



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 15, 2018 – 07:49 PM EST

PDB ID : 3OKJ  
Title : Alpha-keto-aldehyde binding mechanism reveals a novel lead structure motif for proteasome inhibition  
Authors : Groll, M.; Poynor, M.; Gallastegui, P.; Stein, M.; Schmidt, B.; Kloetzel, P.M.; Huber, R.  
Deposited on : 2010-08-25  
Resolution : 2.70 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	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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-20030736

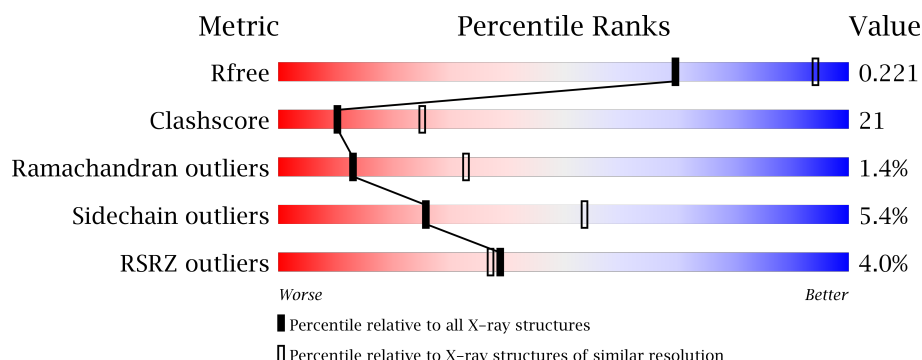
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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.



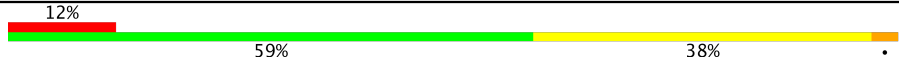
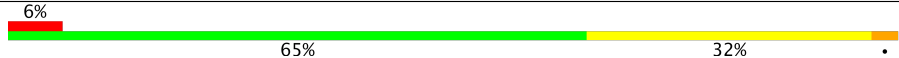







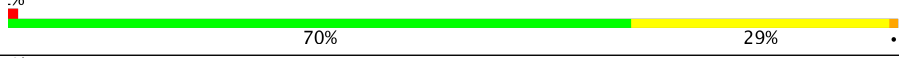

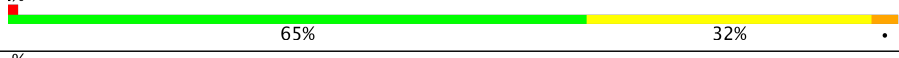
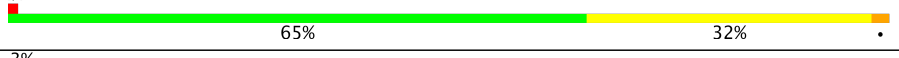

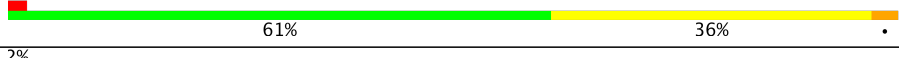
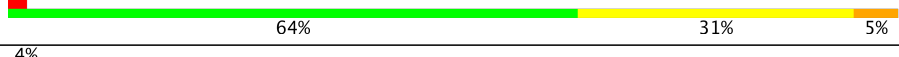







Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	2259 (2.70-2.70)
Clashscore	112137	2590 (2.70-2.70)
Ramachandran outliers	110173	2550 (2.70-2.70)
Sidechain outliers	110143	2550 (2.70-2.70)
RSRZ outliers	101464	2275 (2.70-2.70)

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	250	<div> <div>3%</div> <div>71%</div> <div>28%</div> <div>.</div> </div>
1	O	250	<div> <div>4%</div> <div>69%</div> <div>29%</div> <div>.</div> </div>
2	B	244	<div> <div>5%</div> <div>61%</div> <div>33%</div> <div>5%</div> </div>
2	P	244	<div> <div>6%</div> <div>60%</div> <div>34%</div> <div>5%</div> </div>
3	C	241	<div> <div>6%</div> <div>60%</div> <div>37%</div> <div>.</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	Q	241	
4	D	242	
4	R	242	
5	E	233	
5	S	233	
6	F	244	
6	T	244	
7	G	243	
7	U	243	
8	H	222	
8	V	222	
9	I	204	
9	W	204	
10	J	198	
10	X	198	
11	K	212	
11	Y	212	
12	L	222	
12	Z	222	
13	1	233	
13	M	233	
14	2	196	
14	N	196	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
15	EP9	K	300	X	-	-	-
15	EP9	Y	300	X	-	-	X

## 2 Entry composition

There are 16 unique types of molecules in this entry. The entry contains 50851 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 Proteasome component Y7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	250	Total	C	N	O	S	0	0	0
			1915	1219	315	377	4			
1	O	250	Total	C	N	O	S	0	0	0
			1915	1219	315	377	4			

- Molecule 2 is a protein called Proteasome component Y13.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	244	Total	C	N	O	S	0	0	0
			1905	1201	321	380	3			
2	P	244	Total	C	N	O	S	0	0	0
			1905	1201	321	380	3			

- Molecule 3 is a protein called Proteasome component PRE6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	241	Total	C	N	O	S	0	0	0
			1891	1181	331	375	4			
3	Q	241	Total	C	N	O	S	0	0	0
			1891	1181	331	375	4			

- Molecule 4 is a protein called Proteasome component PUP2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	242	Total	C	N	O	S	0	0	0
			1862	1162	314	379	7			
4	R	242	Total	C	N	O	S	0	0	0
			1862	1162	314	379	7			

- Molecule 5 is a protein called Proteasome component PRE5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	233	Total	C	N	O	S	0	0	0
			1795	1129	312	350	4			
5	S	233	Total	C	N	O	S	0	0	0
			1795	1129	312	350	4			

- Molecule 6 is a protein called Proteasome component C1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	244	Total	C	N	O	S	0	0	0
			1897	1205	330	358	4			
6	T	244	Total	C	N	O	S	0	0	0
			1897	1205	330	358	4			

- Molecule 7 is a protein called Proteasome component C7-alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	243	Total	C	N	O	S	0	0	0
			1921	1221	322	370	8			
7	U	243	Total	C	N	O	S	0	0	0
			1921	1221	322	370	8			

- Molecule 8 is a protein called Proteasome component PUP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	H	222	Total	C	N	O	S	0	0	0
			1685	1061	293	324	7			
8	V	222	Total	C	N	O	S	0	0	0
			1685	1061	293	324	7			

- Molecule 9 is a protein called Proteasome component PUP3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	I	204	Total	C	N	O	S	0	0	0
			1581	1010	258	305	8			
9	W	204	Total	C	N	O	S	0	0	0
			1581	1010	258	305	8			

- Molecule 10 is a protein called Proteasome component C11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	J	198	Total	C	N	O	S	0	0	0
			1585	1005	269	305	6			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	X	198	Total	C	N	O	S	0	0	0
			1585	1005	269	305	6			

- Molecule 11 is a protein called Proteasome component PRE2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	K	212	Total	C	N	O	S	0	0	0
			1644	1045	280	312	7			
11	Y	212	Total	C	N	O	S	0	0	0
			1644	1045	280	312	7			

- Molecule 12 is a protein called Proteasome component C5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	L	222	Total	C	N	O	S	0	0	0
			1757	1115	303	335	4			
12	Z	222	Total	C	N	O	S	0	0	0
			1731	1101	295	331	4			

