Full wwPDB X-ray Structure Validation Report

Feb 1, 2016 – 08:18 AM GMT

PDB ID	:	3E3M
Title	:	Crystal structure of a LacI family transcriptional regulator from Silicibacter pomeroyi
Authors	:	Bonanno, J.B.; Freeman, J.; Bain, K.T.; Iizuka, M.; Ozyurt, S.; Smith, D.; Wasserman, S.; Sauder, J.M.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2008-08-07
Resolution	:	1.60 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

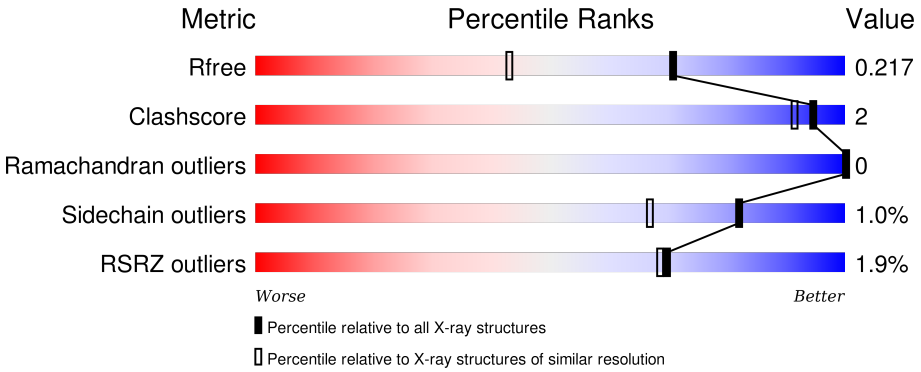
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 1.60 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	2475 (1.60-1.60)
Clashscore	102246	2732 (1.60-1.60)
Ramachandran outliers	100387	2654 (1.60-1.60)
Sidechain outliers	100360	2653 (1.60-1.60)
RSRZ outliers	91569	2479 (1.60-1.60)

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 <=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	355	2% 74% 24%
1	B	355	% 72% 24%
1	C	355	2% 72% 5% 24%
1	D	355	% 71% 5% 24%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 9477 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 Transcriptional regulator, LacI family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	270	Total	C	N	O	S	0	2	0
			2105	1336	369	393	7			
1	B	270	Total	C	N	O	S	0	8	0
			2136	1359	375	394	8			
1	C	270	Total	C	N	O	S	0	1	0
			2100	1333	366	394	7			
1	D	270	Total	C	N	O	S	0	0	0
			2100	1332	369	392	7			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	expression tag	UNP Q5LKY7
A	0	SER	-	expression tag	UNP Q5LKY7
A	1	LEU	-	expression tag	UNP Q5LKY7
A	346	GLU	-	expression tag	UNP Q5LKY7
A	347	GLY	-	expression tag	UNP Q5LKY7
A	348	HIS	-	expression tag	UNP Q5LKY7
A	349	HIS	-	expression tag	UNP Q5LKY7
A	350	HIS	-	expression tag	UNP Q5LKY7
A	351	HIS	-	expression tag	UNP Q5LKY7
A	352	HIS	-	expression tag	UNP Q5LKY7
A	353	HIS	-	expression tag	UNP Q5LKY7
B	-1	MET	-	expression tag	UNP Q5LKY7
B	0	SER	-	expression tag	UNP Q5LKY7
B	1	LEU	-	expression tag	UNP Q5LKY7
B	346	GLU	-	expression tag	UNP Q5LKY7
B	347	GLY	-	expression tag	UNP Q5LKY7
B	348	HIS	-	expression tag	UNP Q5LKY7
B	349	HIS	-	expression tag	UNP Q5LKY7
B	350	HIS	-	expression tag	UNP Q5LKY7
B	351	HIS	-	expression tag	UNP Q5LKY7
B	352	HIS	-	expression tag	UNP Q5LKY7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	353	HIS	-	expression tag	UNP Q5LKY7
C	-1	MET	-	expression tag	UNP Q5LKY7
C	0	SER	-	expression tag	UNP Q5LKY7
C	1	LEU	-	expression tag	UNP Q5LKY7
C	346	GLU	-	expression tag	UNP Q5LKY7
C	347	GLY	-	expression tag	UNP Q5LKY7
C	348	HIS	-	expression tag	UNP Q5LKY7
C	349	HIS	-	expression tag	UNP Q5LKY7
C	350	HIS	-	expression tag	UNP Q5LKY7
C	351	HIS	-	expression tag	UNP Q5LKY7
C	352	HIS	-	expression tag	UNP Q5LKY7
C	353	HIS	-	expression tag	UNP Q5LKY7
D	-1	MET	-	expression tag	UNP Q5LKY7
D	0	SER	-	expression tag	UNP Q5LKY7
D	1	LEU	-	expression tag	UNP Q5LKY7
D	346	GLU	-	expression tag	UNP Q5LKY7
D	347	GLY	-	expression tag	UNP Q5LKY7
D	348	HIS	-	expression tag	UNP Q5LKY7
D	349	HIS	-	expression tag	UNP Q5LKY7
D	350	HIS	-	expression tag	UNP Q5LKY7
D	351	HIS	-	expression tag	UNP Q5LKY7
D	352	HIS	-	expression tag	UNP Q5LKY7
D	353	HIS	-	expression tag	UNP Q5LKY7

