Integrated Framework for Process and Product Synthesis/Design

by

Susilpa Bommareddy

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 13, 2013

Keywords: Computer Aided Product Design, Computer Aided Process Synthesis & Design, Integrated Framework, Group Contribution

Approved by

Mario R. Eden, Chair, Department Chair and McMillan Professor of Chemical Engineering
Christopher B. Roberts, Dean of Engineering and Uthlaut Professor of Chemical Engineering
Allan E. David, John W. Brown Assistant Professor of Chemical Engineering
Nishanth G. Chemmangattuvalappil, Associate Professor of Chemical & Environmental Engineering, The University of Nottingham

Abstract

Future growth within the chemical process industries depends on various factors such as raw material and energy availability, sustainability etc. A systematic process synthesis and design framework integrated with molecular design is needed to synthesize processes that perform this efficiently. Hence, this dissertation describes the development of a novel hybrid method for Computer Aided Flowsheet Design (CAFD) and its effective integration with molecular design. The interactions among process synthesis, process design and molecular design is through a common set of properties that are employed to analyze the processes as well as external agents involved in the process. Knowledge of these specific properties is needed to establish the feasibility of a unit operation in a process and the corresponding conditions of operation. The same information is needed for design of a component as an appropriate external agent. This forms the very basis of the proposed hybrid methodology for flowsheet synthesis/design integrated with molecular design. Both the Computer Aided Flowsheet Design (CAFD) and Computer Aided Molecular Design (CAMD) frameworks developed are group contribution (GC) based approaches. CAFD makes use of functional process groups, characterized by the type of unit operation/process and their corresponding driving force, to generate and represent flowsheets; process group contribution based property models to predict flowsheet properties from a-priori regressed contributions of process groups; a notation system (called SFILES) for storing the flowsheet structural information; and a synthesis method to generate and identify the feasible flowsheets.

The identified candidate flowsheets are ranked based on flowsheet properties (like energy consumption, amount (mass) of external agents used and/or cost/profit) representing flowsheet performance in a quantitative sense. Once the promising flowsheet structures are identified, the flowsheet design parameters that describe the process will be estimated. The reverse simulation method is used to calculate the design variables of the unit operations involved in the process. This also gives a good estimate of the important design parameters. Some alternatives may involve unit operations that require an external agent. Conventional agents may not always meet the property constraints set by the reverse simulation design problem of such operations. Novel agents can be identified by solving a product design problem satisfying the property constraints. This is done by integrating the flowsheet design problem with a molecular design problem. Depending on the type of unit operation in the process where an external agent is required, the CAMD problem is formulated accordingly and the effect of the solution alternatives from the CAMD problem on the process is evaluated by the process models. CAMD includes building blocks (atoms and functional groups) to generate and represent molecules; group contribution based property models to predict target properties; a standard molecular structure notation system to store and visualize the molecular structure information; and a synthesis method to generate and screen molecules that match the target (design) properties. Once a set of near optimal flowsheet alternatives have been identified, rigorous simulation is used to verify the predicted performance and select the best flowsheet. The framework also aims at maintaining a good accuracy of solutions and large application range. A completely automated tool to perform the above tasks is also developed.

Acknowledgments

I would like to express my profound thanks to my research advisor, Dr. Mario R. Eden, for his guidance and advice throughout the research work. His encouragement both on academic and personal front has tremendously contributed towards the completion of my dissertation. My knowledge and enthusiasm in the area of process and product design has increased significantly while working under his guidance and am truly fortunate to have had worked with him.

I would like to thank my research committee, Dr. Christopher B. Roberts, Dr. Allan E. David, Dr. Nishanth Chemmangattuvalappil and Dr. Steven Taylor for their suggestions and critical reading of my dissertation. Their comments helped me improve and complete this dissertation. I would also like to express my gratitude to Professor Rafiqul Gani at the Technical University of Denmark for his suggestions.

My sincere thanks to my friends, co-workers and others at Auburn University. Special thanks to my parents, Seeta Reddy Bommareddy and Siva Leela Bommareddy, my husband Mallikarjun Reddy Nagolu and my sister Sudeepthi Bommareddy. This would have been impossible without all the support and encouragement they have given me. Thanks to the entire Auburn University Chemical Engineering Department.

Table of Contents

List of Tables

List of Figures

1. Introduction

Design of processes and products is among the most creative of engineering activities with many opportunities to invent imaginative new products and processes. In simpler terms, it can be viewed as the effective production of products that meet customer needs through an efficient process. The question that is addressed in this dissertation is:

"Given the product requirements, determine the optimal process flowsheet to manufacture it."

Once the identity of the chemical to be manufactured is known, process design involves determining if it can be manufactured and how. This is the process synthesis and design step. Again, for the designed process, we also need to determine the likely raw materials to be processed in order to manufacture the desired product. Apart from the known raw materials there may be cases where some external agents like mass separating agents etc. are needed in the process. Specific properties governing the process direct the selection of these agents. These agents can be identified by a database search but this approach limits the selection of novel compounds. Hence, this scenario leads to invoking a product design problem.

It is clear that, in most cases process and product design problems cannot be easily decoupled and hence, a framework capable of effectively integrating process and product design is needed. So, to be precise the question addressed in this dissertation is:

"Given the product requirements, develop a framework for solving process and product synthesis/design problems and their integration to ultimately identify the optimal process flowsheet to manufacture it"

Also, traditionally, process/product synthesis and design has been performed in an iterative fashion. Though this evolutionary method that is forward in nature led to a more sophisticated design, it could not guarantee that no even better solution exists. If an efficient way to identify targets beforehand exists, then the iterative nature of solving design problems could be relieved and process/product specifics resulting in these targets could be identified in a reverse fashion. The targets enable a performance assessment of the identified solutions. Since it is not possible to specify the optimum solution apriori from a design perspective, specifying the targets in terms of desired process/product performance is quite beneficial.

While it should be clear from the discussion above that process and product synthesis/design should be solved simultaneously as a single problem in order to achieve an optimal solution, the question of how to handle the complexity of such design problems arises. This issue is solved by insightfully decoupling the process and product design problems and solving them piecewise based on a reverse solution methodology (Eden, Jørgensen, Gani, & El-Halwagi, 2004) to achieve their respective new targets. These new targets are surrogates of the overall design performance target. The performance specification of a design is often in terms of certain measurable properties/functionalities rather than the chemical species involved. For example, in the design of a blanket wash solvent, the primary quality parameters for the designer are the solubility parameter, flammability, vapor pressure etc. of the solvent. Therefore, a methodology capable of

systematically tracking these properties is called for and thus the concept of property clustering (Shelley & El-Halwagi, 2000) is utilized in this work.

This dissertation introduces a novel hybrid method for Computer Aided Flowsheet Design (CAFD) effectively integrated with Computer Aided Molecular Design (CAMD). The developed algorithm systematically identifies feasible process flowsheets in a computationally efficient manner by combining physical insights with algorithmic reverse design approaches. In the reverse approach, the flowsheets meeting the desired process performance are identified. Then, the design variables, which facilitate the desired process performance and the molecules that satisfy the property targets identified by solution of the process design problem are found. Both CAFD and CAMD methodologies are based on group contribution (GC) approaches. In these approaches, various groups (molecular fragments/flowsheet fragments) are tabulated along with their contributions towards a property of the molecule/flowsheet that includes these fragments. These contributions are estimated through regression of large amounts of experimental data. The property model equations for a set of tabulated data depend on how these values are regressed and are unique with respect to each set of GC data. Hence utilizing these tabulated data and respective models, methodologies to solve process and product design problems in a reverse fashion are developed. Doing so, the evaluation of solution alternatives of each with reference to their respective targets is straightforward given the models and the group contributions. A simple notation system, SFILES is employed for efficient storage and transfer of flowsheet information. The design variables for the selected flowsheet(s) are identified through a reverse simulation approach. Once the design parameters of an optimal flowsheet alternative have been identified, rigorous simulation is used to verify the predicted performance.

By solving the integrated process and product synthesis/design problems this way, the effect of any changes in the products on the process as well as the effect of changes in the process on the products can be rapidly evaluated. As an additional benefit the process and product design problems are selectively decoupled, so the solution is achieved with less complexity.

The solution methodology for process and product design problems and their simultaneous solution presented in Chapters 3 & 4. The work was published in *Computer Aided Chemical Engineering* as well as *Computers & Chemical Engineering* (Bommareddy, Chemmangattuvalappil, Solvason, & Eden, 2009b, 2010b). Complete description of the CAMD framework with respect to how the first- and higher- order groups from the group contribution data are incorporated in the different stages of framework was published in the *Brazilian Journal of Chemical Engineering* and the book *Design for Energy & the Environment* (Bommareddy, Chemmangattuvalappil, Solvason, & Eden, 2009a, 2010a). Detailed discussion of the CAFD framework and a completely automated tool to perform the above tasks was developed and published in *Computer Aided Chemical Engineering* (Bommareddy, Eden, & Gani, 2011). Also the integrated framework for process and product synthesis/design was published in *Computer Aided Chemical Engineering* (Bommareddy, Chemmangattuvalappil, & Eden, 2012)

The dissertation has been distributed in five chapters: Chapter 2 covers most of the background information including a discussion of the nature of process and product synthesis/design problems, current product design techniques, and basics of group contribution methods. Chapter 2 also covers the role of property models, the concept of reverse problem formulations and the state of art of process and product synthesis/design problems. Chapter 3 describes the development of a new CAMD framework and its application examples. Chapter 4 starts with the basics of property clustering and process design by both visualization and mathematical methods. Then, it covers the

systematic procedure for integrating process and product design and gives a case study to explain its application. Chapter 5 provides the process synthesis and design (CAFD) framework and its integration with CAMD followed by the software implementation of the developed methodology. Finally, two application examples for the algorithms developed in the chapter are given. The last chapter describes the major achievements and conclusions from this work followed by the next steps to be undertaken as part of this work. Finally the dissertation is appended with the data used from literature.

2. Theoretical Background

2.1 Chemical Process and Product Synthesis/ Design

Chemical product design, as quoted by Moggridge and Cussler (2000) is comprised of four steps: "The first step is the identification of customer needs and the translation of the needs into product specifications. The second step involves generating and winnowing ideas to fill these needs. In the third step the best ideas are chosen for commercial development. The last step requires product prototyping, decisions on manufacturing route and estimation of economic boundaries". Although this design scheme is simplified and thus applicable to many product design problems, it needs to be more clearly defined to adapt it to different cases of problems.

The second and third steps comprise the product design problem, the first step is a pre-design step or problem formulation step and the last step is part of a process design problem. Traditionally process and product design problems have been treated as two separate problems, with little or no feedback between them. Product design has been carried out based on heuristics and expert knowledge to provide options that match the requirements. The compounds identified in this step are analyzed to decide on their suitability and if an efficient manufacturing route to make it exists. If none of the options are practical, the product design problem is redone with looser constraints. Therefore, this is an iterative process as described in Figure 2.1.

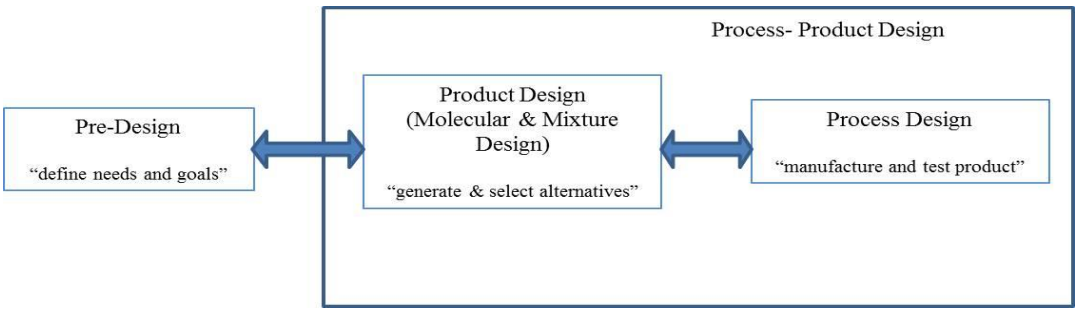


Figure 2.1: The product design process (Gani, 2004).

