



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 13, 2016 – 11:35 AM EDT

PDB ID : 5G1Q  
Title : Compressed conformation of Francisella tularensis ClpP at 2.84 Å  
Authors : Diaz-Saez, L.; Hunter, W.N.  
Deposited on : 2016-03-29  
Resolution : 2.84 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
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)	:	rb-20027939

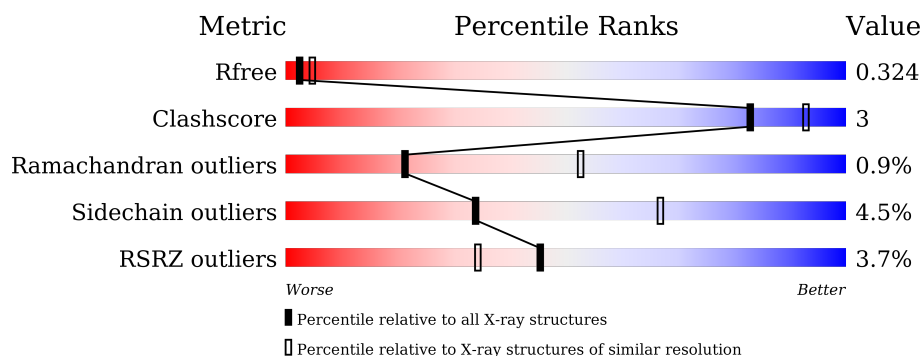
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.84 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))
$R_{free}$	91344	3170 (2.88-2.80)
Clashscore	102246	3658 (2.88-2.80)
Ramachandran outliers	100387	3591 (2.88-2.80)
Sidechain outliers	100360	3594 (2.88-2.80)
RSRZ outliers	91569	3184 (2.88-2.80)

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	201	<div> <div>2%</div> <div>82% 8% 10%</div> </div>
1	B	201	<div> <div>%</div> <div>77% 7% • 14%</div> </div>
1	C	201	<div> <div>6%</div> <div>81% 5% • 13%</div> </div>
1	D	201	<div> <div>6%</div> <div>69% 14% • 15%</div> </div>
1	E	201	<div> <div>%</div> <div>75% 10% 15%</div> </div>
1	F	201	<div> <div>%</div> <div>75% 9% • 15%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	201	<div><div></div><div>4%</div><div>82%</div><div>• • 12%</div></div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 9316 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.

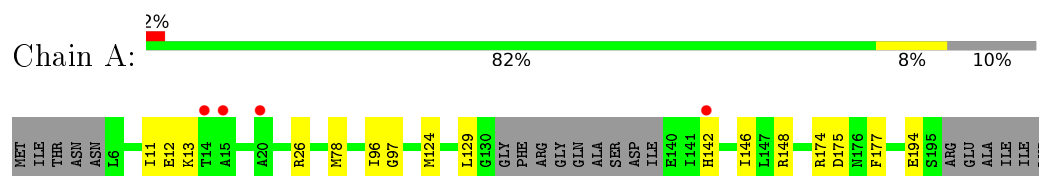
- Molecule 1 is a protein called CLP PROTEASE PROTEOLYTIC SUBUNIT P.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	181	Total	C	N	O	S	0	0	0
			1384	878	232	266	8			
1	B	172	Total	C	N	O	S	0	0	0
			1327	845	222	252	8			
1	C	175	Total	C	N	O	S	5	0	0
			1337	849	225	255	8			
1	D	170	Total	C	N	O	S	0	0	0
			1276	806	214	248	8			
1	E	171	Total	C	N	O	S	0	0	0
			1302	826	217	251	8			
1	F	171	Total	C	N	O	S	5	0	0
			1312	836	218	250	8			
1	G	177	Total	C	N	O	S	3	1	0
			1378	876	236	258	8			

