



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 09:01 AM GMT

PDB ID : 3GV2  
Title : X-ray Structure of Hexameric HIV-1 CA  
Authors : Kelly, B.N.  
Deposited on : 2009-03-30  
Resolution : 7.00 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 : 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

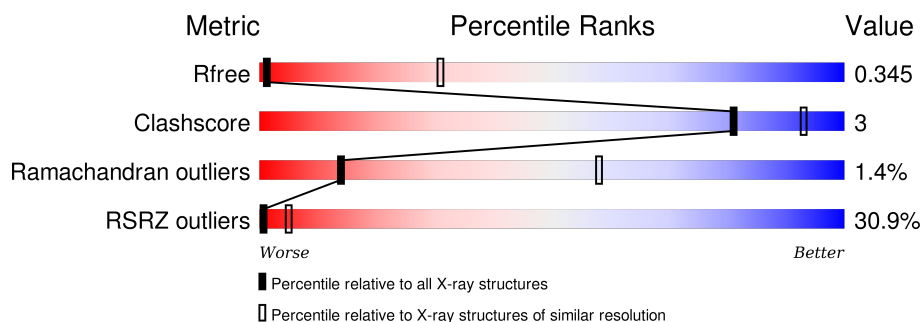
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 7.00 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	1014 (9.50-3.66)
Clashscore	102246	1063 (10.00-3.70)
Ramachandran outliers	100387	1035 (9.50-3.66)
RSRZ outliers	91569	1013 (9.50-3.66)

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	342	<div> <div>18%</div> <div>61%</div> <div>36%</div> </div>
1	B	342	<div> <div>17%</div> <div>63%</div> <div>36%</div> </div>
1	C	342	<div> <div>19%</div> <div>62%</div> <div>36%</div> </div>
1	D	342	<div> <div>21%</div> <div>61%</div> <div>36%</div> </div>
1	E	342	<div> <div>21%</div> <div>63%</div> <div>36%</div> </div>
1	F	342	<div> <div>23%</div> <div>62%</div> <div>36%</div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 5256 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 Fusion protein consisting of Capsid protein p24, Linker and Carbon dioxide-concentrating mechanism protein ccmK homolog 4.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	219	Total	C	N	O	0	0	0
			876	438	219	219			
1	B	219	Total	C	N	O	0	0	0
			876	438	219	219			
1	C	219	Total	C	N	O	0	0	0
			876	438	219	219			
1	D	219	Total	C	N	O	0	0	0
			876	438	219	219			
1	E	219	Total	C	N	O	0	0	0
			876	438	219	219			
1	F	219	Total	C	N	O	0	0	0
			876	438	219	219			

There are 78 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	184	ALA	TRP	ENGINEERED	UNP P12497
A	185	ALA	MET	ENGINEERED	UNP P12497
A	224	THR	-	LINKER	UNP P12497
A	225	ARG	-	LINKER	UNP P12497
A	226	PRO	-	LINKER	UNP P12497
A	227	GLU	-	LINKER	UNP P12497
A	228	LEU	-	LINKER	UNP P12497
A	332	TYR	GLU	ENGINEERED	UNP P73407
A	338	GLU	-	EXPRESSION TAG	UNP P73407
A	339	VAL	-	EXPRESSION TAG	UNP P73407
A	340	LEU	-	EXPRESSION TAG	UNP P73407
A	341	PHE	-	EXPRESSION TAG	UNP P73407
A	342	GLN	-	EXPRESSION TAG	UNP P73407
B	184	ALA	TRP	ENGINEERED	UNP P12497
B	185	ALA	MET	ENGINEERED	UNP P12497
B	224	THR	-	LINKER	UNP P12497