The other way of analyzing the need for integration of process and product synthesis/design problems is listed below. If we arrive at a stage, where we know the product satisfying the target specifications and what raw materials could be treated physically or chemically to produce the product, the remaining task is to design the process to manufacture it. This task involves synthesizing the process alternatives and designing the selected process. One of the important factors that decide the selection of one of the synthesized processes is the choice of chemicals used as external agents (if appropriate) in the process. The steps of synthesizing and designing the process and design of chemicals to be used are fairly interlinked. As cited in the introduction of this dissertation, if the two steps are solved to together, the problem becomes fairly complex. If they are decoupled and solved without analyzing, each of the problems as shown in Figure 2.1 falls prey to the iterative nature of the problem. Hence these problems have to be carefully decoupled such that targets for each of the decoupled problems are functions of the ultimate performance targets and the methodologies developed for each of them serve to provide solutions that match their respective targets.

It is clear that when process and product design problems are solved together, each benefit from the other to yield truly optimal solutions that meet the performance target(s) but the identification of such optimal solutions is challenging. Hence methods to address this issue are presented in this dissertation.

2.1.1 Mathematical Formulation of the Process and Product Synthesis/Design Problem

This dissertation addresses some of the issues in process and product synthesis/design by setting up a mathematical model describing the problem and identifying the solutions. All different types of product-process design problems can be represented using the following set of mathematical expressions (Gani, 2004).

$$F_{obj} = max[C^T y + f(x)] \quad \textbf{2.1}$$

Subject to:

$$h_1(x) = 0 \quad \textbf{2.2}$$

$$h_2(x) = 0 \quad \textbf{2.3}$$

$$h_3(x, y) = 0 \quad \textbf{2.4}$$

$$l_1 \leq g_1(x) \leq u_1 \quad \textbf{2.5}$$

$$l_2 \leq g_2(x, y) \leq u_2 \quad \textbf{2.6}$$

$$l_3 \leq By + Cx \leq u_3 \quad \textbf{2.7}$$

where,

x	:	Vector of continuous variables like fraction of a mixture, flow rates etc.
y	:	Vector representing the presence or absence of a group, compound, operation, etc.
$h_1(x)$	:	Set of equality constraints corresponding to process design specifications.
$h_2(x)$	:	Set of equality constraints corresponding to process model equations.
$h_3(x, y)$	:	Set of equality constraints related to molecular structure generation, mixing rules for properties, etc.
$g_1(x)$	:	Set of inequality constraints related to process design specifications.

$g_2(x, y)$: Set of inequality expressions corresponding to specific problems related to the product design.

$f(x)$: Vector of objective functions.

Many variations of the above mathematical formulation may be derived to represent problems and their corresponding solution methodologies (Gani, 2004). Some examples are:

1. Solve all the equations. This represents an integrated process-product design problem. The combined problem represents a complex mixed integer non-linear programming problem.
2. Only satisfy the constraints in Equations 2.2 – 2.7. This generates a feasible set of products and their corresponding process. Aspects of product-process design are considered simultaneously.
3. Solve a mathematical programming problem that includes Equations 2.1, 2.4 and 2.6. This is optimal product design of the molecule and/or mixture.
4. Satisfying the constraints in Equations 2.4 and 2.6. This is a chemical product design problem that generates molecular structures (or mixtures of molecules) and identifies a set of feasible candidates.
5. Satisfying only constraint 2.6. This represents a product design problem solution based on a database search.

2.2 Integrated Process and Product Design

The traditional solution methods to identify optimum solutions to integrated product and process design problems are forward and iterative in nature and thus may be cumbersome. Hence, identifying process/product performance targets beforehand and matching the solution alternatives

with the targets would make the solution methodology more efficient. This methodology can be stated as a reverse problem formulation. Also, as discussed in section 2.1, when process design and molecular design are handled separately, each of them have inherent limitations due to the nature of their input data. Solving process synthesis/design problems separately would require committing to specific raw materials well in advance in order to reach a solution. On the other hand, in molecular design problems, the desired target properties (dependent on the process) are required input to the solution algorithm. These decisions regarding the input data to the respective problems are made ahead of design and are usually based on experience and thus could possibly yield a sub-optimal design.

To overcome the limitations encompassed by decoupling the process and molecular design problems, a simultaneous approach as outlined in Figure 2.2 has been proposed (Eden, 2003). Using this approach, the molecular building blocks and the desired process performance are given as input to the integrated design problem. The final outputs of the algorithm are the design variables, which facilitate the desired process performance target(s) and the molecules that satisfy the property targets identified by solution of the process design problem.

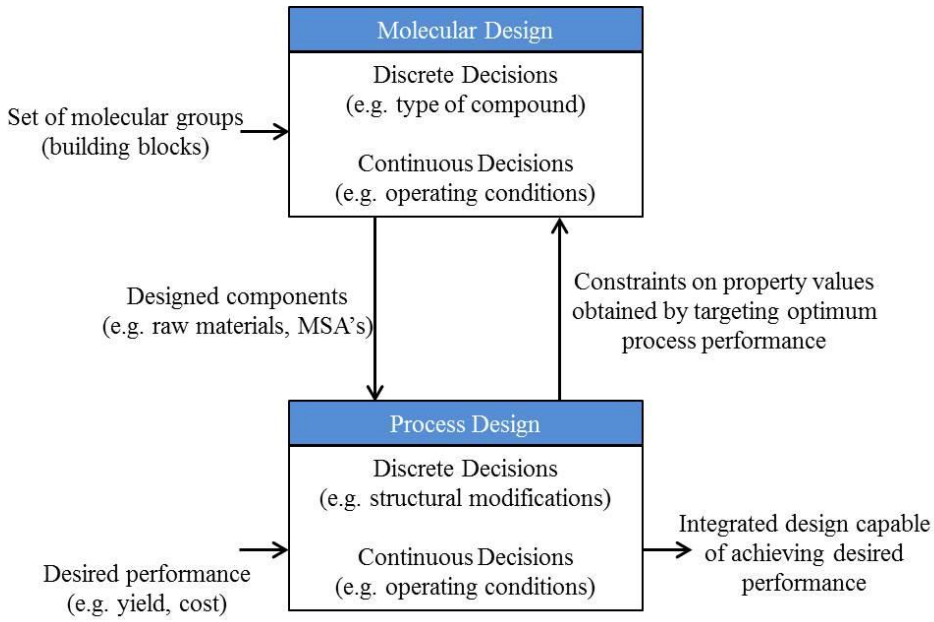


Figure 2.2: Simultaneous solution of process and molecular design problems (adapted from Eden (2003))

As explained in section 2.1.1, when process and product design problems are solved simultaneously, the models involved tend to be highly non-linear. The concept of reverse problem formulation (RPF) has helped formulate integrated process-product design problems without leading to MINLP formulations by insightful decoupling of the constitutive equations (property models) from the process model (Eden, Jørgensen, Gani, & El-Halwagi, 2003a). Reverse problem formulation enables design of novel molecules and solution of process design problems without commitment to specific components during the solution step. One of the challenges in applying such a method is that, the process design problem is solved in terms of the properties and not in terms of components. A systematic way to track properties is presented in the chapter 4.

2.2.1 Property Models

Any mathematical model for a product or process consists of three types of equations, i.e. balance equations, constitutive equations and constraint equations (Russel, Henriksen, Jørgensen, & Gani,

2002). The constitutive equations consist of a set of selected property models which play different roles in the simulation and design calculations (Gani & O'Connell, 2001). The service role by property models is when model parameters are given and the process/product model requests the property values. These types of models are used primarily in process simulators. The service and advice role is played by property models in process design and synthesis problems. Process design and synthesis problems are solved in two steps – (a) a step where alternatives are generated – the property models attempt to identify constraints on feasible conditions of operation and optimum values of process conditions thus providing advice to the synthesis and design algorithms in terms of eliminating infeasible solutions; and (b) a step where properties are determined and alternatives are verified – the property models play only a service role here. Since the property model can provide design targets along with constraints on feasible property values, it is possible to include the property model as a part of the solution routine, thus adding a solve role to the service and advice roles.

When property models (constitutive equations) are used in the solve role, they are decoupled from the process model and solved separately (Eden et al., 2004). Furthermore it must be emphasized that by performing this decoupling, the information flow to and from the property model is also reversed, i.e. the process model is solved for the values of the constitutive variables (properties), and then the property model is solved to yield the corresponding intensive variables, e.g. process conditions, process flowsheets or products (including molecular structures). Also, by setting up the problem to use the property model in the solve mode, it is possible to use different property models for same variable at different stages of the solution.

Knowledge of some common specific properties (constitutive variables) is needed to establish the feasibility of a unit operation in a process and the corresponding conditions of operation. The same

information is needed for design of a component as an appropriate external agent. Hence, the constitutive variables that are used to analyze the processes and products in a system allow process synthesis, process design and molecular design problems to interact with each other.

2.2.2 Reverse Problem Formulation

A mathematical model consisting of balance equations, constraint equations and constitutive equations may be a mixed integer non-linear problem. Though several techniques to handle these kinds of problems are available, in practice, these problems tend to be really hard to solve: they combine the combinatorial nature of mixed integer programming and intrinsic difficulty of nonlinear programs. Decoupling the equations involved in the model and solving them piecewise in an integrated fashion to achieve a common constitutive variable would be an efficient way of solving these models (Eden et al., 2004).

The procedure developed by (Eden, 2003) as illustrated in Figure 2.3 for decoupling the constitutive equations assists in solving the MINLP formulations. The decoupling of the constitutive equations as illustrated provides the foundation for two reverse problem formulations:

1. Given input stream(s) variables, equipment parameters and known output stream(s) variables, determine the constitutive variables.
2. Given values of the constitutive variables, determine the unknown intensive variables (from the set of temperature, pressure and composition) and/or flowsheet structure and/or product.

