



Full wwPDB X-ray Structure Validation Report ⓘ

Jan 17, 2023 – 08:53 PM EST

PDB ID : 2G8L
Title : Crystal structure of a duf89 family protein (ph1575) from pyrococcus horikoshii at 2.04 Å resolution
Authors : Joint Center for Structural Genomics (JCSG)
Deposited on : 2006-03-02
Resolution : 2.04 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

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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 : 2.31.2
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) : 2.31.2

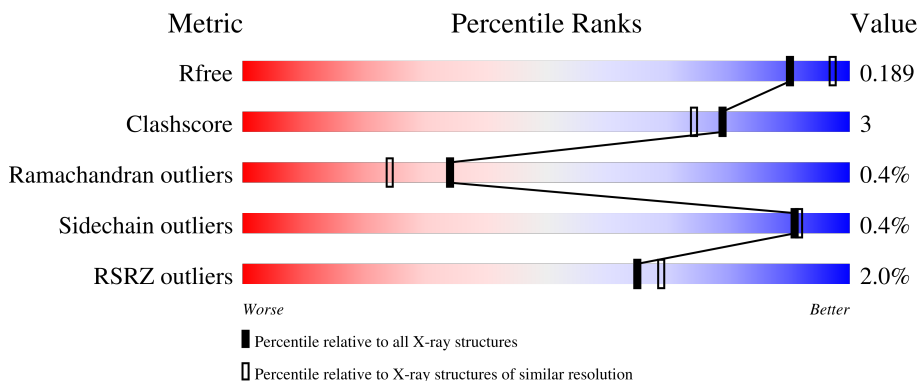
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

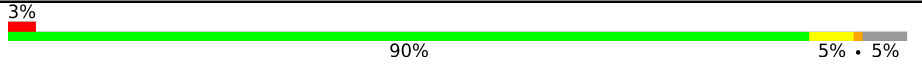

The reported resolution of this entry is 2.04 Å.

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	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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 of the lower bar indicate the fraction of residues that contain outliers for ≥ 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	299	 3% 90% 5% • 5%
1	B	299	 90% 6% •

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
2	UNL	A	288	-	-	X	-
2	UNL	B	288	-	-	X	-

2 Entry composition i

There are 4 unique types of molecules in this entry. The entry contains 4855 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 287aa long hypothetical protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	285	2228	1441	361	410	7	9	0	6	0
1	B	286	2285	1478	372	420	7	8	0	11	0

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	MSE	-	expression tag	GB 3258004
A	-10	GLY	-	expression tag	GB 3258004
A	-9	SER	-	expression tag	GB 3258004
A	-8	ASP	-	expression tag	GB 3258004
A	-7	LYS	-	expression tag	GB 3258004
A	-6	ILE	-	expression tag	GB 3258004
A	-5	HIS	-	expression tag	GB 3258004
A	-4	HIS	-	expression tag	GB 3258004
A	-3	HIS	-	expression tag	GB 3258004
A	-2	HIS	-	expression tag	GB 3258004
A	-1	HIS	-	expression tag	GB 3258004
A	0	HIS	-	expression tag	GB 3258004
A	1	MSE	MET	modified residue	GB 3258004
A	11	MSE	MET	modified residue	GB 3258004
A	21	MSE	MET	modified residue	GB 3258004
A	26	MSE	MET	modified residue	GB 3258004
A	33	MSE	MET	modified residue	GB 3258004
A	81	MSE	MET	modified residue	GB 3258004
A	128	MSE	MET	modified residue	GB 3258004
A	228	MSE	MET	modified residue	GB 3258004
A	281	MSE	MET	modified residue	GB 3258004
B	-11	MSE	-	expression tag	GB 3258004
B	-10	GLY	-	expression tag	GB 3258004
B	-9	SER	-	expression tag	GB 3258004
B	-8	ASP	-	expression tag	GB 3258004

Continued on next page...

