Property Model-Based Chemical Substitution and Chemical Formulation Design

Spardha Jhamb

Supervisors: Xiaodong Lianga, Georgios Kontogeorgisa, Amol Hukkerikarb, Kim Dam Johansena, Rafiqul Gani

aKT Consortium, Department of Chemical and Biochemical Engineering, DTU
bAlfa Laval Copenhagen A/S, Maskinvej 3, 2860 Søborg
Outline

- Motivation and Objectives

- Role of Property Models in Chemical Substitution and Chemical Formulation Design
  - Proposed Framework for Chemical Substitution
  - Framework for Model-based Chemical Formulation Design

- Group Contribution Models for Pure Component, Primary Property Prediction

- Examples employing the Framework for Chemical Substitution
  - Example 1: Single Molecule Product Substitution
  - Example 2: Substitution of an Additive from a Formulation

- Conclusion and Future Work
## Motivation and Objectives

### Motivation for Chemical Substitution: Why Substitute?

<table>
<thead>
<tr>
<th>Hindrance in Process Operability and/or Process Economics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-compliance with the Regulatory Standards</td>
</tr>
<tr>
<td>The Desirable Properties of a Product are Not Achieved</td>
</tr>
<tr>
<td>High Product Price</td>
</tr>
<tr>
<td>Low Product Quality</td>
</tr>
</tbody>
</table>

=> Based on ‘Reason for Substitution’, the ‘Substitution Problem’ is categorised into different types

- **Objective**: Propose a **systematic framework for property model-based chemical substitution and** apply it to various **product and process-centered** problems
Motivation and Objectives

Motivation for Model-based Chemical Product (Formulation) Design

- Chemical Product: Consumer-oriented, chemical-based, structured products
  - Single-species (molecule) products
  - Multiple-species products (Formulations, Blends)
  - Devices
- Chemical Product Design
  - Perceived to be the third chemical engineering paradigm
- Traditional Chemical Product Design
  - Based on trial and error approach
  - Inefficient: time and resource consuming
  - Final, feasible product design maybe more due to luck than the ability of the designer

=> Speed-up the Design Procedure and Spare the Resources

**Objective**: Apply the framework for model-based chemical formulation design (by Conte *et. al*) to design a paint formulation for varied applications
Role of Property Models in Chemical Substitution and Chemical Formulation Design

- Translation of needs to properties
- Screening of alternatives
- Verification of Product

Service Role

- Generation of feasible alternatives/ candidate molecules

Advice Role

Primary Properties: M&G GC Method

Secondary and Functional Properties: Co-relations and Functions

Mixture Properties: KT-UNIFAC

Polymer Models: By Kontogeorgios
Proposed Framework for Chemical Substitution

Reason for Substitution: Non-compliance with Regulatory Standards

**INPUTS Required**

- Structure of the Compound to be Substituted
- Set of Desired Target Properties and Constraints; Knowledge base; Information on Type of Compound (Substitute)
- Structure of the Compound to be Substituted
- Group Description/Assignments from ProPred

**Methodology**

**Task 1: Problem Definition**
- ST.1.1. Identify the Problem Type
- ST.12. Translate the qualitative problem description to quantitative target properties
- ST.1.3. Identify the Needs of the Substitute, Translation to Target Material Properties and Set Constraints on them

**Task 2: Generate candidate compounds (possible substitutes) that satisfy the constraints**

**Task 3: Check if shortlisted possible substitutes are better in terms of EH&S Properties**

**Task 4: Verify the stability of possible substitutes**

**Methods Required**

- GC Property Model
- Computer-Aided Molecular Design
- GC Property Model
- Molecular Modeling

**Tools Required**

- ProPred
- ProCAMD
- ProPred
- Chem3D

**Service Role**

**Advice Role**
Framework for Chemical Formulation Design (by Conte et al.)

**Inputs Required**
- Information about the product, List of performance criteria, List of target properties
- List of feasible AIs, activity/function of the product
- Solubility/Miscibility information, Pure Compound Properties; List of Target properties, Feasible Mixture properties
- AI/AIs + Optimal solvent mixture, List of properties to enhance, List of possible additives, Solubility Parameters