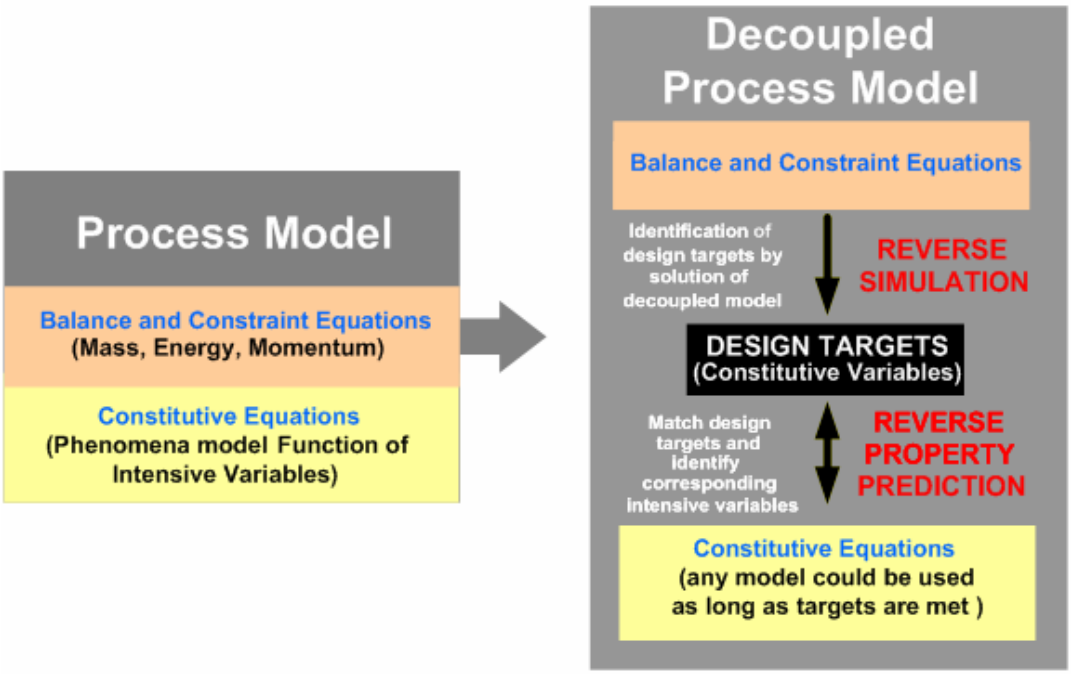


Figure 2.3: Reverse problem formulation (adapted from Eden (2003))

As the complex constitutive equations are separated from the model, the solution step to the first reverse problem is easy. In addition, for the second reverse problem, any number of property models can be used (as needed to describe entire processes) as long as the target constitutive variable values identified by the first reverse problem are matched. It is possible to have more than one solution since the algorithm involves a matching procedure. Therefore, a performance index can be defined and evaluated for all identified solutions to determine the optimal solution.

2.3 Process Synthesis and Design

Process synthesis and design deals with the determination of an optimal flowsheet configuration including the required tasks, appropriate equipment capable of converting the feed streams to product streams. In addition, the design of the equipment and their operating conditions need to be determined. Once a feasible flowsheet has been identified, it is analyzed/tested to make sure the process objectives are met. Finally, to gain this detailed understanding of how the process behaves

and whether the process objectives are met, process analysis tools such as ASPEN Plus, PRO II, and HYSYS are often utilized (El-Halwagi, 2006).

Approaches for process synthesis and design include:

a. Heuristic and Knowledge-based Approaches: These approaches are based on a set of rules developed through experience and available data. In such methods, basically the available knowledge (e.g., known operation tasks or processes for achieving a particular task) is first captured in a systematic manner and mined appropriately for a specific problem based on certain rules and procedures and finally this knowledge is applied to the problem. Hence, such methods, when automated, mimic the human approach to solving these problems, where humans search for relevant existing data and apply useful information from it to the current problem. These rules help in fixing some discrete variables in advance, leading to a reduction in the size of the solution search space. Without these rules, design problems can often be too difficult to converge and/or too large to search, however here again the optimality of the generated solution may not guaranteed (Westerberg, 2004). Also the rules sometimes may be contradictory as the context in which they can be applied is not necessarily fully defined and this approach is useful only in cases when the problem to be solved is similar to previously solved problems (El-Halwagi, 1997). One of the methods meeting this criterion is the one developed by Douglas (1985). This framework is for separation system design and the framework is divided into two parts, namely vapor and liquid recovery, and each part is governed by a set of heuristic rules for the selection of separation tasks. Several solution techniques along the same lines have also been developed (Barnicki & Fair, 1990, 1992; Chen & Fan, 1993). In a strategic process synthesis method developed by Siirola (1996), a library of

various sets of unit operations called “islands” are made; critical unit operations are selected from these islands and are interconnected to obtain the final process flowsheets.

b. Mathematical optimization approaches: Here the process synthesis and design problem is solved using optimization techniques. These methods usually need to obtain a superstructure of all possible alternative flowsheets. Hence, the optimality of the solution solely depends on the comprehensiveness of the mathematical superstructure. Usually, representation of such large optimization problems is in the form of Mixed Integer Non-Linear Programs (MINLPs) which are computationally intensive, requiring efficient solvers to obtain a global optimal solution. The MINLP problem as described by Grossmann, Aguirre, and Barttfeld (2005) involves discrete linear variables *(y)* and continuous non-integer variables *(x,)* as shown below. The goal of the MINLP formulations is to maximize/minimize one or more of the process specifications, e.g. minimizing cost, maximizing throughput, and/or efficiency etc.

$$F_{obj} = min[C^T y + f(x)] \quad \textbf{2.8}$$

Subject to:

$$h(x) = 0 \quad \textbf{2.9}$$

$$By + g(x) \leq 0 \quad \textbf{2.10}$$

$$x \in X, y \in \{0,1\}^m \quad \textbf{2.11}$$

The mathematical superstructure determination has been addressed by Friedler, Tarjan, Huang, and Fan (1993) using a graphical approach. Shah and Kokossis (2002) proposed a task based approach, where tasks represent simple and/or complex distillation column configurations, to generate the superstructure and the corresponding MINLP formulation. McCarthy, Fraga, and Ponton (1998) introduced an automated procedure for product separation synthesis. First, the

procedure performs an in-depth tree search to locate solutions and unit operations, applying design variable discretization to reduce the search space. This methodology has the benefit of avoiding mapping into an apriori generated superstructure. In all these methods, the algorithm generates a set of good, feasible solutions which may be further optimized by continuous means.

c. Hybrid methods: The hybrid approaches combine functionalities of the different approaches described above into one. Often these methods combine the physical insights of knowledge based methods with mathematical programming techniques to formulate and solve process synthesis and design problems. While the simplicity of the knowledge-based methods is carried into these hybrid techniques, rather than heuristics, fixed rules and guidelines based on physico-chemical properties of the components involved in the process are used for process synthesis and design.

In this section, thermodynamic insights based process synthesis of separation processes, the driving force based synthesis and design of separation processes and attainable region analysis for reactor network design are discussed.

Thermodynamic insights based flowsheet synthesis:

Jaksland, Gani, and Lien (1995) developed a method that uses thermodynamic insights for synthesis of separation processes rather than relying on heuristics. The knowledge about a process is retrieved from the physico-chemical properties of the components in the mixture. This method is hierarchical and consists of two main levels: a) the first level calculates the difference in component properties as ratios over a wide range of properties, which in turn are used as a screening criteria to identify the feasible separation techniques. The separation technique is selected in such a way that it exploits the largest property differences between

components of the mixture to be separated. b) in the second level, a detailed mixture analysis is done for further screening. If any external mass separating agents are required in the process, the process design problem is integrated with a molecular design problem.

Driving force based process synthesis and design:

Gani and Bek-Pedersen (2000) introduced the concept of driving force based separation design. The method developed enables fast and easy identification of near optimal design without having to resort to computationally intensive calculations. The Driving Force (DF) for any separation task to be carried out by a given separation technique is the difference in technique specific chemical/physical properties between two coexisting phases that may or may not be in equilibrium. Hence, when the DF is used as selection and sequencing criterion, such that the maximum driving force is utilized at all stages of the process, the most efficient separation system is quickly found. Also, the design of each separation unit (e.g. the number of plates in a column, feed location, solvent requirement and its properties etc.) is evaluated as a function of the maximum driving force. By targeting each unit operation at the largest possible driving force, a near optimal separation sequence can be obtained.

Attainable region for reactor networks:

Horn (1964) introduced a concept called attainable region (AR) analysis which enables simpler, easier, and more robust reactor design and optimization. In attainable region analysis, all possible output concentrations in the stoichiometric subspace from different reactor configurations are determined apriori and the optimal reactor network is found from it. Approaching the problem from this direction ensures that all reactor systems are included in the analysis. There are several examples of methods for the construction of

such regions, e.g. the geometric approach by Glasser, Crowe, and Hildebrandt (1987) and the algorithmic method presented by Hildebrandt and Biegler (1995). Once the attainable region is identified, graphical analysis and solution of simple problems is relatively easy and in the case of more complex problems, the AR can assist in the formulation of the constraints in a mathematical optimization problem.

The integrated approach by Hostrup (2002) combines the thermodynamic insights of Jaksland et al. (1995) and Gani and Bek-Pedersen (2000) with the formulation of structural optimization problems, thus allowing for efficient screening among the alternative routes. d'Anterroches and Gani (2005) provided the basis for a computer aided flowsheet design (CAFD) framework based on group contribution approaches. In this group contribution approach, process groups with their apriori regressed property contributions are used as building blocks. CAFD is solved in a reverse fashion which enables their easy integration with CAMD. In the hybrid methods, the flowsheet synthesis problem is solved as a reverse property prediction problem. Here, given the property target values and/or their functions, the unknown process configurations that match the property targets are identified. The flowsheet design problems are solved by reverse simulation formulations. Here, the design variables are back calculated from the simulation models.

2.3.1 Process Integration

Since chemical processes are integrated systems of interconnected units and streams, an effective process is possible only by accounting for process integration. Process integration is a holistic approach to process design, retrofitting and operation which emphasizes the unity of the process (El-Halwagi, 1997). Based on the two main commodities consumed and processed in a typical facility, namely mass and energy, process integration is categorized into mass integration and

energy integration. Mass integration is a systematic methodology that provides the fundamental understanding of the global flow of mass within the process and employs this understanding in identifying the performance targets and routing the species in a process. Energy integration, on the other hand provides an understanding of energy utilization within the process, thus using it to identify energy targets and optimize heat-recovery and energy-utility systems. There is a rich volume of information available in literature that covers the development and uses of energy and mass integration tools (Cerda, Westerberg, Mason, & Linnhoff, 1983; Dunn & Bush, 2001; El-Halwagi, 1997; Gundersen & Naess, 1988; Linnhoff & Hindmarsh, 1983; Shenoy, 1995). Many processes are driven and governed by properties or functionalities of the streams and not by their chemical constituency. Constraints on process units that can accept recycled/reused process streams and wastes are not limited to compositions of components but are also based on the properties of the feeds to processing units (El-Halwagi, 2006). Since properties (or functionalities) form the basis of performance of many processes, design procedures based on key properties instead of key compounds are used. But, unlike mass, properties are not conserved and cannot be tracked among units without undertaking component material balances. Therefore, to resolve these limitations, conserved property-based clusters are used (Shelley & El-Halwagi, 2000).

Section 2.3 gave an overview of currently utilized process synthesis and design methods. The next sections will concentrate on the methodologies in molecular design algorithms, and the importance of property models for molecular design.

2.4 Computer Aided Molecular Design

Traditionally the search for solvents or products for specific applications has been carried out by looking for them in a database of known compounds. A more systematic way to finding a solution to such problems is computer aided molecular design (CAMD). However, both approaches need thorough experimental validation before they are put to use. By following a systematic approach one would be able to look for novel compounds and also trim the list to be experimentally tested by an exponential factor in comparison to the traditional methods. By definition, a CAMD problem is (Brignole & Cismondi, 2002): Given a set property constraints, determine the molecule or molecular structure that matches these desired physico-chemical and/or environmental properties. The structures of the compounds are represented using descriptors along with an algorithm that identifies these descriptors. This means the property evaluation methods would also be based on these descriptors.

The general approach to solving a CAMD problem is to first generate feasible molecular structures using the set of descriptors (also called building blocks) and then testing them by estimating their desired properties. These properties are estimated based on the apriori calculated values for each descriptor participating in a molecular structure. The set of feasible compounds are identified as those that match the property specifications. The optimal among them is obtained through a problem specific selection criterion. The principal differences between various CAMD methodologies are how the various steps in CAMD are performed, the type of descriptors used and how the necessary property values are obtained. In the method developed in this dissertation (see Chapter 3), CAMD includes building blocks (first- and higher-order groups) used to generate and represent molecules; group contribution based property models to predict target properties

(Marrero & Gani, 2001); and a synthesis method to generate and screen molecules that match the target (design) properties.

