



# Full wwPDB X-ray Structure Validation Report ⓘ

May 13, 2020 – 03:39 am BST

PDB ID : 4NV8  
Title : Crystal Structure of Mesorhizobium Loti Arylamine N-acetyltransferase F42W Mutant  
Authors : Xu, X.M.; Haouz, A.; Weber, P.; Li de la sierra-gallay, I.; Kubiak, X.; Dupret, J.-M.; Rodrigues-lima, F.  
Deposited on : 2013-12-05  
Resolution : 1.84 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

<https://www.wwpdb.org/validation/2017/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
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

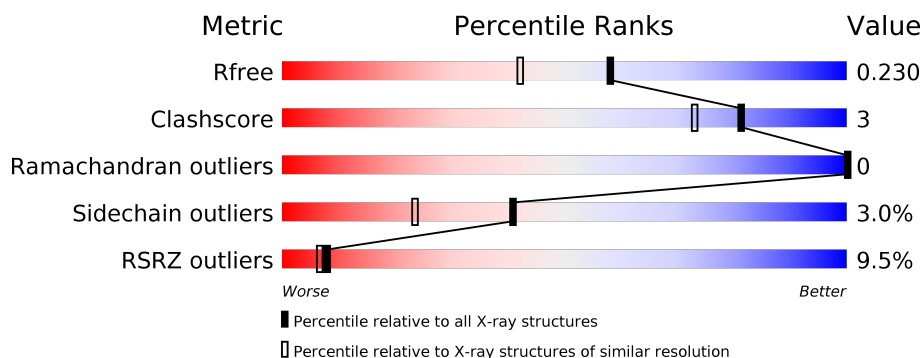
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.84 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))
$R_{free}$	130704	4003 (1.86-1.82)
Clashscore	141614	4233 (1.86-1.82)
Ramachandran outliers	138981	4185 (1.86-1.82)
Sidechain outliers	138945	4186 (1.86-1.82)
RSRZ outliers	127900	3957 (1.86-1.82)

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 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  $\leq 5\%$ . 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	312	<div> <div>12%</div> <div> <div></div> <div>79%</div> <div>7%</div> <div>13%</div> </div> </div>
1	B	312	<div> <div>5%</div> <div> <div></div> <div>78%</div> <div>8%</div> <div>14%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4449 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 Arylamine N-acetyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	270	Total	C	N	O	S	0	1	0
			2128	1359	382	384	3			
1	B	268	Total	C	N	O	S	0	0	0
			2104	1345	376	380	3			

There are 70 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-33	MET	-	INITIATING METHIONINE	UNP Q98D42
A	-32	GLY	-	EXPRESSION TAG	UNP Q98D42
A	-31	SER	-	EXPRESSION TAG	UNP Q98D42
A	-30	SER	-	EXPRESSION TAG	UNP Q98D42
A	-29	HIS	-	EXPRESSION TAG	UNP Q98D42
A	-28	HIS	-	EXPRESSION TAG	UNP Q98D42
A	-27	HIS	-	EXPRESSION TAG	UNP Q98D42
A	-26	HIS	-	EXPRESSION TAG	UNP Q98D42
A	-25	HIS	-	EXPRESSION TAG	UNP Q98D42
A	-24	HIS	-	EXPRESSION TAG	UNP Q98D42
A	-23	SER	-	EXPRESSION TAG	UNP Q98D42
A	-22	SER	-	EXPRESSION TAG	UNP Q98D42
A	-21	GLY	-	EXPRESSION TAG	UNP Q98D42
A	-20	LEU	-	EXPRESSION TAG	UNP Q98D42
A	-19	VAL	-	EXPRESSION TAG	UNP Q98D42
A	-18	PRO	-	EXPRESSION TAG	UNP Q98D42
A	-17	ARG	-	EXPRESSION TAG	UNP Q98D42
A	-16	GLY	-	EXPRESSION TAG	UNP Q98D42
A	-15	SER	-	EXPRESSION TAG	UNP Q98D42
A	-14	HIS	-	EXPRESSION TAG	UNP Q98D42
A	-13	MET	-	EXPRESSION TAG	UNP Q98D42
A	-12	ALA	-	EXPRESSION TAG	UNP Q98D42
A	-11	SER	-	EXPRESSION TAG	UNP Q98D42
A	-10	ASN	-	EXPRESSION TAG	UNP Q98D42
A	-9	THR	-	EXPRESSION TAG	UNP Q98D42

