



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 10:16 am BST

PDB ID : 3KW0  
Title : Crystal structure of Cysteine peptidase (NP\_982244.1) from BACILLUS CEREUS ATCC 10987 at 2.50 Å resolution  
Authors : Joint Center for Structural Genomics (JCSG)  
Deposited on : 2009-11-30  
Resolution : 2.50 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

<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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
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

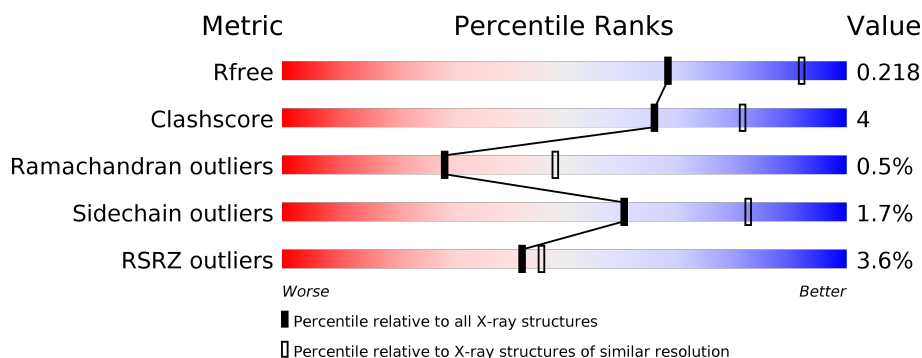
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	214	<div> <div>3%</div> <div> <div></div> <div>78%</div> <div>11%</div> <div>12%</div> </div> </div>
1	B	214	<div> <div>4%</div> <div> <div></div> <div>81%</div> <div>11%</div> <div>8%</div> </div> </div>
1	C	214	<div> <div>3%</div> <div> <div></div> <div>80%</div> <div>12%</div> <div>8%</div> </div> </div>
1	D	214	<div> <div>3%</div> <div> <div></div> <div>75%</div> <div>14%</div> <div>11%</div> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6102 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 Cysteine peptidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	189	Total	C	N	O	S	Se	0	1	0
			1473	942	246	279	3	3			
1	B	197	Total	C	N	O	S	Se	0	0	0
			1523	970	257	289	3	4			
1	C	197	Total	C	N	O	S	Se	0	1	0
			1539	982	260	290	3	4			
1	D	190	Total	C	N	O	S	Se	0	1	0
			1477	941	250	279	3	4			

There are 76 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-18	MSE	-	leader sequence	UNP Q74NK7
A	-17	GLY	-	leader sequence	UNP Q74NK7
A	-16	SER	-	leader sequence	UNP Q74NK7
A	-15	ASP	-	leader sequence	UNP Q74NK7
A	-14	LYS	-	leader sequence	UNP Q74NK7
A	-13	ILE	-	leader sequence	UNP Q74NK7
A	-12	HIS	-	leader sequence	UNP Q74NK7
A	-11	HIS	-	leader sequence	UNP Q74NK7
A	-10	HIS	-	leader sequence	UNP Q74NK7
A	-9	HIS	-	leader sequence	UNP Q74NK7
A	-8	HIS	-	leader sequence	UNP Q74NK7
A	-7	HIS	-	leader sequence	UNP Q74NK7
A	-6	GLU	-	leader sequence	UNP Q74NK7
A	-5	ASN	-	leader sequence	UNP Q74NK7
A	-4	LEU	-	leader sequence	UNP Q74NK7
A	-3	TYR	-	leader sequence	UNP Q74NK7
A	-2	PHE	-	leader sequence	UNP Q74NK7
A	-1	GLN	-	leader sequence	UNP Q74NK7
A	0	GLY	-	leader sequence	UNP Q74NK7
B	-18	MSE	-	leader sequence	UNP Q74NK7
B	-17	GLY	-	leader sequence	UNP Q74NK7