2.4.1 Formulation of property constraints

A set of properties with specific goal values or lower/upper bounds are identified here and are problem specific, e.g. if a given chemical must be liquid at certain conditions it should be translated into constraints on melting and boiling temperature. The property values can be directly determined through a property model. Some properties however, cannot be explicitly described in this way, e.g. smell, taste, etc. These properties can in some cases be represented as a function of explicit properties.

While formulating these constraints, the questions below could help to define the design boundaries (Harper, 2000):

a. Is the compound a replacement for another compound?

   If yes, the constraints can be selected similar to or better than those of the existing compound based on its drawbacks.

b. What would be the operational limits?

   These limits help in defining the upper and lower limit of the constraints on the phase and the phase transition related properties.

c. What criteria should be used to evaluate the performance of the desired product?

   The performance criteria are related to the function of the desired product in the process for which it is being designed. Sometimes, models for evaluation of performance may be very complex.

d. Are there any downstream processing considerations?

When compounds are designed to play a role in downstream processes, in order to obtain a global solution to the CAMD problem, the operational limits of the compounds need to be extended to cover additional possible operations and consequently, other properties may also have to be considered. The possible utilization of available process streams to be mixed with the new compound can be studied here. Due to the evident link between process and product, the molecular design and process design problems should be integrated.

Having the property constraints in hand from the above set of questions and estimation methods to predict the selected properties, an appropriate molecular synthesis algorithm is needed to obtain a CAMD solution.

2.4.2 Molecular Design Algorithms

All CAMD algorithms reported in literature, fall into three main categories: mathematical programming, stochastic optimization, and enumeration techniques (Harper, 2000).

a. Mathematical programming: In solving CAMD problems using optimization (mathematical programming) techniques, the property constraints identified are used as mathematical bounds and the performance requirements are defined by an objective function. Solutions techniques to such optimization problems in general involve solving Mixed Integer Non-linear Programming (MINLP) models. Although widely used and proven to be effective, MINLP methods suffer from a large computational load and lack the guarantee of finding a globally optimal solution. (Duvedi & Achenie, 1996; Pistikopoulos & Stefanis, 1998; Vaidyanathan & El-Halwagi, 1994).

b. Stochastic optimization: Using this method, the solution alternatives are generated by trying random variations of the current solution. Analogous to general optimization problems, this method also aims at finding the optimal value for the objective function, but

the technique it uses varies. The nature of the solution methodology involved here gives the freedom to specify discontinuous properties as the involved optimization methods do not require any gradient information. There are two forms of stochastic optimization based CAMD algorithms: a) A simulated annealing approach that has the ability to easily deal with highly non-linear models (e.g. predictive property models) and large numbers of decision variables (e.g. numerous alternative molecular structures). "The algorithm runs as an iterative process in which, possible parameter modifications generate new parameter values, according to a set of perturbation probabilities" (Marcoulaki & Kokossis, 1998). The generated parameters are tested against previous values in each iteration to satisfy a probability criterion. b) A genetic algorithm approach in which a population of possible solutions (called individuals) is evolved toward better solutions. The evolution usually starts from a set of randomly generated individuals and is an iterative process, with the population in each iteration being called a generation, where the individuals exist based on "survival of the fittest"; i.e. the more fit individuals, usually based on objective function value, are selected from the current population and each individual is modified to form a new generation. Because of the stochastic nature, both approaches are capable of handling non-linear models, although as the problem complexity increases, the genetic algorithm approach reports challenges in terms of computational time (Marcoulaki & Kokossis, 1998; Venkatasubramanian, Chan, & Caruthers, 1994).

c. Enumeration techniques: Here the structurally feasible molecular structures based on group contribution methods are first generated using a combinatorial approach and are then tested against the specifications, where molecules that fail to satisfy the constraints are eliminated. As with stochastic optimization, no gradient information is needed here but the

disadvantage is that solving a CAMD by simple enumeration may lead to combinatorial explosion (Constantinou, Bagherpour, Gani, Klein, & Wu, 1996; Friedler, Fan, Kalotai, & Dallos, 1998; Gani, Nielsen, & Fredenslund, 1991; Joback & Stephanopoulos, 1995). Using some rules, however, this method can be made more effective. This dissertations aims at framing such rules and solving CAMD by rule based enumeration and test methods. A new "generate and test" method was introduced by Harper (2000). Here, the feasible formulations are generated from molecular building blocks using a rule based combinatorial approach. This method uses a multi-level CAMD approach that controls the generation and testing of molecules. Chemmangattuvalappil, Eljack, Solvason, and Eden (2009) also developed an enumeration and test CAMD algorithm which considers proximity effects of the atoms participating in a molecule.

2.4.3 Group Contribution Methods and Property Models

Many CAMD techniques use group contribution methods (GCM) to synthesize molecules and verify whether the generated molecules exhibit the specified set of desirable properties. These techniques prove to be powerful tools for primary estimation of reasonably accurate results for many property values when experimental data is not readily available. Generally in these kinds of methods, various groups (molecular fragments) are tabulated along with their contributions towards a property of the molecule possessing these fragments. These contributions do not depend on the position of the fragment in the molecule or nature of the molecule in which it exists. These contributions are estimated through regression of large amounts of experimental data. The property model equations for a set of tabulated data depends on how these values are regressed and are unique with respect to each set of GCM data.

In the case of simple compounds, GCM can provide accurate trends. However, as the complexity of the molecule increases, the accuracy of first order GCM becomes less reliable. They generally cannot capture proximity effects or differentiate between isomers (Kehiaian, 1983; Wu & Sandler, 1989, 1991). So, several attempts have been made to make the GCM more general and reliable (Constantinou, Prickett, & Mavrovouniotis, 1993; Fedors, 1982). The ABC method introduced by Constantinou et al. (1993), though computationally challenging, provided the basis for future GC methods. The ABC method is based on the contributions of atoms and bonds towards the properties of different conjugate forms of a molecular structure. Here, the property of a molecule has been estimated as the linear combination of contributions from all the conjugate forms of the molecule.

Group Contribution models with higher levels:

In the GC approach by Constantinou and Gani (1994), where, property estimation is done in two stages, two types of molecular building blocks have been defined: first- and higher-order groups. The higher-order groups give an idea about different types of interactions among the first-order groups and the effects of certain molecular group combinations to the property of the final molecule and could possibly differentiate among isomers. The higher-order groups enable a good representation of poly ring compounds and open-chain polyfunctional compounds (Marrero & Gani, 2001). The molecular groups from Marrero and Gani (2001) are used in the methodology developed in this dissertation and their definition and classifications/contributions are provided in Appendix A.

The property estimation model suggested in this approach has the following form:

$$f(X) = \sum_i N_i C_i + w \sum_j M_j D_j + z \sum_k O_k E_k \qquad \textbf{2.12}$$

where,

f(X) is a function of the actual property *X*, C_i is the contribution of first order group *i* that occurs N_i times, D_j the contribution of second order group *j* that occurs M_j times and E_k the contribution of third order group *k* that occurs O_k times in the molecule. The constants *w* and *z* can have values of zero or unity depending on how many levels of estimation are of interest.

The primary properties and the corresponding property functions when using Marrero and Gani groups are listed in Table 2.1. The universal constants for each property function are given in Table 2.2. There are several secondary properties like vapor pressure, flash point, etc. that can be estimated as functions of primary properties. It should be noted that the application of group contribution based CAMD techniques rely on the availability of molecular groups and the estimated property contributions corresponding to each group.

Table 2.1: Group Contribution Models

Property	Property function	Group contribution terms
Normal melting point, T_m	$\exp\left(\frac{T_m}{T_{m0}}\right)$	$\sum_i N_i T_{m1i} + \sum_j M_j T_{m2i} + \sum_k O_k T_{m3i}$
Normal boiling point, T_b	$\exp\left(\frac{T_b}{T_{b0}}\right)$	$\sum_i N_i T_{b1i} + \sum_j M_j T_{b2i} + \sum_k O_k T_{b3i}$
Critical temperature, T_c	$\exp\left(\frac{T_c}{T_{c0}}\right)$	$\sum_i N_i T_{c1i} + \sum_j M_j T_{c2i} + \sum_k O_k T_{c3i}$
Critical pressure, P_c	$\frac{1}{\sqrt{(P_c - P_{c1})}} - P_{c2}$	$\sum_i N_i P_{c1i} + \sum_j M_j P_{c2i} + \sum_k O_k P_{c3i}$

Critical volume, V_c	$V_c - V_{c0}$	$\sum_i N_i V_{c1i} + \sum_j M_j V_{c2i} + \sum_k O_k V_{c3i}$
Standard Gibbs Free energy, G_f	$G_f - G_{f0}$	$\sum_i N_i G_{f1i} + \sum_j M_j G_{f2i} + \sum_k O_k G_{f3i}$
Standard enthalpy of formation, H_f	$H_f - H_{f0}$	$\sum_i N_i H_{f1i} + \sum_j M_j H_{f2i} + \sum_k O_k H_{f3i}$
Standard enthalpy of vaporization, H_v	$H_v - H_{v0}$	$\sum_i N_i H_{v1i} + \sum_j M_j H_{v2i} + \sum_k O_k H_{v3i}$
Standard enthalpy of fusion, H_{fus}	$H_{fus} - H_{fus0}$	$\sum_i N_i H_{fus1i} + \sum_j M_j H_{fus2i} + \sum_k O_k H_{fus3i}$

Table 2.2: Adjustable parameters in Group Contribution Models

Adjustable parameter	Value
$\boldsymbol{T_{m0}}$	147.45 K
$\boldsymbol{T_{b0}}$	222.543 K
$\boldsymbol{T_{c0}}$	231.239 K
$\boldsymbol{P_{c1}}$	5.9827 bar

P_{c2}	0.108998 $bar^{-0.5}$
V_{c0}	7.95 cm^3/mol
G_{f0}	-34.967 kJ/mol
H_{f0}	5.549 kJ/mol
H_{v0}	11.733 kJ/mol
H_{fus0}	-2.806 kJ/mol

2.5 Summary

This chapter provides an overview of chemical process and product synthesis/design. Because of the huge amount of data involved and the non-linear nature of the mathematical formulations involved in process and product design problems, computer aided solution techniques prove to be convenient ways of solving these problems. Process and product synthesis/design problems are explained along with the necessity to integrate them for efficient solutions to a given task. The three different roles of property models are described and the concept of reverse problem formulation (RPF) has been explained to illustrate the advantages of utilizing RPF in process and product design. Finally, the applications of process integration and RPF in the simultaneous consideration of process and product design problems are introduced along with the targeting method to decouple the property models from the design equations.

There is a definite need for solving process and molecular synthesis/design problems together, as the solution space is limited if these problems are solved separately due to the amount of

information that is required prior to invoking the design algorithm. To overcome these limitations, a hybrid method for Computer Aided Flowsheet Design (CAFD) and its effective integration with molecular design (CAMD) is proposed by incorporating the benefits of the principal concepts outlined in this chapter. Using this approach, the process along with its design variables and the molecules, which facilitate the desired process performance target, are identified.

3. Computer Aided Molecular Design

3.1 Molecular Design by decomposition based approach

Molecular design involves identifying a compound or a collection of compounds having specified properties while the structure of these compounds (molecules) is represented using appropriate molecular descriptors. Hence the objective here is; given the building blocks (descriptors) and a specified set of target properties; to find an algorithm that identifies the given input (descriptors and property targets), processes them subject to structural and property constraints and finally determines the molecule that matches these properties.