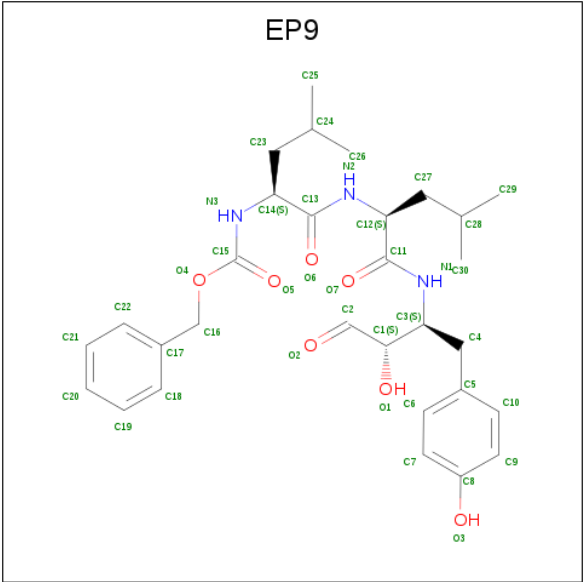
- Molecule 13 is a protein called Proteasome component PRE4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
13	M	233	Total	C	N	O	S	0	0	0
			1824	1154	312	351	7			
13	1	233	Total	C	N	O	S	0	0	0
			1807	1143	308	349	7			

- Molecule 14 is a protein called Proteasome component PRE3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
14	N	196	Total	C	N	O	S	0	0	0
			1512	955	250	300	7			
14	2	196	Total	C	N	O	S	0	0	0
			1487	941	245	294	7			

- Molecule 15 is N-[(benzyloxy)carbonyl]-L-leucyl-N-[(2S,3S)-3-hydroxy-1-(4-hydroxyphenyl)-4-oxobutan-2-yl]-L-leucinamide (three-letter code: EP9) (formula: C<sub>30</sub>H<sub>41</sub>N<sub>3</sub>O<sub>7</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
15	K	1	Total	C	N	O	0	0
			39	30	3	6		
15	Y	1	Total	C	N	O	0	0
			39	30	3	6		

- Molecule 16 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
16	A	52	Total	O	0	0
			52	52		
16	B	38	Total	O	0	0
			38	38		
16	C	39	Total	O	0	0
			39	39		
16	D	42	Total	O	0	0
			42	42		
16	E	20	Total	O	0	0
			20	20		
16	F	47	Total	O	0	0
			47	47		
16	G	60	Total	O	0	0
			60	60		
16	H	49	Total	O	0	0
			49	49		
16	I	62	Total	O	0	0
			62	62		
16	J	49	Total	O	0	0
			49	49		

Continued on next page...

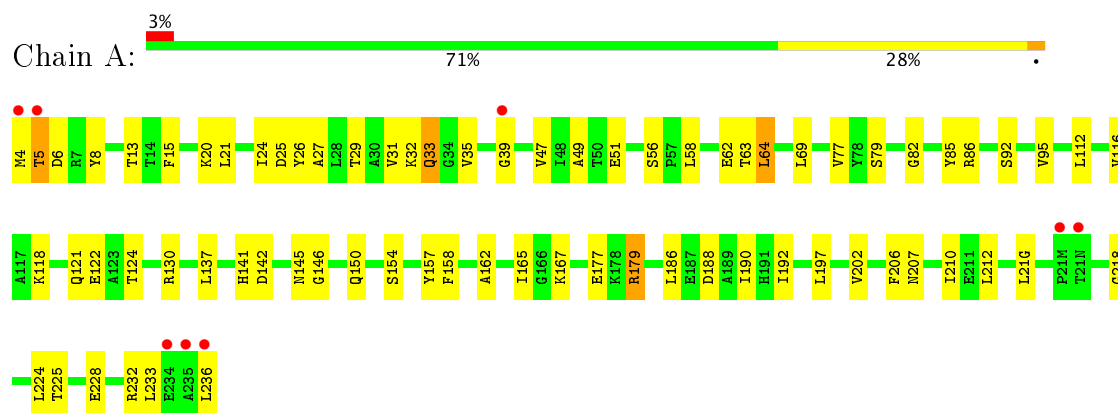
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
16	K	44	Total 44	O 44	0	0
16	L	57	Total 57	O 57	0	0
16	M	68	Total 68	O 68	0	0
16	N	58	Total 58	O 58	0	0
16	O	34	Total 34	O 34	0	0
16	P	28	Total 28	O 28	0	0
16	Q	28	Total 28	O 28	0	0
16	R	32	Total 32	O 32	0	0
16	S	18	Total 18	O 18	0	0
16	T	40	Total 40	O 40	0	0
16	U	61	Total 61	O 61	0	0
16	V	44	Total 44	O 44	0	0
16	W	58	Total 58	O 58	0	0
16	X	41	Total 41	O 41	0	0
16	Y	47	Total 47	O 47	0	0
16	Z	47	Total 47	O 47	0	0
16	1	74	Total 74	O 74	0	0
16	2	56	Total 56	O 56	0	0

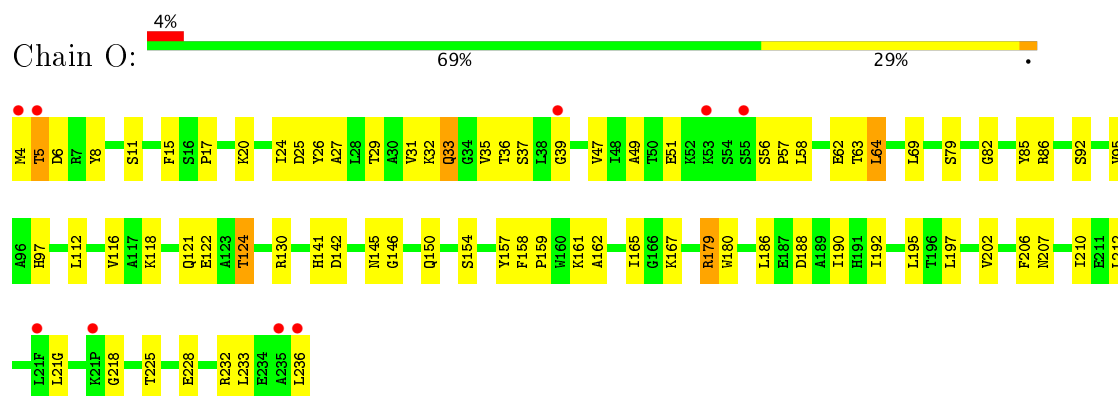
### 3 Residue-property plots [i](#)

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.

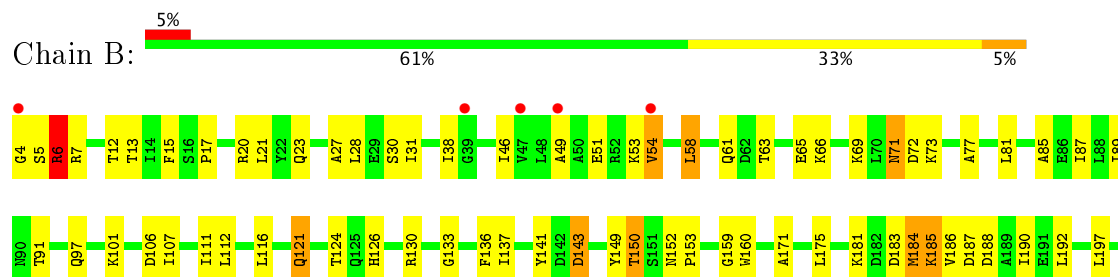
#### • Molecule 1: Proteasome component Y7

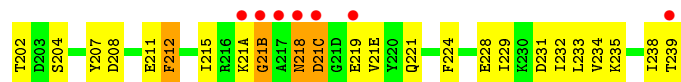


#### • Molecule 1: Proteasome component Y7

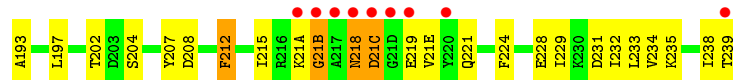
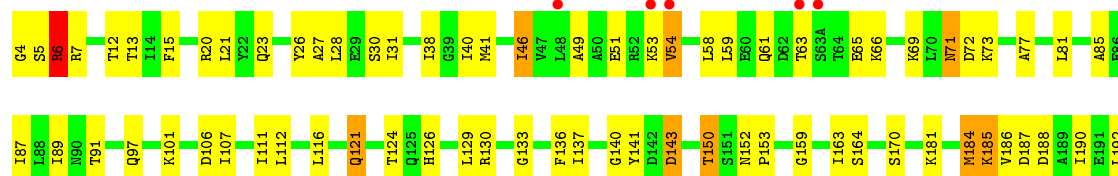