- Molecule 2 is water.

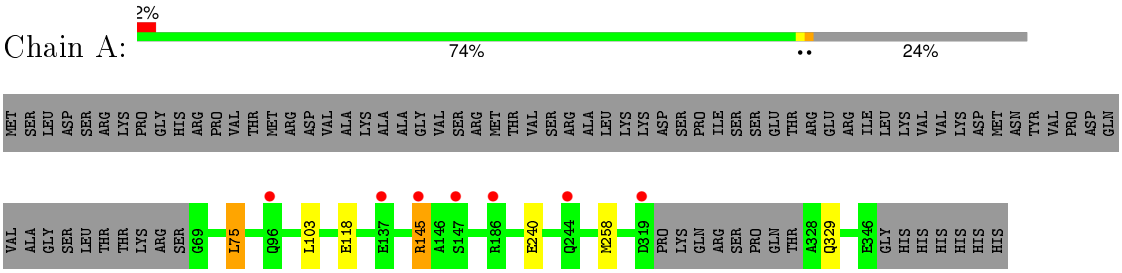
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	266	Total O 266 266	0	0
2	B	272	Total O 272 272	0	0
2	C	229	Total O 229 229	0	0
2	D	269	Total O 269 269	0	0

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 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.

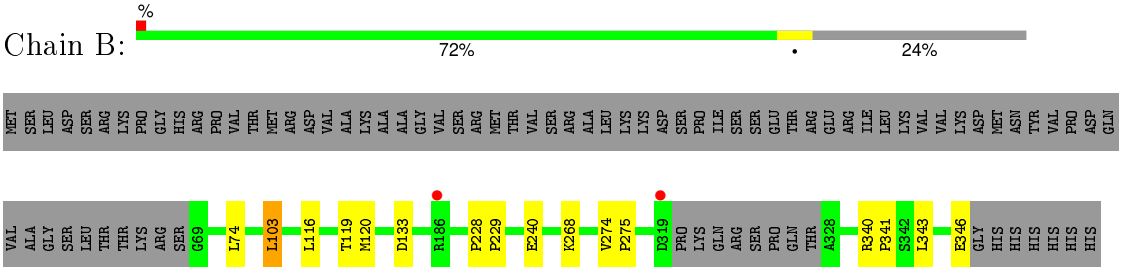
- Molecule 1: Transcriptional regulator, LacI family

Chain A:



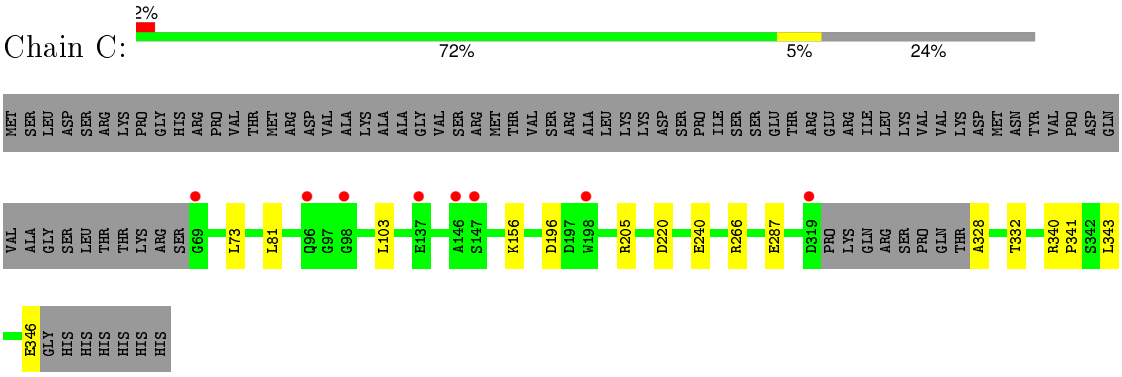
- Molecule 1: Transcriptional regulator, LacI family