The methodology developed in this dissertation can be termed as a solution to a reverse property prediction problem. Typically a forwardly formulated problem would be designing various molecules and testing if they exhibit the targeted performance. This kind of treatment of the molecular design problem would obviously suffer from problems owing to its iterative nature. But here, product performance targets in terms of properties are identified beforehand and solution alternatives with the targeted performance are designed. This is shown in Figure 3.1.

It is evident from the Figure 3.1 that a forward problem has to be solved firsthand to have the molecular groups and property models in hand. The molecular groups by Marrero and Gani (2001) consisting of descriptors along with the property evaluation methods based on these descriptors is a result of such forward problem and is used in the developed methodology. These are provided in

appendix. How efficiently the reverse problem shown in Figure 3.1 is addressed is proportional to obtaining solution molecules with less computational load.

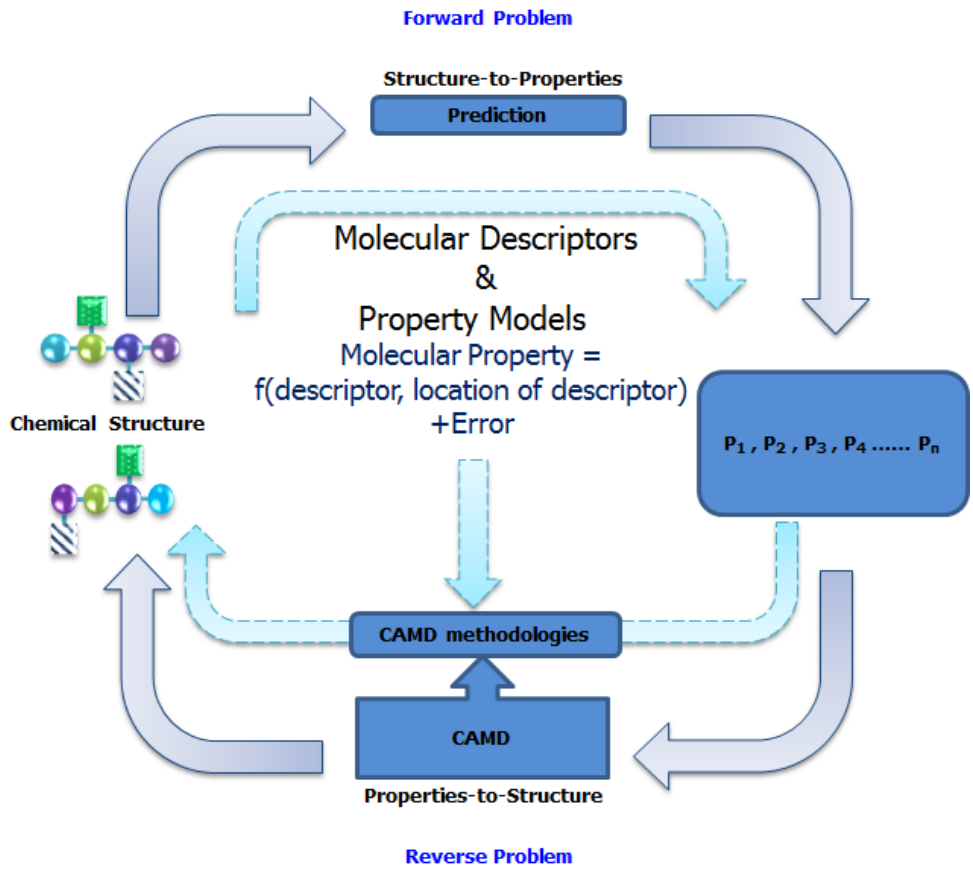


Figure 3.1: Reverse Problem Formulation of Molecular Design.

Numerous contributions have been made in the field of Computer-Aided Molecular Design (CAMD). Many of these methods include the use of Group Contribution Methods (GCM) which utilize tables comprising of various molecular fragments/groups and their contribution towards a property in the molecule. Higher order groups are also given in these tables to better explain the change in the contribution of a group towards a property due to its neighboring groups. Employing a systematic methodology to design molecules based on GCM decreases the permutations of groups that need to be checked if they have a valid molecule hidden in them. Algorithms to identify

the molecules that meet the process targets have been developed by many research groups, including (Marcoulaki & Kokossis, 1998), (Harper & Gani, 2000), (Eljack & Eden, 2008), and (Chemmangattuvalappil et al., 2009). In this contribution, the focus is to introduce methods that improve the efficiency of solving the molecular design problem as part of the integrated solution of process and product synthesis/design problems. Earlier methods either did not efficiently incorporate higher groups within the algorithm or did not incorporate the contribution of higher order groups during the initial stage thus increasing the number of combinations which need to be checked whether the molecule leading to given process performance is structurally sound. Incorporating higher order groups at a later stage in the algorithm may also lead to a situation where some potential combinations are omitted without being considered in further stages of the algorithm.

3.2 Processing of Molecular Descriptors

The molecular groups (descriptors) describing a compound is a collection of three types of groups: first-order groups, second-order groups and third-order groups. The first-orders groups are intended to describe a wide variety of organic compounds, while the role of the second and third-order groups is to provide more structural information about molecular fragments of compounds whose description is insufficient through the first-order groups. This kind of segregation ensures unique representation of a wide number of compounds.

Constraints by the virtue of the nature of molecular groups

Based on the fundamental principles behind the formation of molecular groups (Marrero & Gani, 2001), the following rules are articulated:

***Rule 1*:** The molecule must be described entirely by first-order groups.

Rule 2: There must be no overlap between first-order groups.

Rule 3: If the same fragment of a given compound is related to more than one first-order group, the heavier group must be chosen to represent it instead of the lighter groups.

Rule 4: The decision on whether groups are to be part of a ring or aromatic ring compound should be made ahead of design because the property contributions of the same group is different in aromatic, cyclic and acyclic compounds.

Rule 5: A detailed first-order representation of aromatic compounds should be provided at a first level of estimation, i.e. for an aromatic substituent, aC-R group should be considered over an aC group.

Rule 6: The entire molecule does not need to be described with higher-order groups (second and first-order groups).

Rule 7: Higher-order groups have first order groups as building blocks.

Rule 8: Double and triple bonds are included within the first order groups, i.e. two first order groups connect by only a single bond.

Rule 9: Higher-order groups are allowed to overlap with each other.

Rule 10: If any of the higher order groups completely embodies some other higher order group, only the larger group must be chosen in order to prevent redundant description of the same molecular fragment.

Structural Constraints:

In graph theory, a branch of mathematics (Bondy & Murty, 2008), the handshaking lemma is the statement that every finite undirected graph has an even number of vertices with odd degree. This establishes the following relationship between the sum of the node degree, deg p, and the number of graph edges, q.

$$\sum \deg p = 2 \sum q \qquad \textbf{3.1}$$

In the case of molecular design each first-order group is represented by a node in a graph

$$\sum \deg p = \sum_{f=1}^{N_f} n_f \, FBN_f \qquad \textbf{3.2}$$

where, *f* stands for the first-order groups ranging from *1* to N_f, n_f is the number of first-order group, *f*, FBN_f is the free bond number of first-order group *f* (valence of a group *f*)

In the case of acyclic molecules, these molecules can be viewed analogous to trees in conventional graph theory. Hence, according to basic graph theory (Bondy & Murty, 2008):

$$q = \sum_{f=1}^{N_f} (n_f) - 1 \qquad \textbf{3.3}$$

In case of cyclic molecules or mixed structures involving cyclic and acyclic fragments:

$$q = \left(\sum_{f=1}^{N_f} (n_f) - 1 \right) + N_{rings} \qquad \textbf{3.4}$$

where, N_{rings} is the number of rings in the molecule.

In case of aromatic compounds or mixed structures involving aromatic, cyclic and acyclic fragments:

$$q = \left(\sum_{f=1}^{N_f} (n_f) - 1 \right) + N_{rings} + N_d \qquad \textbf{3.5}$$

where, N_{rings} here considers even the number of aromatic rings along with non-aromatic rings in the molecule, N_d is the number of alternating double bonds in the aromatic rings.

Finally, the Free Bond Number, *FBN* of a molecule is given by:

$$FBN = \sum_{f=1}^{N_f} n_f \, FBN_f - 2\left[\left(\sum_{f=1}^{N_f}(n_f) - 1\right) + N_{rings} + N_d\right] \qquad \textbf{3.6}$$

Rule 11: The Free Bond Number, *FBN* of a molecule is zero. This ensures that there are no free hanging bonds in any molecule.

Rule 12: The number of any first-order group is non-negative.

Rule 13: n_r, the total number of first order groups forming cyclic fragments should be at least three in case a cyclic molecule is allowed and n_{ac}, the total number of first order groups forming aromatic fragments is exactly six or multiples of six. It is also possible that n_{ac} be assigned a value other than multiples of six for fused ring compounds.

Molecular property feasibility constraints:

The formulation of property constraints is a prerequisite for solving molecular design problems. A set of properties is selected as constraints with some combination of specific goal values, lower and upper bounds.

The property values can be directly determined through a property model. Some properties however, cannot be explicitly constrained like smell, taste etc. These properties can in some cases be represented as a function of explicit properties.

A group contribution method is one that uses the principle that some simple aspects of the structures of chemical components are always the same in many different molecules. The smallest common constituents are the atoms and the bonds or more complex building blocks like the molecular groups, which are themselves built of few atoms and bonds. These components behave in a similar fashion irrespective of which molecule they exist in or their position in a given molecule

Rule 14: The contribution of any first-order group towards a molecule's property is independent of the molecule in which the group occurs.

In group contribution methods, to predict properties of pure components and mixtures, group or atom properties are used. This reduces the amount of needed data dramatically. Instead of needing to know the properties of thousands or millions components, only data for a few dozen or few hundred groups have to be known. The property estimation model for the molecular groups considered in this molecular design algorithm is given by Marrero and Gani (2001) and is mathematically represented as:

$$f^M(P_k) = \sum_{f=1}^{N_f} n_f P_{k,f} + \sum_{h=1}^{N_h} n_h P_{k,h} \qquad \textbf{3.7}$$

where,

f stands for the first-order groups ranging from *1* to N_f, n_f is the number of first order group, *f*, $P_{k,f}$ is the contribution of first-order group, *f* towards molecular property, P_k ;

h stands for the higher-order groups ranging from *1* to N_h, n_h is the number of higher-order group, *h*, $P_{k,h}$ is the contribution of first-order group, *h* towards molecular property, P_k (the higher order groups constitute the second- and third-order groups*;* and $f^M(P_k)$ is a function of property, P_k and $f^M(P_k)$ is obtained by applying a respective property operator/property function to a property value (as given in Table 2.1).

For some of the groups defined by Marrero and Gani (2001), their contribution towards a molecular property is not available. In these cases, to better estimate the properties for molecules comprising these groups, a property model developed by Gani, Harper, and Hostrup (2005) based on connectivity indices described by Kier and Hall (1986) is used. The Connectivity Indices (CI) based property model is given by Equation 3.8.

$$f(Y_g) = \sum_i a_i A_i + b({}^{v}\chi^0) + 2c({}^{v}\chi^1) + d \qquad \textbf{3.8}$$

where $f(Y_g)$ is the property contribution of group g , A_i is the number of i^{th}-atoms occurring in the molecular structure, ${}^{v}\chi^0$ is the zeroth-order (atom) connectivity index given by Equation 3.9, ${}^{v}\chi^1$ is the first-order (bond) connectivity index given by Equation 3.10, a_i is the contribution of atom i, b and c are adjustable parameters, and d is a constant.

$$ {}^{v}\chi^0 = \sum_i \left(\frac{1}{\sqrt{\delta_i^v}}\right) \qquad i = 1, L \qquad \textbf{3.9}$$

where, L is the number of atoms in the group and the values of δ_i^v are the atom indices whose values can be obtained from Table 3.1 for the corresponding atom.

