Electronic Supplementary Material (ESI) for Reaction Chemistry & Engineering.

Supporting Information:

Computer Aided Recipe Design: Optimization of Polydisperse Chemical Mixtures using Molecular Descriptors

*Anja Massolle, Jakob Schneider, Jan Meyer, Christoph Loschen**

Covestro AG, Kaiser-Wilhelm-Allee 60

51373 Leverkusen, Deutschland

Email: christoph.loschen@covestro.com

```
begin model

begin molecule types
        _GLY_(oh~s~p,oh~s~p,oh~s~p)
        _HEXD_(oh~p,oh~p)
        _ISOPS_(cooh,cooh)
        _H2O_(oh,oh)
        _EG_(oh~p,oh~p)
end molecule types

begin reaction rules
        _ISOPS_(cooh) + _GLY_(oh~s) -> _ISOPS_(cooh!1)._GLY_(oh~s!1) +
_H2O_(oh,oh) 0.4
        _ISOPS_(cooh) + _GLY_(oh~p) -> _ISOPS_(cooh!1)._GLY_(oh~p!1) +
_H2O_(oh,oh) 1.0
        _ISOPS_(cooh) + _EG_(oh~p) -> _ISOPS_(cooh!1)._EG_(oh~p!1) +
_H2O_(oh,oh) 1.0
        _ISOPS_(cooh) + _HEXD_(oh~p) -> _ISOPS_(cooh!1)._HEXD_(oh~p!1) +
_H2O_(oh,oh) 1.0
end reaction rules

begin seed species
        _ISOPS_(cooh,cooh) 14056
        _GLY_(oh~p,oh~p,oh~s) 4739
        _HEXD_(oh~p,oh~p) 1402
        _EG_(oh~p,oh~p) 79803
end seed species

end model
```

Listing S1. BNGL input for the reaction rule of optimization example 1.

```
begin model

begin molecule types
  _CPL_(oh~p~s,cooh~AH)
  _BDO_(oh~p,oh~p)
  _DPG_(oh~p,oh~p)
end molecule types

begin reaction rules
  _CPL_(cooh~AH!1,oh~p!1) + _BDO_(oh~p) -> _CPL_(oh~p,cooh~AH!1)._BDO_(oh~p!1)
1.0
  _CPL_(cooh~AH!1,oh~p!1) + _DPG_(oh~p) -> _CPL_(oh~p,cooh~AH!1)._DPG_(oh~p!1)
1.0
  _CPL_(oh~p!1,cooh~AH!1) + _CPL_(oh~p) -> _CPL_(oh~p,cooh~AH!1)._CPL_(oh~p!1)
1.0
end reaction rules

begin seed species
  _DPG_(oh~p,oh~p) 25000
  _BDO_(oh~p,oh~p) 25000
  _CPL_(oh~p!1,cooh~AH!1) 50000
end seed species

end model
```

Listing S2: BNGL input for modelling the CPL polyester polyol species.

```
begin model

begin molecule types
  _HEXD_(oh~p,oh~p)
  _EG_(oh~p,oh~p)
  _ADPS_(cooh,cooh)
  _H2O_(oh,oh)
end molecule types

begin reaction rules
  _ADPS_(cooh) + _EG_(oh~p) -> _ADPS_(cooh!1)._EG_(oh~p!1) + _H2O_(oh,oh) 1.0
  _ADPS_(cooh) + _HEXD_(oh~p) -> _ADPS_(cooh!1)._HEXD_(oh~p!1) + _H2O_(oh,oh)
1.0
end reaction rules

begin seed species
  _ADPS_(cooh,cooh) 20000
  _HEXD_(oh~p,oh~p) 40000
  _EG_(oh~p,oh~p) 40000
end seed species

end model
```