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	-8	GLY	-	EXPRESSION TAG	UNP Q98D42
A	-7	GLY	-	EXPRESSION TAG	UNP Q98D42
A	-6	GLN	-	EXPRESSION TAG	UNP Q98D42
A	-5	GLN	-	EXPRESSION TAG	UNP Q98D42
A	-4	MET	-	EXPRESSION TAG	UNP Q98D42
A	-3	GLY	-	EXPRESSION TAG	UNP Q98D42
A	-2	ARG	-	EXPRESSION TAG	UNP Q98D42
A	-1	GLY	-	EXPRESSION TAG	UNP Q98D42
A	0	SER	-	EXPRESSION TAG	UNP Q98D42
A	42	TRP	PHE	ENGINEERED MUTATION	UNP Q98D42
B	-33	MET	-	INITIATING METHIONINE	UNP Q98D42
B	-32	GLY	-	EXPRESSION TAG	UNP Q98D42
B	-31	SER	-	EXPRESSION TAG	UNP Q98D42
B	-30	SER	-	EXPRESSION TAG	UNP Q98D42
B	-29	HIS	-	EXPRESSION TAG	UNP Q98D42
B	-28	HIS	-	EXPRESSION TAG	UNP Q98D42
B	-27	HIS	-	EXPRESSION TAG	UNP Q98D42
B	-26	HIS	-	EXPRESSION TAG	UNP Q98D42
B	-25	HIS	-	EXPRESSION TAG	UNP Q98D42
B	-24	HIS	-	EXPRESSION TAG	UNP Q98D42
B	-23	SER	-	EXPRESSION TAG	UNP Q98D42
B	-22	SER	-	EXPRESSION TAG	UNP Q98D42
B	-21	GLY	-	EXPRESSION TAG	UNP Q98D42
B	-20	LEU	-	EXPRESSION TAG	UNP Q98D42
B	-19	VAL	-	EXPRESSION TAG	UNP Q98D42
B	-18	PRO	-	EXPRESSION TAG	UNP Q98D42
B	-17	ARG	-	EXPRESSION TAG	UNP Q98D42
B	-16	GLY	-	EXPRESSION TAG	UNP Q98D42
B	-15	SER	-	EXPRESSION TAG	UNP Q98D42
B	-14	HIS	-	EXPRESSION TAG	UNP Q98D42
B	-13	MET	-	EXPRESSION TAG	UNP Q98D42
B	-12	ALA	-	EXPRESSION TAG	UNP Q98D42
B	-11	SER	-	EXPRESSION TAG	UNP Q98D42
B	-10	ASN	-	EXPRESSION TAG	UNP Q98D42
B	-9	THR	-	EXPRESSION TAG	UNP Q98D42
B	-8	GLY	-	EXPRESSION TAG	UNP Q98D42
B	-7	GLY	-	EXPRESSION TAG	UNP Q98D42
B	-6	GLN	-	EXPRESSION TAG	UNP Q98D42
B	-5	GLN	-	EXPRESSION TAG	UNP Q98D42
B	-4	MET	-	EXPRESSION TAG	UNP Q98D42
B	-3	GLY	-	EXPRESSION TAG	UNP Q98D42
B	-2	ARG	-	EXPRESSION TAG	UNP Q98D42

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	GLY	-	EXPRESSION TAG	UNP Q98D42
B	0	SER	-	EXPRESSION TAG	UNP Q98D42
B	42	TRP	PHE	ENGINEERED MUTATION	UNP Q98D42