#### • Molecule 2: Proteasome component Y13

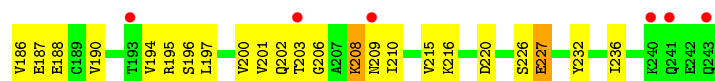
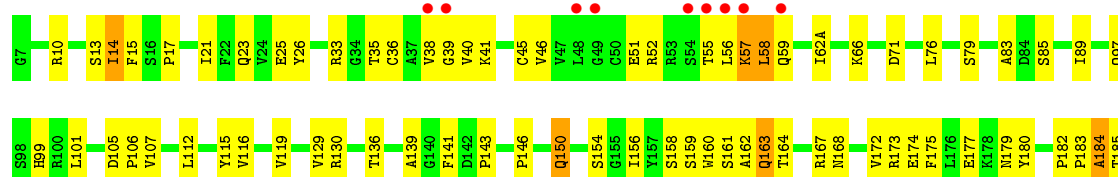




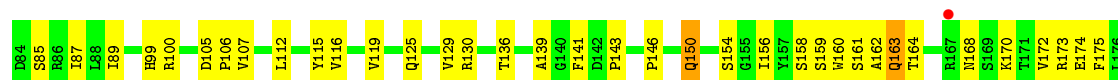
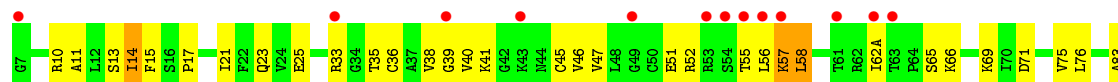
• Molecule 2: Proteasome component Y13