Similarly,

$$ {}^{v}\chi^1 = \sum_k \left(\frac{1}{\sqrt{\beta^k}}\right) \qquad k = 1, M \qquad \textbf{3.10}$$

where M is the number of bonds in the group while the bond index β^k is given by Equation 3.11.

$$\beta^k = \delta_i^v \, \delta_j^v \qquad \textbf{3.11}$$

Equation 3.8 can be treated as the sum of property contribution by the group and its correction due to the effect of its surrounding groups, i.e.

$$f^M(Y_g) = f^M({Y_g}^1) + f^M({Y_g}^2) \qquad \textbf{3.12}$$

where,

$$f^M({Y_g}^1) = \sum_i a_i A_i + b({}^{v}\chi^0) + d \qquad \textbf{3.13}$$

$$f^M({Y_g}^2) = 2c({}^{v}\chi^1) \qquad \textbf{3.14}$$

Depending on the function of the desired product, operational limits in the process that the product is to take part of and downstream processing conditions; the property constraints are framed. Due

to an evident link between process and product design, the scope of integrating the molecular design with process design problems arises.

Rule 15: Having known from above, how to calculate the properties from participating molecular groups in a molecule, compounds satisfying the property constraints are valid final molecules.

Table 3.1: Values of the Atomic Index δ^v for Each Atom/Vertex (n_H is the number of connected hydrogen atoms) (Gani et al., 2005)

	C	Si	N	F	Br	Cl	I	Na	K	O	P	S
acyclic	4-n_H	4-n_H	5-n_H	7	7/27	7/9	7/47	1/10	1/18	6-n_H	5/9	
cyclic	14-n_H	14-n_H	15-n_H							16-n_H		"special" + 9
special			nitro: 6								PH_2: 1/3	SH: 5/9
											PH: 4/9	S: 2/3
												S: 8/3

Molecular structure feasibility constraints:

To limit the compounds generated by the molecular design algorithm to a complexity that the property estimation models described above can handle, or to limit the range of compounds to types more likely to be commercially available, an upper bound is imposed on the number of types of functional groups allowed in a molecule. These rules in a way confirm the stability of a molecule as molecules overly crowded with a variety of functional groups tend to be unstable.

The first-order groups can be divided into multiple subcategories and classes as proposed by Gani et al. (1991). This categorization and classification is given in Table 3.2. Additionally, an upper bound is imposed on the total number of groups from a given sub-class or category that can participate in a single molecule.

Table 3.2: Classification of Groups (Gani et al., 1991).

	Category				
Class	1	2	3	4	5
0	CH3OH CH3SH (CH2OH)2 NMP Diethyl glycol 2-Propanol	CH3NO2 CH3CN CH2CL2 CH3NH2 CCL3F C4H4S	H2O Furfural CHCL3 TCE Pyridine CHCL2F Morpholine	CH3NH2 HCOOH ACRY MFA 1-propanol CHCLF2	DMSO DMF TMS CS2 CCL2F2 CCLF3 CF3
	CH3	CH2CN CH2NO2 CH2NH2 I BR CL F	CH3CO CONH2 CONHCH3 CON(CH3)2 CH3NH CH2NH2	OH CHO COOH CH2CL CH3COO CH3O C2H5O2 CH3S CCL2F CH3NH CHCL2 SH CH2SH COO CCL3 CF3 CCL2F CHCLF HCOO	C4H3S C≡CH SiH3 CH2=CH
2	CH2	CHNO2	CH2CO CH2COO CH2O CONCH3CH2	CHNH2 CH2NH CHCL CONHCH2 C2H4O2 CH2S CH3N CCL2	CH=CH CH2=C C4H2S C≡C SiH2
3	CH		CON(CH2)2	CHNH CH2N CCL CH-O CHS	CH=C SiH SiH2O SiHO
4	C				C=C SiO, Si
5	ACH	ACCH3	ACOH ACNH2 ACCL ACNO2		
6		ACCH2			
7		ACCH			
8		AC			

The "A" denotes aromatic groups.

Rule 16: Selection of the number of first-order groups from each sub-category and class is subject to apriori well defined restrictions.

3.3 Mathematical model of the molecular design problem

Molecular design problem is solved using a decomposition based approach where the problem is divided into smaller sub problems, namely:

Subproblem 1: Maximizing the number of each first-order group that could possibly appear in the molecule.

Subproblem 2: Enumerating all group subsets of available first-order groups that could form at least one molecule.

Subproblem 3: Estimating possible higher order groups.

Subproblem 4: Eliminating property infeasible group subsets.

Subproblem 5: Forming final molecules.

Figure 3.2 gives a flow diagram of the developed CAMD methodology.

Each subproblem as well as its solution is explained below:

3.3.1 Problem prerequisites

Property constraints and set of first order groups are to be given as inputs to the model.

Step 1: The given property constraints are translated into maxima, $f^M(P_k^U)$, minima, $f^M(P_k^L)$ and goals, $f^M\left(P_m^{goal}\right)$ of the molecular property functions/operators using the corresponding functions in GCM.

Step 2: The first-order groups are sub grouped into acyclic, aromatic and cyclic groups. However, acyclic groups can be a part of mixed structures involving aromatic, cyclic and acyclic fragments.

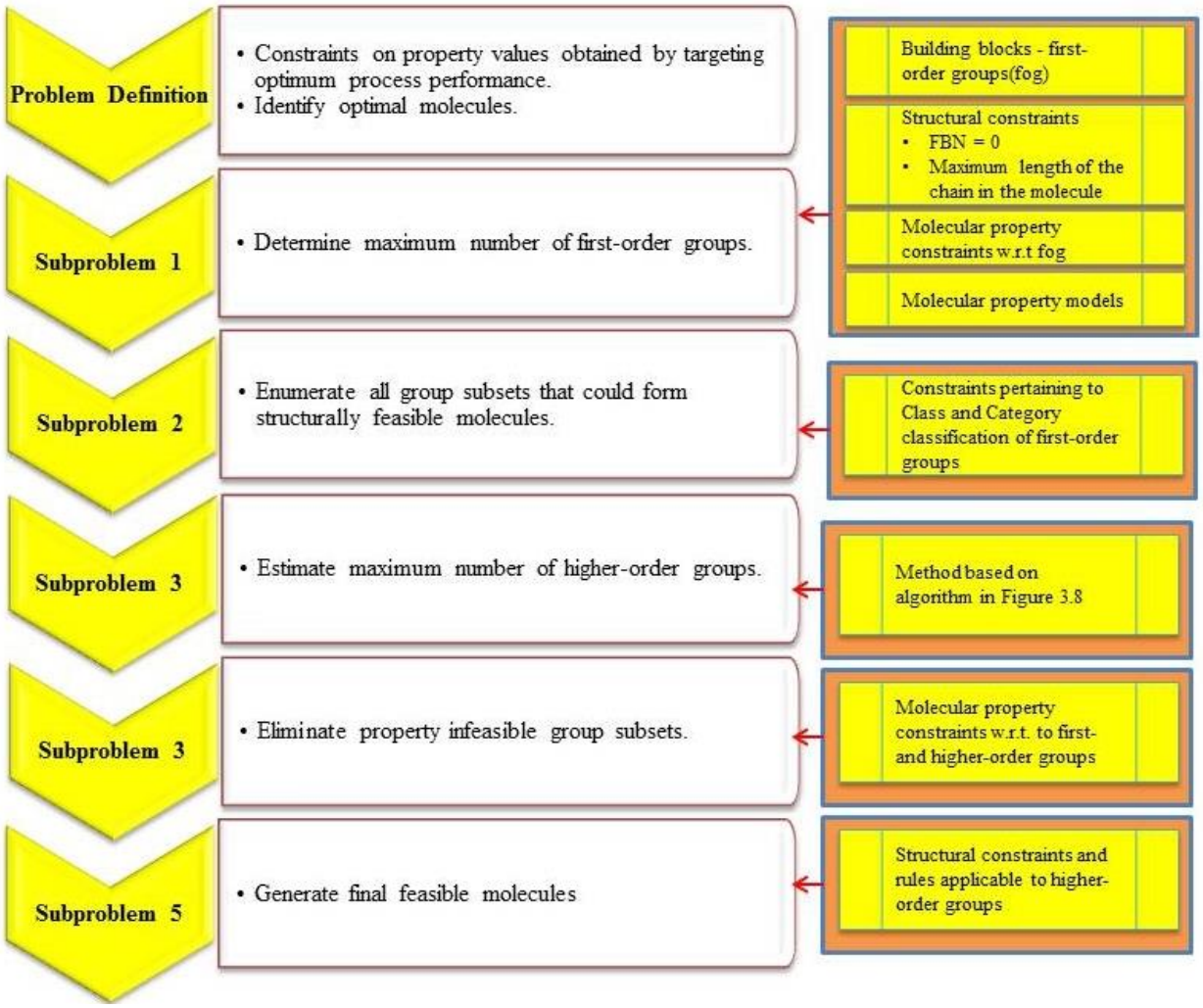


Figure 3.2: CAMD framework

3.3.2 Subproblem 1: Maximizing the number of each first-order group that could possibly appear in the molecule

Model:

Max n_f

Subject to:

Structural constraints

Rule 12:

$$n_f \geq 0 \quad \forall n_f = 1 \; to \; N_f \tag{3.15}$$

Rule 13:

$$n_r \begin{cases} = 0, no \; cyclic \; fragments \\ \geq 3, cyclic \; fragments \; exist \end{cases}$$

$$n_{ac} \begin{cases} = 0, no \; aromatic \; fragments \\ = multiples \; of \; six, polyring \; aromatic \; fragments \; exist \\ \geq 10, fused \; aromatic \; fragments \; exist \end{cases} \tag{3.16}$$

Rule 11:

$$FBN = \sum_{f=1}^{N_f} n_f \, FBN_f - 2\left[\left(\sum_{f=1}^{N_f}(n_f) - 1\right) + N_{rings} + N_d\right] = 0 \tag{3.17}$$

Here the problem needs to be solved differently for each specified set of N_{rings} and N_d.

Additionally, to constrain the length of the chains that the current group contribution model by Gani et al. (1991) can handle

$$\sum_{f=0}^{N_f} n_f \leq 12 \quad \forall n_f = 1 \; to \; N_f \tag{3.18}$$

Rule 15

$$f^M(P_k^L) \leq f^M(P_k) \leq f^M(P_k^U) \quad \forall \; bounded \; properties$$

$$f^M(P_m) = f^M\left(P_m^{goal}\right) \quad \forall \; properties \; with \; goal \; values \tag{3.19}$$

where, P_k is calculated using Equations 2.12 & 3.8.

The model is linear in nature and can be globally solved using Microsoft Excel 2007 - Solver based on the Simplex solution method. The model is fed to the system using Microsoft Excel 2007 - Visual Basic for Applications (VBA). It is also obvious that the minimum number of each group is zero.

3.3.3 Subproblem 2: Enumerating all group subsets of available first-order groups that could form at least one molecule

A group subset is a set of distinct objects with length equal to the number of first-order groups considered and with each object being the numbers of each first order group, *f*.