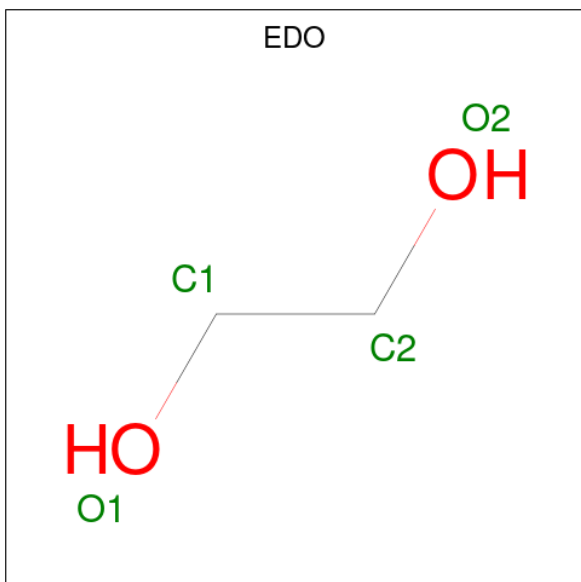
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-7	LYS	-	expression tag	GB 3258004
B	-6	ILE	-	expression tag	GB 3258004
B	-5	HIS	-	expression tag	GB 3258004
B	-4	HIS	-	expression tag	GB 3258004
B	-3	HIS	-	expression tag	GB 3258004
B	-2	HIS	-	expression tag	GB 3258004
B	-1	HIS	-	expression tag	GB 3258004
B	0	HIS	-	expression tag	GB 3258004
B	1	MSE	MET	modified residue	GB 3258004
B	11	MSE	MET	modified residue	GB 3258004
B	21	MSE	MET	modified residue	GB 3258004
B	26	MSE	MET	modified residue	GB 3258004
B	33	MSE	MET	modified residue	GB 3258004
B	81	MSE	MET	modified residue	GB 3258004
B	128	MSE	MET	modified residue	GB 3258004
B	228	MSE	MET	modified residue	GB 3258004
B	281	MSE	MET	modified residue	GB 3258004

- Molecule 2 is UNKNOWN LIGAND (three-letter code: UNL) (formula:).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O 8 8	0	0
2	B	1	Total O 8 8	0	0

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



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

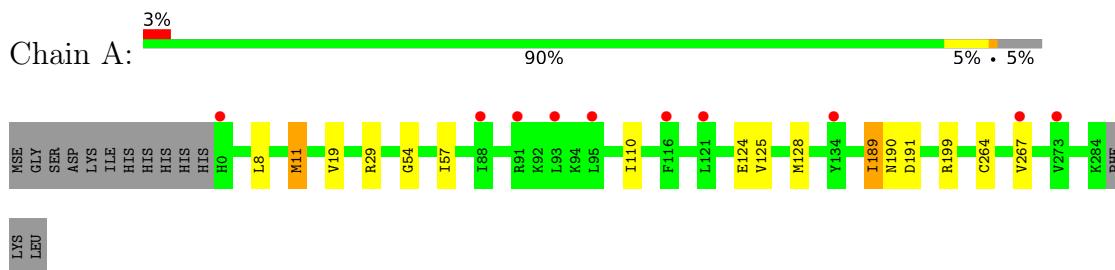
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	145	Total O 145 145	0	0
4	B	161	Total O 161 161	0	0

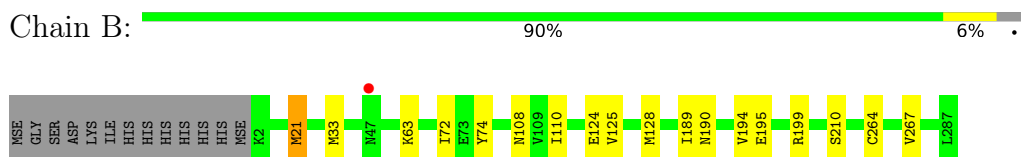
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: 287aa long hypothetical protein



- Molecule 1: 287aa long hypothetical protein



4 Data and refinement statistics i

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α , β , γ	88.86Å 88.86Å 81.50Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	39.01 – 2.04 39.01 – 2.04	Depositor EDS
% Data completeness (in resolution range)	99.7 (39.01-2.04) 99.7 (39.01-2.04)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.60 (at 2.05Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.144 , 0.182 0.155 , 0.189	Depositor DCC
R_{free} test set	2287 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	30.8	Xtriage
Anisotropy	0.354	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 49.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.015 for -h,-k,l 0.036 for h,-h-k,-l 0.022 for -k,-h,-l	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	4855	wwPDB-VP
Average B, all atoms (Å ²)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.52% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: UNL, 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	0.66	1/2264 (0.0%)	0.65	0/3033
1	B	0.73	0/2338	0.70	0/3127
All	All	0.69	1/4602 (0.0%)	0.68	0/6160

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	11	MSE	SE-CE	-7.76	1.49	1.95

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	2228	0	2275	10	0
1	B	2285	0	2372	12	0
2	A	8	0	0	2	0
2	B	8	0	0	3	0
3	A	12	0	18	0	0
3	B	8	0	12	0	0
4	A	145	0	0	1	0
4	B	161	0	0	3	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	4855	0	4677	26	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 3.