Listing S3: BNGL input for modelling the ADPS polyester polyol species.

```
begin model

begin molecule types
  _EO_(epo,oh~0~1)
  _EG_(oh~p,oh~p)
  _PERYT_(oh~p,oh~p,oh~p,oh~p)
  _PO_(epo,oh~0~1)
end molecule types

begin reaction rules
  _EG_(oh~p) + _EO_(epo,oh~0) -> _EG_(oh~p!1)._EO_(epo!1,oh~1)   1.0
  _EG_(oh~p) + _PO_(epo,oh~0) -> _EG_(oh~p!1)._PO_(epo!1,oh~1)   1.0
  _PERYT_(oh~p) + _EO_(epo,oh~0) -> _PERYT_(oh~p!1)._EO_(epo!1,oh~1) 1.0
  _PERYT_(oh~p) + _PO_(epo,oh~0) -> _PERYT_(oh~p!1)._PO_(epo!1,oh~1) 1.0
  _EO_(epo,oh~0) + _EO_(oh~1) -> _EO_(oh~1!1)._EO_(epo!1,oh~1)   1.0
  _PO_(epo,oh~0) + _PO_(oh~1) -> _PO_(oh~1!1)._PO_(epo!1,oh~1)   1.0
end reaction rules

begin seed species
  _PERYT_(oh~p,oh~p,oh~p,oh~p) 1140
  _EO_(epo,oh~0) 580
  _PO_(epo,oh~0) 70907
  _EG_(oh~p,oh~p) 27373
end seed species

end model
```

Listing S4: BNGL input for modelling the polether polyol species.

1.1 Correlation of Graph-based Descriptors

A high correlation can be observed for graph-based descriptors computed on the kMC and molecular graph. This is exemplarily shown for the Wiener index computed on the molecular graph (y-axis) and kMC graph (x-axis) for various molecular species in Figure S1. However, no perfect correlation can be observed. This is caused by the fact that molecular species consisting of different monomers with the same functionality and connectivity, like ethylene oxide and propylene oxide, result in the same kMC graph topology while the molecular graph is different. This can also be seen in Figure S1 where a change in the chain extender molecule yields the same Wiener index for the kMC graph while the value changes for the molecular graph. The same holds true for a change in the starter molecule, if the functionality of the starter does not change.