• Molecule 3: Proteasome component PRE6

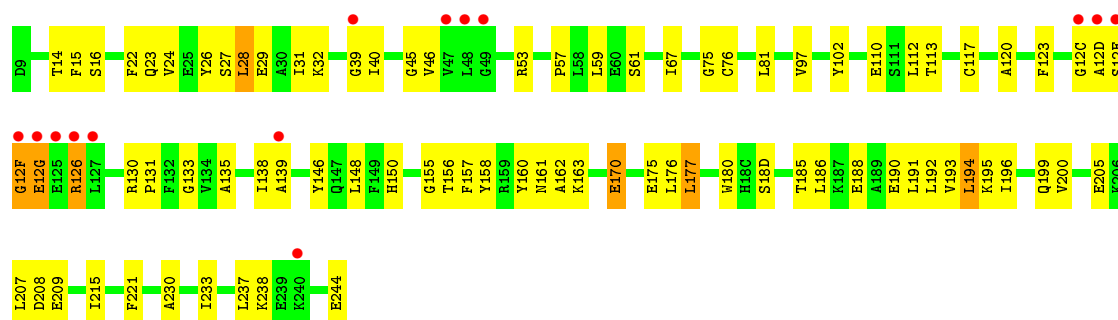


• Molecule 3: Proteasome component PRE6

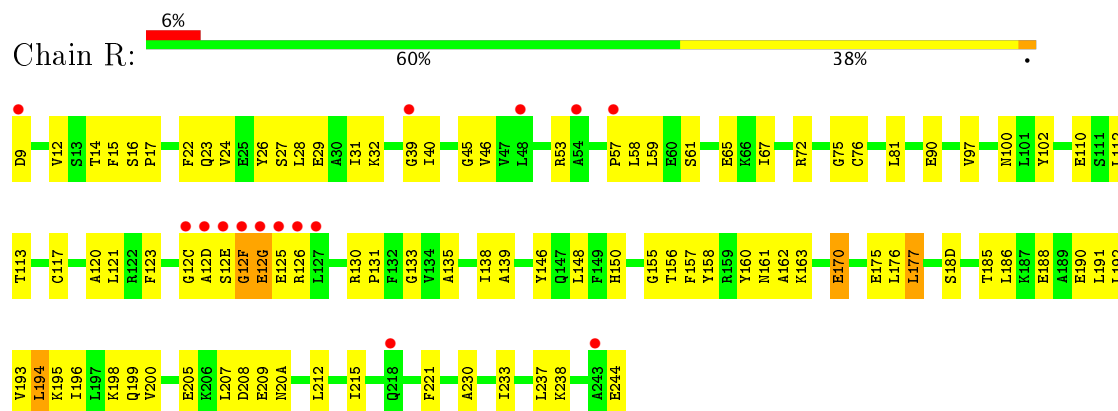


• Molecule 4: Proteasome component PUP2

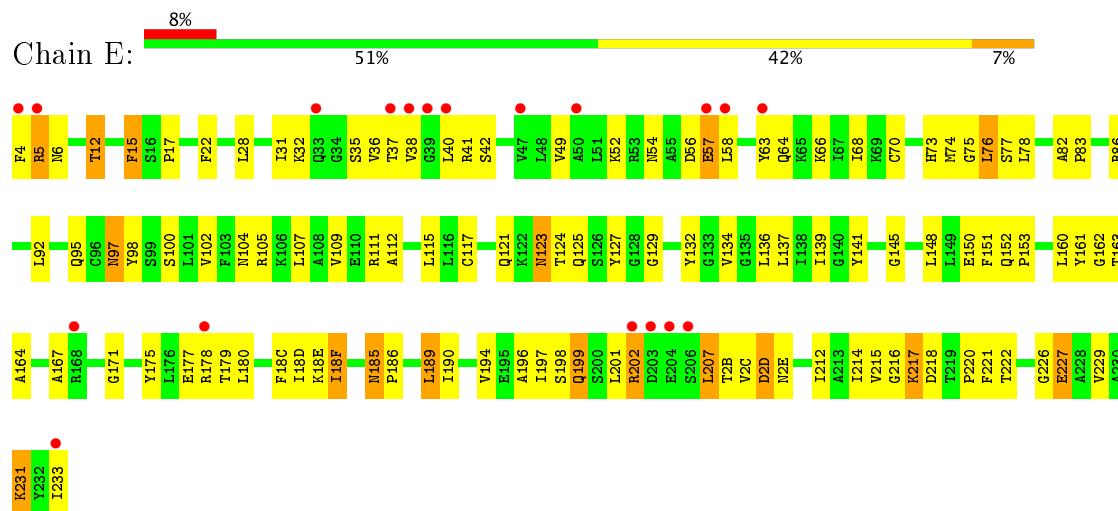




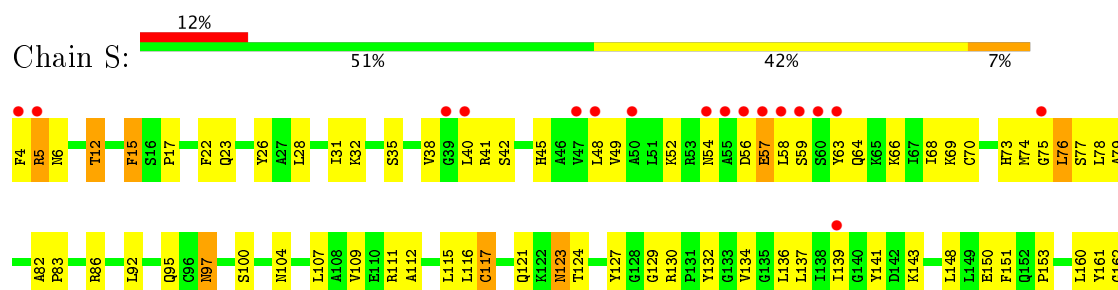
• Molecule 4: Proteasome component PUP2

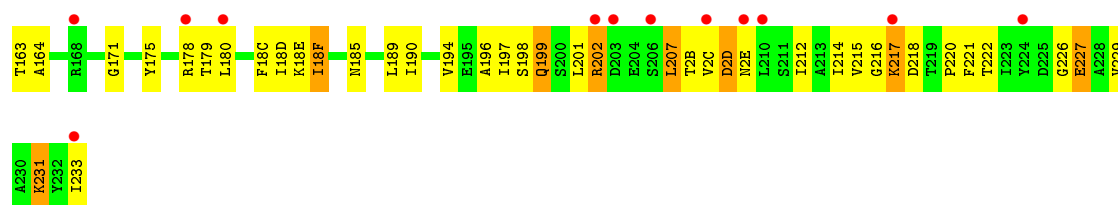


• Molecule 5: Proteasome component PRE5

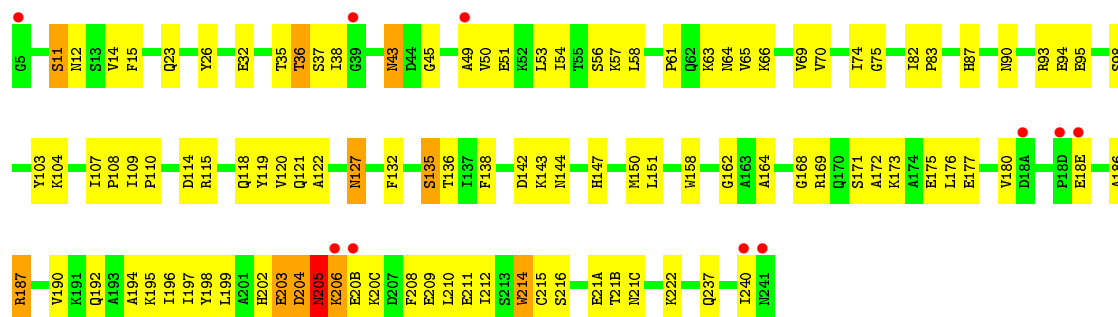


• Molecule 5: Proteasome component PRE5

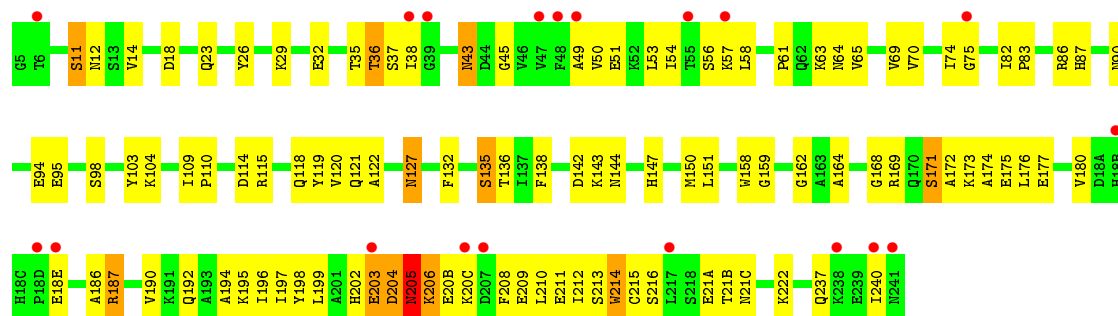




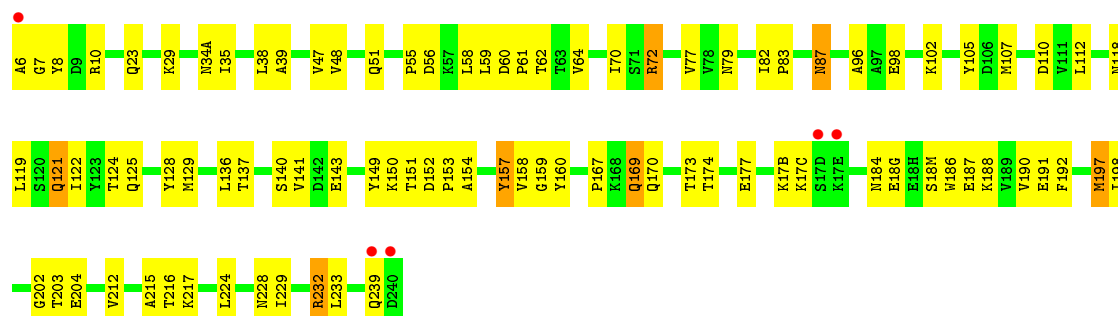
• Molecule 6: Proteasome component C1



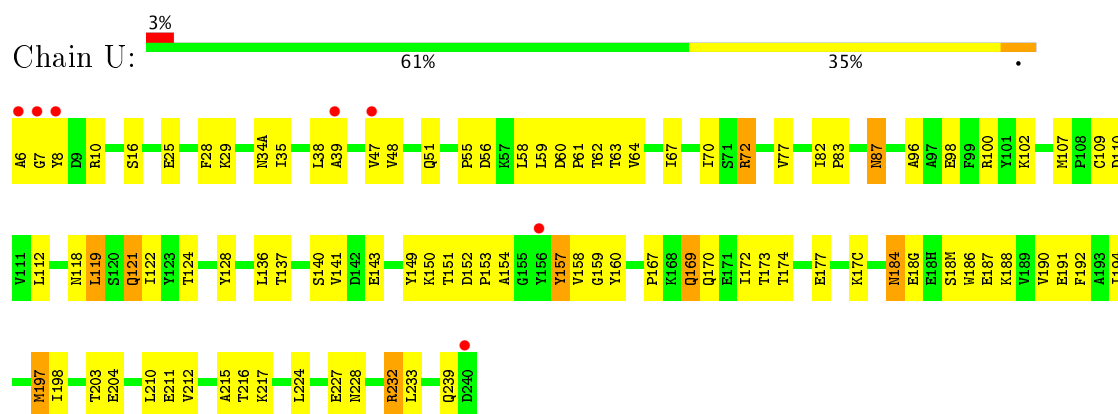
• Molecule 6: Proteasome component C1



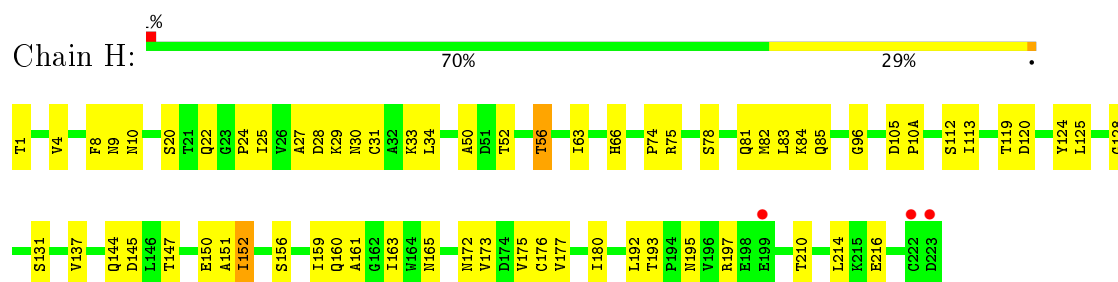
• Molecule 7: Proteasome component C7-alpha