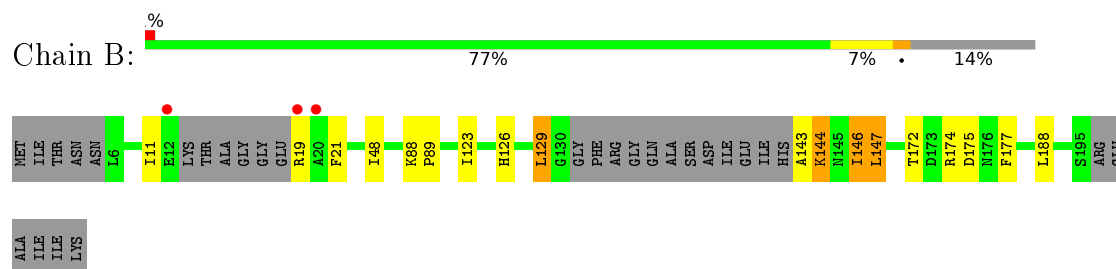
### 3 Residue-property plots [i](#)

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.

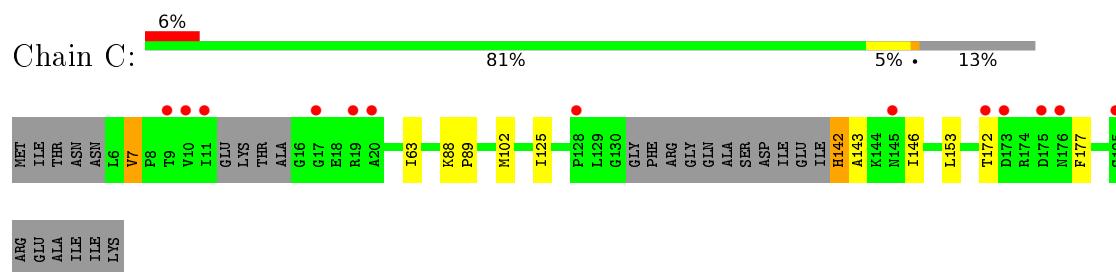
- Molecule 1: CLP PROTEASE PROTEOLYTIC SUBUNIT P



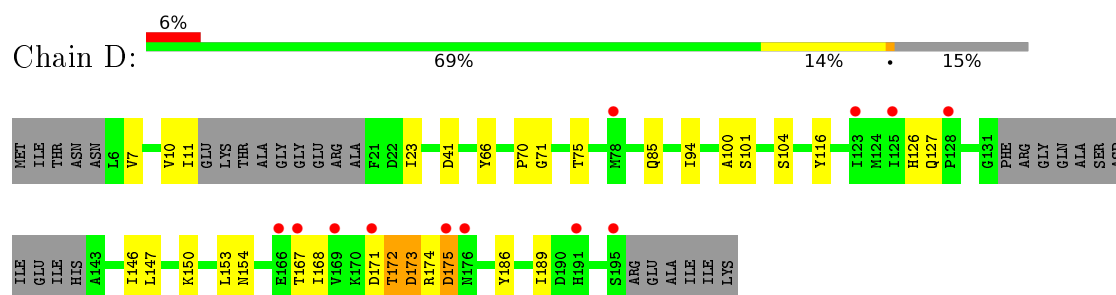
- Molecule 1: CLP PROTEASE PROTEOLYTIC SUBUNIT P



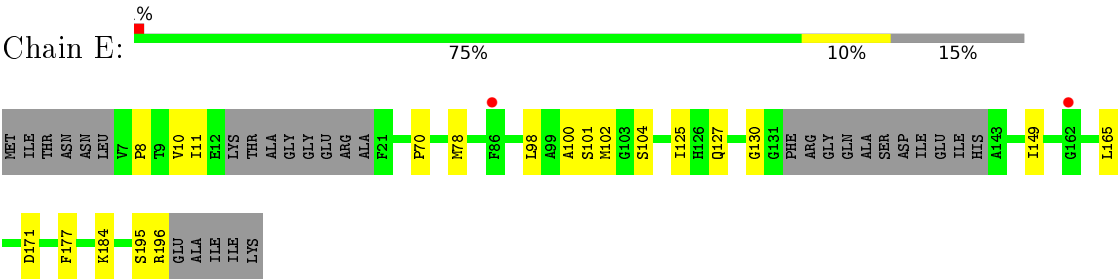
- Molecule 1: CLP PROTEASE PROTEOLYTIC SUBUNIT P



