Systematic Methods for Reaction Solvent Design

Teng Zhou

8th June, 2017

Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany
Technical University of Denmark, Lyngby, Denmark

KT Consortium Annual Meeting, 2017
Solvents are widely used as reaction media and separation agents.

Can we find the best solvent that gives the highest reaction or separation performances?

→ We need **systematic methods for solvent design**!
Computer-Aided Molecular Design (CAMD)

Typical CAMD problems:
Find the best reaction medium for a chemical reaction
working fluid for an organic rankine cycle (ORC) process
refrigerant for a refrigeration process
solvent for an extraction process

Huge number of molecules → difficult to enumerate and test all

CAMD via mathematical optimization is able to identify promising molecules without testing all possible candidates.
A Typical CAMD Program

Given a set of molecular building groups, find the compound with a minimal viscosity.

<table>
<thead>
<tr>
<th>Group j</th>
<th>Group classes (v_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\text{CH}_3)</td>
<td>(G_M) 1</td>
</tr>
<tr>
<td>(\text{CH}_2)</td>
<td>(G_M) 2</td>
</tr>
<tr>
<td>(\text{CH})</td>
<td>(G_M) 3</td>
</tr>
<tr>
<td>(\text{C})</td>
<td>(G_M) 4</td>
</tr>
<tr>
<td>(\text{OH})</td>
<td>(G_F) 1</td>
</tr>
<tr>
<td>(\text{CH}_3\text{CO})</td>
<td>(G_F) 1</td>
</tr>
<tr>
<td>(\text{CH}_2\text{CO})</td>
<td>(G_F) 2</td>
</tr>
<tr>
<td>(\text{CHO})</td>
<td>(G_F) 1</td>
</tr>
<tr>
<td>(\text{CH}_3\text{COO})</td>
<td>(G_F) 1</td>
</tr>
<tr>
<td>(\text{CH}_2\text{COO})</td>
<td>(G_F) 2</td>
</tr>
<tr>
<td>(\text{HCOO})</td>
<td>(G_F) 1</td>
</tr>
<tr>
<td>(\text{OCH}_3)</td>
<td>(G_F) 1</td>
</tr>
<tr>
<td>(\text{OCH}_2)</td>
<td>(G_F) 2</td>
</tr>
<tr>
<td>(\text{OCH})</td>
<td>(G_F) 3</td>
</tr>
<tr>
<td>(\text{COOH})</td>
<td>(G_F) 1</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\min_{j \in G} & \quad \eta_j \\
\text{Subject to:} & \\
(1) \text{Structure-property relationship model:} & \\
& \eta_j = \eta_0 + \sum_{j=1}^{N} n_j \eta_j \\
(2) \text{Structural feasibility constraints:} & \\
& \sum_{j=1}^{N} \left(2 - v_j\right) n_j - 2m = 0 \\
& n_j \left(v_j - 1\right) + 2m - \sum_{k=1}^{N} n_k \leq 0 \quad (j = 1 \ldots N) \\
(3) \text{Structural complexity constraints:} & \\
& n_{\text{low}} \leq \sum_{j=1}^{N} n_j \leq n_{\text{upp}} \\
& 0 \leq n_j \leq n_{\max}(j)
\end{align*}
\]

\[
n_{j_{\text{opt}}} = [1 \ 3 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \quad \rightarrow \quad \text{CH}_3(\text{CH}_2)_3\text{OH} \text{ (Butanol)}
\]
CAMD of separation solvents has been widely studied.

We mainly focus on reaction solvent design.

**Solvent effects on reaction rate**

<table>
<thead>
<tr>
<th>Solvent</th>
<th>C₂H₅OH</th>
<th>CH₃OH</th>
<th>HCONH₂</th>
<th>HCOOH</th>
<th>H₂O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative rate const.</td>
<td>1</td>
<td>9</td>
<td>430</td>
<td>12200</td>
<td>335000</td>
</tr>
</tbody>
</table>

**Solvent effects on chemical equilibrium**

<table>
<thead>
<tr>
<th>Solvent</th>
<th>Diethyl ether</th>
<th>Ethyl acetate</th>
<th>Ethanol</th>
<th>Methanol</th>
<th>Acetone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enol/Keto</td>
<td>6.81</td>
<td>1.98</td>
<td>1.67</td>
<td>0.87</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Our Task: to use CAMD [a] to design solvents to increase reaction rates; and [b] to design solvents to improve reaction equilibrium conversions.
Methods to quantify/predict solvent effects on reaction rates

- **Experimental measurements**
  - + accurate
  - – expensive and time-consuming
  - – not predictive

