



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 10, 2018 – 04:42 am GMT

PDB ID : 2A2O  
Title : CRYSTAL STRUCTURE OF a putativeTenA family transcriptional regulator (BT\_3146) FROM BACTEROIDES THETA IOTAOMICRON VPI-5482 AT 2.16 Å RESOLUTION  
Authors : Joint Center for Structural Genomics (JCSG)  
Deposited on : 2005-06-22  
Resolution : 2.16 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](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 : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

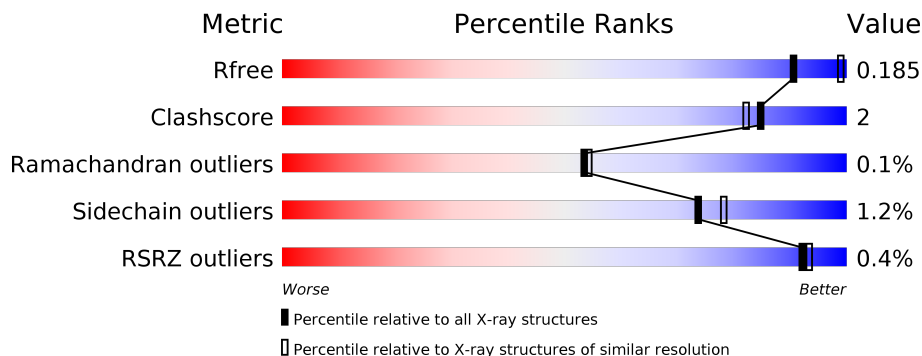
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.16 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))
$R_{free}$	111664	1287 (2.16-2.16)
Clashscore	122126	1390 (2.16-2.16)
Ramachandran outliers	120053	1368 (2.16-2.16)
Sidechain outliers	120020	1367 (2.16-2.16)
RSRZ outliers	108989	1262 (2.16-2.16)

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	258	
1	B	258	
1	C	258	
1	D	258	
1	E	258	
1	F	258	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	G	258	 83% 5% • 11%

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 14881 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 hypothetical protein BT3146.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	230	Total 1863	C 1197	N 299	O 357	S 4	Se 6	0	4	0
1	B	230	Total 1861	C 1194	N 297	O 360	S 4	Se 6	0	5	0
1	C	230	Total 1864	C 1198	N 299	O 357	S 4	Se 6	0	7	0
1	D	229	Total 1847	C 1186	N 295	O 356	S 4	Se 6	0	3	0
1	E	230	Total 1889	C 1212	N 303	O 364	S 4	Se 6	0	10	0
1	F	229	Total 1850	C 1188	N 295	O 357	S 4	Se 6	0	4	0
1	G	230	Total 1872	C 1200	N 300	O 362	S 4	Se 6	0	6	0

There are 133 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	MSE	-	LEADER SEQUENCE	UNP Q8A309
A	-10	GLY	-	LEADER SEQUENCE	UNP Q8A309
A	-9	SER	-	LEADER SEQUENCE	UNP Q8A309
A	-8	ASP	-	LEADER SEQUENCE	UNP Q8A309
A	-7	LYS	-	LEADER SEQUENCE	UNP Q8A309
A	-6	ILE	-	LEADER SEQUENCE	UNP Q8A309
A	-5	HIS	-	LEADER SEQUENCE	UNP Q8A309
A	-4	HIS	-	LEADER SEQUENCE	UNP Q8A309
A	-3	HIS	-	LEADER SEQUENCE	UNP Q8A309
A	-2	HIS	-	LEADER SEQUENCE	UNP Q8A309
A	-1	HIS	-	LEADER SEQUENCE	UNP Q8A309
A	0	HIS	-	LEADER SEQUENCE	UNP Q8A309
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
A	76	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
A	157	MSE	MET	MODIFIED RESIDUE	UNP Q8A309

