

# Full wwPDB X-ray Structure Validation Report (i)

#### Apr 3, 2023 – 07:06 PM JST

PDB ID	:	8I2L
Title	:	E. coli tryptophanyl-tRNA synthetase bound with a chemical fragment at the
		dimerization interface
Authors	:	Xiang, M.; Zhou, H.
Deposited on		
Resolution	:	1.95  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 A user guide is available at https://www.wwpdb.org/validation/2017/XrayValidationReportHelp with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

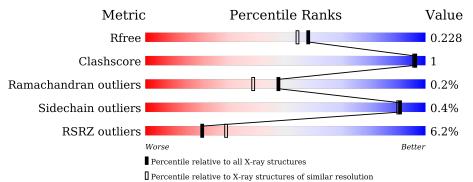
Xtriage (Phenix) EDS buster-report Percentile statistics Refmac CCP4 Ideal geometry (proteins) Ideal geometry (DNA, RNA)	: : : : :	20191225.v01 (using entries in the PDB archive December 25th 2019) 5.8.0158 7.0.044 (Gargrove) Engh & Huber (2001) Parkinson et al. (1996)
Ideal geometry (DNA, RNA) Validation Pipeline (wwPDB-VP)		Parkinson et al. (1996) 2.32.2

# 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:  $X\text{-}RAY \, DIFFRACTION$ 

The reported resolution of this entry is 1.95 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	$\begin{array}{c} \textbf{Whole archive} \\ (\#\textbf{Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
R <sub>free</sub>	130704	2580(1.96-1.96)
Clashscore	141614	2705(1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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	А	340	8%	•••
1	В	340	<u>4%</u> 94%	•••



# 2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 5454 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.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	Δ	325	Total	С	Ν	Ο	$\mathbf{S}$	0	0	0
	A	323	2486	1577	423	473	13	0	0	0
1	р	333	Total	С	Ν	0	S	0	1	0
	D	ანა	2593	1643	444	493	13	0	1	0

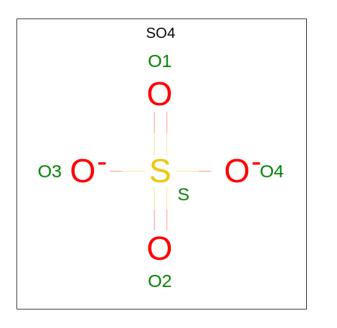
• Molecule 1 is a protein called Tryptophan–tRNA ligase.

Residue	Modelled	Actual	Comment	Reference
334	ARG	-	expression tag	UNP P00954
335	HIS	-	expression tag	UNP P00954
336	HIS	-	expression tag	UNP P00954
337	HIS	-	expression tag	UNP P00954
338	HIS	-	expression tag	UNP P00954
339	HIS	-	expression tag	UNP P00954
340	HIS	-	expression tag	UNP P00954
334	ARG	-	expression tag	UNP P00954
335	HIS	-	expression tag	UNP P00954
336	HIS	-	expression tag	UNP P00954
337	HIS	-	expression tag	UNP P00954
338	HIS	-	expression tag	UNP P00954
339	HIS	-	expression tag	UNP P00954
340	HIS	-	expression tag	UNP P00954
	334         335         336         337         338         339         340         334         335         336         337         338         339         340         334         335         336         337         338         339	334       ARG         335       HIS         336       HIS         337       HIS         338       HIS         339       HIS         340       HIS         334       ARG         335       HIS         334       ARG         335       HIS         336       HIS         337       HIS         336       HIS         337       HIS         338       HIS         339       HIS	334       ARG       -         335       HIS       -         336       HIS       -         337       HIS       -         337       HIS       -         338       HIS       -         339       HIS       -         340       HIS       -         334       ARG       -         335       HIS       -         336       HIS       -         337       HIS       -         336       HIS       -         337       HIS       -         338       HIS       -         337       HIS       -         337       HIS       -         338       HIS       -         339       HIS       -	334ARG-expression tag335HIS-expression tag336HIS-expression tag337HIS-expression tag338HIS-expression tag339HIS-expression tag340HIS-expression tag334ARG-expression tag335HIS-expression tag336HIS-expression tag337HIS-expression tag337HIS-expression tag338HIS-expression tag339HIS-expression tag339HIS-expression tag