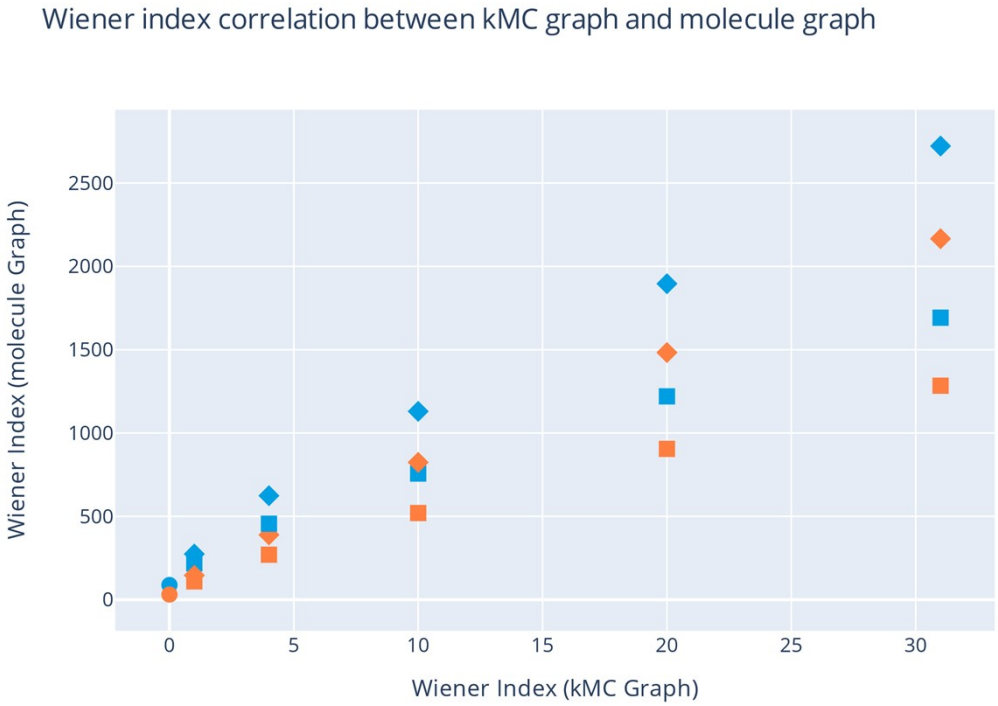

Figure S1: Correlation of the Wiener index between the kMC and molecular graph representation. A blue marker represents a polyether polyol with a trimethylolpropane and an orange marker with a glycerol starter. Species represented with a circle do not contain any epoxides, with a square contain only ethylene oxide and with a diamond contain only propylene oxide as chain extenders.

1.2 OH number distribution

By visual inspection the distribution of hydroxyl numbers is very similar between the target and proposed recipe. Moreover, the peak around the OH number of 1800 coming from the ethylene glycol monomer is represented in both recipes. Furthermore, the maximum of both distributions is in good agreement.

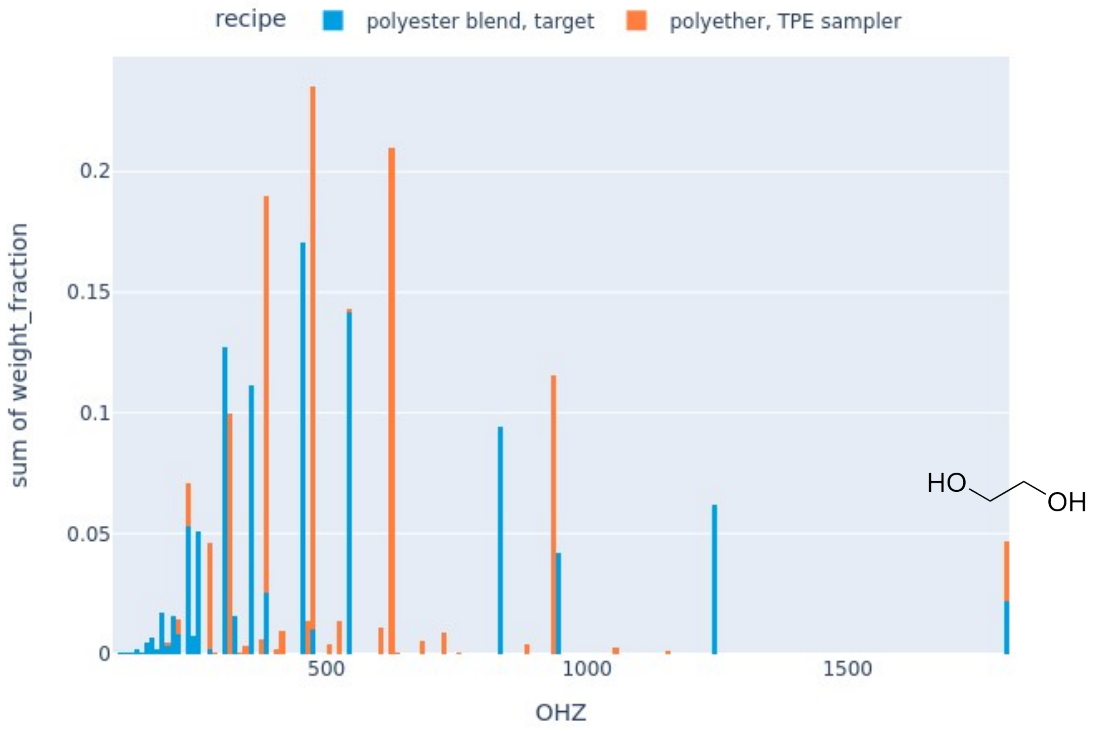

Figure S2: Distribution of hydroxy numbers (OHZ). A binning width of 10 hydroxy numbers is applied.

1.3 Covariance Matrix Adaptation Evolution Strategy (CMA-ES) Sampler

Table S1: Computed descriptors for the polyester reference and the new proposed polyether recipe using the CMA-ES sampler.