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	161	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
A	202	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
A	205	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
A	226	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
B	-11	MSE	-	LEADER SEQUENCE	UNP Q8A309
B	-10	GLY	-	LEADER SEQUENCE	UNP Q8A309
B	-9	SER	-	LEADER SEQUENCE	UNP Q8A309
B	-8	ASP	-	LEADER SEQUENCE	UNP Q8A309
B	-7	LYS	-	LEADER SEQUENCE	UNP Q8A309
B	-6	ILE	-	LEADER SEQUENCE	UNP Q8A309
B	-5	HIS	-	LEADER SEQUENCE	UNP Q8A309
B	-4	HIS	-	LEADER SEQUENCE	UNP Q8A309
B	-3	HIS	-	LEADER SEQUENCE	UNP Q8A309
B	-2	HIS	-	LEADER SEQUENCE	UNP Q8A309
B	-1	HIS	-	LEADER SEQUENCE	UNP Q8A309
B	0	HIS	-	LEADER SEQUENCE	UNP Q8A309
B	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
B	76	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
B	157	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
B	161	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
B	202	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
B	205	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
B	226	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
C	-11	MSE	-	LEADER SEQUENCE	UNP Q8A309
C	-10	GLY	-	LEADER SEQUENCE	UNP Q8A309
C	-9	SER	-	LEADER SEQUENCE	UNP Q8A309
C	-8	ASP	-	LEADER SEQUENCE	UNP Q8A309
C	-7	LYS	-	LEADER SEQUENCE	UNP Q8A309
C	-6	ILE	-	LEADER SEQUENCE	UNP Q8A309
C	-5	HIS	-	LEADER SEQUENCE	UNP Q8A309
C	-4	HIS	-	LEADER SEQUENCE	UNP Q8A309
C	-3	HIS	-	LEADER SEQUENCE	UNP Q8A309
C	-2	HIS	-	LEADER SEQUENCE	UNP Q8A309
C	-1	HIS	-	LEADER SEQUENCE	UNP Q8A309
C	0	HIS	-	LEADER SEQUENCE	UNP Q8A309
C	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
C	76	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
C	157	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
C	161	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
C	202	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
C	205	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
C	226	MSE	MET	MODIFIED RESIDUE	UNP Q8A309

*Continued on next page...*

*Continued from previous page...*

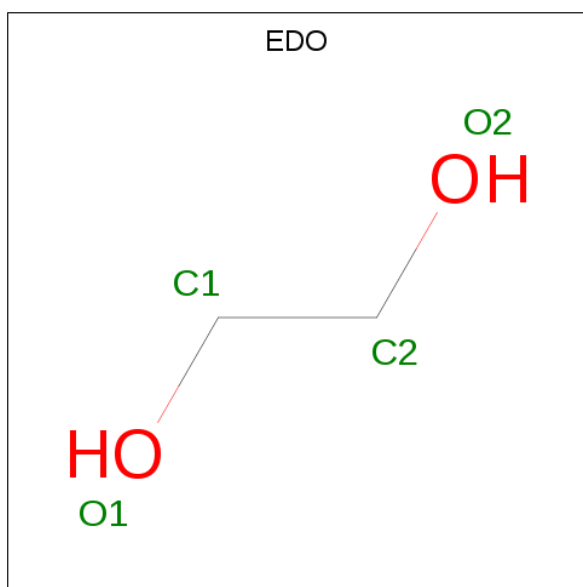
Chain	Residue	Modelled	Actual	Comment	Reference
D	-11	MSE	-	LEADER SEQUENCE	UNP Q8A309
D	-10	GLY	-	LEADER SEQUENCE	UNP Q8A309
D	-9	SER	-	LEADER SEQUENCE	UNP Q8A309
D	-8	ASP	-	LEADER SEQUENCE	UNP Q8A309
D	-7	LYS	-	LEADER SEQUENCE	UNP Q8A309
D	-6	ILE	-	LEADER SEQUENCE	UNP Q8A309
D	-5	HIS	-	LEADER SEQUENCE	UNP Q8A309
D	-4	HIS	-	LEADER SEQUENCE	UNP Q8A309
D	-3	HIS	-	LEADER SEQUENCE	UNP Q8A309
D	-2	HIS	-	LEADER SEQUENCE	UNP Q8A309
D	-1	HIS	-	LEADER SEQUENCE	UNP Q8A309
D	0	HIS	-	LEADER SEQUENCE	UNP Q8A309
D	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
D	76	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
D	157	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
D	161	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
D	202	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
D	205	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
D	226	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
E	-11	MSE	-	LEADER SEQUENCE	UNP Q8A309
E	-10	GLY	-	LEADER SEQUENCE	UNP Q8A309
E	-9	SER	-	LEADER SEQUENCE	UNP Q8A309
E	-8	ASP	-	LEADER SEQUENCE	UNP Q8A309
E	-7	LYS	-	LEADER SEQUENCE	UNP Q8A309
E	-6	ILE	-	LEADER SEQUENCE	UNP Q8A309
E	-5	HIS	-	LEADER SEQUENCE	UNP Q8A309
E	-4	HIS	-	LEADER SEQUENCE	UNP Q8A309
E	-3	HIS	-	LEADER SEQUENCE	UNP Q8A309
E	-2	HIS	-	LEADER SEQUENCE	UNP Q8A309
E	-1	HIS	-	LEADER SEQUENCE	UNP Q8A309
E	0	HIS	-	LEADER SEQUENCE	UNP Q8A309
E	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
E	76	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
E	157	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
E	161	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
E	202	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
E	205	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
E	226	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
F	-11	MSE	-	LEADER SEQUENCE	UNP Q8A309
F	-10	GLY	-	LEADER SEQUENCE	UNP Q8A309
F	-9	SER	-	LEADER SEQUENCE	UNP Q8A309
F	-8	ASP	-	LEADER SEQUENCE	UNP Q8A309