*Continued on next page...*

*Continued from previous page...*

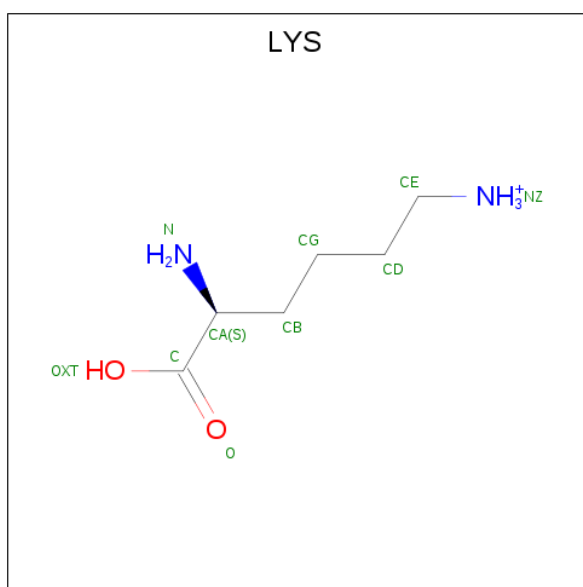
Chain	Residue	Modelled	Actual	Comment	Reference
B	-16	SER	-	leader sequence	UNP Q74NK7
B	-15	ASP	-	leader sequence	UNP Q74NK7
B	-14	LYS	-	leader sequence	UNP Q74NK7
B	-13	ILE	-	leader sequence	UNP Q74NK7
B	-12	HIS	-	leader sequence	UNP Q74NK7
B	-11	HIS	-	leader sequence	UNP Q74NK7
B	-10	HIS	-	leader sequence	UNP Q74NK7
B	-9	HIS	-	leader sequence	UNP Q74NK7
B	-8	HIS	-	leader sequence	UNP Q74NK7
B	-7	HIS	-	leader sequence	UNP Q74NK7
B	-6	GLU	-	leader sequence	UNP Q74NK7
B	-5	ASN	-	leader sequence	UNP Q74NK7
B	-4	LEU	-	leader sequence	UNP Q74NK7
B	-3	TYR	-	leader sequence	UNP Q74NK7
B	-2	PHE	-	leader sequence	UNP Q74NK7
B	-1	GLN	-	leader sequence	UNP Q74NK7
B	0	GLY	-	leader sequence	UNP Q74NK7
C	-18	MSE	-	leader sequence	UNP Q74NK7
C	-17	GLY	-	leader sequence	UNP Q74NK7
C	-16	SER	-	leader sequence	UNP Q74NK7
C	-15	ASP	-	leader sequence	UNP Q74NK7
C	-14	LYS	-	leader sequence	UNP Q74NK7
C	-13	ILE	-	leader sequence	UNP Q74NK7
C	-12	HIS	-	leader sequence	UNP Q74NK7
C	-11	HIS	-	leader sequence	UNP Q74NK7
C	-10	HIS	-	leader sequence	UNP Q74NK7
C	-9	HIS	-	leader sequence	UNP Q74NK7
C	-8	HIS	-	leader sequence	UNP Q74NK7
C	-7	HIS	-	leader sequence	UNP Q74NK7
C	-6	GLU	-	leader sequence	UNP Q74NK7
C	-5	ASN	-	leader sequence	UNP Q74NK7
C	-4	LEU	-	leader sequence	UNP Q74NK7
C	-3	TYR	-	leader sequence	UNP Q74NK7
C	-2	PHE	-	leader sequence	UNP Q74NK7
C	-1	GLN	-	leader sequence	UNP Q74NK7
C	0	GLY	-	leader sequence	UNP Q74NK7
D	-18	MSE	-	leader sequence	UNP Q74NK7
D	-17	GLY	-	leader sequence	UNP Q74NK7
D	-16	SER	-	leader sequence	UNP Q74NK7
D	-15	ASP	-	leader sequence	UNP Q74NK7
D	-14	LYS	-	leader sequence	UNP Q74NK7
D	-13	ILE	-	leader sequence	UNP Q74NK7