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	225	ARG	-	LINKER	UNP P12497
B	226	PRO	-	LINKER	UNP P12497
B	227	GLU	-	LINKER	UNP P12497
B	228	LEU	-	LINKER	UNP P12497
B	332	TYR	GLU	ENGINEERED	UNP P734077
B	338	GLU	-	EXPRESSION TAG	UNP P734077
B	339	VAL	-	EXPRESSION TAG	UNP P734077
B	340	LEU	-	EXPRESSION TAG	UNP P734077
B	341	PHE	-	EXPRESSION TAG	UNP P734077
B	342	GLN	-	EXPRESSION TAG	UNP P734077
C	184	ALA	TRP	ENGINEERED	UNP P12497
C	185	ALA	MET	ENGINEERED	UNP P12497
C	224	THR	-	LINKER	UNP P12497
C	225	ARG	-	LINKER	UNP P12497
C	226	PRO	-	LINKER	UNP P12497
C	227	GLU	-	LINKER	UNP P12497
C	228	LEU	-	LINKER	UNP P12497
C	332	TYR	GLU	ENGINEERED	UNP P734077
C	338	GLU	-	EXPRESSION TAG	UNP P734077
C	339	VAL	-	EXPRESSION TAG	UNP P734077
C	340	LEU	-	EXPRESSION TAG	UNP P734077
C	341	PHE	-	EXPRESSION TAG	UNP P734077
C	342	GLN	-	EXPRESSION TAG	UNP P734077
D	184	ALA	TRP	ENGINEERED	UNP P12497
D	185	ALA	MET	ENGINEERED	UNP P12497
D	224	THR	-	LINKER	UNP P12497
D	225	ARG	-	LINKER	UNP P12497
D	226	PRO	-	LINKER	UNP P12497
D	227	GLU	-	LINKER	UNP P12497
D	228	LEU	-	LINKER	UNP P12497
D	332	TYR	GLU	ENGINEERED	UNP P734077
D	338	GLU	-	EXPRESSION TAG	UNP P734077
D	339	VAL	-	EXPRESSION TAG	UNP P734077
D	340	LEU	-	EXPRESSION TAG	UNP P734077
D	341	PHE	-	EXPRESSION TAG	UNP P734077
D	342	GLN	-	EXPRESSION TAG	UNP P734077
E	184	ALA	TRP	ENGINEERED	UNP P12497
E	185	ALA	MET	ENGINEERED	UNP P12497
E	224	THR	-	LINKER	UNP P12497
E	225	ARG	-	LINKER	UNP P12497
E	226	PRO	-	LINKER	UNP P12497
E	227	GLU	-	LINKER	UNP P12497

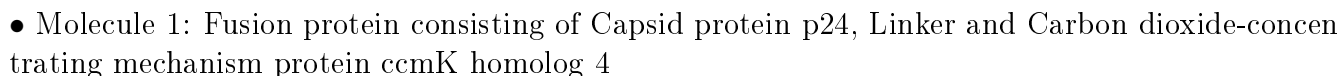
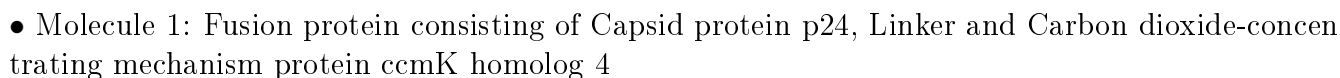
*Continued on next page...*

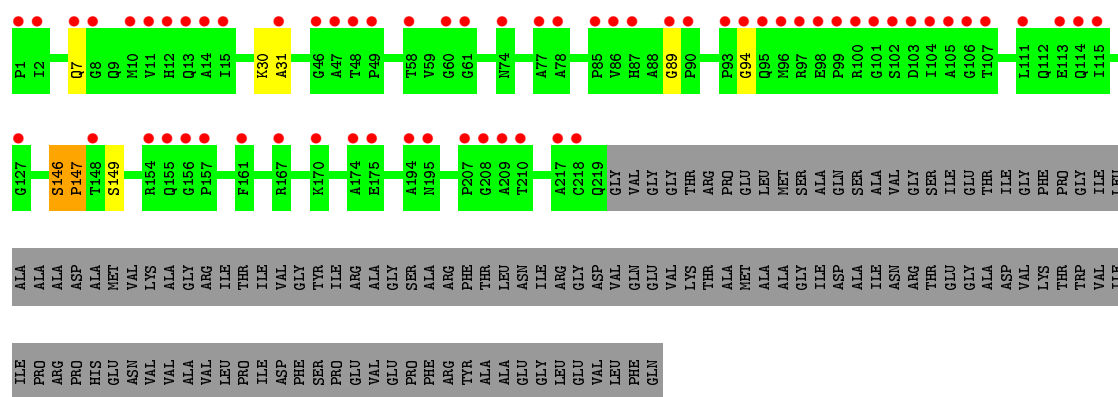
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	228	LEU	-	LINKER	UNP P12497
E	332	TYR	GLU	ENGINEERED	UNP P734077
E	338	GLU	-	EXPRESSION TAG	UNP P734077
E	339	VAL	-	EXPRESSION TAG	UNP P734077
E	340	LEU	-	EXPRESSION TAG	UNP P734077
E	341	PHE	-	EXPRESSION TAG	UNP P734077
E	342	GLN	-	EXPRESSION TAG	UNP P734077
F	184	ALA	TRP	ENGINEERED	UNP P12497
F	185	ALA	MET	ENGINEERED	UNP P12497
F	224	THR	-	LINKER	UNP P12497
F	225	ARG	-	LINKER	UNP P12497
F	226	PRO	-	LINKER	UNP P12497
F	227	GLU	-	LINKER	UNP P12497
F	228	LEU	-	LINKER	UNP P12497
F	332	TYR	GLU	ENGINEERED	UNP P734077
F	338	GLU	-	EXPRESSION TAG	UNP P734077
F	339	VAL	-	EXPRESSION TAG	UNP P734077
F	340	LEU	-	EXPRESSION TAG	UNP P734077
F	341	PHE	-	EXPRESSION TAG	UNP P734077
F	342	GLN	-	EXPRESSION TAG	UNP P734077