There are 14 discrepancies between the modelled and reference sequences:

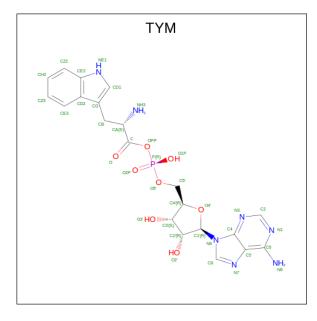
• Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).





Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	А	1	Total 5	0 4	S 1	0	0

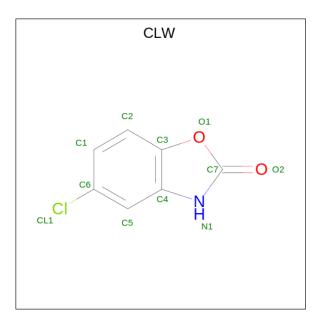
 $\bullet \ \ Molecule \ 3 \ is \ TRYPTOPHANYL-5'AMP \ (three-letter \ code: \ TYM) \ (formula: \ C_{21}H_{24}N_7O_8P).$ 



Mol	Chain	Residues		Ato	oms			ZeroOcc	AltConf
2	р	1	Total	С	Ν	0	Р	0	0
J	D	1	37	21	7	8	1	0	0

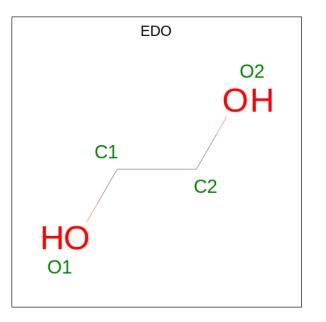
• Molecule 4 is CHLORZOXAZONE (three-letter code: CLW) (formula:  $C_7H_4ClNO_2$ ) (labeled as "Ligand of Interest" by depositor).





Mol	Chain	Residues		At	$\mathbf{oms}$			ZeroOcc	AltConf
4	В	1	Total	С	Cl	Ν	Ο	0	0
4	D	1	11	7	1	1	2	0	0

• Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	В	1	Total 4	${ m C} { m 2}$	O 2	0	0

• Molecule 6 is water.

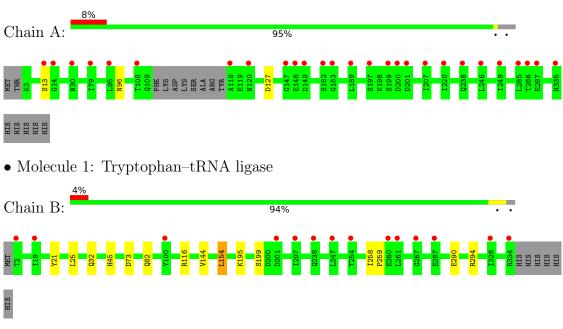


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	А	138	Total O 138 138	0	0
6	В	180	Total O 180 180	0	0



# 3 Residue-property plots (i)

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: Tryptophan–tRNA ligase



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants	61.62Å 80.05Å 78.28Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $105.69^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	50.00 - 1.95	Depositor
Resolution (A)	19.89 - 1.95	EDS
% Data completeness	99.2 (50.00-1.95)	Depositor
(in resolution range)	99.3 (19.89-1.95)	EDS
R <sub>merge</sub>	0.07	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.99 (at 1.94 \text{\AA})$	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
D D.	0.199 , $0.221$	Depositor
$R, R_{free}$	0.205 , $0.228$	DCC
$R_{free}$ test set	2636 reflections $(4.97%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	27.7	Xtriage
Anisotropy	0.856	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.34, 37.6	EDS
L-test for twinning <sup>2</sup>	$ < L >=0.49, < L^2>=0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	5454	wwPDB-VP
Average B, all atoms $(Å^2)$	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 6.38% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 5 Model quality (i)