Chain B:



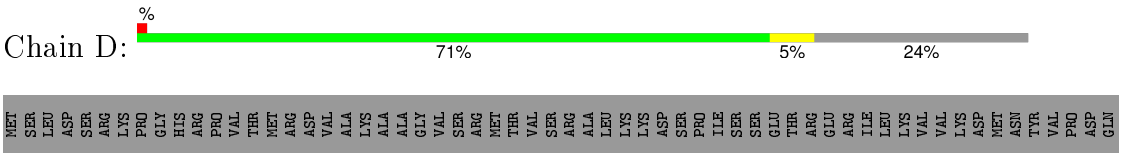
- Molecule 1: Transcriptional regulator, LacI family

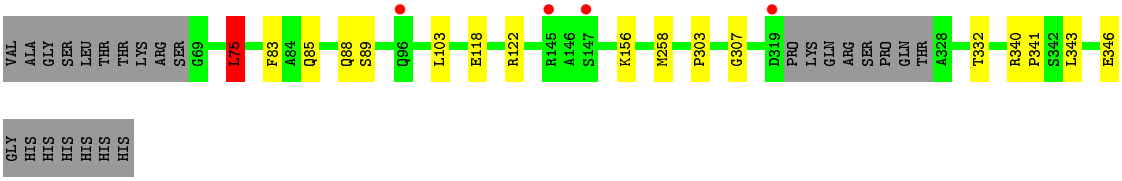
Chain C:



- Molecule 1: Transcriptional regulator, LacI family

Chain D:




VAL
ALA
GLY
SER
LEU
THR
THR
LYS
ARG
SER
G69
L75
F83
A84
Q85
Q88
S89
Q96
L103
E118
R122
R145
A146
S147
K156
M258
P303
G307
D319
PRO
LYS
GLN
ARG
SER
PRO
GLN
THR
A328
T332
R340
P341
S342
L343
E346
GLY
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	60.40Å 60.44Å 90.77Å 82.91° 76.96° 87.22°	Depositor
Resolution (Å)	15.00 – 1.60 18.53 – 1.60	Depositor EDS
% Data completeness (in resolution range)	95.6 (15.00-1.60) 91.9 (18.53-1.60)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	2.00 (at 1.60Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.184 , 0.215 0.187 , 0.217	Depositor DCC
R_{free} test set	7876 reflections (5.31%)	DCC
Wilson B-factor (Å^2)	14.9	Xtriage
Anisotropy	0.117	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.41 , 55.6	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.29$	Xtriage
Outliers	0 of 156396 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	9477	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.34% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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

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	0.53	0/2152	0.64	2/2919 (0.1%)
1	B	0.50	0/2201	0.65	1/2984 (0.0%)
1	C	0.49	0/2144	0.61	0/2909
1	D	0.52	0/2141	0.64	2/2904 (0.1%)
All	All	0.51	0/8638	0.64	5/11716 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	75	LEU	CA-CB-CG	5.64	128.26	115.30
1	D	258	MET	CG-SD-CE	-5.62	91.20	100.20
1	A	258	MET	CG-SD-CE	-5.46	91.47	100.20
1	D	75	LEU	CA-CB-CG	5.30	127.50	115.30
1	B	103	LEU	CA-CB-CG	5.12	127.09	115.30