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
F	-7	LYS	-	LEADER SEQUENCE	UNP Q8A309
F	-6	ILE	-	LEADER SEQUENCE	UNP Q8A309
F	-5	HIS	-	LEADER SEQUENCE	UNP Q8A309
F	-4	HIS	-	LEADER SEQUENCE	UNP Q8A309
F	-3	HIS	-	LEADER SEQUENCE	UNP Q8A309
F	-2	HIS	-	LEADER SEQUENCE	UNP Q8A309
F	-1	HIS	-	LEADER SEQUENCE	UNP Q8A309
F	0	HIS	-	LEADER SEQUENCE	UNP Q8A309
F	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
F	76	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
F	157	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
F	161	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
F	202	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
F	205	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
F	226	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
G	-11	MSE	-	LEADER SEQUENCE	UNP Q8A309
G	-10	GLY	-	LEADER SEQUENCE	UNP Q8A309
G	-9	SER	-	LEADER SEQUENCE	UNP Q8A309
G	-8	ASP	-	LEADER SEQUENCE	UNP Q8A309
G	-7	LYS	-	LEADER SEQUENCE	UNP Q8A309
G	-6	ILE	-	LEADER SEQUENCE	UNP Q8A309
G	-5	HIS	-	LEADER SEQUENCE	UNP Q8A309
G	-4	HIS	-	LEADER SEQUENCE	UNP Q8A309
G	-3	HIS	-	LEADER SEQUENCE	UNP Q8A309
G	-2	HIS	-	LEADER SEQUENCE	UNP Q8A309
G	-1	HIS	-	LEADER SEQUENCE	UNP Q8A309
G	0	HIS	-	LEADER SEQUENCE	UNP Q8A309
G	1	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
G	76	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
G	157	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
G	161	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
G	202	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
G	205	MSE	MET	MODIFIED RESIDUE	UNP Q8A309
G	226	MSE	MET	MODIFIED RESIDUE	UNP Q8A309

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0
2	B	1	Total C O 4 2 2	0	0
2	B	1	Total C O 4 2 2	0	0
2	C	1	Total C O 4 2 2	0	0
2	C	1	Total C O 4 2 2	0	0
2	C	1	Total C O 4 2 2	0	0
2	D	1	Total C O 4 2 2	0	0
2	D	1	Total C O 4 2 2	0	0
2	D	1	Total C O 4 2 2	0	0
2	D	1	Total C O 4 2 2	0	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	E	1	Total C O 4 2 2	0	0
2	E	1	Total C O 4 2 2	0	0
2	E	1	Total C O 4 2 2	0	0
2	E	1	Total C O 4 2 2	0	0
2	E	1	Total C O 4 2 2	0	0
2	F	1	Total C O 4 2 2	0	0
2	F	1	Total C O 4 2 2	0	0
2	F	1	Total C O 4 2 2	0	0
2	F	1	Total C O 4 2 2	0	0
2	F	1	Total C O 4 2 2	0	0
2	F	1	Total C O 4 2 2	0	0
2	G	1	Total C O 4 2 2	0	0
2	G	1	Total C O 4 2 2	0	0
2	G	1	Total C O 4 2 2	0	0
2	G	1	Total C O 4 2 2	0	0

- Molecule 3 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	G	1	Total K 1 1	0	0
3	D	2	Total K 2 2	0	0
3	E	1	Total K 1 1	0	0
3	B	2	Total K 2 2	0	0
3	C	1	Total K 1 1	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	F	1	Total K 1 1	0	0


- Molecule 4 is water.

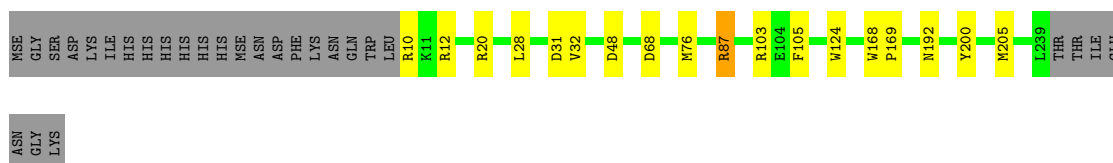
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	243	Total O 243 243	0	0
4	B	232	Total O 232 232	0	0
4	C	266	Total O 266 266	0	0
4	D	234	Total O 234 234	0	0
4	E	251	Total O 251 251	0	0
4	F	248	Total O 248 248	0	0
4	G	241	Total O 241 241	0	0