All (26) 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)
2:B:288:UNL:O2	2:B:288:UNL:O7	1.52	1.27
2:A:288:UNL:O2	2:A:288:UNL:O7	1.53	1.25
2:A:288:UNL:O2	2:A:288:UNL:O3	1.54	1.25
2:B:288:UNL:O4	2:B:288:UNL:O3	1.52	1.23
2:B:288:UNL:O2	2:B:288:UNL:O3	1.54	1.19
1:B:63:LYS:HE3	4:B:350:HOH:O	1.90	0.71
1:B:125:VAL:HA	1:B:128:MSE:HE3	1.80	0.64
1:B:195[B]:GLU:CD	1:B:199[B]:ARG:NH1	2.53	0.62
1:B:108:ASN:CG	4:B:396:HOH:O	2.39	0.61
1:A:19[B]:VAL:HG12	1:A:29:ARG:HG2	1.83	0.59
1:A:11:MSE:CE	1:A:57:ILE:HD12	2.36	0.56
1:A:8:LEU:HD23	1:B:33:MSE:HE3	1.88	0.56
1:A:125:VAL:HA	1:A:128:MSE:HE3	1.89	0.54
1:B:110:ILE:HD11	1:B:125:VAL:HG22	1.91	0.51
1:A:11:MSE:HE3	1:A:54:GLY:HA2	1.93	0.50
1:A:124:GLU:HG3	1:A:128:MSE:HE2	1.95	0.48
1:B:72:ILE:HG22	4:B:368:HOH:O	2.12	0.48
1:B:74:TYR:CD1	1:B:190[B]:ASN:OD1	2.67	0.47
1:A:264[A]:CYS:SG	1:A:267:VAL:HG23	2.54	0.47
1:B:21:MSE:HE2	1:B:21:MSE:HB3	1.81	0.45
1:A:110:ILE:HD11	1:A:128:MSE:HE1	2.00	0.44
1:B:124[A]:GLU:CG	1:B:128:MSE:HE2	2.48	0.44
1:B:264[B]:CYS:SG	1:B:267:VAL:HG23	2.58	0.43
1:B:194[B]:VAL:HG21	1:B:210:SER:HB2	2.01	0.43
1:A:199:ARG:NH2	4:A:405:HOH:O	2.43	0.43
1:A:189:ILE:HG22	1:A:190:ASN:H	1.85	0.42

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	289/299 (97%)	281 (97%)	7 (2%)	1 (0%)	41	31
1	B	295/299 (99%)	288 (98%)	5 (2%)	2 (1%)	22	12
All	All	584/598 (98%)	569 (97%)	12 (2%)	3 (0%)	34	18

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	189[A]	ILE
1	B	189[B]	ILE
1	A	189	ILE

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	230/248 (93%)	229 (100%)	1 (0%)	91	91
1	B	243/248 (98%)	242 (100%)	1 (0%)	91	91
All	All	473/496 (95%)	471 (100%)	2 (0%)	91	91

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	191	ASP
1	B	21	MSE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 2 are unknown - leaving 5 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	$\# Z > 2$
3	EDO	B	289	-	3,3,3	0.61	0	2,2,2	0.03	0
3	EDO	A	289	-	3,3,3	0.54	0	2,2,2	0.12	0
3	EDO	B	290	-	3,3,3	0.78	0	2,2,2	0.32	0
3	EDO	A	291	-	3,3,3	0.54	0	2,2,2	0.50	0
3	EDO	A	290	-	3,3,3	0.53	0	2,2,2	0.08	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
3	EDO	B	289	-	-	0/1/1/1	-
3	EDO	A	289	-	-	1/1/1/1	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	EDO	B	290	-	-	0/1/1/1	-
3	EDO	A	291	-	-	0/1/1/1	-
3	EDO	A	290	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	289	EDO	O1-C1-C2-O2
3	A	290	EDO	O1-C1-C2-O2

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, 95th 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 ²)	Q<0.9
1	A	276/299 (92%)	-0.19	10 (3%) 42 46	24, 37, 52, 81	0
1	B	278/299 (92%)	-0.40	1 (0%) 92 93	20, 31, 46, 60	0
All	All	554/598 (92%)	-0.29	11 (1%) 65 69	20, 34, 50, 81	0

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	91	ARG	3.8
1	A	0	HIS	2.8
1	A	95	LEU	2.8
1	A	116	PHE	2.7
1	A	134	TYR	2.7
1	A	273	VAL	2.5
1	B	47	ASN	2.5
1	A	88	ILE	2.2
1	A	93	LEU	2.1
1	A	121	LEU	2.0
1	A	267	VAL	2.0

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 monosaccharides 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, 95th 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 ²)	Q<0.9
2	UNL	B	288	8/-	0.77	0.19	82,90,96,97	0
2	UNL	A	288	8/-	0.81	0.18	89,93,94,96	0
3	EDO	A	289	4/4	0.84	0.12	65,69,69,71	0
3	EDO	A	290	4/4	0.86	0.30	68,70,73,74	0
3	EDO	A	291	4/4	0.88	0.18	48,51,51,58	0
3	EDO	B	290	4/4	0.92	0.14	37,50,57,60	0
3	EDO	B	289	4/4	0.98	0.06	34,35,36,38	0

6.5 Other polymers [i](#)

There are no such residues in this entry.