## 5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: CLW, SO4, EDO, TYM

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		
Moi Chain		RMSZ	# Z  > 5	RMSZ	# Z  > 5	
1	А	0.37	0/2534	0.60	0/3440	
1	В	0.37	0/2644	0.61	0/3584	
All	All	0.37	0/5178	0.61	0/7024	

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 (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	А	2486	0	2410	1	0
1	В	2593	0	2529	9	0
2	А	5	0	0	0	0
3	В	37	0	23	1	0
4	В	11	0	4	1	0
5	В	4	0	6	0	0
6	А	138	0	0	0	0
6	В	180	0	0	0	0
All	All	5454	0	4972	10	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 1.



ő	8I2	$^{2}L$

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:32:GLN:HE22	1:B:73:ASP:H	1.30	0.77
1:B:290:GLU:OE1	1:B:294:ARG:NH2	2.36	0.55
1:B:45:HIS:CD2	1:B:82:GLN:HE22	2.31	0.48
1:B:144:VAL:HG11	1:B:154:LEU:HD21	1.94	0.48
1:A:127:ASP:OD2	4:B:402:CLW:H51	2.16	0.45
1:B:116:ARG:NH2	1:B:199:SER:O	2.35	0.44
1:B:258:ILE:HB	1:B:259:PRO:HD3	2.01	0.42
1:B:21:TYR:HA	1:B:25:LEU:HB2	2.03	0.40
1:B:144:VAL:CG1	1:B:154:LEU:HD21	2.51	0.40
1:B:195:LYS:NZ	3:B:401:TYM:N7	2.69	0.40

All (10) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

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	nalysed Favoured Allowed		Outliers	Perce	ntiles
1	А	321/340~(94%)	313~(98%)	7 (2%)	1 (0%)	41	30
1	В	332/340~(98%)	321~(97%)	11 (3%)	0	100	100
All	All	653/680~(96%)	634 (97%)	18 (3%)	1 (0%)	47	38

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	А	13	SER

#### 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		
1	А	258/290~(89%)	257~(100%)	1 (0%)	91 90		
1	В	271/290 (93%)	270 (100%)	1 (0%)	91 90		
All	All	529/580~(91%)	527 (100%)	2 (0%)	91 90		

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	А	96	ASN
1	В	154	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	А	272	HIS
1	В	32	GLN
1	В	45	HIS
1	В	297	ASN
1	В	316	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 monosaccharides in this entry.

## 5.6 Ligand geometry (i)

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	Turne	Chain Res Link		Bo	Bond lengths			Bond angles		
IVIOI	Type	Chain	nes	LIIIK	Counts	RMSZ	# Z >2	Counts	RMSZ	# Z >2
4	CLW	В	402	-	8,12,12	1.10	1 (12%)	$9,\!17,\!17$	0.94	1 (11%)
3	TYM	В	401	-	36,41,41	1.08	2 (5%)	39,61,61	1.33	6 (15%)
2	SO4	А	401	-	4,4,4	0.32	0	6,6,6	0.11	0
5	EDO	В	403	-	3,3,3	0.51	0	2,2,2	0.27	0

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
5	EDO	В	403	-	-	1/1/1/1	-
3	TYM	В	401	-	-	0/16/39/39	0/5/5/5
4	CLW	В	402	-	-	-	0/2/2/2

Mol	Chain	Res	Type	Atoms	Ζ	Observed(Å)	$\mathrm{Ideal}(\mathrm{\AA})$
3	В	401	TYM	C5-C4	2.29	1.47	1.40
3	В	401	TYM	C2-N3	2.16	1.35	1.32
4	В	402	CLW	C5-C4	-2.09	1.38	1.41