### 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: hypothetical protein BT3146

Chain A: 




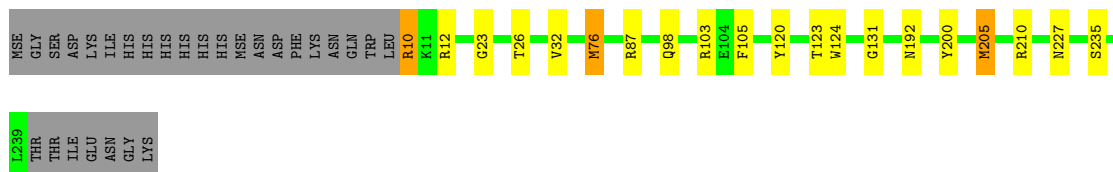
- Molecule 1: hypothetical protein BT3146

Chain B: 




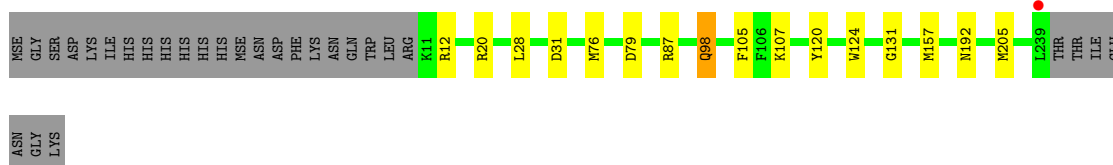
- Molecule 1: hypothetical protein BT3146

Chain C: 




- Molecule 1: hypothetical protein BT3146

Chain D: 




- Molecule 1: hypothetical protein BT3146

Chain E:  82% 7% 11%




GLU  
ASN  
GLY  
LYS

- Molecule 1: hypothetical protein BT3146

Chain F:  84% 11%



- Molecule 1: hypothetical protein BT3146

Chain G:  83% 5% 11%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 63 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	201.24Å 201.24Å 291.03Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.85 – 2.16 48.83 – 2.16	Depositor EDS
% Data completeness (in resolution range)	95.3 (48.85-2.16) 95.3 (48.83-2.16)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.38 (at 2.16Å)	Xtrriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.145 , 0.173 0.161 , 0.185	Depositor DCC
$R_{free}$ test set	8818 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	36.5	Xtrriage
Anisotropy	0.048	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 41.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	14881	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	40.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.13% 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: K, EDO

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	0.72	0/1925	0.77	4/2612 (0.2%)
1	B	0.73	0/1927	0.73	3/2615 (0.1%)
1	C	0.78	1/1938 (0.1%)	0.79	3/2631 (0.1%)
1	D	0.75	0/1905	0.75	3/2586 (0.1%)
1	E	0.76	0/1975	0.76	3/2681 (0.1%)
1	F	0.76	0/1912	0.75	4/2596 (0.2%)
1	G	0.79	1/1942 (0.1%)	0.75	4/2635 (0.2%)
All	All	0.76	2/13524 (0.0%)	0.76	24/18356 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	96	CYS	CB-SG	-7.98	1.68	1.82
1	C	205	MSE	SE-CE	-5.02	1.65	1.95

All (24) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	103	ARG	NE-CZ-NH2	-7.54	116.53	120.30
1	A	87	ARG	NE-CZ-NH2	-7.43	116.58	120.30
1	F	103	ARG	NE-CZ-NH2	-6.82	116.89	120.30
1	B	103	ARG	NE-CZ-NH1	6.67	123.64	120.30
1	G	87	ARG	NE-CZ-NH2	-6.16	117.22	120.30
1	D	20	ARG	NE-CZ-NH1	6.15	123.37	120.30
1	G	20	ARG	NE-CZ-NH1	6.08	123.34	120.30
1	E	87	ARG	NE-CZ-NH2	-6.04	117.28	120.30
1	F	87	ARG	NE-CZ-NH2	-5.97	117.32	120.30
1	F	20	ARG	NE-CZ-NH1	5.62	123.11	120.30
1	A	103	ARG	NE-CZ-NH1	5.61	123.11	120.30
1	C	210	ARG	NE-CZ-NH2	-5.60	117.50	120.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	20	ARG	NE-CZ-NH2	-5.60	117.50	120.30
1	G	68	ASP	CB-CG-OD1	5.60	123.34	118.30
1	E	68	ASP	CB-CG-OD1	5.58	123.32	118.30
1	C	76	MSE	CG-SE-CE	5.52	111.05	98.90
1	G	76	MSE	CG-SE-CE	5.48	110.95	98.90
1	A	68	ASP	CB-CG-OD1	5.45	123.20	118.30
1	D	157	MSE	CG-SE-CE	-5.40	87.03	98.90
1	B	76	MSE	CG-SE-CE	5.31	110.58	98.90
1	A	20	ARG	NE-CZ-NH1	5.19	122.89	120.30
1	F	87	ARG	NE-CZ-NH1	5.12	122.86	120.30
1	E	103	ARG	NE-CZ-NH1	5.06	122.83	120.30
1	B	87	ARG	NE-CZ-NH2	-5.03	117.79	120.30