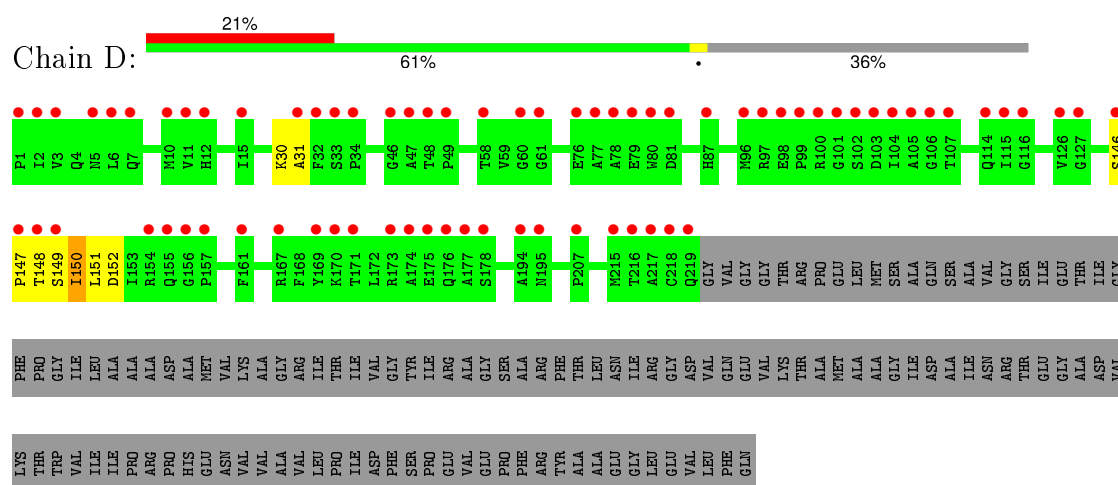


- Molecule 1: Fusion protein consisting of Capsid protein p24, Linker and Carbon dioxide-concentrating mechanism protein ccmK homolog 4

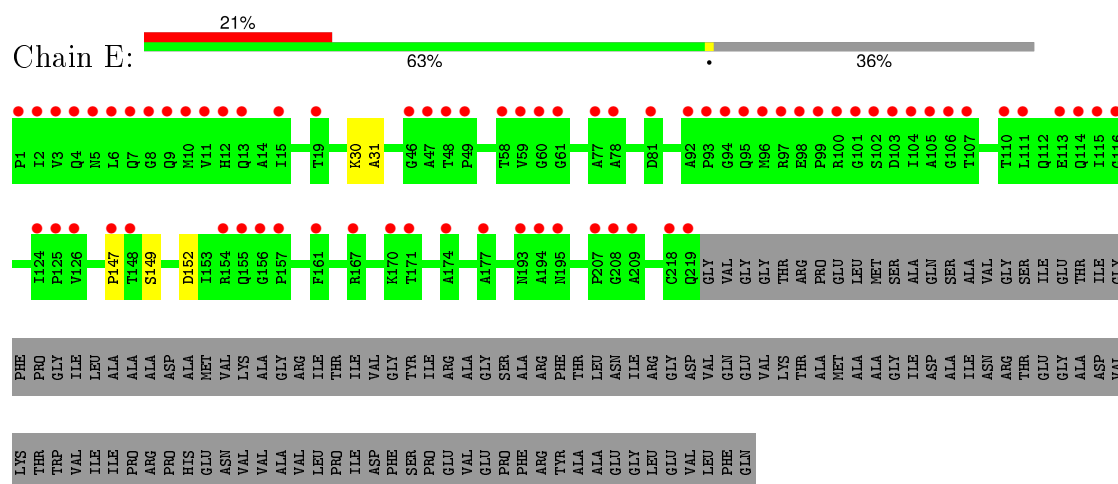




- Molecule 1: Fusion protein consisting of Capsid protein p24, Linker and Carbon dioxide-concentrating mechanism protein ccmK homolog 4