All (3) bond length outliers are listed below:

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Ζ	$Observed(^{o})$	$Ideal(^{o})$
3	В	401	TYM	N3-C2-N1	-3.61	123.03	128.68
3	В	401	TYM	OPP-P-O5'	-2.72	94.98	102.92
3	В	401	TYM	CE3-CD2-CE2	2.33	121.26	118.17
4	В	402	CLW	C6-C5-C4	-2.29	117.56	119.50
3	В	401	TYM	C4-C5-N7	-2.22	107.08	109.40
3	В	401	TYM	CB-CG-CD1	-2.20	125.25	127.97
3	В	401	TYM	O1P-P-OPP	2.10	110.98	104.14

There are no chirality outliers.



All (1) torsion outliers are listed below:

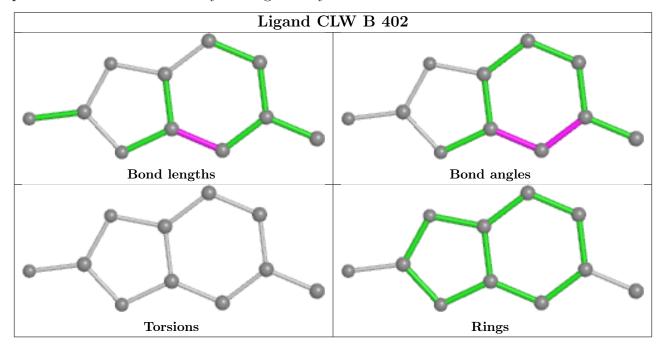
Mol	Chain	Res	Type	Atoms
5	В	403	EDO	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

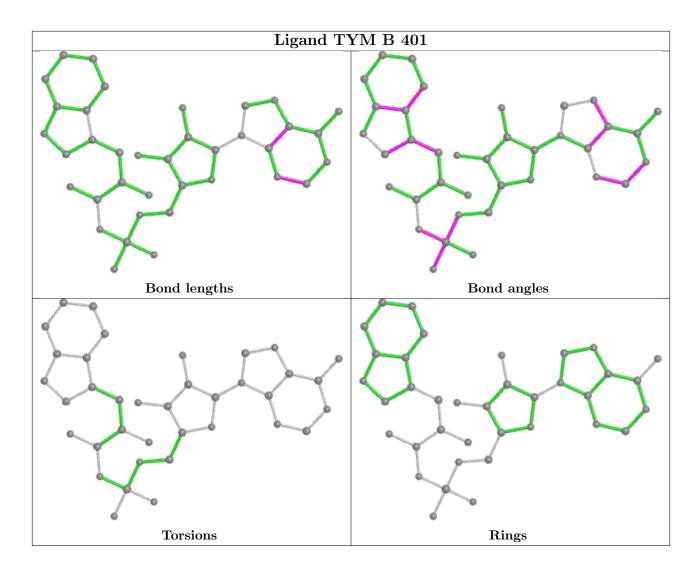
Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	В	402	CLW	1	0
3	В	401	TYM	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.









## 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,  $95^{th}$  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	А	325/340~(95%)	0.51	27 (8%) 11 17	19, 36, 54, 71	0
1	В	333/340~(97%)	0.30	14 (4%) 36 45	21, 32, 49, 57	0
All	All	658/680~(96%)	0.40	41 (6%) 20 29	19, 34, 51, 71	0

All (41) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	108	THR	4.6
1	А	201	ASP	4.5
1	А	246	LEU	4.3
1	А	199	SER	4.2
1	В	254	THR	4.1
1	А	335	HIS	4.0
1	А	183	GLY	3.8
1	А	30	ASN	3.7
1	В	201	ASP	3.6
1	А	118	ALA	3.5
1	А	249	ILE	3.5
1	А	189	LEU	3.4
1	А	120	ASN	3.2
1	А	285	LEU	3.0
1	А	200	ASP	3.0
1	А	14	GLY	2.8
1	А	287	GLU	2.8
1	А	286	THR	2.7
1	А	197	SER	2.7
1	А	148	GLU	2.6
1	В	334	ARG	2.6
1	В	267	GLY	2.6
1	В	261	LEU	2.6
1	A	147	GLY	2.6

Continued on next page...