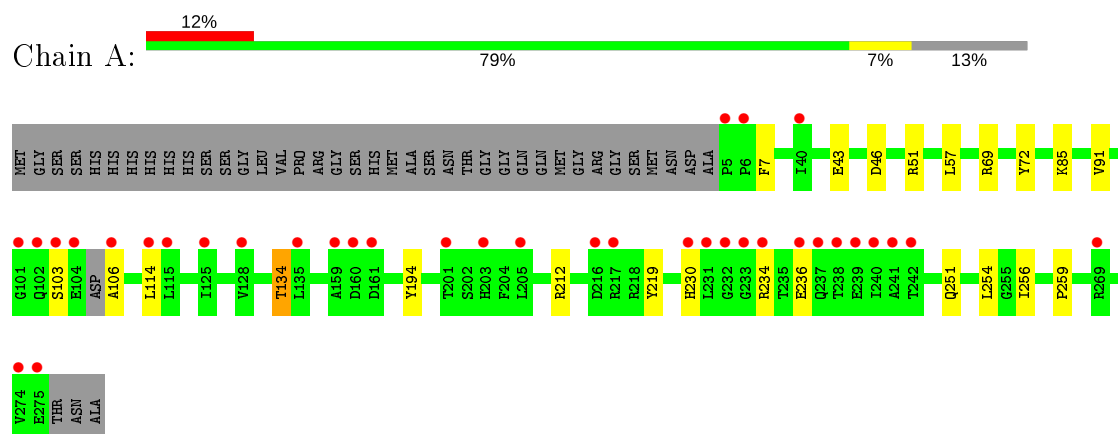
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	99	Total O 99 99	0	0
2	B	118	Total O 118 118	0	0

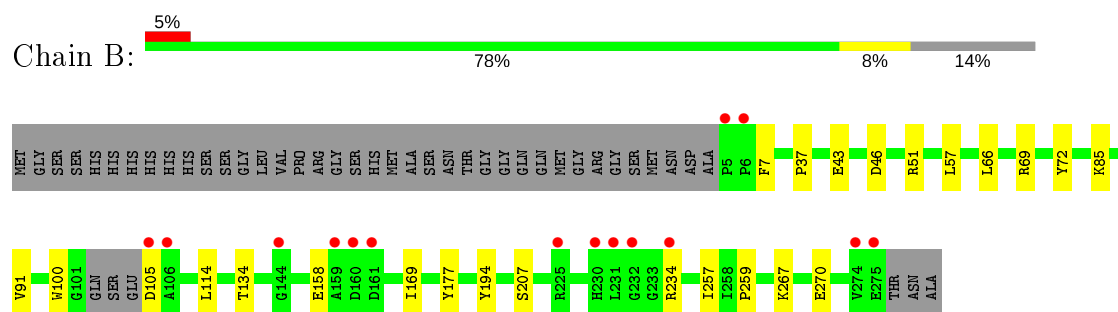
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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: Arylamine N-acetyltransferase



#### • Molecule 1: Arylamine N-acetyltransferase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.86Å 115.08Å 115.57Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.01 – 1.84 32.01 – 1.84	Depositor EDS
% Data completeness (in resolution range)	99.6 (32.01-1.84) 99.6 (32.01-1.84)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.05 (at 1.84Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, $R_{free}$	0.204 , 0.228 0.208 , 0.230	Depositor DCC
$R_{free}$ test set	3126 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.4	Xtriage
Anisotropy	0.154	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 44.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.016 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	4449	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.23% 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, 0.333 respectively for untwinned datasets, and 0.375, 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	A	0.57	0/2182	0.74	1/2962 (0.0%)
1	B	0.68	1/2155 (0.0%)	0.77	0/2927
All	All	0.63	1/4337 (0.0%)	0.75	1/5889 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	177	TYR	CE1-CZ	-5.24	1.31	1.38

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	69	ARG	NE-CZ-NH2	-5.61	117.49	120.30

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	A	2128	0	2117	13	0
1	B	2104	0	2089	10	0
2	A	99	0	0	1	0
2	B	118	0	0	0	0
All	All	4449	0	4206	22	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 3.