System	OH#	std(OH#)	Mn	Wiener Index	logP	Monomers
Reference Polyester	528	441	212	0.57	0.73	I: BDO, DPG, CPL II: EG, HEXD, ADPS
Proposed Polyether	816	266	275	16.32	-2.00	EG, PO

Table S2: Computed observables for the polyester reference and the new proposed polyether recipe using the CMA-ES sampler.

System	std(M)	Mw	Mw/Mn	std(logP)	Monomers
Reference Polyester	123	284	1.34	1.17	I: BDO, DPG, CPL II: EG, HEXD, ADPS
Proposed Polyether	78	297	1.08	0.03	EG, PO

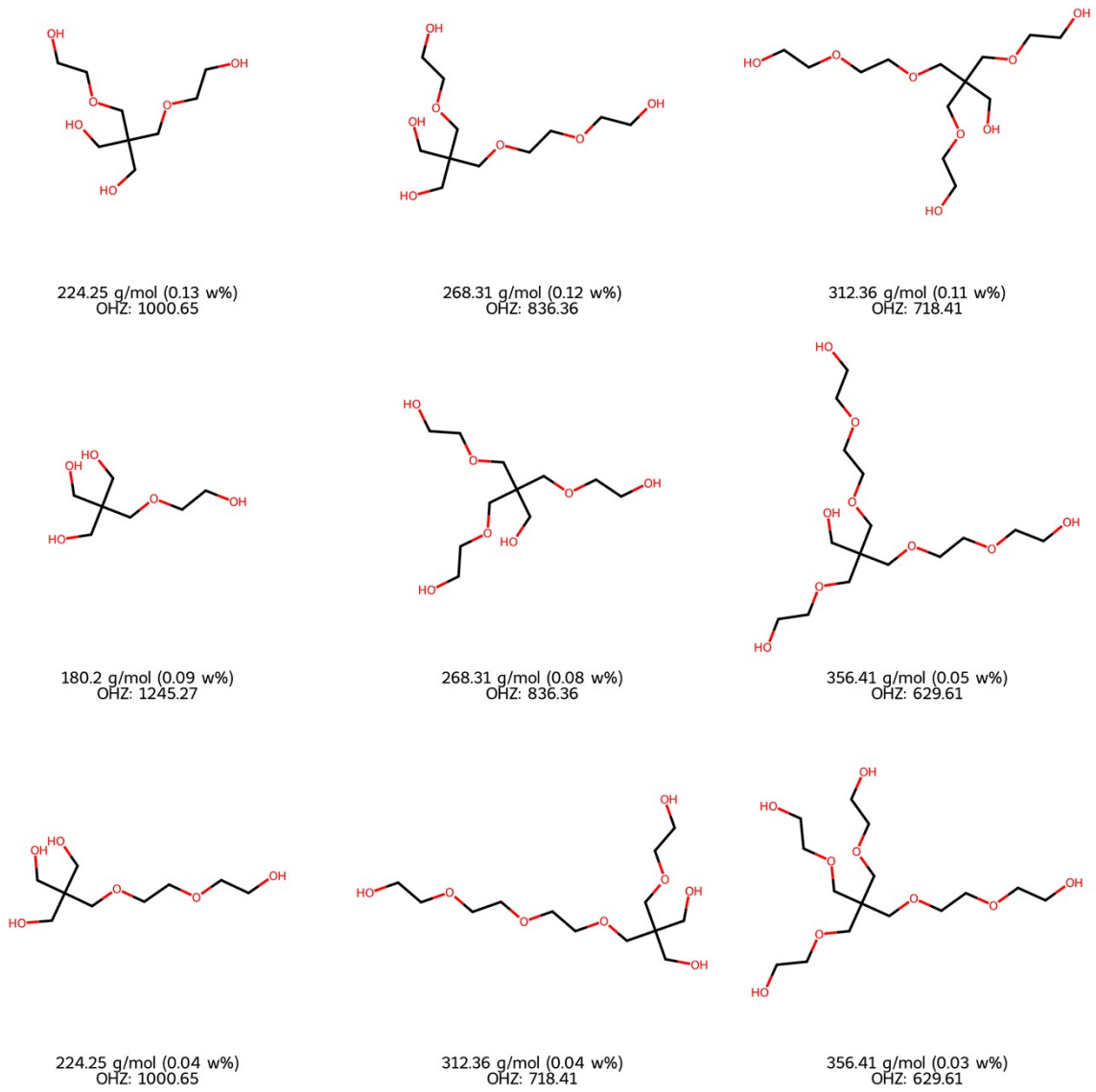

Figure S3: Most frequent oligomeric species (ordered by weight fraction) as created by the kinetic Monte Carlo simulation for the optimized polyether polyol using the CMA-ES sampler. Molar mass and weight fraction of the individual species are shown below each species. Molecular structures have been obtained by mapping the polymer graph into molecular graphs after the simulation.

1.4 Optimizer Benchmarking

The python library utilizes the optuna package for iterative optimization, which offers a wide variety of samplers for different use cases. To determine the optimal sampler choice, we benchmarked different samplers on an optimization problem close to typical use cases: The reference recipe consists of three different alcohols and one epoxide. The set of monomers to optimize is identical to the reference recipe so that an optimal solution is guaranteed. Allowed

are two reactions leading to a polyether formation: a chain start reaction and a chain extension (with equal probability). A fingerprint of seven descriptors was used to provide a measure do compare the properties of different simulated poly-/oligomers: Mn, Mw, polydispersity, number of hydroxy groups per molecule, OH-number and the Wiener index of the kMC graph. Four samplers were compared: CMA-ES, random, TPE, QMC. 1,000 optimization steps were run for each case. The result of the benchmarking can be found in Figure S4.

