



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 09:52 AM GMT

PDB ID : 3JZI  
Title : Crystal structure of biotin carboxylase from E. Coli in complex with benzimidazole series  
Authors : Cheng, C.; Shipps, G.W.; Yang, Z.; Sun, B.; Kawahata, N.; Soucy, K.; Soriano, A.; Orth, P.; Xiao, L.; Mann, P.; Black, T.  
Deposited on : 2009-09-23  
Resolution : 2.31 Å(reported)

This is a Full wwPDB X-ray Structure 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/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

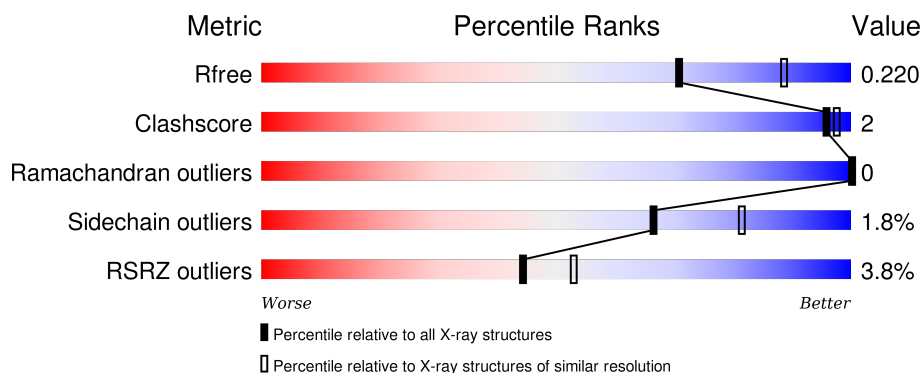
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*



The reported resolution of this entry is 2.31 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	4425 (2.34-2.30)
Clashscore	102246	5057 (2.34-2.30)
Ramachandran outliers	100387	5008 (2.34-2.30)
Sidechain outliers	100360	5007 (2.34-2.30)
RSRZ outliers	91569	4432 (2.34-2.30)

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. 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	A	486	 7% 88% • 8%
1	B	486	 88% • 8%

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7280 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 Biotin carboxylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	445	Total	C	N	O	S	0	0	0
			3376	2126	597	632	21			
1	B	446	Total	C	N	O	S	0	0	0
			3442	2167	618	635	22			

There are 74 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-20	MET	-	EXPRESSION TAG	UNP P24182
A	-19	GLY	-	EXPRESSION TAG	UNP P24182
A	-18	HIS	-	EXPRESSION TAG	UNP P24182
A	-17	HIS	-	EXPRESSION TAG	UNP P24182
A	-16	HIS	-	EXPRESSION TAG	UNP P24182
A	-15	HIS	-	EXPRESSION TAG	UNP P24182
A	-14	HIS	-	EXPRESSION TAG	UNP P24182
A	-13	HIS	-	EXPRESSION TAG	UNP P24182
A	-12	HIS	-	EXPRESSION TAG	UNP P24182
A	-11	HIS	-	EXPRESSION TAG	UNP P24182
A	-10	HIS	-	EXPRESSION TAG	UNP P24182
A	-9	HIS	-	EXPRESSION TAG	UNP P24182
A	-8	SER	-	EXPRESSION TAG	UNP P24182
A	-7	SER	-	EXPRESSION TAG	UNP P24182
A	-6	GLY	-	EXPRESSION TAG	UNP P24182
A	-5	HIS	-	EXPRESSION TAG	UNP P24182
A	-4	ILE	-	EXPRESSION TAG	UNP P24182
A	-3	GLU	-	EXPRESSION TAG	UNP P24182
A	-2	GLY	-	EXPRESSION TAG	UNP P24182
A	-1	ARG	-	EXPRESSION TAG	UNP P24182
A	0	HIS	-	EXPRESSION TAG	UNP P24182
A	450	SER	-	EXPRESSION TAG	UNP P24182
A	451	ASP	-	EXPRESSION TAG	UNP P24182
A	452	PRO	-	EXPRESSION TAG	UNP P24182
A	453	ASN	-	EXPRESSION TAG	UNP P24182