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	1863	0	1758	7	0
1	B	1861	0	1742	7	0
1	C	1864	0	1754	14	0
1	D	1847	0	1736	11	0
1	E	1889	0	1778	10	0
1	F	1850	0	1739	9	0
1	G	1872	0	1761	7	0
2	A	20	0	30	0	0
2	B	8	0	12	0	0
2	C	12	0	18	0	0
2	D	16	0	24	0	0
2	E	20	0	30	0	0
2	F	20	0	30	3	0
2	G	16	0	24	1	0
3	B	2	0	0	0	0
3	C	1	0	0	0	0
3	D	2	0	0	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	E	1	0	0	0	0
3	F	1	0	0	0	0
3	G	1	0	0	0	0
4	A	243	0	0	0	0
4	B	232	0	0	0	0
4	C	266	0	0	5	0
4	D	234	0	0	2	0
4	E	251	0	0	1	0
4	F	248	0	0	4	0
4	G	241	0	0	1	1
All	All	14881	0	12436	64	1

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 (64) 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:23:GLY:O	1:C:26:THR:HG23	1.75	0.86
1:D:28:LEU:HD22	1:D:205:MSE:HE2	1.76	0.67
1:E:76:MSE:HA	1:E:76:MSE:HE3	1.80	0.62
1:C:227[B]:ASN:CG	4:C:382:HOH:O	2.39	0.60
1:B:76:MSE:HG2	1:B:124:TRP:CE2	2.39	0.58
1:E:76:MSE:HG2	1:E:124:TRP:CE2	2.40	0.56
1:F:98:GLN:NE2	4:F:379:HOH:O	2.11	0.56
1:G:52:GLN:HG2	4:G:473:HOH:O	2.07	0.54
1:A:32[B]:VAL:HG22	1:A:200:TYR:CE2	2.43	0.54
1:A:105:PHE:CD1	1:A:205:MSE:HE3	2.43	0.54
1:D:105:PHE:CD1	1:D:205:MSE:HE3	2.43	0.54
1:C:227[B]:ASN:ND2	4:C:382:HOH:O	2.40	0.53
1:C:105:PHE:CD1	1:C:205:MSE:HE3	2.44	0.53
1:C:32[B]:VAL:HG22	1:C:200:TYR:CE2	2.44	0.52
1:D:76:MSE:HG2	1:D:124:TRP:CE2	2.44	0.52
1:A:28:LEU:HD22	1:A:205:MSE:HE2	1.91	0.52
1:F:76:MSE:HG2	1:F:124:TRP:CE2	2.45	0.52
1:B:28:LEU:HD22	1:B:205:MSE:HE2	1.92	0.51
1:A:12:ARG:HD3	1:A:192:ASN:OD1	2.11	0.50
1:B:12:ARG:HD3	1:B:192:ASN:OD1	2.10	0.50
1:B:87:ARG:C	1:B:87:ARG:HD3	2.32	0.50
1:D:76:MSE:HE3	1:D:76:MSE:HA	1.93	0.50
1:F:235:SER:HB3	2:F:252:EDO:H21	1.94	0.50

Continued on next page...