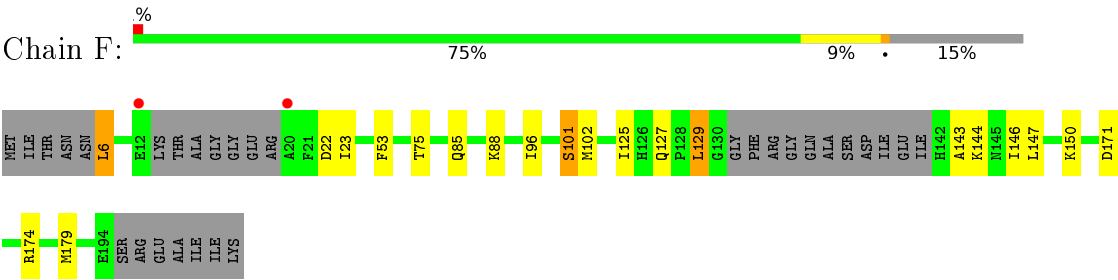
- Molecule 1: CLP PROTEASE PROTEOLYTIC SUBUNIT P



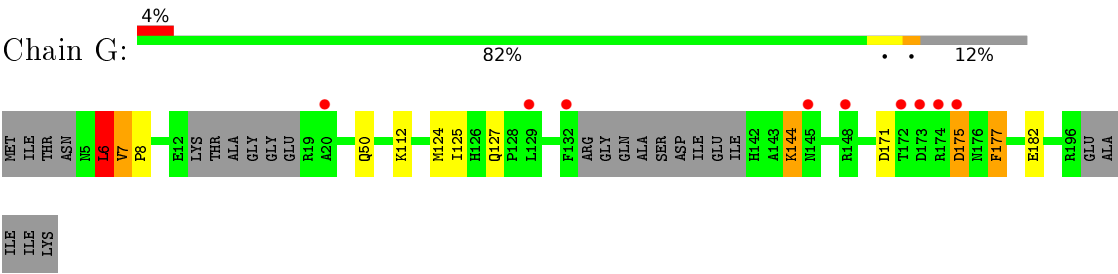
- Molecule 1: CLP PROTEASE PROTEOLYTIC SUBUNIT P



• Molecule 1: CLP PROTEASE PROTEOLYTIC SUBUNIT P



• Molecule 1: CLP PROTEASE PROTEOLYTIC SUBUNIT P



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	113.46 Å   125.94 Å   96.95 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	96.95 – 2.84 55.06 – 2.84	Depositor EDS
% Data completeness (in resolution range)	97.4 (96.95-2.84) 97.4 (55.06-2.84)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.00 (at 2.86 Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, $R_{free}$	0.249   ,   0.326 0.250   ,   0.324	Depositor DCC
$R_{free}$ test set	1629 reflections (5.26%)	DCC
Wilson B-factor (Å <sup>2</sup> )	31.2	Xtriage
Anisotropy	0.378	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 33.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.52$ , $\langle L^2 \rangle = 0.36$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.88	EDS
Total number of atoms	9316	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.08% 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 ⓘ

### 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	0.44	0/1406	0.64	0/1900
1	B	0.43	0/1347	0.66	0/1817
1	C	0.57	3/1358 (0.2%)	0.70	3/1834 (0.2%)
1	D	0.43	0/1294	0.61	0/1751
1	E	0.40	0/1322	0.62	0/1788
1	F	0.58	2/1333 (0.2%)	0.71	3/1802 (0.2%)
1	G	0.41	0/1404	0.68	1/1892 (0.1%)
All	All	0.47	5/9464 (0.1%)	0.66	7/12784 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	D	0	1
1	F	0	1
1	G	0	2
All	All	0	4

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	7	VAL	CB-CG1	-14.13	1.23	1.52
1	F	150	LYS	CD-CE	-11.76	1.21	1.51
1	F	88	LYS	CG-CD	-8.39	1.24	1.52
1	C	7	VAL	CB-CG2	5.73	1.64	1.52
1	C	88	LYS	CG-CD	-5.00	1.35	1.52

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	7	VAL	CA-CB-CG1	14.29	132.34	110.90