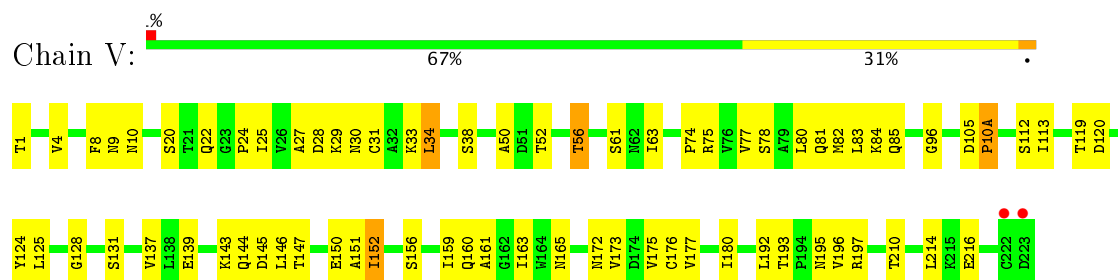
• Molecule 7: Proteasome component C7-alpha



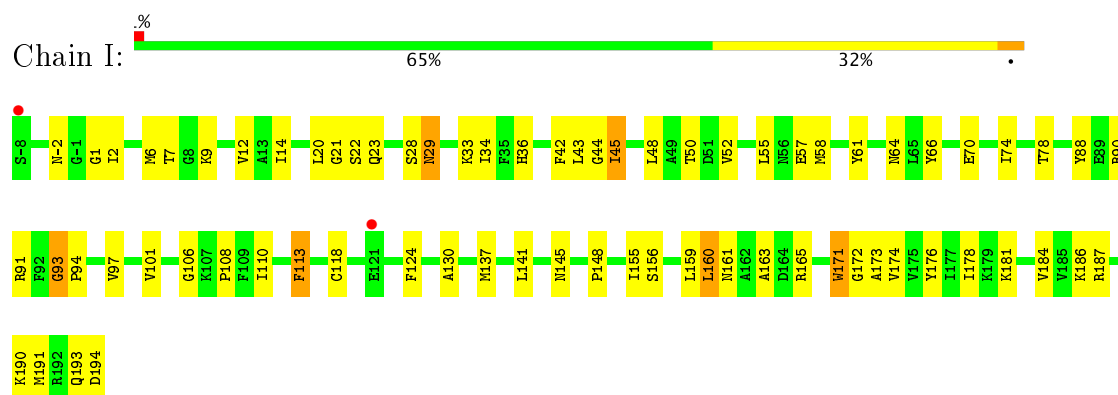
• Molecule 8: Proteasome component PUP1



• Molecule 8: Proteasome component PUP1

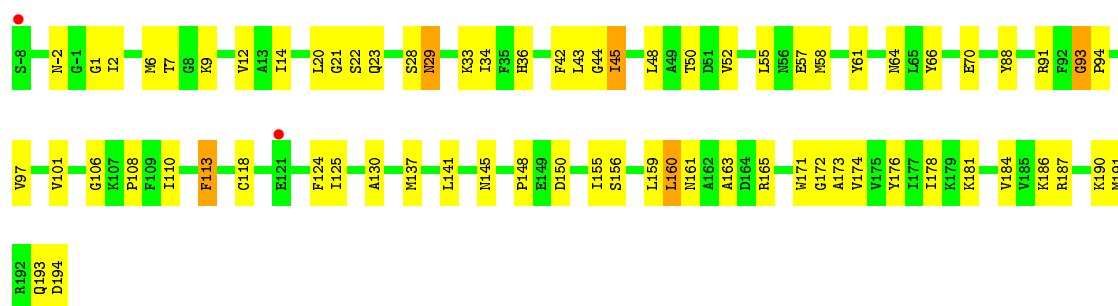


• Molecule 9: Proteasome component PUP3

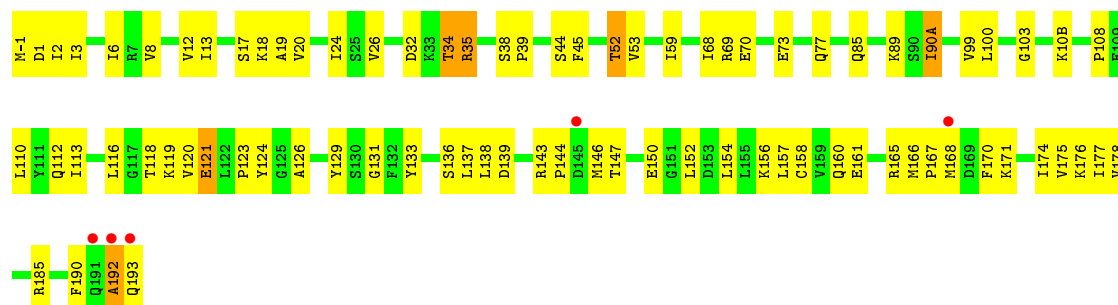


• Molecule 9: Proteasome component PUP3

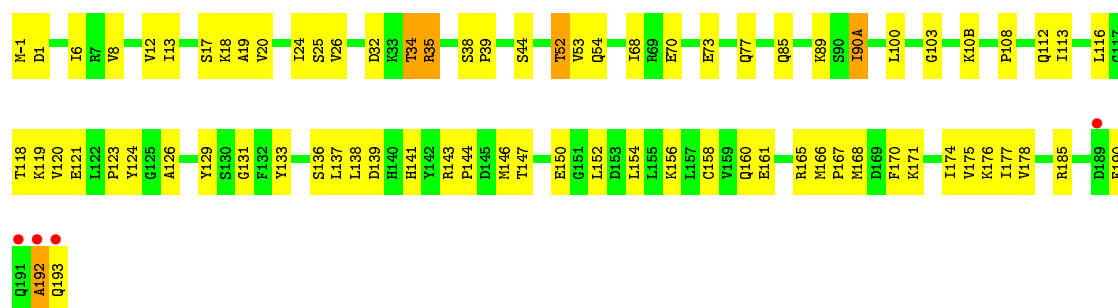




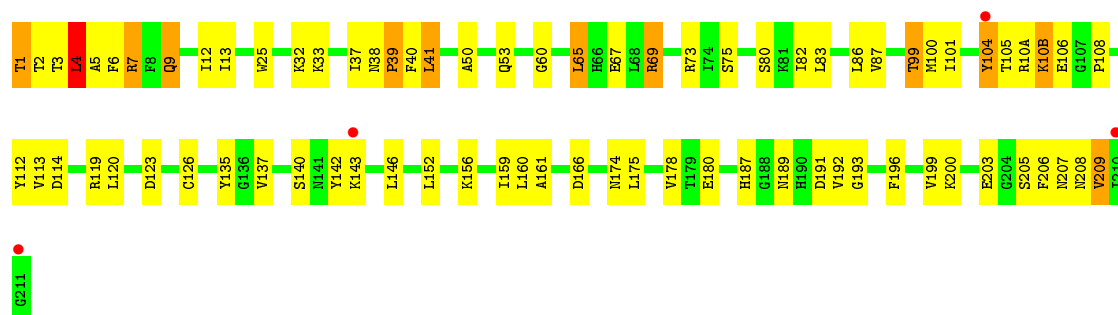
• Molecule 10: Proteasome component C11



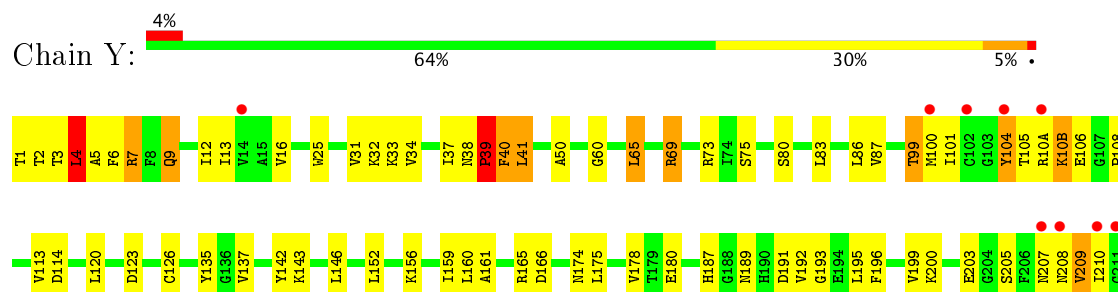