- Molecule 1: Fusion protein consisting of Capsid protein p24, Linker and Carbon dioxide-concentrating mechanism protein ccmK homolog 4



- Molecule 1: Fusion protein consisting of Capsid protein p24, Linker and Carbon dioxide-concentrating mechanism protein ccmK homolog 4







## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.62Å 156.44Å 196.64Å 90.00° 100.35° 90.00°	Depositor
Resolution (Å)	49.75 – 7.00 49.75 – 6.95	Depositor EDS
% Data completeness (in resolution range)	92.7 (49.75-7.00) 81.2 (49.75-6.95)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.10 (at 6.68Å)	Xtriage
Refinement program	PHENIX (phenix.refine)	Depositor
R, $R_{free}$	0.282 , 0.323 0.301 , 0.345	Depositor DCC
$R_{free}$ test set	409 reflections (10.08%)	DCC
Wilson B-factor (Å <sup>2</sup> )	161.2	Xtriage
Anisotropy	1.361	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 341.8	EDS
Estimated twinning fraction	0.116 for h,-k,-h-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	2 of 4325 reflections (0.046%)	Xtriage
$F_o, F_c$ correlation	0.71	EDS
Total number of atoms	5256	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.99% 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, 0.375 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	0.57	0/875	1.02	0/1092
1	B	0.56	0/875	1.02	0/1092
1	C	0.57	0/875	1.02	0/1092
1	D	0.57	0/875	1.03	0/1092
1	E	0.57	0/875	1.03	0/1092
1	F	0.57	0/875	1.03	0/1092
All	All	0.57	0/5250	1.02	0/6552

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

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	876	0	230	7	0
1	B	876	0	230	1	0
1	C	876	0	230	3	2
1	D	876	0	230	5	0
1	E	876	0	230	2	0
1	F	876	0	230	4	0
All	All	5256	0	1380	22	2

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 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:A:147:PRO:CA	1:A:149:SER:N	2.40	0.84
1:F:147:PRO:O	1:F:148:THR:O	2.02	0.78
1:D:149:SER:O	1:D:151:LEU:N	2.24	0.69
1:A:143:ARG:C	1:A:145:TYR:H	1.96	0.68
1:A:147:PRO:CA	1:A:149:SER:H	2.13	0.62

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:94:GLY:O	1:C:94:GLY:O[2_756]	1.21	0.99
1:C:7:GLN:O	1:C:89:GLY:CA[2_756]	1.82	0.38

## 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	217/342 (64%)	204 (94%)	8 (4%)	5 (2%)	8	48
1	B	217/342 (64%)	207 (95%)	9 (4%)	1 (0%)	34	77
1	C	217/342 (64%)	206 (95%)	8 (4%)	3 (1%)	14	58
1	D	217/342 (64%)	205 (94%)	8 (4%)	4 (2%)	11	53
1	E	217/342 (64%)	206 (95%)	10 (5%)	1 (0%)	34	77
1	F	217/342 (64%)	206 (95%)	7 (3%)	4 (2%)	11	53
All	All	1302/2052 (64%)	1234 (95%)	50 (4%)	18 (1%)	14	58

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	144	MET
1	A	147	PRO
1	C	146	SER
1	C	147	PRO
1	C	149	SER

### 5.3.2 Protein sidechains [i](#)

There are no protein residues with a non-rotameric sidechain to report in this entry.

### 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 [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	219/342 (64%)	1.47	61 (27%) <b>1</b> <b>5</b>	19, 33, 81, 100	0
1	B	219/342 (64%)	1.39	59 (26%) <b>1</b> <b>5</b>	19, 33, 81, 100	0
1	C	219/342 (64%)	1.64	64 (29%) <b>1</b> <b>5</b>	19, 33, 81, 100	0
1	D	219/342 (64%)	1.60	72 (32%) <b>0</b> <b>5</b>	19, 33, 81, 100	0
1	E	219/342 (64%)	1.59	71 (32%) <b>1</b> <b>5</b>	19, 33, 81, 100	0
1	F	219/342 (64%)	1.78	79 (36%) <b>0</b> <b>4</b>	19, 33, 81, 100	0
All	All	1314/2052 (64%)	1.58	406 (30%) <b>1</b> <b>5</b>	19, 33, 81, 100	0

The worst 5 of 406 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	94	GLY	12.8
1	A	99	PRO	11.5
1	B	148	THR	9.2
1	A	98	GLU	9.2
1	F	7	GLN	9.1

### 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.