All (22) 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:B:51:ARG:HH12	1:B:259:PRO:HG2	1.56	0.69
1:A:51:ARG:HH12	1:A:259:PRO:HG2	1.58	0.69
1:B:46:ASP:OD1	1:B:51:ARG:NH1	2.26	0.69
1:B:51:ARG:NH1	1:B:259:PRO:HG2	2.09	0.68
1:A:46:ASP:OD1	1:A:51:ARG:NH1	2.28	0.67
1:A:51:ARG:NH1	1:A:259:PRO:HG2	2.13	0.63
1:A:134:THR:HG21	2:A:382:HOH:O	2.06	0.56
1:A:7:PHE:HD1	1:A:57:LEU:HD11	1.72	0.55
1:A:219:TYR:HB3	1:A:254:LEU:HD21	1.89	0.54
1:A:194:TYR:CE2	1:B:194:TYR:CE2	2.98	0.52
1:B:37:PRO:HB3	1:B:134:THR:HG21	1.94	0.49
1:B:267:LYS:HA	1:B:270:GLU:HG2	1.94	0.48
1:A:230:HIS:HE1	1:A:236:GLU:OE2	1.97	0.47
1:B:85:LYS:HG3	1:B:91:VAL:CG1	2.45	0.47
1:B:7:PHE:HD1	1:B:57:LEU:HD11	1.80	0.46
1:B:69:ARG:HG2	1:B:257:ILE:HB	1.98	0.46
1:A:251:GLN:HG2	1:A:256:ILE:O	2.18	0.44
1:A:85:LYS:HG3	1:A:91:VAL:CG1	2.47	0.44
1:A:230:HIS:CE1	1:A:236:GLU:OE2	2.71	0.43
1:A:103:SER:O	1:A:106:ALA:N	2.52	0.43
1:A:212:ARG:NE	1:A:254:LEU:O	2.45	0.42
1:B:100:TRP:HH2	1:B:169:ILE:CD1	2.35	0.40

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	
1	A	267/312 (86%)	258 (97%)	9 (3%)	0	100	100
1	B	264/312 (85%)	255 (97%)	9 (3%)	0	100	100
All	All	531/624 (85%)	513 (97%)	18 (3%)	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	218/249 (88%)	213 (98%)	5 (2%)	50	34
1	B	215/249 (86%)	207 (96%)	8 (4%)	34	16
All	All	433/498 (87%)	420 (97%)	13 (3%)	41	23

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

Mol	Chain	Res	Type
1	A	43	GLU
1	A	72	TYR
1	A	114	LEU
1	A	134	THR
1	A	234	ARG
1	B	43	GLU
1	B	66	LEU
1	B	72	TYR
1	B	105	ASP
1	B	114	LEU
1	B	158	GLU
1	B	207	SER
1	B	234	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	230	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 ⓘ

### 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	270/312 (86%)	0.56	36 (13%) 3 2	19, 33, 60, 80	0
1	B	268/312 (85%)	0.16	15 (5%) 24 22	17, 28, 54, 76	0
All	All	538/624 (86%)	0.36	51 (9%) 8 7	17, 30, 58, 80	0

All (51) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	232	GLY	7.2
1	A	5	PRO	6.8
1	A	106	ALA	6.1
1	B	159	ALA	5.3
1	A	231	LEU	4.4
1	B	161	ASP	4.2
1	A	103	SER	4.1
1	A	234	ARG	4.0
1	A	201	THR	4.0
1	B	160	ASP	4.0
1	B	231	LEU	3.8
1	A	274	VAL	3.8
1	B	234	ARG	3.7
1	A	161	ASP	3.7
1	B	105	ASP	3.7
1	A	128	VAL	3.6
1	A	216	ASP	3.5
1	A	233	GLY	3.4
1	B	232	GLY	3.3
1	A	160	ASP	3.3
1	A	104	GLU	3.1
1	A	135	LEU	3.0
1	B	274	VAL	3.0
1	B	5	PRO	3.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	237	GLN	2.8
1	B	106	ALA	2.8
1	B	144	GLY	2.8
1	B	275	GLU	2.7
1	A	114	LEU	2.7
1	A	275	GLU	2.6
1	A	6	PRO	2.6
1	A	230	HIS	2.5
1	A	240	ILE	2.5
1	A	236	GLU	2.5
1	A	241	ALA	2.4
1	A	217	ARG	2.4
1	A	40	ILE	2.4
1	B	225	ARG	2.4
1	A	239	GLU	2.4
1	A	205	LEU	2.4
1	A	242	THR	2.3
1	A	238	THR	2.3
1	A	203	HIS	2.2
1	A	125	ILE	2.2
1	B	230	HIS	2.1
1	A	101	GLY	2.1
1	A	159	ALA	2.1
1	A	269	ARG	2.1
1	B	6	PRO	2.1
1	A	115	LEU	2.1
1	A	102	GLN	2.0

## 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.