- **Computational chemistry methods**
  - + classical theory based
  - – inaccurate on absolute reaction rate values
  - – computationally expensive

- **QSPR methods**
  - $k = k_0 + aA + bB + \ldots$ (A, B: solvent parameters or descriptors)
  - + fast and predictive
  - + accurate based on a limited set of reliable experimental data

1) Conventional QSPR methods use empirical solvent parameters (e.g. Folić et al., 2007)
2) We use theoretical-descriptor-based QSPR methods
  - + molecular structure or charge related
  - + easy to generate, do not require any experiment

---

QSPR: Quantitative Structure–Property Relationship
Conductor-like Screening Model (COSMO)

The 3D distribution of Screening Charge Density is converted into a 2D composition function, so called “σ-profile”, $P(σ)$.

Surface Screening Charge Density

$$\sigma_i = \frac{q_i}{a_i}$$

σ-profile is molecule-specific. It is a good “molecular fingerprint”, based on which we derive solvent descriptors to correlate solvent effects on reaction rates.
Solvent Structure – Reaction Rate Relationship

Group contribution model

\[ S_i = s_{i0} + \sum_{j=1}^{N} n_j s_{ij} \]

Solvent descriptors \( S_1 \sim S_6 \)

QSPR correlation with exp. rate constants in a few solvents

\[ \log k_{cal} = \log k_0 + \sum_{i=1}^{6} h_i S_i \]

Solvent-sensitive rate constant
Computer-Aided Solvent Design

<table>
<thead>
<tr>
<th>Group j</th>
<th>Group ID, Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH₃</td>
<td>G1, G_M</td>
</tr>
<tr>
<td>CH₂</td>
<td>G2, G_M</td>
</tr>
<tr>
<td>CH</td>
<td>G3, G_M</td>
</tr>
<tr>
<td>C</td>
<td>G4, G_M</td>
</tr>
<tr>
<td>OH</td>
<td>G5, G_F</td>
</tr>
<tr>
<td>CH₃CO</td>
<td>G6, G_F</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>OCH</td>
<td>G14, G_F</td>
</tr>
<tr>
<td>COOH</td>
<td>G15, G_F</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{max } \log k \\
&= \sum_{j \in G} n_j \log \log \left( \sum_{i=1}^{6} h_i S_i \right) \\
S_i &= s_{i0} + \sum_{j=1}^{N} n_j s_{ij} \\
\log k &= \log k_0 + \sum_{i=1}^{N} h_i S_i
\end{align*}
\]

(solvent structure – solvent descriptor relation)

(solvent descriptor – rate constant relation)

Structural feasibility rules:

\[
\sum_{j=1}^{N} (2 - v_j) n_j - 2m = 0 \\
0 \leq n_j \leq n_{max}(j)
\]

Structural complexity constraints:

\[
\sum_{j=1}^{N} n_j (v_j - 1) + 2m - \sum_{k=1}^{N} n_k \leq 0 \quad (j = 1 \ldots N) \\
n_{low} \leq \sum_{j=1}^{N} n_j \leq n_{upp}
\]

Results: Top 5 solvents designed

<table>
<thead>
<tr>
<th>Group combinations</th>
<th>Solvent structures</th>
<th>Log (k_{pred})</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 CH₃, 1 C, 1 OH, 1 COOH</td>
<td>(CH₃)₂C(OH)COOH</td>
<td>-1.875</td>
</tr>
<tr>
<td>3 CH₂, 1 OH, 1 COOH</td>
<td>OHCH₂CH₂CH₂COOH</td>
<td>-1.934</td>
</tr>
<tr>
<td>1 CH₃, 1 CH₂, 1 CH, 1 OH, 1 COOH</td>
<td>CH₃CH(OH)CH₂COOH</td>
<td>-1.942</td>
</tr>
<tr>
<td>2 CH₂, 1 OH, 1 COOH</td>
<td>OHCH₂CH₂COOH</td>
<td>-1.962</td>
</tr>
<tr>
<td>1 CH₃, 1 CH, 1 OH, 1 COOH</td>
<td>CH₃CH(OH)COOH</td>
<td>-1.971</td>
</tr>
</tbody>
</table>

Best solvent found experimentally: acetic acid (Log \(k = -2.491\))

Example: Diels-Alder reaction

1,3-Cyclopentadiene (Cyclo) + Acrolein (Acro) → 5-Norbornane-2-carboxaldehyde (Norbo)

The best reaction solvent also best for the whole process?

✓ Solvents affect not only reactions, but also other process properties, such as heat utilities, solvent-product separations.

✓ In order to find the best solvent for the whole process, it is necessary to perform "integrated solvent and process design" where solvent molecular structure and process operating conditions are simultaneously optimized.