• Molecule 10: Proteasome component C11



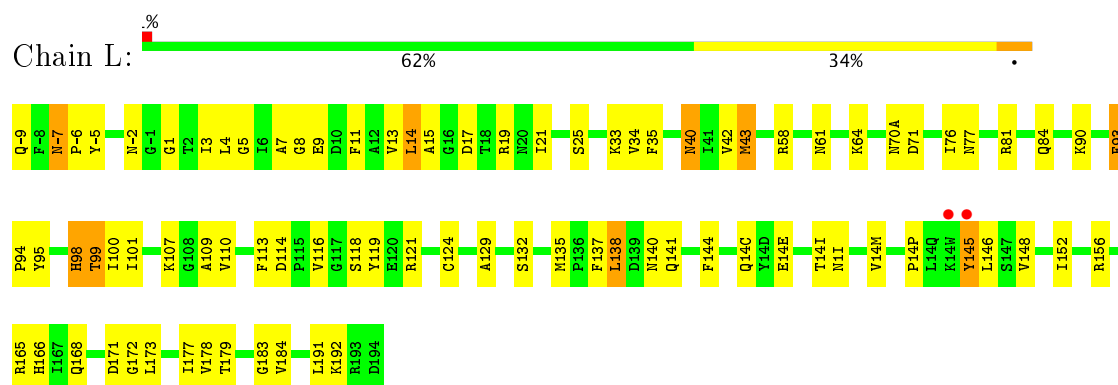
• Molecule 11: Proteasome component PRE2



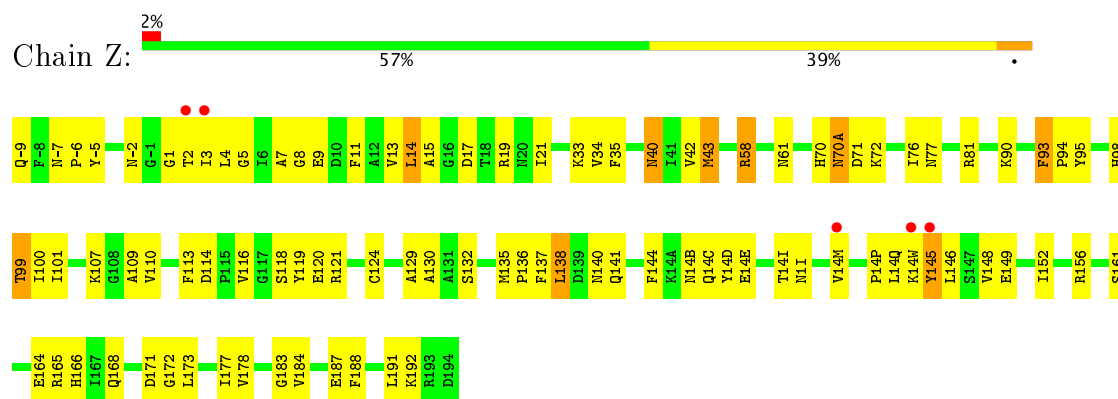
- Molecule 11: Proteasome component PRE2



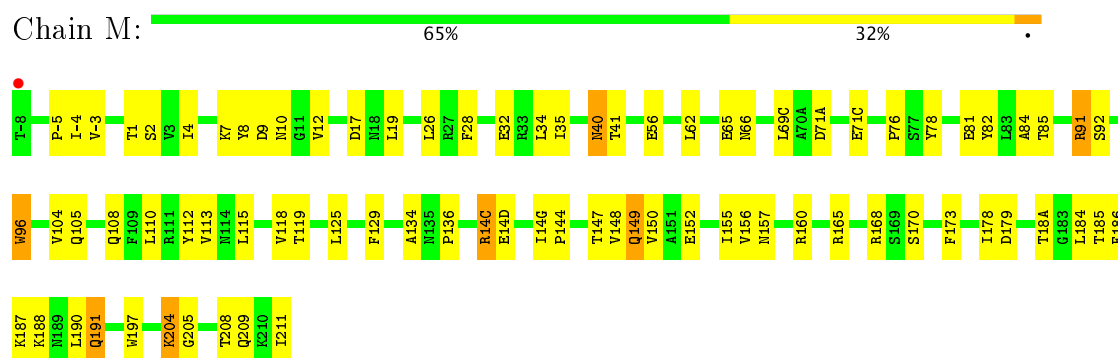
- Molecule 12: Proteasome component C5



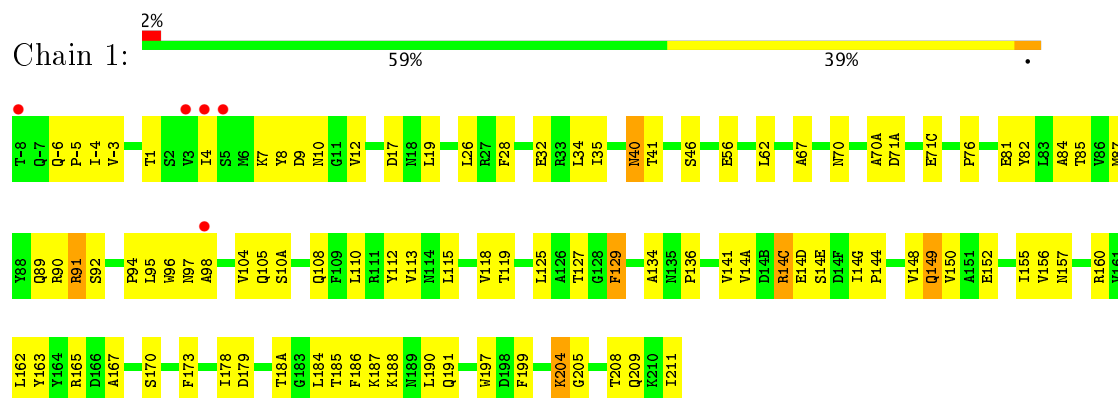
- Molecule 12: Proteasome component C5



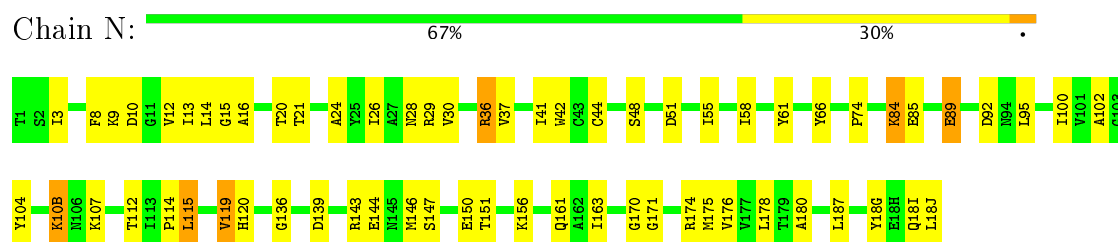
- Molecule 13: Proteasome component PRE4



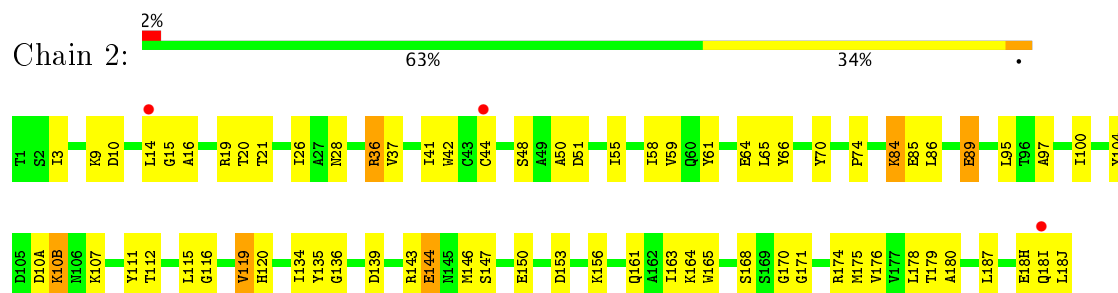
• Molecule 13: Proteasome component PRE4