*Continued on next page...*

*Continued from previous page...*

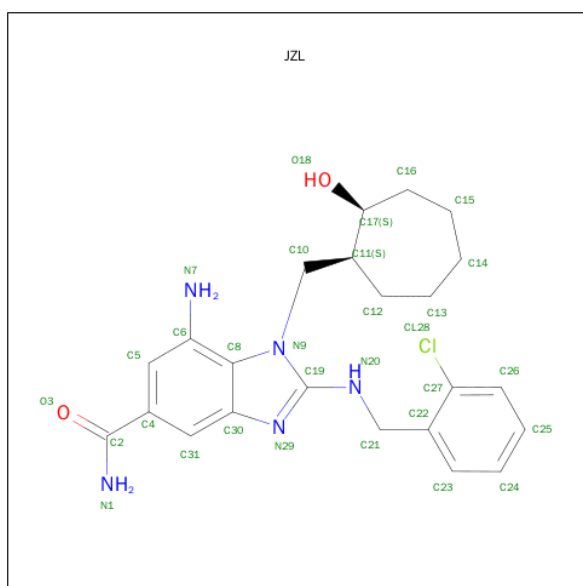
Chain	Residue	Modelled	Actual	Comment	Reference
A	454	SER	-	EXPRESSION TAG	UNP P24182
A	455	SER	-	EXPRESSION TAG	UNP P24182
A	456	SER	-	EXPRESSION TAG	UNP P24182
A	457	VAL	-	EXPRESSION TAG	UNP P24182
A	458	ASP	-	EXPRESSION TAG	UNP P24182
A	459	LYS	-	EXPRESSION TAG	UNP P24182
A	460	LEU	-	EXPRESSION TAG	UNP P24182
A	461	ALA	-	EXPRESSION TAG	UNP P24182
A	462	ALA	-	EXPRESSION TAG	UNP P24182
A	463	ALA	-	EXPRESSION TAG	UNP P24182
A	464	LEU	-	EXPRESSION TAG	UNP P24182
A	465	GLU	-	EXPRESSION TAG	UNP P24182
B	-20	MET	-	EXPRESSION TAG	UNP P24182
B	-19	GLY	-	EXPRESSION TAG	UNP P24182
B	-18	HIS	-	EXPRESSION TAG	UNP P24182
B	-17	HIS	-	EXPRESSION TAG	UNP P24182
B	-16	HIS	-	EXPRESSION TAG	UNP P24182
B	-15	HIS	-	EXPRESSION TAG	UNP P24182
B	-14	HIS	-	EXPRESSION TAG	UNP P24182
B	-13	HIS	-	EXPRESSION TAG	UNP P24182
B	-12	HIS	-	EXPRESSION TAG	UNP P24182
B	-11	HIS	-	EXPRESSION TAG	UNP P24182
B	-10	HIS	-	EXPRESSION TAG	UNP P24182
B	-9	HIS	-	EXPRESSION TAG	UNP P24182
B	-8	SER	-	EXPRESSION TAG	UNP P24182
B	-7	SER	-	EXPRESSION TAG	UNP P24182
B	-6	GLY	-	EXPRESSION TAG	UNP P24182
B	-5	HIS	-	EXPRESSION TAG	UNP P24182
B	-4	ILE	-	EXPRESSION TAG	UNP P24182
B	-3	GLU	-	EXPRESSION TAG	UNP P24182
B	-2	GLY	-	EXPRESSION TAG	UNP P24182
B	-1	ARG	-	EXPRESSION TAG	UNP P24182
B	0	HIS	-	EXPRESSION TAG	UNP P24182
B	450	SER	-	EXPRESSION TAG	UNP P24182
B	451	ASP	-	EXPRESSION TAG	UNP P24182
B	452	PRO	-	EXPRESSION TAG	UNP P24182
B	453	ASN	-	EXPRESSION TAG	UNP P24182
B	454	SER	-	EXPRESSION TAG	UNP P24182
B	455	SER	-	EXPRESSION TAG	UNP P24182
B	456	SER	-	EXPRESSION TAG	UNP P24182
B	457	VAL	-	EXPRESSION TAG	UNP P24182
B	458	ASP	-	EXPRESSION TAG	UNP P24182