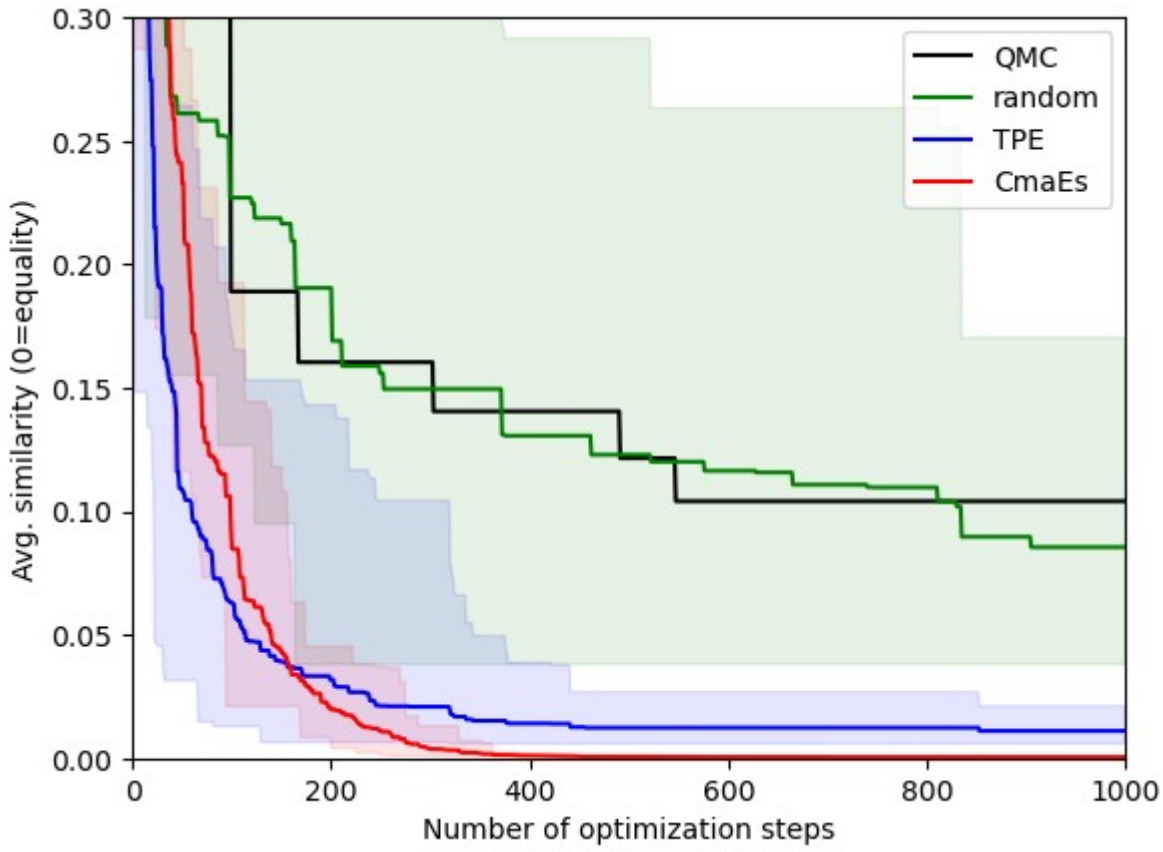

Figure S4: Benchmarking results to find the optimal sampler for a specific optimization of an alkoxylation relipe.

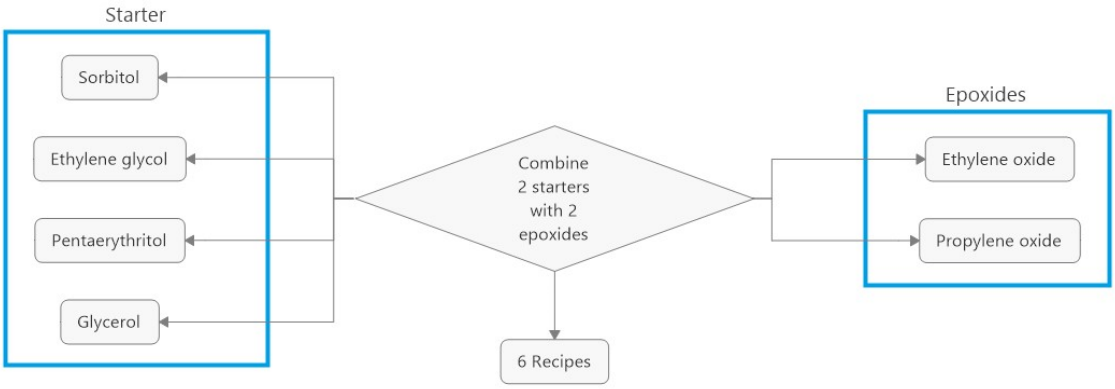