✓ Process performance instead of reaction performance should be maximized.
Integrated Molecular and Process Design

Computer Aided Molecular Design: molecular structure – property relation

Integrated Molecular and Process Design: molecular structure – process performance relation

- **Property models**: group contribution (GC) models, QSPR models, UNIFAC

- **Process models**: mass balance, energy balance, etc.

\[
T_b = 222.543 \ln \left( \sum_{j=1}^{N} n_j t_{b,j} \right)
\]

\[
\log k = \log k_0 + \sum_{i=1}^{6} h_i S_i
\]

\[
\ln(\gamma_i) = \ln(\gamma_i^C) + \ln(\gamma_i^R)
\]
Integrated Molecular and Process Design

- **Property models**
  - **Molecular structure** (Discrete variables)

- **Molecular properties**

- **Process models**
  - **Process operating conditions** (Continuous variables)

- **Objective + Constraints**

- **Optimization Algorithms (MINLP Solvers)**
  - **Optimal molecular structure**
  - **Optimal process operating conditions**

MINLP: Mixed-Integer Nonlinear Programming
MINLP Formulation

**Maximize:** Process economic profit $F$

**Variables:** Integer $n_j$ (group $j = 1, 2, \ldots, N$)
Reaction conversion $X$ and product recovery $\xi_{nor}$

**Subject to:**

**Variable boundaries:**

\[
\begin{align*}
0.50 \leq X & \leq 0.99 \\
0 \leq \xi_{nor} & \leq 0.2 \\
0 \leq n_j & \leq n_j^{upp}
\end{align*}
\]

continuous process variables

(discrete molecular variables)

**Solvent structure constraints:**

Structural feasibility
Structural complexity

**Solvent property models:**

Reaction rate constant (previously mentioned method)
Physical property (existing GC models)

**Process and cost models:**

Reactor; Heat exchanger; Distillation column
Capital investment; Utility cost
### Integrated solvent and process design results

<table>
<thead>
<tr>
<th></th>
<th>Integrated Design (MINLP)</th>
<th>Reference Case (NLP)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Optimal solvent</strong></td>
<td>Isopropanol</td>
<td>Acetic acid (Fixed)</td>
</tr>
<tr>
<td>Opt. react. conversion</td>
<td>0.755</td>
<td>0.872</td>
</tr>
<tr>
<td>Opt. product recovery</td>
<td>99.45%</td>
<td>97.77%</td>
</tr>
<tr>
<td><strong>Total profit</strong> ($/year)</td>
<td><strong>27918.1</strong></td>
<td><strong>22776.4</strong></td>
</tr>
<tr>
<td>$EP$ ($/year)</td>
<td>65319.7</td>
<td>74118.8</td>
</tr>
<tr>
<td>$CI_R$ ($/year)</td>
<td>10145.0</td>
<td>5750.3</td>
</tr>
<tr>
<td>$CI_{column}$ ($/year)</td>
<td>7662.7</td>
<td>17133.2</td>
</tr>
<tr>
<td>$UC_{reb}$ ($/year)</td>
<td>13399.3</td>
<td>20816.1</td>
</tr>
<tr>
<td>log $k$</td>
<td>$-3.294$</td>
<td>$-2.551$</td>
</tr>
</tbody>
</table>

**Tradeoffs** between different solvent properties **are fully captured** via integrated solvent and process design.

Design Task: Given a liquid reaction mixture under a specified $T$ and $P$, determine which solvent and by which amount it should be added to the mixture in order to achieve a maximum equilibrium conversion of the reaction.
To design solvents for extractive reaction, we need:

1. A **thermodynamic method** to determine combined chemical and phase equilibrium under specified solvents.

2. A **CAMD method** to design solvents to maximize reaction equilibrium conversion determined by the thermodynamic method.

**Variables**

- Given solvent structure and composition
- Genetic Algorithm (GA) based CAMD method to optimize solvent structure and composition
- Rate-based dynamic method to calculate combined chemical and phase equilibrium

**Objective function**

- Equilibrium conversion of the reaction
Rate-based Dynamic Method

Mass transfer between phases:

\[ J_{i}^{k,\alpha} = k_{\text{trans},i} \left( x_{i}^{k} Y_{i}^{k} - x_{i}^{\alpha} Y_{i}^{\alpha} \right) \]

Reaction rate of the \( m \)th reaction in phase \( \alpha \):

\[ r_{m}^{\alpha} = k_{\text{react},m} \left[ \prod_{i \in NC \atop v_{i,m} < 0} (x_{i}^{\alpha} Y_{i}^{\alpha})^{-v_{i,m}} - \frac{1}{K_{th,m}^{\alpha}} \prod_{i \in NC \atop v_{i,m} > 0} (x_{i}^{\alpha} Y_{i}^{\alpha})^{v_{i,m}} \right] \]