Mol	Chain	Res	Type	RSRZ
1	В	100	TYR	2.4
1	А	238	GLN	2.4
1	В	328	ILE	2.4
1	А	182	SER	2.3
1	А	207	ILE	2.3
1	В	247	LEU	2.3
1	В	207	ILE	2.3
1	В	287	GLU	2.3
1	А	95	LEU	2.2
1	А	13	SER	2.2
1	В	2	THR	2.2
1	А	220	ILE	2.2
1	А	149	ASP	2.1
1	В	18	ILE	2.1
1	А	79	ILE	2.1
1	В	238	GLN	2.0
1	В	260	GLU	2.0

Continued from previous page...

### 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 monosaccharides in this entry.

#### 6.4 Ligands (i)

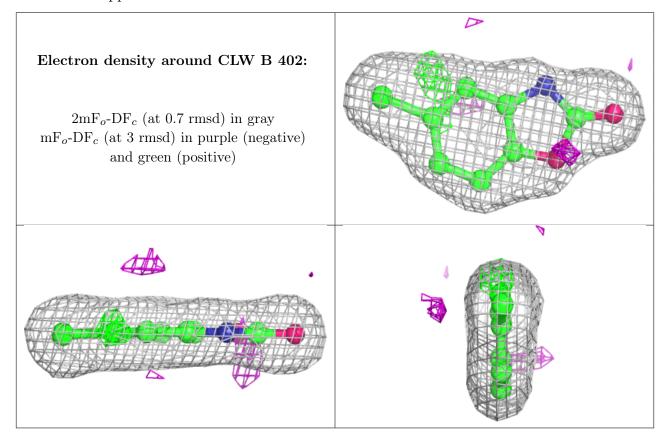
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median,  $95^{th}$  percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-factors}(\mathbf{A}^2)$	Q < 0.9
5	EDO	В	403	4/4	0.83	0.29	42,43,43,43	0
4	CLW	В	402	11/11	0.88	0.13	26,32,33,38	0
2	SO4	А	401	5/5	0.92	0.25	75,76,77,78	0
3	TYM	В	401	37/37	0.94	0.10	22,30,33,34	0

The following is a graphical depiction of the model fit to experimental electron density of all

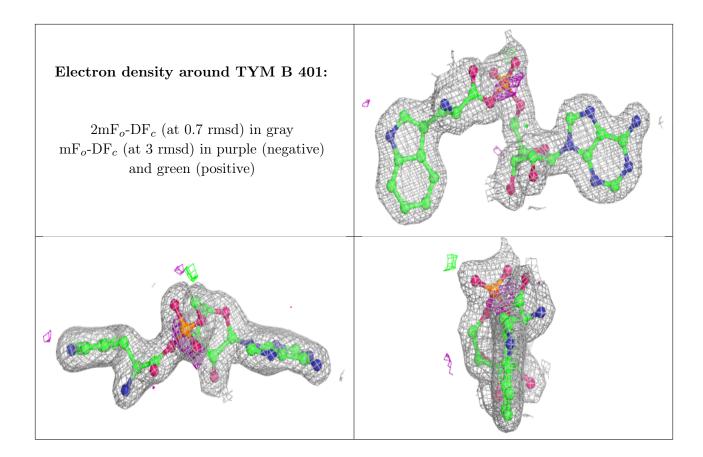


instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.









## 6.5 Other polymers (i)

There are no such residues in this entry.