For example, consider the first-order groups: CH_3, CH_2, CH and OH, an example of a group subset generated is shown in Table 3.3.:

Table 3.3: Example of enumerated group subset.

Group *(f)*	**CH3**	**CH2**	**CH**	**OH**	**FBN**	**Molecule**
Sub group 1	2	1	1	1	0	✓

Subproblem 2 deals with the task of enumerating combinations of all first order groups while ensuring these combinations are constrained by Rule 11 and Rule 16.

Rule 11 as given by Equation 3.17 ensures the *FBN* of all group subsets as zero. Additionally, the group subsets are constrained according to Rule 16 for removing non-feasible subsets. The constraints representing Rule 16 are derived as follows:

The first order groups are classified into four classes and five categories (Gani et al., 1991) as shown in Table 3.2 and hence,

$$\sum_{X} n_X \leq C^{X}_{T,L,n_L}$$

$$X = \begin{cases} Category\ 3 & Category\ 3+4+5 \\ Category\ 4 & Category\ 4+5 \\ Category\ 5 \end{cases} \qquad \textbf{3.20}$$

where, n_X is the number of groups in category X; L is the largest possible class; n_L is number of groups in class L; T, the total number of first order groups in a generated group subset is given by:

$$T = \sum_{f=0}^{N_f} n_f \qquad \textbf{3.21}$$

and C^X_{T,L,n_L} is the maximum allowable number of groups from category X as given in Equation 3.21 and these limits are given in Appendix A.

Subproblem 2 is an enumeration problem while the solutions are also subject to a few constraints. The task is performed in Microsoft Excel 2007 with the code being fed to the system using Microsoft Excel 2007 - Visual Basic for Applications (VBA). Hence the solution to subproblem 2 yields group subsets that could possibly form at least one structurally and functionally feasible molecule.

3.3.4 Subproblem 3: Estimating possible higher order groups

Earlier methods (Chemmangattuvalappil et al., 2009; Eljack & Eden, 2008) either did not incorporate the contribution of higher order groups or did not efficiently incorporate higher-order groups within the algorithm during the initial stage thus carrying a number of infeasible group subsets until the last stage of molecular design. Incorporating higher order groups at a later stage in the algorithm may also lead to a situation where some potential combinations are omitted without being considered in further stages of the algorithm. Algebraically enumerating higher order groups beforehand would be an efficient extension to these methods, particularly for large problems that normally are prone to combinatorial explosion.

From Rule 7 and Rule 9, it is evident that higher order groups are built from first order groups and they overlap with each other. Additionally, owing to the knowledge of the connection between the

first order groups forming higher order groups, available free bonds in a higher order group and possible maximum occurrence of a first order group in a molecule, an algebraic expression can be generated to estimate the upper bound on the possible occurrence of a given higher order group in a molecule.

Rule 7 leads to the following expression for identifying the possible number of maximum groups; if (k : n) is the set of first order groups that are the building blocks of one higher order group, h, then (n_k : n_n) is the number of those first order groups present in the molecule, η is the number of occurrences of one particular first order group in a selected higher order group, n_h is the number of possible higher order groups from those first order groups, then according to Chemmangattuvalappil et al. (2009):

$$n_h = Min\left(\frac{n_k}{\eta_k} : \frac{n_n}{\eta_n}\right) \qquad \textbf{3.22}$$

According to Rule 7, for instance, to form the higher order group CH (CH3) CH (CH3), there must be two CH and two (CH3) groups. It is not possible to consider a CH (CH3) group as half of a higher order group. Hence, n_h must be rounded down to the nearest integer number.

Moreover, according to Rule 9, two higher order groups of the same kind can even share first order groups just like higher order groups of different kind. For instance, 2 OH and 1 CH group can form 2CHOH groups. Hence, possibility of the sharing of available first order groups participating in the given higher order group is considered. To account for these additions, Equation 3.22 needs to be corrected depending on the nature of higher order groups if $n_h > 0$. Upper bounds on the possible existence of higher-order groups are achieved by identifying closely connected confirmations of their respective building blocks and counting the higher-order groups in that confirmation. Also, while identifying these confirmations, it is ensured that the free bonds of each involved first-order

group are maximally utilized. Once this is identified, the correction factors to Equation 3.22 are carefully explored for respective higher order groups.

For example, in the higher-order group, CH (CH3) CH (CH3), Equation 3.22 becomes:

$$n_h = Min\left(\frac{n_{CH}}{2} : \frac{n_{CH_3}}{2}\right) \qquad \textbf{3.23}$$

But on account of the possibility of sharing of first-order groups, the upper bound of its existence in a molecule is possible by the closely connected sequential confirmation of first order groups CH3 and CH as -CH (CH3) CH (CH3) CH (CH3) CH (CH3)-

The correction to Equation 3.22 can now be given by the equation set in Equation 3.24.

$$k = Min\left(\frac{n_{CH}}{1} : \frac{n_{CH_3}}{1}\right) \quad and$$
$$n_h = k - 1 \qquad \textbf{3.24}$$

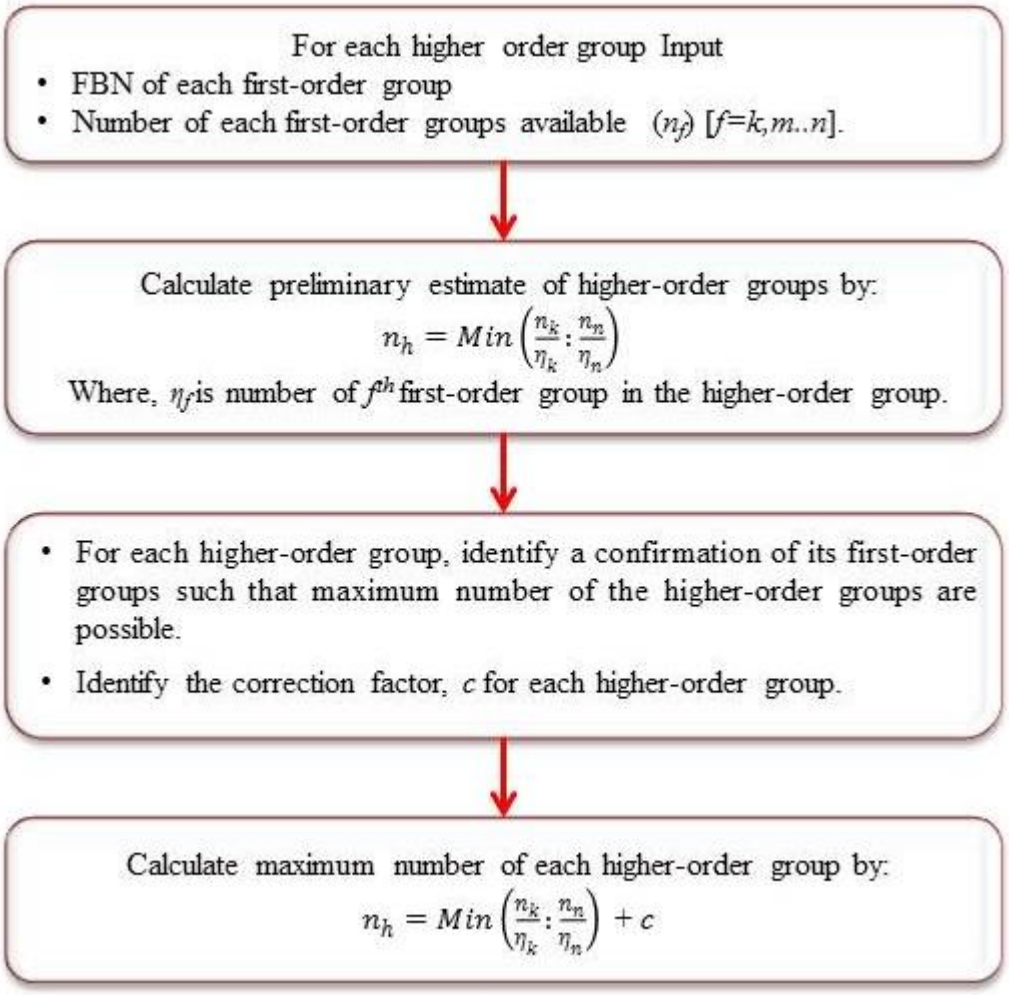


Figure 3.3: Method to estimate maximum number of higher-order groups.

3.3.5 Subproblem 4: Eliminating property infeasible group subsets

The property P_k of a molecule can be estimated using equation 2.12. The contribution towards a property P_k from the higher order groups in each group subset obtained in subproblem 2 is not a unique value. It fluctuates between an upper and lower bound as the different confirmations of first-order groups in each group subset may lead to formation of different higher order groups. The maximum and minimum of this value is obtained by solving the following linear model for each group subset.

$$\text{Max} \sum_{h=1}^{N_h} n_h P_{k,h}$$

Subject to:

$$0 \leq n_h \leq n_h^{max} \quad \textbf{3.25}$$

Where, n_h^{max} is obtained from subproblem 3 for each higher order group and $P_{k,h}$ is taken from Marrero and Gani (2001)

Once the maximum and minimum property values for each group subsets are known, the group subsets are checked against the property constraints given by Equation 3.19. If the estimated property range of a group subset falls completely outside the targeted property range of molecules, the group subset is excluded from being considered further.

3.3.6 Subproblem 5: Forming final molecules.

First, all possible combinations of higher order groups for each group subset capable of forming structurally and functionally feasible molecule are enumerated. This ensures the identification of structural isomers as the possibility of nonexistence of each higher order groups is considered which in turn leads to different confirmations of first order groups. As the number of higher order groups are estimated by considering that all its building first order blocks are used by it alone and that since these higher order groups overlap with each other, some of the enumerated combinations can be excluded prior to calculation of their properties. For example: From first order groups 2C, 4CH3, the number of possible higher order groups estimated are n[C(CH_3)(CH_3)C(CH_3)(CH_3)] = 1 and n[C(CH_3)(CH_3) (CH_3)]=1. It is clearly seen that both these higher order groups cannot coexist because for these higher order confirmations to coexist the minimum number of CH_3 groups needed is 5. Again Rule 10 indicates that a higher order group is not completely overlapped

by any other higher order group. Hence, the combinations which could form a molecule only by allowing a complete overlap of higher-order groups are also eliminated. Finally, for the remaining combinations of first- and higher-order groups, if the combination of the groups forms possible molecules satisfying all structural constraints without any floating free bonds, knowing the number of existing first order and higher order groups in each enumerated combination, the property of the molecule is calculated by Equations 3.7 & 3.8. Finally, the feasibility of all designed molecules with respect to their properties is checked. This is done by checking if all the property constraints from the process targeting model are met.

3.4 Summary

This chapter covers techniques to solve product design problems from a property platform. Molecular design by a decomposition approach using Marrero and Gani (2001) GC groups is discussed. In cases where the property contributions of groups are not available in literature, the contribution is estimated using a CI based property model. The method developed in this dissertation is a "generate and test" method, but taking into account the higher order groups and handling them efficiently, the developed method does not suffer from combinatorial explosion – an inherent drawback of conventional generate and test methods.

3.5 Case Studies – Computer Aided Molecular Design

3.5.1 Case Study – Design of blanket wash solvent

The application of the developed approach for product design is illustrated by reworking the design of a blanket wash solvent for a phenolic resin printing ink. Sinha and Achenie (2001) originally solved this design as a mixed-integer non-linear programming problem (MINLP). Eljack and Eden

(2008) solved the problem visually using molecular property clusters. An algebraic molecular design approach using higher order groups has been developed by Chemmangattuvalappil et al. (2009) but its application range could be improved if the accuracy of the property prediction is enhanced by the improved techniques for enumerating higher order groups developed in this dissertation. The method aims at removing most of the infeasible combinations of groups capable of forming molecules before the final molecules can be further considered for experimental tests. Also, this new method is significantly less computationally intensive compared to the previous geometric and non-linear programming methods of molecular design.