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	-12	HIS	-	leader sequence	UNP Q74NK7
D	-11	HIS	-	leader sequence	UNP Q74NK7
D	-10	HIS	-	leader sequence	UNP Q74NK7
D	-9	HIS	-	leader sequence	UNP Q74NK7
D	-8	HIS	-	leader sequence	UNP Q74NK7
D	-7	HIS	-	leader sequence	UNP Q74NK7
D	-6	GLU	-	leader sequence	UNP Q74NK7
D	-5	ASN	-	leader sequence	UNP Q74NK7
D	-4	LEU	-	leader sequence	UNP Q74NK7
D	-3	TYR	-	leader sequence	UNP Q74NK7
D	-2	PHE	-	leader sequence	UNP Q74NK7
D	-1	GLN	-	leader sequence	UNP Q74NK7
D	0	GLY	-	leader sequence	UNP Q74NK7

- Molecule 2 is LYSINE (three-letter code: LYS) (formula:  $C_6H_{15}N_2O_2$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			10	6	2	2		
2	B	1	Total	C	N	O	0	0
			10	6	2	2		
2	C	1	Total	C	N	O	0	0
			10	6	2	2		
2	D	1	Total	C	N	O	0	0
			10	6	2	2		

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

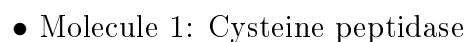
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	3	Total Cl 3 3	0	0
3	A	3	Total Cl 3 3	0	0
3	D	1	Total Cl 1 1	0	0
3	C	2	Total Cl 2 2	0	0

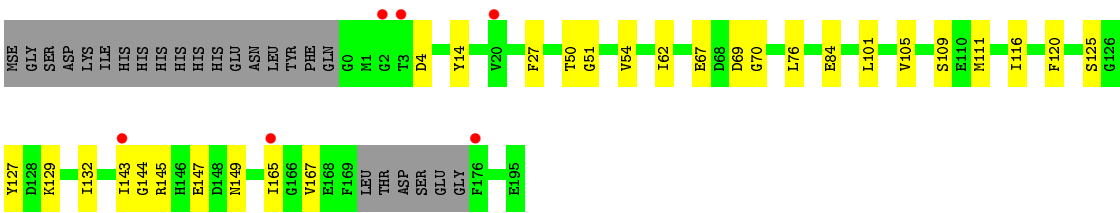
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	13	Total O 13 13	0	0
4	B	6	Total O 6 6	0	0
4	C	9	Total O 9 9	0	0
4	D	13	Total O 13 13	0	0



- Molecule 1: Cysteine peptidase







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.96 Å 64.96 Å 407.80 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	37.77 – 2.50 37.77 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.7 (37.77-2.50) 99.7 (37.77-2.50)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.84 (at 2.51 Å)	Xtriage
Refinement program	REFMAC 5.5.0102, PHENIX	Depositor
R, $R_{free}$	0.192 , 0.219 0.195 , 0.218	Depositor DCC
$R_{free}$ test set	1689 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	63.5	Xtriage
Anisotropy	0.239	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 44.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	0.194 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	6102	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.76% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section:  
CL

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	0.65	2/1496 (0.1%)	0.71	0/2014
1	B	0.67	2/1544 (0.1%)	0.71	0/2077
1	C	0.68	2/1565 (0.1%)	0.72	0/2105
1	D	0.64	0/1500	0.71	0/2017
All	All	0.66	6/6105 (0.1%)	0.71	0/8213

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	77	GLU	CB-CG	6.43	1.64	1.52
1	C	77	GLU	CG-CD	5.81	1.60	1.51
1	C	77	GLU	CB-CG	5.70	1.62	1.52
1	A	77	GLU	CB-CG	5.69	1.62	1.52
1	B	77	GLU	CG-CD	5.59	1.60	1.51

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	A	1473	0	1424	13	0
1	B	1523	0	1460	10	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	1539	0	1481	15	0
1	D	1477	0	1436	16	0
2	A	10	0	12	0	0
2	B	10	0	12	0	0
2	C	10	0	12	0	0
2	D	10	0	12	1	0
3	A	3	0	0	0	0
3	B	3	0	0	0	0
3	C	2	0	0	0	0
3	D	1	0	0	0	0
4	A	13	0	0	0	0
4	B	6	0	0	0	0
4	C	9	0	0	0	0
4	D	13	0	0	0	0
All	All	6102	0	5849	48	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 4.