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	459	LYS	-	EXPRESSION TAG	UNP P24182
B	460	LEU	-	EXPRESSION TAG	UNP P24182
B	461	ALA	-	EXPRESSION TAG	UNP P24182
B	462	ALA	-	EXPRESSION TAG	UNP P24182
B	463	ALA	-	EXPRESSION TAG	UNP P24182
B	464	LEU	-	EXPRESSION TAG	UNP P24182
B	465	GLU	-	EXPRESSION TAG	UNP P24182

- Molecule 2 is 7-AMINO-2-[(2-CHLOROBENZYL)AMINO]-1-[(1S,2S)-2-HYDROXYCYCLOHEPTYL]METHYL}-1H-BENZIMIDAZOLE-5-CARBOXAMIDE (three-letter code: JZL) (formula:  $C_{23}H_{28}ClN_5O_2$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	Cl	N	O	0	0
			31	23	1	5	2		
2	B	1	Total	C	Cl	N	O	0	0
			31	23	1	5	2		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	176	Total	O	0	0
			176	176		
3	B	224	Total	O	0	0
			224	224		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	84.76Å 107.82Å 121.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.97 – 2.31 19.97 – 2.31	Depositor EDS
% Data completeness (in resolution range)	(Not available) (19.97-2.31) 96.4 (19.97-2.31)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.62 (at 2.30Å)	Xtriage
Refinement program	BUSTER 2.9.2	Depositor
R, $R_{free}$	0.187 , 0.223 0.183 , 0.220	Depositor DCC
$R_{free}$ test set	2386 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	33.8	Xtriage
Anisotropy	0.268	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 37.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	1 of 47545 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7280	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.65% of the height of the origin peak. No significant pseudotranslation is detected.*

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

<sup>2</sup>Theoretical 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: JZL

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	0.42	0/3439	0.61	0/4650
1	B	0.43	0/3506	0.61	0/4731
All	All	0.43	0/6945	0.61	0/9381

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	3376	0	3343	10	0
1	B	3442	0	3465	9	0
2	A	31	0	27	7	0
2	B	31	0	28	5	0
3	A	176	0	0	0	0
3	B	224	0	0	0	0
All	All	7280	0	6863	21	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 2.

All (21) 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:A:209:HIS:CD2	2:A:466:JZL:H16A	2.40	0.57
1:B:276:GLU:OE2	1:B:290:ASN:ND2	2.37	0.57
1:A:236:HIS:HB3	2:A:466:JZL:H12	1.92	0.51
1:B:236:HIS:CD2	2:B:466:JZL:H12	2.46	0.51
1:A:233:GLN:OE1	2:A:466:JZL:H14A	2.12	0.49
1:B:5:ILE:HG22	1:B:77:ALA:HB3	1.94	0.49
1:A:209:HIS:HD2	2:A:466:JZL:H16A	1.77	0.48
1:A:298:PRO:O	1:A:302:MET:HG2	2.15	0.46
1:A:165:GLY:O	2:A:466:JZL:H10A	2.16	0.46
1:B:236:HIS:HD2	2:B:466:JZL:H12	1.80	0.45
1:A:157:ILE:HD11	1:A:169:MET:HB3	1.99	0.44
1:B:223:ILE:HG12	1:B:321:LEU:HD13	1.99	0.44
1:A:138:LEU:HD11	1:A:200:MET:HB2	2.00	0.43
1:B:298:PRO:O	1:B:302:MET:HG2	2.19	0.43
1:B:131:VAL:HG22	1:B:285:TYR:HB3	2.00	0.43
1:B:169:MET:CE	2:B:466:JZL:H10	2.49	0.43
2:A:466:JZL:HN7A	2:A:466:JZL:H10	1.84	0.42
2:B:466:JZL:H11	2:B:466:JZL:HN7A	1.85	0.41
1:B:165:GLY:O	2:B:466:JZL:H10A	2.21	0.41
1:A:209:HIS:HD2	2:A:466:JZL:C16	2.33	0.41
1:A:5:ILE:HG22	1:A:77:ALA:HB3	2.02	0.41