• Molecule 14: Proteasome component PRE3



• Molecule 14: Proteasome component PRE3



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	134.59Å 299.83Å 143.60Å 90.00° 112.60° 90.00°	Depositor
Resolution (Å)	15.00 – 2.70 24.88 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.6 (15.00-2.70) 99.6 (24.88-2.70)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.14 (at 2.72Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.232 , 0.259 0.217 , 0.221	Depositor DCC
$R_{free}$ test set	14203 reflections (5.26%)	DCC
Wilson B-factor (Å <sup>2</sup> )	50.6	Xtriage
Anisotropy	0.890	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 73.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	50851	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.06% 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: EP9

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	0.37	0/1952	0.63	0/2642
1	O	0.37	0/1952	0.63	0/2642
2	B	0.37	0/1935	0.63	0/2618
2	P	0.37	0/1935	0.63	0/2618
3	C	0.35	0/1920	0.62	0/2598
3	Q	0.37	0/1920	0.62	0/2598
4	D	0.37	0/1887	0.63	0/2541
4	R	0.37	0/1887	0.63	0/2541
5	E	0.35	0/1823	0.61	0/2463
5	S	0.37	0/1823	0.61	0/2463
6	F	0.39	0/1937	0.61	1/2614 (0.0%)
6	T	0.39	0/1937	0.62	1/2614 (0.0%)
7	G	0.39	0/1959	0.63	0/2652
7	U	0.39	0/1959	0.63	0/2652
8	H	0.37	0/1716	0.65	0/2326
8	V	0.37	0/1716	0.65	0/2326
9	I	0.39	0/1611	0.67	0/2174
9	W	0.41	0/1611	0.67	0/2174
10	J	0.41	0/1613	0.65	0/2173
10	X	0.40	0/1613	0.65	0/2173
11	K	0.45	0/1681	0.69	2/2274 (0.1%)
11	Y	0.44	0/1681	0.70	1/2274 (0.0%)
12	L	0.39	0/1795	0.68	0/2420
12	Z	0.39	0/1747	0.69	0/2328
13	1	0.39	0/1816	0.65	0/2430
13	M	0.39	0/1855	0.66	0/2514
14	2	0.39	0/1495	0.64	0/1996
14	N	0.40	0/1541	0.64	0/2087
All	All	0.39	0/50317	0.64	5/67925 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
11	Y	4	LEU	CA-CB-CG	5.75	128.54	115.30
11	K	4	LEU	CA-CB-CG	5.73	128.49	115.30
6	F	135	SER	N-CA-C	-5.18	97.01	111.00
6	T	135	SER	N-CA-C	-5.16	97.07	111.00
11	K	1	THR	CB-CA-C	-5.01	98.06	111.60

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	1915	0	1926	58	0
1	O	1915	0	1926	59	0
2	B	1905	0	1901	93	0
2	P	1905	0	1901	92	0
3	C	1891	0	1900	96	0
3	Q	1891	0	1900	97	0
4	D	1862	0	1836	63	0
4	R	1862	0	1836	86	0
5	E	1795	0	1797	106	0
5	S	1795	0	1797	122	0
6	F	1897	0	1886	83	0
6	T	1897	0	1886	81	0
7	G	1921	0	1910	82	0
7	U	1921	0	1910	86	0
8	H	1685	0	1688	46	0
8	V	1685	0	1688	52	0
9	I	1581	0	1574	67	0
9	W	1581	0	1574	63	0
10	J	1585	0	1590	100	0
10	X	1585	0	1590	94	0
11	K	1644	0	1592	100	0
11	Y	1644	0	1592	94	0
12	L	1757	0	1711	76	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
12	Z	1731	0	1642	99	0
13	1	1807	0	1785	104	0
13	M	1824	0	1832	69	0
14	2	1487	0	1411	75	0
14	N	1512	0	1481	54	0
15	K	39	0	39	1	0
15	Y	39	0	39	2	0
16	1	74	0	0	6	0
16	2	56	0	0	5	0
16	A	52	0	0	0	0
16	B	38	0	0	3	0
16	C	39	0	0	3	0
16	D	42	0	0	0	0
16	E	20	0	0	4	0
16	F	47	0	0	6	0
16	G	60	0	0	5	0
16	H	49	0	0	1	0
16	I	62	0	0	3	0
16	J	49	0	0	11	0
16	K	44	0	0	6	0
16	L	57	0	0	3	0
16	M	68	0	0	5	0
16	N	58	0	0	4	0
16	O	34	0	0	0	0
16	P	28	0	0	3	0
16	Q	28	0	0	5	0
16	R	32	0	0	7	0
16	S	18	0	0	4	0
16	T	40	0	0	2	0
16	U	61	0	0	4	0
16	V	44	0	0	4	0
16	W	58	0	0	3	0
16	X	41	0	0	3	0
16	Y	47	0	0	4	0
16	Z	47	0	0	1	0
All	All	50851	0	49140	2057	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 21.

The worst 5 of 2057 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)
12:L:43:MET:HB2	12:L:101:ILE:HG22	1.31	1.08
12:Z:43:MET:HB2	12:Z:101:ILE:HG22	1.32	1.08
11:Y:10(B):LYS:HD2	11:Y:10(B):LYS:H	1.10	1.08
11:K:10(B):LYS:H	11:K:10(B):LYS:HD2	1.08	1.06
14:2:135:TYR:CZ	14:2:135:TYR:CD1	2.41	1.03

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	248/250 (99%)	228 (92%)	16 (6%)	4 (2%)	11	28
1	O	248/250 (99%)	228 (92%)	16 (6%)	4 (2%)	11	28
2	B	242/244 (99%)	217 (90%)	19 (8%)	6 (2%)	6	17
2	P	242/244 (99%)	219 (90%)	17 (7%)	6 (2%)	6	17
3	C	239/241 (99%)	219 (92%)	15 (6%)	5 (2%)	8	21
3	Q	239/241 (99%)	220 (92%)	15 (6%)	4 (2%)	11	27
4	D	240/242 (99%)	220 (92%)	13 (5%)	7 (3%)	5	13
4	R	240/242 (99%)	219 (91%)	14 (6%)	7 (3%)	5	13
5	E	231/233 (99%)	203 (88%)	21 (9%)	7 (3%)	5	12
5	S	231/233 (99%)	201 (87%)	23 (10%)	7 (3%)	5	12
6	F	242/244 (99%)	222 (92%)	16 (7%)	4 (2%)	11	27
6	T	242/244 (99%)	222 (92%)	16 (7%)	4 (2%)	11	27
7	G	241/243 (99%)	227 (94%)	12 (5%)	2 (1%)	22	49
7	U	241/243 (99%)	228 (95%)	10 (4%)	3 (1%)	15	37
8	H	220/222 (99%)	206 (94%)	12 (6%)	2 (1%)	20	46
8	V	220/222 (99%)	206 (94%)	11 (5%)	3 (1%)	13	33

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
9	I	202/204 (99%)	191 (95%)	10 (5%)	1 (0%)	32 60
9	W	202/204 (99%)	190 (94%)	11 (5%)	1 (0%)	32 60
10	J	196/198 (99%)	183 (93%)	11 (6%)	2 (1%)	18 43
10	X	196/198 (99%)	184 (94%)	10 (5%)	2 (1%)	18 43
11	K	210/212 (99%)	196 (93%)	13 (6%)	1 (0%)	32 60
11	Y	210/212 (99%)	195 (93%)	13 (6%)	2 (1%)	18 43
12	L	220/222 (99%)	205 (93%)	13 (6%)	2 (1%)	20 46
12	Z	207/222 (93%)	195 (94%)	10 (5%)	2 (1%)	18 43
13	1	212/233 (91%)	197 (93%)	15 (7%)	0	100 100
13	M	231/233 (99%)	213 (92%)	17 (7%)	1 (0%)	38 66
14	2	183/196 (93%)	171 (93%)	12 (7%)	0	100 100
14	N	194/196 (99%)	183 (94%)	11 (6%)	0	100 100
All	All	6269/6368 (98%)	5788 (92%)	392 (6%)	89 (1%)	13 33