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	2105	0	2118	3	0
1	B	2136	0	2177	10	0
1	C	2100	0	2106	9	0
1	D	2100	0	2111	8	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	266	0	0	3	0
2	B	272	0	0	1	0
2	C	229	0	0	1	0
2	D	269	0	0	1	0
All	All	9477	0	8512	28	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 (28) 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:C:156:LYS:NZ	1:C:332:THR:OG1	2.14	0.73
1:C:240:GLU:OE2	1:C:266:ARG:NE	2.33	0.60
1:D:122:ARG:HD2	2:D:387:HOH:O	2.07	0.55
1:A:103:LEU:HB2	1:B:103:LEU:HB2	1.90	0.54
1:C:343:LEU:HD23	1:C:346:GLU:HG3	1.90	0.54
1:C:103:LEU:HB2	1:D:103:LEU:HB2	1.91	0.53
1:B:133:ASP:HB2	2:B:410:HOH:O	2.08	0.52
2:A:360:HOH:O	1:B:268:LYS:HE3	2.12	0.50
1:D:343:LEU:HD23	1:D:346:GLU:HG3	1.94	0.49
1:A:145:ARG:HG2	2:A:439:HOH:O	2.11	0.49
1:D:156:LYS:HE2	1:D:332:THR:OG1	2.13	0.48
1:B:116:LEU:HA	1:B:119[B]:THR:HG22	1.95	0.48
1:A:329:GLN:HG3	2:A:413:HOH:O	2.13	0.48
1:C:196:ASP:OD2	1:C:205:ARG:NH1	2.37	0.47
1:B:343:LEU:HD23	1:B:346:GLU:HG3	1.97	0.47
1:C:81:LEU:HD22	1:C:287[B]:GLU:HG3	1.98	0.45
1:C:328:ALA:N	2:C:385:HOH:O	2.49	0.45
1:B:74:LEU:HD21	1:B:120[A]:MET:CE	2.46	0.45
1:D:75:LEU:HD23	1:D:83:PHE:HB3	1.99	0.45
1:B:74:LEU:HD21	1:B:120[A]:MET:HE1	2.00	0.44
1:B:274:VAL:HA	1:B:275:PRO:HA	1.84	0.44
1:D:85:GLN:OE1	1:D:303:PRO:HG2	2.18	0.43
1:D:89:SER:OG	1:D:307:GLY:HA3	2.19	0.43
1:B:340:ARG:HB3	1:B:341:PRO:CD	2.51	0.41
1:C:340:ARG:HB3	1:C:341:PRO:CD	2.51	0.41
1:B:228:PRO:HA	1:B:229:PRO:HA	1.92	0.40
1:D:340:ARG:HB3	1:D:341:PRO:CD	2.51	0.40
1:C:73:LEU:HD23	1:C:103:LEU:HD22	2.01	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	268/355 (76%)	262 (98%)	6 (2%)	0	100	100
1	B	274/355 (77%)	270 (98%)	4 (2%)	0	100	100
1	C	267/355 (75%)	262 (98%)	5 (2%)	0	100	100
1	D	266/355 (75%)	262 (98%)	4 (2%)	0	100	100
All	All	1075/1420 (76%)	1056 (98%)	19 (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	226/301 (75%)	222 (98%)	4 (2%)	66	41
1	B	232/301 (77%)	231 (100%)	1 (0%)	93	88
1	C	225/301 (75%)	224 (100%)	1 (0%)	93	88
1	D	225/301 (75%)	222 (99%)	3 (1%)	76	56
All	All	908/1204 (75%)	899 (99%)	9 (1%)	82	67

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	75	LEU
1	A	118	GLU
1	A	145	ARG
1	A	240	GLU
1	B	240	GLU
1	C	220	ASP
1	D	75	LEU
1	D	88	GLN
1	D	118	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	115	GLN
1	C	115	GLN
1	D	277	GLN

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

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^{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	A	270/355 (76%)	-0.19	7 (2%)	59	57	12, 18, 28, 37	0
1	B	270/355 (76%)	-0.15	2 (0%)	89	89	13, 19, 28, 32	0
1	C	270/355 (76%)	0.02	8 (2%)	54	51	13, 22, 31, 39	0
1	D	270/355 (76%)	-0.15	4 (1%)	76	75	13, 20, 30, 39	0
All	All	1080/1420 (76%)	-0.12	21 (1%)	70	68	12, 20, 29, 39	0

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	198	TRP	4.9
1	D	147	SER	3.6
1	C	147	SER	3.3
1	C	319	ASP	3.2
1	A	186	ARG	3.1
1	A	147	SER	3.0
1	C	96	GLN	2.8
1	D	96	GLN	2.7
1	C	69	GLY	2.6
1	A	145	ARG	2.6
1	D	319	ASP	2.6
1	A	96	GLN	2.4
1	C	98	GLY	2.4
1	B	186	ARG	2.4
1	A	244	GLN	2.3
1	C	137	GLU	2.2
1	D	145	ARG	2.2
1	C	146	ALA	2.2
1	A	319	ASP	2.1
1	A	137	GLU	2.0
1	B	319	ASP	2.0

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.