**Methodology**

**Stage I**
- **Task 1: Problem Definition**
- **Task 2: AIs Identification**
- **Task 3: Solvent Mixture Design**
- **Task 4: Additives Identification**

**Stage II**
- Planning of Experiments

**Stage III**
- Experimental Verification

**Methodology**

- Knowledge Base
- Property Models
- Computer-Aided Mixture Design, Linear / Non-linear models for mixture properties, Rigorous mixture property models
- Property Models

**Tools Required**

- [+] ProPred
- [+] VPPD Lab
- [+] ProPred

**Service Role**

**Advice Role**
Group Contribution Models for Pure Component, Primary Property Prediction

- Existing GC-Models for Pure Component Property Prediction by Marrero and Gani (MG) Method

- Properties that can be predicted by ‘GC-Model (Marrero and Gani)’
  - Organic compounds (C=3 - 60)

  - \( T_m \)
  - \( T_b \)

  - \( T_c \)
  - \( P_c \)
  - \( V_c \)
  - \( \Delta H_f^0 \)
  - \( \Delta H_{fus}^0 \)
  - \( \Delta H_{vap}^0 \)
  - \( \Delta G_f \)

- Properties that can be predicted by ‘GC-Model (Hukkerikar et al.)’

  - Environmental Impact on Air
    - Oral rat LD50
    - Permissible exposure limit (OSHA-TWA)
    - Photochemical oxidation potential
    - Global Warming Potential
    - Ozone Depletion Potential
    - Acidification Potential
    - Emission to urban air (carcinogenic and noncarcinogenic)

  - Environmental Impact on Water
    - Fathead minnow 96-h LC50
    - Daphnia magna 48-h LC50
    - Aqueous solubility
    - Bioconcentration Factor
    - Emission to Continental Fresh Water (carcinogenic and noncarcinogenic)
    - Emission to Continental Seawater (carcinogenic and noncarcinogenic)

  - Environmental Impact on Earth (Soil)
    - Emission to Continental Natural Soil (Carcinogenic and Noncarcinogenic)
    - Emission to Continental Agricultural Soil (Carcinogenic and Noncarcinogenic)
Group Contribution Models for Pure Component, Primary Property Prediction

- **Marrero and Gani (MG) Method**
  - Primary Property of Pure Compound
    - Function of structurally dependent parameters
    - Determined as a function of the frequency of the groups representing the pure compound and their contributions
  - Molecular Structure: Collection of three types of groups

**First Order Groups**
- Simple and monofunctional groups
- Partially captures the proximity effects

**Second Order Groups**
- Polyfunctional, polar/nonpolar compounds of medium size (C =3-6); aromatic or cycloaliphatic compounds with only one ring and several substituents
- Differentiation between isomers

**Third Order Groups**
- Complex heterocyclic and large (C=7-60), polyfunctional acyclic compounds
- For molecular fragments of compounds whose description is insufficient through the first and second level groups

\[ f(X) = \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k \] 

=> Computationally undemanding
Group Contribution Models for Pure Component, Primary Property Prediction

- Extensive Use of Amino Acids in various Industrial Sectors
  - Aspartyl-phenylalanine methyl ester: Low Calorie Sweetener
  - L-Cysteine: Antioxidant for preservation of fruit juices

- Scarcity of accurate predictive models for their physicochemical properties
- Insufficient description of these molecules through existing groups of GC models

- Need for Introduction of New Structural Parameters for Amino Acids
  - To account for the complex interactions due to zwitterionic structures

- Food Industry
  - Aspartyl-phenylalanine methyl ester: Low Calorie Sweetener
  - L-Cysteine: Antioxidant for preservation of fruit juices

- Cosmetics and personal care
  - N-Acylglycinate: Non-ionic surfactant
  - L-Glutamine: Neutralizer in soap (alternative acid)

- Pharmaceutical
  - Dihydroxyphenylalanine: Treatment of Parkinson’s disease
  - 5-hydroxytryptophan: Treatment of depression, and sleeping disorders

- As a Solvent
  - Potassium salt solution of L-Lysine: Solvent for CO$_2$ capture in flue gas

- As a Biodegradable polymer
  - Polyaspartate: Disposable diapers

- Non-ionic form
  - Zwitterionic form
Group Contribution Models for Pure Component, Primary Property Prediction

