohydrates in this entry.

### 6.4 Ligands

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, 95<sup>th</sup> 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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	LBT	D	401	23/23	0.11	0.81	27,35,50,66	0
2	LBT	B	401	23/23	0.10	-0.21	23,38,52,63	0
2	LBT	A	401	23/23	0.09	-0.67	23,38,62,84	0
2	LBT	C	401	23/23	0.09	-0.92	27,41,58,82	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.