5 of 89 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	54	VAL
2	B	184	MET
2	B	21(C)	ASP
3	C	58	LEU
4	D	12(G)	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	209/209 (100%)	204 (98%)	5 (2%)	54 83
1	O	209/209 (100%)	203 (97%)	6 (3%)	48 77
2	B	203/203 (100%)	192 (95%)	11 (5%)	26 54
2	P	203/203 (100%)	192 (95%)	11 (5%)	26 54

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	C	213/213 (100%)	203 (95%)	10 (5%)	30	60
3	Q	213/213 (100%)	203 (95%)	10 (5%)	30	60
4	D	198/198 (100%)	187 (94%)	11 (6%)	25	51
4	R	198/198 (100%)	187 (94%)	11 (6%)	25	51
5	E	192/192 (100%)	174 (91%)	18 (9%)	10	23
5	S	192/192 (100%)	175 (91%)	17 (9%)	11	26
6	F	201/201 (100%)	186 (92%)	15 (8%)	16	36
6	T	201/201 (100%)	186 (92%)	15 (8%)	16	36
7	G	207/207 (100%)	196 (95%)	11 (5%)	26	54
7	U	207/207 (100%)	196 (95%)	11 (5%)	26	54
8	H	181/181 (100%)	176 (97%)	5 (3%)	49	79
8	V	181/181 (100%)	176 (97%)	5 (3%)	49	79
9	I	172/172 (100%)	165 (96%)	7 (4%)	35	66
9	W	172/172 (100%)	165 (96%)	7 (4%)	35	66
10	J	175/175 (100%)	166 (95%)	9 (5%)	28	56
10	X	175/175 (100%)	166 (95%)	9 (5%)	28	56
11	K	169/169 (100%)	157 (93%)	12 (7%)	17	39
11	Y	169/169 (100%)	156 (92%)	13 (8%)	15	34
12	L	185/185 (100%)	174 (94%)	11 (6%)	23	49
12	Z	171/185 (92%)	161 (94%)	10 (6%)	23	50
13	1	193/199 (97%)	184 (95%)	9 (5%)	30	60
13	M	199/199 (100%)	189 (95%)	10 (5%)	28	57
14	2	152/162 (94%)	144 (95%)	8 (5%)	26	54
14	N	162/162 (100%)	154 (95%)	8 (5%)	29	58
All	All	5302/5332 (99%)	5017 (95%)	285 (5%)	26	54

5 of 285 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
13	M	71(A)	ASP
2	P	218	ASN
12	Z	70(A)	ASN
13	M	149	GLN
1	O	33	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 201 such sidechains are listed below:

Mol	Chain	Res	Type
12	L	1(I)	ASN
2	P	177	GLN
11	Y	207	ASN
13	M	40	ASN
14	N	161	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](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
15	EP9	K	300	11	39,40,41	3.63	25 (64%)	50,53,54	2.36	12 (24%)
15	EP9	Y	300	11	39,40,41	3.82	24 (61%)	50,53,54	2.05	11 (22%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
15	EP9	K	300	11	1/1/9/13	0/41/41/43	0/2/2/2
15	EP9	Y	300	11	1/1/9/13	0/41/41/43	0/2/2/2

The worst 5 of 49 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
15	K	300	EP9	C9-C10	2.03	1.42	1.38
15	K	300	EP9	C30-C28	2.05	1.63	1.51
15	K	300	EP9	C27-C12	2.08	1.60	1.53
15	Y	300	EP9	C12-C11	2.35	1.59	1.52
15	Y	300	EP9	C27-C12	2.38	1.61	1.53

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
15	K	300	EP9	C4-C3-N1	-6.08	103.05	110.17
15	K	300	EP9	O4-C15-O5	-6.03	112.09	124.22
15	Y	300	EP9	O4-C15-O5	-5.53	113.11	124.22
15	K	300	EP9	O5-C15-N3	-5.38	115.61	124.87
15	Y	300	EP9	O5-C15-N3	-5.05	116.17	124.87

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
15	K	300	EP9	C1
15	Y	300	EP9	C1

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
15	K	300	EP9	1	0
15	Y	300	EP9	2	0

## 5.7 Other polymers ⓘ

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<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	250/250 (100%)	-0.07	8 (3%)	48	48	49, 69, 99, 124	0
1	O	250/250 (100%)	-0.09	9 (3%)	43	42	50, 70, 100, 123	0
2	B	244/244 (100%)	0.09	12 (4%)	30	29	51, 70, 107, 130	0
2	P	244/244 (100%)	0.10	14 (5%)	24	23	51, 71, 108, 130	0
3	C	241/241 (100%)	0.22	15 (6%)	21	20	53, 76, 125, 143	0
3	Q	241/241 (100%)	0.34	30 (12%)	4	4	56, 78, 126, 143	0
4	D	242/242 (100%)	0.23	14 (5%)	24	22	51, 76, 109, 140	0
4	R	242/242 (100%)	0.25	15 (6%)	21	20	52, 77, 109, 140	0
5	E	233/233 (100%)	0.18	19 (8%)	12	10	56, 82, 110, 126	0
5	S	233/233 (100%)	0.39	29 (12%)	4	4	57, 84, 110, 125	0
6	F	244/244 (100%)	-0.02	10 (4%)	38	36	51, 73, 109, 120	0
6	T	244/244 (100%)	0.23	19 (7%)	14	12	52, 73, 110, 120	0
7	G	243/243 (100%)	-0.13	5 (2%)	64	65	48, 68, 97, 129	0
7	U	243/243 (100%)	-0.05	7 (2%)	52	52	50, 69, 97, 128	0
8	H	222/222 (100%)	-0.27	3 (1%)	75	76	49, 64, 83, 125	0
8	V	222/222 (100%)	-0.30	2 (0%)	84	85	51, 65, 84, 125	0
9	I	204/204 (100%)	-0.30	2 (0%)	82	82	44, 62, 81, 98	0
9	W	204/204 (100%)	-0.29	2 (0%)	82	82	46, 62, 82, 98	0
10	J	198/198 (100%)	-0.25	5 (2%)	58	58	46, 60, 81, 137	0
10	X	198/198 (100%)	-0.29	4 (2%)	65	66	46, 60, 80, 137	0
11	K	212/212 (100%)	-0.22	4 (1%)	67	68	42, 60, 83, 105	0
11	Y	212/212 (100%)	-0.21	9 (4%)	37	35	43, 61, 84, 107	0
12	L	222/222 (100%)	-0.32	2 (0%)	84	85	46, 64, 89, 111	0
12	Z	222/222 (100%)	-0.10	5 (2%)	61	61	49, 69, 94, 115	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
13	1	233/233 (100%)	-0.12	5 (2%) 64 65	49, 69, 86, 90	0
13	M	233/233 (100%)	-0.29	1 (0%) 92 93	46, 64, 82, 87	0
14	2	196/196 (100%)	-0.05	3 (1%) 74 75	51, 65, 85, 99	0
14	N	196/196 (100%)	-0.35	0 100 100	48, 61, 82, 95	0
All	All	6368/6368 (100%)	-0.05	253 (3%) 39 37	42, 68, 104, 143	0

The worst 5 of 253 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	12(D)	ALA	11.1
4	D	12(E)	SER	10.0
4	D	12(F)	GLY	9.9
3	C	55	THR	9.9
4	R	12(D)	ALA	9.4

## 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
15	EP9	Y	300	39/40	0.91	0.23	2.21	43,50,69,70	0
15	EP9	K	300	39/40	0.91	0.23	1.81	40,47,69,71	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.