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:76:MSE:HA	1:A:76:MSE:HE3	1.95	0.49
1:D:98:GLN:HG3	4:D:350:HOH:O	2.12	0.49
1:B:76:MSE:HA	1:B:76:MSE:HE3	1.95	0.48
1:D:87:ARG:NH1	4:D:357:HOH:O	2.39	0.48
1:G:76:MSE:HG2	1:G:124:TRP:CE2	2.48	0.48
1:F:28:LEU:HD22	1:F:205:MSE:HE2	1.95	0.48
1:F:60:LYS:NZ	4:F:376:HOH:O	2.47	0.48
1:F:76:MSE:HE1	2:F:251:EDO:H22	1.96	0.47
1:F:76:MSE:CE	2:F:251:EDO:H22	2.44	0.47
1:A:76:MSE:HG2	1:A:124:TRP:CE2	2.50	0.47
1:C:10:ARG:HH12	1:C:123:THR:HG21	1.80	0.47
1:G:76:MSE:HE2	1:G:120:TYR:OH	2.16	0.46
1:C:76:MSE:HE2	1:C:120:TYR:OH	2.16	0.46
1:G:76:MSE:HE2	1:G:120:TYR:CE1	2.51	0.46
1:C:76:MSE:HG2	1:C:124:TRP:CE2	2.50	0.45
1:G:163:PRO:HG2	1:G:225:ALA:HB3	1.98	0.45
1:E:10:ARG:HD2	1:E:191:TRP:CH2	2.52	0.44
1:D:76:MSE:HE2	1:D:120:TYR:CE1	2.53	0.44
1:D:12:ARG:HD3	1:D:192:ASN:OD1	2.17	0.44
1:E:163:PRO:HG2	1:E:225:ALA:HB3	2.00	0.44
1:F:26[A]:THR:CG2	4:F:472:HOH:O	2.66	0.44
1:A:168:TRP:HB2	1:A:169:PRO:HD3	2.00	0.43
1:E:76:MSE:CE	1:E:79:ASP:HB2	2.48	0.43
1:G:227[B]:ASN:HD22	2:G:251:EDO:H22	1.82	0.43
1:E:10:ARG:HD2	1:E:191:TRP:CZ2	2.54	0.43
1:C:12:ARG:HD3	1:C:192:ASN:OD1	2.19	0.43
1:E:118[B]:GLU:OE1	1:E:122:GLN:NE2	2.43	0.43
1:E:28:LEU:HD22	1:E:205:MSE:HE2	2.02	0.42
1:E:12:ARG:HD3	1:E:192:ASN:OD1	2.19	0.42
1:C:235:SER:OG	4:C:315:HOH:O	2.22	0.42
1:D:76:MSE:HE3	1:D:79:ASP:HB2	2.01	0.42
1:C:32[B]:VAL:HG23	4:C:427:HOH:O	2.21	0.41
1:C:98:GLN:N	4:C:345:HOH:O	2.44	0.41
1:B:76:MSE:HE2	1:B:120:TYR:OH	2.20	0.41
1:D:76:MSE:HE2	1:D:120:TYR:OH	2.19	0.41
1:C:76:MSE:HE2	1:C:120:TYR:CE1	2.55	0.41
1:C:131:GLY:HA3	1:D:131:GLY:HA3	2.03	0.41
1:B:105:PHE:CD1	1:B:205:MSE:HE3	2.56	0.41
1:E:76:MSE:HE1	4:E:433:HOH:O	2.20	0.40
1:F:165:GLU:OE2	4:F:275:HOH:O	2.22	0.40
1:G:28:LEU:HD13	1:G:205:MSE:HE2	2.02	0.40

All (1) 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)
4:G:437:HOH:O	4:G:466:HOH:O[10_775]	2.13	0.07

## 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	232/258 (90%)	231 (100%)	1 (0%)	0	100	100
1	B	233/258 (90%)	233 (100%)	0	0	100	100
1	C	235/258 (91%)	235 (100%)	0	0	100	100
1	D	230/258 (89%)	229 (100%)	0	1 (0%)	36	31
1	E	238/258 (92%)	236 (99%)	2 (1%)	0	100	100
1	F	231/258 (90%)	231 (100%)	0	0	100	100
1	G	234/258 (91%)	233 (100%)	1 (0%)	0	100	100
All	All	1633/1806 (90%)	1628 (100%)	4 (0%)	1 (0%)	53	54

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	98	GLN

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

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	194/210 (92%)	190 (98%)	4 (2%)	56	60
1	B	194/210 (92%)	193 (100%)	1 (0%)	90	93
1	C	195/210 (93%)	193 (99%)	2 (1%)	78	83
1	D	192/210 (91%)	190 (99%)	2 (1%)	78	83
1	E	201/210 (96%)	197 (98%)	4 (2%)	58	62
1	F	193/210 (92%)	192 (100%)	1 (0%)	90	93
1	G	197/210 (94%)	194 (98%)	3 (2%)	67	72
All	All	1366/1470 (93%)	1349 (99%)	17 (1%)	74	78

All (17) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	10	ARG
1	A	31	ASP
1	A	48	ASP
1	A	87	ARG
1	B	31	ASP
1	C	10	ARG
1	C	87	ARG
1	D	31	ASP
1	D	107	LYS
1	E	31	ASP
1	E	87	ARG
1	E	98[A]	GLN
1	E	98[B]	GLN
1	F	87	ARG
1	G	10	ARG
1	G	31	ASP
1	G	87	ARG

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

Mol	Chain	Res	Type
1	C	55	GLN
1	D	98	GLN
1	G	208	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](#)

Of 36 ligands modelled in this entry, 8 are monoatomic - leaving 28 for Mogul analysis.