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	443/486 (91%)	432 (98%)	11 (2%)	0	100	100
1	B	444/486 (91%)	434 (98%)	10 (2%)	0	100	100
All	All	887/972 (91%)	866 (98%)	21 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	346/392 (88%)	340 (98%)	6 (2%)	68	83
1	B	359/392 (92%)	352 (98%)	7 (2%)	65	80
All	All	705/784 (90%)	692 (98%)	13 (2%)	66	81

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	5	ILE
1	A	72	ILE
1	A	140	ASP
1	A	326	GLU
1	A	382	ASP
1	A	442	LYS
1	B	71	GLU
1	B	111	ARG
1	B	206	ASN
1	B	281	ASN
1	B	292	ARG
1	B	323	ILE
1	B	382	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	209	HIS
1	A	237	GLN
1	A	290	ASN
1	A	404	ASN
1	B	180	GLN
1	B	236	HIS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	340	ASN
1	B	404	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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

2 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	JZL	A	466	-	29,34,34	1.47	4 (13%)	30,48,48	1.40	5 (16%)
2	JZL	B	466	-	29,34,34	1.28	3 (10%)	30,48,48	1.64	5 (16%)

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	JZL	A	466	-	-	0/7/25/25	0/4/4/4
2	JZL	B	466	-	-	0/7/25/25	0/4/4/4

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	466	JZL	C10-N9	-4.33	1.43	1.48
2	B	466	JZL	C10-N9	-3.95	1.44	1.48
2	B	466	JZL	C23-C22	2.11	1.43	1.39
2	A	466	JZL	C19-N20	2.17	1.36	1.33
2	A	466	JZL	C23-C22	2.51	1.44	1.39
2	B	466	JZL	C31-C4	2.60	1.41	1.37
2	A	466	JZL	C31-C4	2.75	1.42	1.37

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	466	JZL	C4-C2-N1	-3.33	114.17	117.82
2	B	466	JZL	C13-C12-C11	-3.10	110.50	116.90
2	A	466	JZL	C13-C12-C11	-2.99	110.73	116.90
2	A	466	JZL	C4-C31-C30	-2.12	118.60	120.80
2	B	466	JZL	C4-C31-C30	-2.08	118.65	120.80
2	A	466	JZL	C5-C6-C8	-2.05	115.14	117.09
2	A	466	JZL	C21-C22-C27	-2.01	118.26	121.19
2	B	466	JZL	O18-C17-C16	2.26	114.02	108.93
2	A	466	JZL	C8-C6-N7	4.03	126.60	121.14
2	B	466	JZL	C8-C6-N7	4.84	127.69	121.14

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 12 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	466	JZL	7	0
2	B	466	JZL	5	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 ⓘ

### 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	445/486 (91%)	-0.03	32 (7%) 18 26	22, 35, 92, 105	0
1	B	446/486 (91%)	-0.42	2 (0%) 93 95	21, 33, 53, 81	0
All	All	891/972 (91%)	-0.22	34 (3%) 44 53	21, 33, 81, 105	0

All (34) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	187	ALA	5.6
1	A	152	ILE	5.1
1	A	140	ASP	4.7
1	A	193	PHE	4.3
1	A	157	ILE	4.2
1	B	-1	ARG	4.0
1	A	199	TYR	3.8
1	A	158	ILE	3.8
1	A	176	ALA	3.8
1	A	177	GLU	3.7
1	A	180	GLN	3.7
1	A	138	LEU	3.5
1	A	184	MET	3.5
1	A	194	SER	3.5
1	A	192	ALA	3.4
1	A	198	VAL	3.1
1	A	182	ILE	3.0
1	A	172	VAL	2.8
1	A	200	MET	2.8
1	A	191	ALA	2.6
1	A	174	GLY	2.5
1	A	347	PHE	2.4
1	A	141	ASP	2.4
1	A	154	TYR	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	322	SER	2.4
1	B	345	ASN	2.3
1	A	183	SER	2.3
1	A	167	ARG	2.2
1	A	196	ASP	2.1
1	A	179	ALA	2.1
1	A	139	GLY	2.1
1	A	181	SER	2.1
1	A	148	ILE	2.1
1	A	445	GLY	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](#)

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	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	JZL	A	466	31/31	0.93	0.12	-0.88	50,52,61,63	0
2	JZL	B	466	31/31	0.96	0.09	-0.88	25,33,40,41	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.