The worst 5 of 48 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:125:SER:HB3	1:D:69:ASP:OD1	1.90	0.69
1:D:111:MSE:HE2	1:D:165:ILE:HD12	1.76	0.68
1:C:69:ASP:OD1	1:D:125:SER:HB3	1.94	0.68
1:C:59[B]:HIS:NE2	1:C:77:GLU:OE2	2.32	0.62
1:A:110[B]:GLU:OE2	1:A:110[B]:GLU:HA	2.00	0.60

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	186/214 (87%)	180 (97%)	6 (3%)	0	100	100
1	B	193/214 (90%)	182 (94%)	9 (5%)	2 (1%)	15	28
1	C	194/214 (91%)	183 (94%)	10 (5%)	1 (0%)	29	48
1	D	187/214 (87%)	178 (95%)	8 (4%)	1 (0%)	29	48
All	All	760/856 (89%)	723 (95%)	33 (4%)	4 (0%)	29	48

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	170	LEU
1	B	1	MSE
1	B	170	LEU
1	D	147	GLU

### 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	157/187 (84%)	155 (99%)	2 (1%)	69	87
1	B	160/187 (86%)	157 (98%)	3 (2%)	57	80
1	C	163/187 (87%)	161 (99%)	2 (1%)	71	88
1	D	158/187 (84%)	156 (99%)	2 (1%)	69	87
All	All	638/748 (85%)	629 (99%)	9 (1%)	60	86

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	110	GLU
1	D	84	GLU
1	C	84	GLU
1	B	4	ASP
1	C	4	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such

sidechains are listed below:

Mol	Chain	Res	Type
1	B	8	ASN
1	C	8	ASN

### 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](#)

Of 13 ligands modelled in this entry, 9 are monoatomic - leaving 4 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	186/214 (86%)	0.31	7 (3%) 40 43	27, 38, 71, 82	0
1	B	193/214 (90%)	0.34	8 (4%) 37 40	27, 39, 74, 82	0
1	C	193/214 (90%)	0.32	6 (3%) 49 52	27, 39, 74, 82	0
1	D	186/214 (86%)	0.29	6 (3%) 47 51	26, 38, 72, 90	0
All	All	758/856 (88%)	0.32	27 (3%) 42 46	26, 39, 74, 90	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	176	PHE	7.0
1	B	170	LEU	5.5
1	B	195	GLU	5.1
1	C	172	ASP	5.0
1	C	-3	TYR	4.9

### 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 [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<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	B-factors( $\text{\AA}^2$ )	Q<0.9
3	CL	C	203	1/1	0.75	0.15	82,82,82,82	0
3	CL	B	203	1/1	0.92	0.08	56,56,56,56	0
2	LYS	A	201	10/10	0.94	0.18	29,37,40,42	0
2	LYS	B	201	10/10	0.95	0.14	30,34,37,39	0
3	CL	D	202	1/1	0.96	0.09	54,54,54,54	0
3	CL	A	203	1/1	0.97	0.08	53,53,53,53	0
3	CL	A	204	1/1	0.97	0.08	56,56,56,56	0
2	LYS	C	201	10/10	0.97	0.16	27,36,39,43	0
3	CL	B	204	1/1	0.97	0.14	61,61,61,61	0
3	CL	B	202	1/1	0.97	0.30	63,63,63,63	0
2	LYS	D	201	10/10	0.97	0.17	36,39,46,49	0
3	CL	A	202	1/1	0.98	0.16	65,65,65,65	0
3	CL	C	202	1/1	0.99	0.08	53,53,53,53	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.