- Need for Introduction of New Structural Parameters for Amino Acids
  - To account for ionic interactions in amino acid salts

- Trend in the $T_m$ property for different classes of amino acids
  - ‘Nonpolar amino acids with aliphatic side chain’
  - ‘Fmoc-L amino acids with aliphatic side chain’
  - ‘Bmoc L-amino acids with aliphatic side chain’

$\Rightarrow$ requires special treatment to achieve reliable predictions of such molecules
Group Contribution Models for Pure Component, Primary Property Prediction

- Experimental Property Dataset* and Models used for Amino Acid Property Prediction

\[ \exp \left( \frac{T_m}{T_{m0} + \alpha MW} \right) = \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k \]

691 Natural + Modified Amino Acids

- Melting Point \( (T_m) \): 239 Compounds
- Water Solubility \( (\log W_s) \): 210 Compounds
- Octanol Water Partition Coefficient \( (\log K_{ow}) \): 334 Compounds

\[ \log K_{ow} = A_{LogKow} + \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k \]

\[ \log W_s = A_{LogWS} + B_{LogWS} MW + \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k \]

*Collected from Sigma Aldrich and database available at KT-Consortium
New Set of Groups to Define Amino Acid Structures (2 out of 36)

1. $O=C-NH-CH-COO-(CH_n)_m$ - (n in 0, …, 2) (m in 0, 1, 2…)

2. $CH_n(NH_3^+Cl^-)-COO-(CH_2)_m$ - n in (1, 2); m in (0, 1, 2…)

- N-(tert-Butoxycarbonyl)-L-phenylalanine methyl ester
- N-Boc-3-(3-methyl-4-nitrobenzyl)-L-histidine methyl ester
- L-Tryptophan ethyl ester hydrochloride
- L-tert-Leucine methyl ester hydrochloride
Group Contribution Models for Pure Component, Primary Property Prediction

Model Performance Statistics for the Developed GC-Models

**Absolute Error for Property 1: Melting Point (Tm)** Coefficient (log $K_{ow}$)

Summary of Statistics for all three properties

<table>
<thead>
<tr>
<th>Property</th>
<th>N</th>
<th>R²</th>
<th>Residual distribution plot</th>
<th>SD</th>
<th>AAE</th>
<th>AEₘₐₓ</th>
<th>Main type of compounds giving large deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_m$</td>
<td>239</td>
<td>0.94</td>
<td><img src="residual_distribution.png" alt="" /></td>
<td>15.03</td>
<td>10.81</td>
<td>42.82</td>
<td>Ester hydrochlorides of amino acids with aliphatic side chain</td>
</tr>
<tr>
<td>$\log W_s$</td>
<td>210</td>
<td>0.94</td>
<td><img src="residual_distribution.png" alt="" /></td>
<td>0.29</td>
<td>0.19</td>
<td>0.97</td>
<td>Amino benzene sulfonic acids</td>
</tr>
<tr>
<td>$\log K_{ow}$</td>
<td>334</td>
<td>0.99</td>
<td><img src="residual_distribution.png" alt="" /></td>
<td>0.25</td>
<td>0.16</td>
<td>1.18</td>
<td>Aminobenzoic acids</td>
</tr>
</tbody>
</table>
Examples to compare model prediction with experimental values

**Property 1: Melting Point (T_m) – Modified Amino Acid**

<table>
<thead>
<tr>
<th>Compound-1</th>
<th>First-order groups</th>
<th>Occurrences (N_i)</th>
<th>Contribution (C_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Tryptophan (CAS No. 73-22-3)</td>
<td>CH₂</td>
<td>1</td>
<td>-0.0687</td>
</tr>
<tr>
<td></td>
<td>aCH</td>
<td>4</td>
<td>0.4154</td>
</tr>
<tr>
<td></td>
<td>aC fused with non-aromatic ring</td>
<td>2</td>
<td>-0.3506</td>
</tr>
<tr>
<td></td>
<td>CHNH₂</td>
<td>1</td>
<td>4.0020</td>
</tr>
<tr>
<td></td>
<td>COOH</td>
<td>1</td>
<td>2.2226</td>
</tr>
<tr>
<td></td>
<td>CH=CH (cyclic)</td>
<td>1</td>
<td>0.8447</td>
</tr>
<tr>
<td></td>
<td>NH (cyclic)</td>
<td>1</td>
<td>1.4318</td>
</tr>
</tbody>
</table>