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	$\# Z  > 2$
2	EDO	A	247	-	3,3,3	0.66	0	2,2,2	0.39	0
2	EDO	A	248	-	3,3,3	0.66	0	2,2,2	0.30	0
2	EDO	A	249	-	3,3,3	0.91	0	2,2,2	0.75	0
2	EDO	A	250	-	3,3,3	0.57	0	2,2,2	0.66	0
2	EDO	A	251	-	3,3,3	0.60	0	2,2,2	0.23	0
2	EDO	B	249	-	3,3,3	0.75	0	2,2,2	0.13	0
2	EDO	B	250	-	3,3,3	0.77	0	2,2,2	0.69	0
2	EDO	C	248	-	3,3,3	0.66	0	2,2,2	0.25	0
2	EDO	C	249	-	3,3,3	0.82	0	2,2,2	0.76	0
2	EDO	C	250	-	3,3,3	0.40	0	2,2,2	0.46	0
2	EDO	D	249	-	3,3,3	0.60	0	2,2,2	0.43	0
2	EDO	D	250	-	3,3,3	0.95	0	2,2,2	0.91	0
2	EDO	D	251	-	3,3,3	0.64	0	2,2,2	0.30	0
2	EDO	D	252	-	3,3,3	0.37	0	2,2,2	0.11	0
2	EDO	E	248	-	3,3,3	0.64	0	2,2,2	0.39	0
2	EDO	E	249	-	3,3,3	0.75	0	2,2,2	0.61	0
2	EDO	E	250	-	3,3,3	0.87	0	2,2,2	0.50	0
2	EDO	E	251	-	3,3,3	0.26	0	2,2,2	0.99	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	EDO	E	252	-	3,3,3	0.76	0	2,2,2	0.23	0
2	EDO	F	248	-	3,3,3	0.73	0	2,2,2	0.48	0
2	EDO	F	249	-	3,3,3	0.86	0	2,2,2	0.58	0
2	EDO	F	250	-	3,3,3	0.50	0	2,2,2	0.44	0
2	EDO	F	251	-	3,3,3	0.44	0	2,2,2	0.13	0
2	EDO	F	252	-	3,3,3	0.74	0	2,2,2	0.49	0
2	EDO	G	248	-	3,3,3	0.65	0	2,2,2	0.28	0
2	EDO	G	249	-	3,3,3	0.85	0	2,2,2	0.91	0
2	EDO	G	250	-	3,3,3	0.43	0	2,2,2	0.19	0
2	EDO	G	251	-	3,3,3	0.63	0	2,2,2	0.51	0

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
2	EDO	A	247	-	-	0/1/1/1	0/0/0/0
2	EDO	A	248	-	-	0/1/1/1	0/0/0/0
2	EDO	A	249	-	-	0/1/1/1	0/0/0/0
2	EDO	A	250	-	-	0/1/1/1	0/0/0/0
2	EDO	A	251	-	-	0/1/1/1	0/0/0/0
2	EDO	B	249	-	-	0/1/1/1	0/0/0/0
2	EDO	B	250	-	-	0/1/1/1	0/0/0/0
2	EDO	C	248	-	-	0/1/1/1	0/0/0/0
2	EDO	C	249	-	-	0/1/1/1	0/0/0/0
2	EDO	C	250	-	-	0/1/1/1	0/0/0/0
2	EDO	D	249	-	-	0/1/1/1	0/0/0/0
2	EDO	D	250	-	-	0/1/1/1	0/0/0/0
2	EDO	D	251	-	-	0/1/1/1	0/0/0/0
2	EDO	D	252	-	-	0/1/1/1	0/0/0/0
2	EDO	E	248	-	-	0/1/1/1	0/0/0/0
2	EDO	E	249	-	-	0/1/1/1	0/0/0/0
2	EDO	E	250	-	-	0/1/1/1	0/0/0/0
2	EDO	E	251	-	-	0/1/1/1	0/0/0/0
2	EDO	E	252	-	-	0/1/1/1	0/0/0/0
2	EDO	F	248	-	-	0/1/1/1	0/0/0/0
2	EDO	F	249	-	-	0/1/1/1	0/0/0/0
2	EDO	F	250	-	-	0/1/1/1	0/0/0/0
2	EDO	F	251	-	-	0/1/1/1	0/0/0/0
2	EDO	F	252	-	-	0/1/1/1	0/0/0/0
2	EDO	G	248	-	-	0/1/1/1	0/0/0/0
2	EDO	G	249	-	-	0/1/1/1	0/0/0/0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	EDO	G	250	-	-	0/1/1/1	0/0/0/0
2	EDO	G	251	-	-	0/1/1/1	0/0/0/0