Sources/sinks of component \( i \) in phase \( \alpha \) induced by the \( m \)th reaction:

\[ \sigma_{i,m}^{\alpha} = \nu_{i,m} r_{m}^{\alpha} \]

Mass balance of component \( i \) in phase \( \alpha \):

\[ \frac{dn_{i}^{\alpha}}{dt} = \sum_{k=1}^{NP} J_{i}^{k,\alpha} + \sum_{m=1}^{NR} \sigma_{i,m}^{\alpha} \quad (i = 1, 2, \ldots, N_{C}, \alpha = 1, 2, \ldots, N_{P}) \]

The steady state of the \((N_{C} \times N_{P})\) ODEs indicates the equilibrium state of the multiphasic reactive system.
Each GA individual includes a solvent initial composition and a solvent molecular structure.

- **Solvent initial composition**, a continuous variable, is encoded as a **binary array**.
- **Solvent molecules** are encoded as **tree structures** with UNIFAC groups as the tree nodes.

<table>
<thead>
<tr>
<th>Group ID</th>
<th>Valence</th>
<th>Size</th>
<th>*Child</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group information fields</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural information fields</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Structure of a group node:

Overall Design Framework

- Start program
  - Algorithm parameters
  - Structural group candidates

Randomly initialize the first generation

- Solvent structure
- Solvent composition

Rate-based dynamic method

**Objective function:**
reaction equilibrium
conversion

Create new generation

- **Binary array genetic operations**
- **Molecular tree genetic operations**

Roulette Wheel
Selection of individuals

Reach maximum generation number?

No

Yes

Take the best individual as final solution

Start program

1 0 1 0 0 1 0 1
1 0 1 0 0 1 0 1
1 1 0 0 1 0 1 0

Mutation
1 0 1 1 0 1 0 1

Crossover
1 0 1 0 0 0 1 0
1 1 0 0 1 1 0 1

Mutation with groups with the same valency
Insert node

Crossover
Exchange pointer addresses

Insert leaf
Case study on an esterification reaction

Evolution of average group weights in a generation

- CH₃
- CH
- CHCl₂
- CH₃CO
- OCH₂
- CCl

Case study on an esterification reaction
### Results of 10 optimization runs

1-Hexanol (A) + Acetic acid (B) [→] Hexyl acetate (C) + Water (D)

<table>
<thead>
<tr>
<th>No. of run</th>
<th>Optimal solvent structure</th>
<th>Optimal solvent initial composition</th>
<th>Reaction equilibrium conversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CH₃(CH₂)₂CHCl₂</td>
<td>0.5524</td>
<td>0.918000</td>
</tr>
<tr>
<td>2</td>
<td>CH₃(CH₂)₂CHCl₂</td>
<td>0.5714</td>
<td>0.918019</td>
</tr>
<tr>
<td>3</td>
<td>CH₃(CH₂)₂CHCl₂</td>
<td>0.5619</td>
<td>0.918025</td>
</tr>
<tr>
<td>4</td>
<td>CH₃(CH₂)₂CHCl₂</td>
<td>0.5619</td>
<td>0.918025</td>
</tr>
<tr>
<td>5</td>
<td>(CH₃)₂CHCHCl₂</td>
<td>0.5524</td>
<td>0.918042</td>
</tr>
<tr>
<td>6</td>
<td>(CH₃)₂CHCHCl₂</td>
<td>0.5810</td>
<td>0.918043</td>
</tr>
<tr>
<td>7</td>
<td>(CH₃)₂CHCHCl₂</td>
<td>0.5714</td>
<td>0.918066</td>
</tr>
<tr>
<td>8</td>
<td>(CH₃)₂CHCHCl₂</td>
<td>0.5619</td>
<td>0.918068</td>
</tr>
<tr>
<td>9</td>
<td>(CH₃)₂CHCHCl₂</td>
<td>0.5619</td>
<td>0.918068</td>
</tr>
<tr>
<td>10</td>
<td>(CH₃)₂CHCHCl₂</td>
<td>0.5619</td>
<td>0.918068</td>
</tr>
</tbody>
</table>

**Small deviation** in the reaction conversions obtained from 10 random optimization runs

Near-global solutions

Solvents can influence both reaction kinetics and thermodynamics, CAMD methods are developed for finding optimal solvents to increase reaction rates and equilibrium conversions.

When designing solvents for chemical reactions, the subsequent separation cost should also be considered. For that reason, integrated solvent and process design is necessary.
Thank you for your attention!