Second-order groups

| | Occurrences (M_j) | Contribution (D_j) |
| | | |
| CHₘ(NHₙ)-COOH (m.n in 0..2) | 1 | 3.3531 |
| (CHₙ=C)cyc-CH₂ (n in 0..2) | 1 | 0.8447 |

Third-order groups

| | Occurrences (O_k) | Contribution (E_k) |
| | | |
| aC-(CHₙ=CHₘ)cyc (fused rings) (n.m in 0..1) | 1 | 0.6558 |
| aC-NHₙcyc (fused rings) (n in 0..1) | 1 | -0.7514 |
| AROM.FUSED[2] | 1 | 0.3252 |
| (CHₙ=C)cyc-CH₂-CH(NH₂)-COOH (n in 0..2) | 1 | 1.0612 |

\[
\exp\left(\frac{T_m}{T_{m0} + \alpha \cdot MW}\right) = \sum_i N_i C_i + \sum_j M_j D_j + \sum_k N_k C_k \Rightarrow T_m = 548.8 K (T_{m0} = 217.1 K; \alpha = -0.0623)
\]
Group Contribution Models for Pure Component, Primary Property Prediction

- Examples to compare model prediction with experimental values

Property 2: Water Solubility (log \( W_s \))

<table>
<thead>
<tr>
<th>Compound-3</th>
<th>L-Tryptophan (CAS No. 73-22-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log W_s \text{,experimental} = 4.0569 )</td>
<td></td>
</tr>
</tbody>
</table>

\[
\log W_s = A_{\log W_s} + B_{\log W_s} \times MW + \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k = 4.0569
\]

The calculated 95\% confidence interval of the predicted property value = 1.4492

Property 3: Octanol Water Partition Coefficient (log \( K_{ow} \))

<table>
<thead>
<tr>
<th>Compound-4</th>
<th>L-Tryptophan (CAS No. 73-22-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log K_{ow,\text{experimental}} = -1.06 )</td>
<td></td>
</tr>
</tbody>
</table>

\[
\log K_{ow} = A_{\log K_{ow}} + \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k = -1.06
\]

The calculated 95\% confidence interval of the predicted property value = 0.8526
Examples employing the Framework for Chemical Substitution
Examples employing the Framework for Chemical Substitution

Example 1: Single Molecule Product Substitution in a Process-Centered Problem

Problem description*:

Objective:
To substitute a single molecule product (1,2,4-Trichlorobenzene) used in a process because it is environmentally harmful i.e. does not comply with the regulatory norms (REACH Regulations).

Gel-Spinning Manufacturing Process of UHMW-PE fibers

## Example 1: Single Molecule Product Substitution in a Process-Centered Problem

### Procedure:

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Methods Used</th>
<th>Tools Used</th>
<th>Inputs Required</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task 1: Problem Definition</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST.1. Identification of Problem Type:</td>
<td></td>
<td>MG GC</td>
<td>Structure of the compound to be substituted</td>
</tr>
<tr>
<td>➢ Non-compliance with regulatory standards (i.e. 1,2,4-</td>
<td></td>
<td>Property</td>
<td></td>
</tr>
<tr>
<td>Trichlorobenzene appears in the Restricted Substances List compiled by ECHA)</td>
<td></td>
<td>Model</td>
<td></td>
</tr>
<tr>
<td>ST.2. Translate the qualitative problem type description to quantitative</td>
<td></td>
<td>Pro-Pred</td>
<td></td>
</tr>
<tr>
<td>target property</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>➢ (-\log (LC_{50})<em>{FM, 1,2,4\text{-Trichlorobenzene}} = 4.63 (\text{Log mol/L})) ➢ (-\log (LC</em>{50})_{\text{substitute}} &lt; 4.63 (\text{Log mol/L}))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST.3. Identification of Needs, Translation to Target Material Properties</td>
<td></td>
<td>Knowledge</td>
<td>Solubility Parameter of UHMW-PE at 405.15 K</td>
</tr>
<tr>
<td>and Setting Constraints on them</td>
<td></td>
<td>Base</td>
<td></td>
</tr>
<tr>
<td>1.3.1. Pure Component Target Properties of the substitute</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Need 1: Dissolves UHMW-PE at temperatures 405.15 K</td>
<td></td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>➢ Hildebrand Solubility Parameter of substitute at 405.15 K</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((\delta_{\text{substitute}}, 405.15 K) \approx \text{Hildebrand Solubility Parameter of UHMW-PE at 405.15 K} (\delta_{\text{UHMW-PE, 405.15 K}})) ➢ \delta_{\text{substitute}, 405.15 K} \approx 21 \text{MPa}^{1/2} ➢ \delta_{\text{substitute}, 298.15 K} &lt; 21 \text{MPa}^{1/2}</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Example 1: Single Molecule Product Substitution in a Process-Centered Problem