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.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	F	251	EDO	2	0
2	F	252	EDO	1	0
2	G	251	EDO	1	0

## 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	224/258 (86%)	-0.42	0 <a href="#">100</a>   <a href="#">100</a>	31, 37, 50, 76	0
1	B	224/258 (86%)	-0.57	1 (0%) <a href="#">92</a>   <a href="#">93</a>	30, 37, 50, 82	0
1	C	224/258 (86%)	-0.60	0 <a href="#">100</a>   <a href="#">100</a>	31, 37, 51, 77	0
1	D	223/258 (86%)	-0.53	1 (0%) <a href="#">92</a>   <a href="#">93</a>	31, 36, 50, 65	0
1	E	224/258 (86%)	-0.54	1 (0%) <a href="#">92</a>   <a href="#">93</a>	31, 37, 50, 85	0
1	F	223/258 (86%)	-0.50	2 (0%) <a href="#">84</a>   <a href="#">87</a>	31, 37, 50, 66	0
1	G	224/258 (86%)	-0.40	1 (0%) <a href="#">92</a>   <a href="#">93</a>	30, 37, 50, 80	0
All	All	1566/1806 (86%)	-0.51	6 (0%) <a href="#">92</a>   <a href="#">93</a>	30, 37, 50, 85	0

All (6) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	239	LEU	3.5
1	B	239	LEU	2.9
1	E	191	TRP	2.5
1	D	239	LEU	2.4
1	G	239	LEU	2.4
1	F	191	TRP	2.1

### 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

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(Å <sup>2</sup> )	Q<0.9
3	K	D	248	1/1	0.79	0.09	96,96,96,96	0
2	EDO	F	252	4/4	0.81	0.26	58,60,63,68	0
2	EDO	E	250	4/4	0.81	0.19	60,67,67,74	0
2	EDO	F	249	4/4	0.82	0.17	53,57,59,64	0
2	EDO	A	247	4/4	0.82	0.14	66,72,74,77	0
3	K	C	247	1/1	0.84	0.09	93,93,93,93	0
3	K	B	247	1/1	0.85	0.09	81,81,81,81	0
3	K	F	247	1/1	0.86	0.10	92,92,92,92	0
2	EDO	G	251	4/4	0.87	0.18	60,63,65,67	0
2	EDO	A	251	4/4	0.88	0.43	72,75,75,78	0
2	EDO	E	252	4/4	0.89	0.16	45,51,58,62	0
3	K	D	247	1/1	0.90	0.06	85,85,85,85	0
2	EDO	G	248	4/4	0.91	0.13	35,43,52,55	0
2	EDO	E	249	4/4	0.91	0.12	45,50,50,58	0
2	EDO	A	250	4/4	0.92	0.12	48,52,53,54	0
2	EDO	F	248	4/4	0.92	0.11	43,52,56,62	0
2	EDO	D	251	4/4	0.93	0.10	43,52,53,61	0
2	EDO	F	251	4/4	0.94	0.15	36,60,61,62	0
2	EDO	A	249	4/4	0.94	0.12	43,50,55,64	0
2	EDO	G	249	4/4	0.94	0.12	44,48,53,57	0
2	EDO	C	249	4/4	0.94	0.12	41,48,49,56	0
2	EDO	D	249	4/4	0.94	0.15	30,45,52,55	0
2	EDO	B	250	4/4	0.94	0.12	46,48,51,63	0
2	EDO	D	252	4/4	0.94	0.20	50,54,56,61	0
3	K	B	248	1/1	0.95	0.07	82,82,82,82	0
2	EDO	D	250	4/4	0.95	0.11	39,40,46,50	0
2	EDO	E	248	4/4	0.95	0.16	37,51,51,55	0
2	EDO	C	250	4/4	0.96	0.18	48,48,52,55	0
2	EDO	E	251	4/4	0.96	0.11	50,50,51,58	0
3	K	E	247	1/1	0.96	0.06	82,82,82,82	0
2	EDO	G	250	4/4	0.96	0.20	56,59,63,64	0
2	EDO	C	248	4/4	0.96	0.12	34,48,50,54	0
2	EDO	F	250	4/4	0.97	0.14	48,53,54,55	0
2	EDO	A	248	4/4	0.97	0.14	35,45,48,51	0
2	EDO	B	249	4/4	0.97	0.17	32,49,50,58	0
3	K	G	247	1/1	0.99	0.15	78,78,78,78	1



## 6.5 Other polymers [i](#)

There are no such residues in this entry.