Computer aided solvent design for extraction of a pharmaceutical intermediate

Madakashira Harini[†], Jhumpa Adhikari[†], Santosh B. Noronha[†] and K. Yamuna Rani[*]

[†]*Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai-400076*
[*]*Chemical Engineering Division, Indian Institute of Chemical Technology, Hyderabad-500007*

A majority of common solvents that are extensively used in chemical industry for reaction, formulation and separation have potential health hazards. The use of volatile organic solvents in industry can be assessed by Sheldon E-factor, (kg of waste per kg of product) and ranges from 25 to 100 for pharmaceutical industry with a production of 10^3 tonnes/year.[1] To reduce the E-factor, the need to minimize the usage of solvent in a process arises with regard to health and environmental concerns. In addition, the usage of a green solvent that can replace these hazardous conventionally used solvents will be able to reduce the impact on environment to a larger extent. Computer aided molecular design[2] (CAMD) approach is generally adopted to design a solvent reducing the need for carrying out experiments performed on a trial and error basis. The CAMD approach employs structure property correlations and an optimization technique to find a potential solvent for a physical separation process. In this study, CAMD approach has been applied to a case study where an aliphatic/ aromatic solvent has been designed for efficient extraction of R-phenylacetylcarbinol (PAC), a pharmaceutical intermediate from an aqueous phase. Current industrial practice involves the use of the solvent toluene.

The properties which have been considered in our CAMD approach for the design of a solvent to replace toluene include distribution coefficient, solvent loss, solvent power, selectivity, octanol water partition coefficient, viscosity, boiling point and toxicity. These properties have been predicted using group contribution models and the constraints on these properties are built by choosing toluene as a reference solvent. A variant of the genetic algorithm, namely non-dominated sorting genetic algorithm[3] (NSGA-II) along with the UNIFAC[4] (universal functional activity coefficients) model has been employed for optimization. A total of 64 UNIFAC groups (including repetitions) have been considered for the current study and the combinations of these groups lead to the design of different types of aliphatic solvent molecules such as aldehydes, ketones,

ethers, carboxylates and esters. In the case of the design of an aromatic solvent, only the feasible combinations of functional groups have been attached to a fixed aromatic ring. UNIFAC based structural constraints are further employed to ensure the design of a structurally feasible molecule. A multi objective optimization, minimizing solvent loss while maximizing distribution coefficient, solvent power and selectivity, has been performed to arrive at globally best task specific solvents. The CAMD approach yielded a pareto set of optimal aliphatic solvents including, amyl acetate, 2-methyl penta 1,3-diene, methyl hexene and ethyl pentene; whereas the optimal aromatic solvents obtained are propyl benzene, 1-phenyl propanone, ethoxy benzene and propenyl benzene.

Gibbs ensemble Monte Carlo simulations at constant temperature of 298.15 K and a constant pressure of 101.325 k Pa, have been employed to study the extraction of PAC from the aqueous phase by the solvents identified above, using the MCCCS *towhee* package[5]. Simulation results confirm the property estimation predictions and hence, the short-list of solvents chosen for further investigations. Finally, the suitability of the solvents designed using CAMD for PAC recovery are confirmed using experiments. Accounting for economic considerations, the CAMD derived aromatic solvents are found cheaper when compared to the aliphatic solvents.

References:

1. Grodowska, K. & Parczewski, A. Organic solvents in the pharmaceutical industry. *Acta poloniae pharmaceutica* **67**, 3–12 (2010).
2. Harper, P. M. , Gani, R., Kolar, P., Ishikawa, T. Computer-aided molecular design with combined molecular modeling and group contribution. *Fluid Phase Equilibria* **158-160**, 337–347 (1999).
3. Deb, K., Member, A., Pratap, A., Agarwal, S. & Meyarivan, T. A Fast and Elitist Multiobjective Genetic Algorithm. *IEEE transactions on evolutionary computation* **6**, 182–197 (2002).
4. Magnussen, T. UNIFAC Parameter Table for Prediction of Liquid-Liquid Equilibria. *Industrial Engineering Chemistry Process Design and Development* **20**, 331–339 (1981).
5. MCCCS Towhee. Available at: http://towhee.sourceforge.net.