#### Methodology

<table>
<thead>
<tr>
<th>Task 1: ST.3. Identification of Needs, Translation to Target Material Properties and Setting Constraints on them 1.3.1. Pure Component Target Properties of the Substitute</th>
<th>Methods Used</th>
<th>Tools Used</th>
<th>Inputs Required</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Need 2:</strong> Should allow easy crystallization of the polymer at 298.15 K</td>
<td>Knowledge Base</td>
<td>--</td>
<td>Solubility Parameter of UHMW-PE at 298.15 K</td>
</tr>
<tr>
<td>Hildebrand Solubility Parameter of substitute at 298.15 K $(\delta_{\text{substitute}, 298.15 \text{ K}}) \neq \text{Hildebrand Solubility Parameter of UHMW-PE at 298.15 K} (\delta_{\text{UHMW-PE, 298.15 \text{ K}}})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Rightarrow \delta_{\text{UHMW-PE, 298.15 K}} = 17.2 \text{ MPa}^{1/2}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Rightarrow \delta_{\text{substitute, 298.15 K}} &gt; 17.2 \text{ MPa}^{1/2} \text{ OR } &lt; 17.2 \text{ MPa}^{1/2}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Need 3:</strong> Should be removed by an extraction solvent (decalin) post-extrusion</td>
<td>MG GC Property Model</td>
<td>Pro-Pred</td>
<td>Structure of the compound</td>
</tr>
<tr>
<td>Hildebrand Solubility Parameter of substitute at 298.15 K $(\delta_{\text{substitute, 298.15 K}}) \approx \text{Hildebrand Solubility Parameter of extraction solvent (decalin) at 298.15 K} (\delta_{\text{decalin, 298.15 K}})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Rightarrow \delta_{\text{decalin, 298.15 K}} \approx 17.5 \text{ MPa}^{1/2}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Rightarrow \delta_{\text{substitute, 298.15 K}} \approx 17.5 \text{ MPa}^{1/2}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From Needs 1, 2 and 3, $17.2 \text{ MPa}^{1/2} &lt; \delta_{\text{substitute, 298.15 K}} &lt; 21 \text{ MPa}^{1/2}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Example 1: Single Molecule Product Substitution in a Process-Centered Problem

**Procedure:**

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Methods Used</th>
<th>Tools Used</th>
<th>Inputs Required</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task 1:</strong> ST.3. Identification of Needs, Translation to Target Material Properties and Setting Constraints on them</td>
<td>MG GC Property Model</td>
<td>Pro-Pred</td>
<td>Structure of the compound to be substituted</td>
</tr>
<tr>
<td>1.3.1. Pure Component Target Properties of the Substitute</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Need 4:</strong> Should be in liquid phase at room as well as operating temperature of extrusion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>➢ Melting Point of substitute ((\text{MP}_{\text{substitute}})) &lt; 285.15 K</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>➢ Boiling Point ((\text{BP}_{\text{substitute}})) &gt; 485.15 K</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3.2. Mixture Target Properties of the Substitute</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Need 5:</strong> Should be miscible with the extraction solvent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>➢ Miscibility of the substitute with the extraction solvent (decalin): Partial to Total Miscibility</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Examples employing the Framework for Chemical Substitution