Figure S5: Generation of possible polyether polyol ingredient lists of example 2.

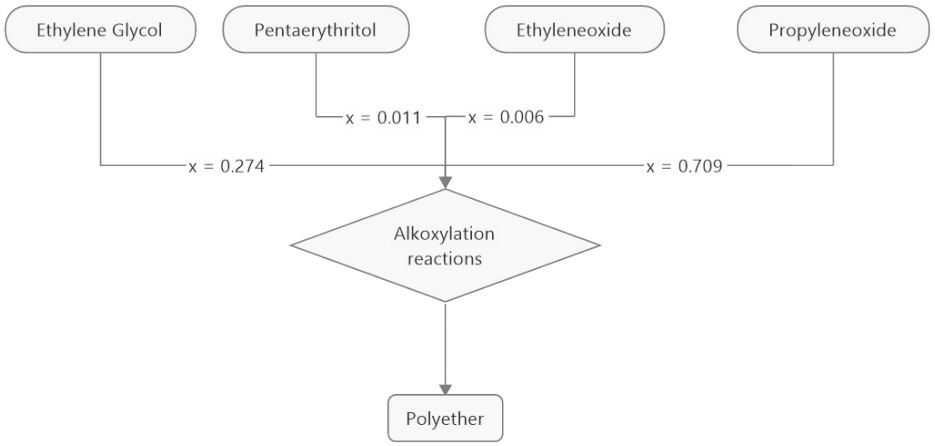

Figure S6: One-pot polyether synthesis route as proposed by the optimizer for example 2.