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	F	88	LYS	CB-CG-CD	7.46	130.98	111.60
1	C	7	VAL	CA-CB-CG2	-7.15	100.18	110.90
1	G	112	LYS	CB-CG-CD	-6.18	95.53	111.60
1	C	7	VAL	CG1-CB-CG2	6.13	120.71	110.90

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	147	LEU	Peptide
1	F	147	LEU	Peptide
1	G	6	LEU	Peptide
1	G	7	VAL	Peptide

## 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	1384	0	1378	9	0
1	B	1327	0	1338	13	0
1	C	1337	0	1329	9	0
1	D	1276	0	1252	19	0
1	E	1302	0	1281	6	0
1	F	1312	0	1305	8	0
1	G	1378	0	1385	6	0
All	All	9316	0	9268	57	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.

The worst 5 of 57 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:D:126:HIS:ND1	1:D:175:ASP:OD1	1.83	1.11
1:D:168:ILE:O	1:D:172:THR:HG23	1.81	0.80

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:171:ASP:OD2	1:D:186:TYR:OH	2.09	0.70
1:D:172:THR:O	1:D:174:ARG:N	2.27	0.68
1:B:172:THR:C	1:B:174:ARG:H	1.98	0.67

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	177/201 (88%)	165 (93%)	11 (6%)	1 (1%)	30	63
1	B	166/201 (83%)	154 (93%)	9 (5%)	3 (2%)	11	32
1	C	169/201 (84%)	160 (95%)	8 (5%)	1 (1%)	30	63
1	D	164/201 (82%)	149 (91%)	13 (8%)	2 (1%)	16	44
1	E	165/201 (82%)	150 (91%)	14 (8%)	1 (1%)	30	63
1	F	165/201 (82%)	157 (95%)	7 (4%)	1 (1%)	30	63
1	G	172/201 (86%)	160 (93%)	10 (6%)	2 (1%)	16	44
All	All	1178/1407 (84%)	1095 (93%)	72 (6%)	11 (1%)	21	52

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	144	LYS
1	D	173	ASP
1	G	8	PRO
1	B	146	ILE
1	F	129	LEU

### 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	149/168 (89%)	143 (96%)	6 (4%)	38	72
1	B	145/168 (86%)	140 (97%)	5 (3%)	44	77
1	C	144/168 (86%)	143 (99%)	1 (1%)	88	97
1	D	136/168 (81%)	129 (95%)	7 (5%)	29	61
1	E	140/168 (83%)	130 (93%)	10 (7%)	18	44
1	F	142/168 (84%)	133 (94%)	9 (6%)	22	51
1	G	150/168 (89%)	143 (95%)	7 (5%)	32	65
All	All	1006/1176 (86%)	961 (96%)	45 (4%)	34	67

5 of 45 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	11	ILE
1	E	177	PHE
1	G	127	GLN
1	E	104	SER
1	E	184	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	154	ASN
1	G	45	ASN
1	E	127	GLN
1	C	142	HIS
1	E	45	ASN

### 5.3.3 RNA ⓘ

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 [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<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	181/201 (90%)	0.06	4 (2%) 65 56	10, 19, 67, 83	0
1	B	172/201 (85%)	0.02	3 (1%) 73 65	15, 28, 58, 70	0
1	C	175/201 (87%)	0.46	13 (7%) 17 10	23, 40, 75, 99	2 (1%)
1	D	170/201 (84%)	0.55	12 (7%) 19 11	28, 47, 79, 92	0
1	E	171/201 (85%)	0.37	2 (1%) 81 74	25, 44, 74, 82	0
1	F	171/201 (85%)	-0.04	2 (1%) 81 74	13, 24, 49, 77	2 (1%)
1	G	177/201 (88%)	0.05	9 (5%) 32 22	11, 23, 66, 77	1 (0%)
All	All	1217/1407 (86%)	0.21	45 (3%) 45 34	10, 34, 71, 99	5 (0%)

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	173	ASP	4.6
1	D	128	PRO	4.2
1	C	176	ASN	4.0
1	B	20	ALA	4.0
1	A	15	ALA	3.9

### 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

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.