### Example 1: Single Molecule Product Substitution in a Process-Centered Problem

**Procedure:**

**Methodology**

<table>
<thead>
<tr>
<th>Task 2: Generate candidate compounds (possible substitutes) satisfying the constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>➢ Number of Compounds satisfying the Constraints considering Acyclic Alcohols, Ethers and Hydrocarbons = <strong>10</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Possible Substitute-1</th>
<th>Possible Substitute-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ = 20.90 MPa$^{1/2}$</td>
<td>δ = 20.89 MPa$^{1/2}$</td>
</tr>
<tr>
<td>$T_m = 256.09$ K</td>
<td>$T_m = 265.08$ K</td>
</tr>
<tr>
<td>$T_b = 496.50$ K</td>
<td>$T_b = 495.80$ K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Possible Substitute-3</th>
<th>Possible Substitute-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ = 20.68 MPa$^{1/2}$</td>
<td>δ = 20.67 MPa$^{1/2}$</td>
</tr>
<tr>
<td>$T_m = 235.65$ K</td>
<td>$T_m = 263.06$ K</td>
</tr>
<tr>
<td>$T_b = 490.85$ K</td>
<td>$T_b = 503.45$ K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Possible Substitute-5</th>
<th>Possible Substitute-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ = 20.66 MPa$^{1/2}$</td>
<td>δ = 20.64 MPa$^{1/2}$</td>
</tr>
<tr>
<td>$T_m = 262.85$ K</td>
<td>$T_m = 275.44$ K</td>
</tr>
<tr>
<td>$T_b = 490.13$ K</td>
<td>$T_b = 493.78$ K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Possible Substitute-7</th>
<th>Possible Substitute-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ = 20.50 MPa$^{1/2}$</td>
<td>δ = 20.25 MPa$^{1/2}$</td>
</tr>
<tr>
<td>$T_m = 247.05$ K</td>
<td>$T_m = 268.02$ K</td>
</tr>
<tr>
<td>$T_b = 490.06$ K</td>
<td>$T_b = 487.26$ K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Possible Substitute-9</th>
<th>Possible Substitute-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ = 17.87 MPa$^{1/2}$</td>
<td>δ = 17.76 MPa$^{1/2}$</td>
</tr>
<tr>
<td>$T_m = 261.35$ K</td>
<td>$T_m = 267.99$ K</td>
</tr>
<tr>
<td>$T_b = 491.03$ K</td>
<td>$T_b = 498.17$ K</td>
</tr>
</tbody>
</table>

**Methods Used**

- Computer-Aided Molecular Design

**Tools Used**

- ProCAMD

**Inputs Required**

- Set of Desired Target properties and Constraints;
- Knowledge base;
- Information on Type of Compound (Substitute)

=> Only pure component target properties corresponding to Needs 1-4 have been considered here
# Examples employing the Framework for Chemical Substitution

## Example 1: Single Molecule Product Substitution in a Process-Centered Problem

**Procedure:**

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Methods Used</th>
<th>Tools Used</th>
<th>Inputs Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 3: Check if shortlisted possible substitutes are better in terms of EH&amp;S Properties</td>
<td>GC Property Model</td>
<td>ProPred</td>
<td>Structure of the possible substitute</td>
</tr>
<tr>
<td>➢ 0.94 ( \text{Log mol/L} ) &lt; (-\text{log (LC}<em>{50})</em>{\text{FM,substitute-1,..,10}} ) &lt; 2.40 ( \text{Log mol/L} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task 4: Verify the stability of possible substitutes</td>
<td>Molecular modeling</td>
<td>Chem3D</td>
<td>Group Description/Assignments from ProPred</td>
</tr>
</tbody>
</table>

### Future Work:

➢ To Check if Mixture Target Properties of the Substitute are met and select the final substitute

**i.e. Need 5:** Should be miscible with the extraction solvent

➢ Miscibility of the substitute with the extraction solvent (decalin): **Partial to Total Miscibility**
Examples employing the Framework for Chemical Substitution

Example 2: Substitution of Additive from a Chemical Product Formulation

Problem description:

Objective:

To substitute an additive (TRIDECETH-9) from a liquid soap formulation.
Example 2: Substitution of Additive from a Chemical Product Formulation

Procedure:

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Methods Used</th>
<th>Tools Used</th>
<th>Inputs Required</th>
</tr>
</thead>
</table>
| Task 1: **Problem Definition**<br>ST.1. Identification of Problem Type:  
➢ Non-compliance with regulatory standards (i.e. Classified as ‘expected to be toxic’ in ‘Environment Canada Domestic Substance List’)  
ST.2. Translate the qualitative problem type description to quantitative target property  
➢ \[-\log (LC_{50})_{FM, \text{Trideceth-9}} = 5.05 \text{ (Log mol/L)}\]  
⇒ \(-\log (LC_{50})_{\text{substitute}} < 5.05 \text{ (Log mol/L)}\)  
ST.3. Identification of Needs, Translation to Target Material Properties and Setting Constraints on them  
1.3.1. Target Properties of the substitute  
**Need 1:** HLB Parameter of the substitute (\(HLB_{\text{substitute}}\)) should be \(12 < HLB_{\text{substitute}} < 16\) to form an O/W emulsion and nearly equal to \(HLB_{\text{Trideceth-9}}\)  
➢ \(HLB_{\text{substitute}} \approx HLB_{\text{Trideceth-9}}\)  
⇒ \(HLB_{\text{substitute}} \approx 13.3\) | MG GC Property Model | ProPred | Structure of the compound to be substituted |
| | GC method by Davis-Rideal | -- | Group parameters for HLB calculation; Structure of the compound |
Examples employing the Framework for Chemical Substitution

Example 2: Substitution of Additive from a Chemical Product Formulation

Procedure:

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Methods Used</th>
<th>Tools Used</th>
<th>Inputs Required</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task 2:</strong> Generate candidate compounds (possible substitutes) satisfying the constraints</td>
<td>--</td>
<td>--</td>
<td>Set of desired Target properties and Constraints; Knowledge base; Information on Type of Compound (Substitute)</td>
</tr>
<tr>
<td>➢ A mixture of Oleic Acid (OA) and N-[3-lauryloxy-2-hydroxypropyl]-L-Arginine L-glutamate (C12HEA-Glu) has been reported* as a candidate with comparable emulsifying potency</td>
<td><strong>GC Property Model</strong></td>
<td><strong>ProPred</strong></td>
<td>Structure of the substitute mixture</td>
</tr>
<tr>
<td>[ \sum_{i} w_i HLB_i = 13.3 \Rightarrow w_{C_{12HEA-Glu}} = 31.25 % ]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Task 3:</strong> Check if shortlisted possible substitute mixtures are better in terms of EH&amp;S Properties</td>
<td><strong>GC Property Model</strong></td>
<td><strong>ProPred</strong></td>
<td>Structure of the substitute mixture</td>
</tr>
<tr>
<td>(-log (LC_{50})<em>{C</em>{12HEA-Glu}}) values are not available</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>However qualitative description of easy biodegradability of C12HEA-Glu has been reported</strong>.*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Examples employing the Framework for Chemical Substitution

Example 2: Substitution of Additive from a Chemical Product Formulation

Future Work:

1. To complete this example we are working on:
   - Getting Log LC$_{50}$ values for amino acids (possible also to develop a model)
   - Quantify the biodegradability property of the substitutes

2. To generate and test more candidates that satisfy the problem needs
   - Collect data for HLB parameters of more amino acids
Conclusion and Future Work

- The developed GC-property prediction models perform well
  - Define structures of natural as well as modified amino acids
  - Performance statistics are better in comparison to currently used property prediction models
  - Motivated by these results, current and future work is focussed on:
    - Modelling more properties for amino acids
    - Using them to develop chemical product formulations with amino acid as additive

- A framework for Chemical Substitution is proposed and used in two conceptual examples
  - The framework needs to be expanded to include problem types apart from non-compliance with regulatory norms

- Paint formulation for a wider range of applications needs to be developed using the framework by Conte et al.
  - Expansion of the database for AI (different pigments and binders)
  - Study of other target properties of a paint formulation for applications besides home decor
    - Example, ‘resistance to corrosion and abrasion’ incase of marine paints
  - Testing the applicability of rigorous thermodynamic models for property prediction
Thank-you for your attention!

Name: Spardha Jhamb
Email ID: spajha@kt.dtu.dk

Supervisors: Xiaodong Liang\textsuperscript{a}, Georgios Kontogeorgis\textsuperscript{a}, Amol Hukkerikar\textsuperscript{b}, Kim Dam Johansen\textsuperscript{a}, Rafiqul Gani\textsuperscript{a}

Any Questions?

\textsuperscript{a}KT Consortium, Department of Chemical and Biochemical Engineering, DTU
\textsuperscript{b}Alfa Laval Copenhagen A/S, Maskinvej 3, 2860 Søborg
Extra Slides
**Examples to compare model prediction with experimental values**

**Property 1: Melting Point \(T_m\) – Modified Amino Acid Group Contribution Models for Pure Component, Primary Property Prediction**

### Compound-1: N-(Carboxybenzyloxy)-L-glutamic acid (CAS No. 1155-62-0)

<table>
<thead>
<tr>
<th>First-order groups</th>
<th>Occurrences (N_i)</th>
<th>Contribution (C_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH(_2)</td>
<td>2</td>
<td>-0.0687</td>
</tr>
<tr>
<td>CH</td>
<td>1</td>
<td>-0.6259</td>
</tr>
<tr>
<td>aCH</td>
<td>5</td>
<td>0.4154</td>
</tr>
<tr>
<td>aC-CH(_2)</td>
<td>1</td>
<td>-0.1866</td>
</tr>
<tr>
<td>COOH</td>
<td>2</td>
<td>2.2226</td>
</tr>
<tr>
<td>NHCO except as above</td>
<td>1</td>
<td>0.5721</td>
</tr>
<tr>
<td>-O-</td>
<td>1</td>
<td>-0.8632</td>
</tr>
</tbody>
</table>

### Second-order groups

<table>
<thead>
<tr>
<th>Occurrences (M_j)</th>
<th>Contribution (D_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH(_m)(NH(_n))\text{COOH} ((m.n) in 0..2)</td>
<td>1</td>
</tr>
<tr>
<td>aC-CH(_n)OOC ((n) in 1..2)</td>
<td>1</td>
</tr>
</tbody>
</table>

### Third-order groups

<table>
<thead>
<tr>
<th>Occurrences (O_k)</th>
<th>Contribution (E_k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aC-CH(_2)O-O=C-NH-CH(_n)\text{COOH} ((n) in 1,2)</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
\exp \left( \frac{T_m}{T_{m0} + \alpha MW} \right) = \sum N_iC_i + \sum M_jD_j + \sum N_kC_k \Rightarrow T_m = 387.3 \text{ K}
\]

\(T_{m,\text{experimental}} = 389.2 \text{ K}\)

### Comparison of other methods with the current work:

<table>
<thead>
<tr>
<th>Method</th>
<th>Predicted Value (K)</th>
<th>Absolute Error (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present work</td>
<td>384.6</td>
<td>1.8</td>
</tr>
<tr>
<td>MPBPWIN (EPI® suite)</td>
<td>462.5</td>
<td>75.2</td>
</tr>
</tbody>
</table>
Examples to compare model prediction with experimental values

Property 2: Water Solubility (log $W_s$)

$\text{Compund-3}
\begin{align*}
\text{2-amino 5-bromo benzenesulfonic acid} \\
\text{(CAS No. 1576-59-6)}
\end{align*}

\[
\log W_s = A_{\log W_s} + B_{\log W_s} MW + \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k = 3.3629
\]

The calculated 95% confidence interval of predicted property value = 1.1521

$\log W_{s,\text{experimental}} = 2.7745$

Property 3: Octanol Water Partition Coefficient (log $K_{ow}$)

$\text{Compound-4}
\begin{align*}
\text{N$\alpha$-acetyl-L-Asparagine} \\
\text{(CAS No. 4033-40-3)}
\end{align*}

\[
\log K_{ow} = A_{\log K_{ow}} + \sum_i N_i C_i + \sum_j M_j D_j + \sum_k O_k E_k = -2.36
\]

The calculated 95% confidence interval of predicted property value = 0.59

$\log K_{\text{ow,experimental}} = -2.60$