Problem Statement:

Solvents are extensively used as a major component of ink in the printing industry. In letterpress and offset lithographic printing processes, the inked image on a printing plate is printed on a rubber cylinder commonly known as "blankets" and then transferred to paper or other material. The produced quality images are greatly dependent on the cleanliness of the blanket. Paper fibers, ink residue, paper coating and dried ink etc. must be removed from the rubber blankets. Blanket washes are specially formulated to clean ink and other residues from rubber blankets. They are generally petroleum-based solvents that consist of volatile organic compounds (VOCs). Reasonably, there is a lot of concern regarding the effects of such solvents on the environment as well as the direct effect on human health.

The goal of this study is to design optimal solvents to be used as a blanket wash. These solvents should (a) have minimal drying time (b) be liquid at room temperature (c) have low vapor pressure (*VP*), and (d) dissolve the ink. Hence, solubility (R_{ij}) of the solvent is an important factor, the

drying time is related to the heat of vaporization (H_v), and the state of the solvent at room temperature is directly related to melting (T_m) and boiling (T_b) temperatures.

The property constraints for the solvents are listed in Table 3.4

Table 3.4: Property targets for Blanket wash

Property Targets	Lower Limit	Upper Limit
H_v (kJ/mol)	20	60
T_b (K)	350	400
T_m (K)	150	300
H_{fus} (kJ/mol)	10	20
VP (mmHg)		100
R^{ij}	0	19.8

The property models for estimating H_v, T_b , T_m and H_{fus} are given by (Marrero & Gani, 2001):

$$f^M(H_v) = H_v - h_{v0} \ , h_{v0} = 11.733 \quad \textbf{3.26}$$

$$f^M(T_b) = exp\left(\frac{T_b}{t_{bo}}\right) \ , t_{bo} = 222.543 \quad \textbf{3.27}$$

$$f^M(T_m) = exp\left(\frac{T_m}{t_{mo}}\right) \ , t_{mo} = 147.45 \quad \textbf{3.28}$$

$$f^M(H_{fus}) = H_{fus} - h_{fus0} \ , h_{fus0} = -2.806 \quad \textbf{3.29}$$

Vapor pressure is predicted using the McGowon Hovarth Equation, as a function of boiling and operating temperatures (Sinha & Achenie, 2001)

$$\log VP(mmHg) = 5.58 - 2.7\left(\frac{T_b}{T}\right)^{1.7} \quad \textbf{3.30}$$

The effectiveness of the designed solvent is greatly dependent on its ability to dissolve the ink, i.e. it is dependent on the solubility power of the designed solvent. The interactions between phenolic resin molecules (solute) with the solvent molecules are very important in this design problem. Solubility, R_{ij} is determined by using (Sinha & Achenie, 2001):

$$R_{ij} = \sqrt{4(\delta_{di} - \delta_{dj})^2 + (\delta_{pi} - \delta_{pj})^2 + (\delta_{hi} - \delta_{hj})^2} \quad \textbf{3.31}$$

where *i* correspond to the solvent while *j* corresponds to the solute and each δ parameter can be estimated using the following equations (Van Krevelen and Hoftyzer, 1976):

$$\delta_d = \frac{\sum_i F_{di}}{V} \qquad \delta_p = \frac{\sum_i F_{pi}^2}{V} \qquad \delta_h = \frac{\sum_i E_{hi}}{V} \quad \textbf{3.32}$$

where F_d is the dispersion component, F_p the polar component and E_h the contribution of hydrogen bonding forces. These parameters can also be calculated by the group contribution method proposed by Pistikopoulos and Stefanis (1998).

Phenolic resins are commonly used in printing inks. The dried ink (solute) is assumed to be phenolic resins, specifically 'Super Bakacite® 1001, Reichhold'. The solubility parameter components of the resin are non-polar, δ_{dj} = 23.3, polar, δ_{pj} = 6.6 and hydrogen bonding, δ_{hj} = 8.3 $MPa^{1/2}$ (Barton, 1985). The molecular property targets based on Equations 3.26- 3.29 is given in Table 3.5.

Table 3.5: Molecular property targets for blanket wash.

Molecular Property Targets	Lower Limit	Upper Limit
H_v (kJ/mol)	8.268	48.268
T_b (K)	4.8195	6.034

T_m (K)	2.7657	7.6489
H_{fus} (kJ/mol)	12.806	22.806

The following first order groups have been considered for molecular design:

1	CH_3	5	CH-O
2	CH_2	6	CH_3-CO
3	CH	7	CH_2-CO
4	OH		

Subproblem 1:

Property data for these selected groups is taken from Marrero and Gani (2001), which are reproduced in Appendix A. Now, inequality expressions for each property are formulated. The number of preselected first order molecular groups is maximized subject to the specific constraints mentioned in Table 3.5. The reason behind maximizing these groups is to ensure that no potential molecule is left out. The variations in the property values caused by inclusion of the higher order groups will be considered in the later stages.

The maximum values are as follows:

	max		**max**
CH_3	3	CH-O	2
CH_2	6	CH_3-CO	1

CH OH	2 1	CH_2-CO	1

Subproblem 2:

The class and category of the groups considered above are obtained from the group classification tables by Gani et al. (1991).

Group	**Class**	**Category**
CH_3	1	1
CH_2	2	1
CH	3	1
OH	1	4
CH-O	1	4
CH_3-CO	1	3
CH_2-CO	2	3

350 group subsets whose FBN is zero are generated. For each of these subsets, the number of groups, n_X in each of category X such that X= *3, 4, 5, 3+4+5, 4+5*, is identified and checked against C^X_{T,L,n_L} such that:

$$\sum_X n_X \leq C^X_{T,L,n_L} \qquad \textbf{3.33}$$

where, L is the largest possible class; n_L is the number of groups in class L; T is the total number of first order groups in the group subset and C^X_{T,L,n_L} is identified using Table A.15 .

269 of the tested group subsets satisfy the class and category constraints and these are considered for further subproblems.

Subproblem 3:

The possible higher-order groups are identified in Table 3.6.

Table 3.6: Possible higher order groups for blanket wash case study.

1	$(CH_3)_2CH$
2	$CH(CH_3)CH(CH_3)$
3	CH-CHO
4	CH_2-CH_3CO
5	CH-CH_3CO
6	CHOH
7	OH-CH-CH_3CO
8	OH-CH_2-CH_3CO

The maximum possible number of each higher-group is computed algebraically by using the methodology explained in the previous chapter as shown below.

1. $(CH_3)_2CH$

   If the group subsets comprises of only 3 CH3 and 1CH groups, $n_h = 2$;

   Else, $n_h = int\left(\min\left(\frac{n_{CH_3}}{2}; n_{CH}\right)\right)$

2. $CH(CH_3)CH(CH_3)$

   Maximum number of groups of this kind can exist when the groups $CH(CH_3)$ are positioned in a series, Hence, if *a* is the number of $CH(CH_3)$ groups i.e.,

$If\ a = int\left(\min\left(n_{CH_3}; n_{CH}\right)\right) \quad and \quad a > 1;$

$n_h = a - 1 \quad Else \quad n_h = 0$

3. CH-CHO

   If *a* is the number of $CH(CHO)_2$ groups, *b* is the number of CH-CHO groups from balance CH and CHO groups, i.e.

   $If\ a = int\left(\min\left(\frac{n_{CHO}}{2}; n_{CH}\right)\right) \ and \ \ b = int(\min(n_{CHO}; n_{CH}))$

   $n_h = 2a + b$

4. CH_2-CH_3CO

   If the group subsets comprises of only 2 CH_3CO and 1CH_2 groups, $n_h = 2$;

   Else, $n_h = int\left(\min\left(n_{CH_3CO}; n_{CH_2}\right)\right)$

5. CH- CH_3CO

   If *a* is the number of $CH(CH_3CO)_2$ groups, *b* is the number of CH- CH_3CO groups from balance CH and CH_3CO groups, i.e.

   $If\ a = int\left(\min\left(\frac{n_{CH_3CO}}{2}; n_{CH}\right)\right) \ and \ \ b = int\left(\min\left(n_{CH_3CO}; n_{CH}\right)\right)$

   $n_h = 2a + b$

6. CHOH

   If *a* is the number of $CH(OH)_2$ groups, *b* is the number of CHOH groups from balance CH and OH groups, i.e.

   $If\ a = int\left(\min\left(\frac{n_{OH}}{2}; n_{CH}\right)\right) \ and \ \ b = int(\min(n_{OH}; n_{CH}))$

   $n_h = 2a + b$

7. OH-CH-CH_3CO

If the group subsets comprises of 1 CH and total of 3 [OH + CH_3CO] groups alone $n_h = 1$;

Else, $n_h = int\left(\min\left(n_{CH_3CO}; n_{CH}; n_{OH}\right)\right)$

8. OH-CH_2-CH_3CO

   It is whole molecule, so, if the group subsets comprise of 1 CH_2 and total of 1 OH and 1 CH_3CO groups alone $n_h = 1$;

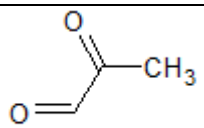
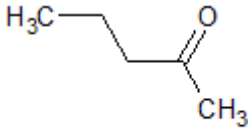
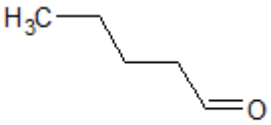
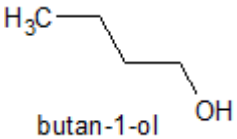
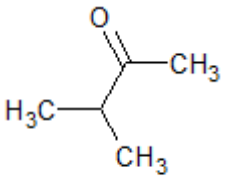
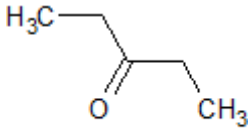
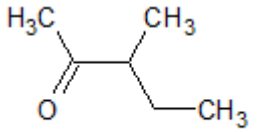
Subproblem 4:

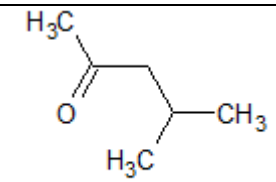
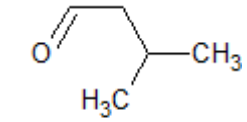
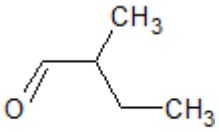
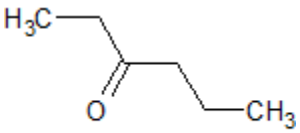
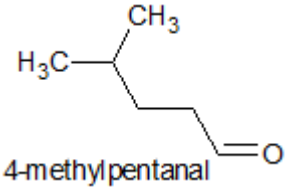
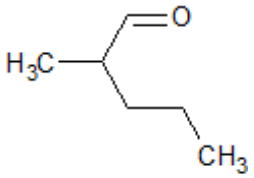
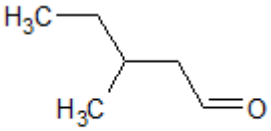
For each group subset to be tested in this subproblem, the maximum and minimum contributions from possible higher-order groups are estimated. Based on the molecular property constraints listed in Table 3.5, group subsets whose property range falls completely outside the targeted property range of molecules are excluded from being considered further. 18 group subsets obey the property constraints and are hence considered for the next subproblem.

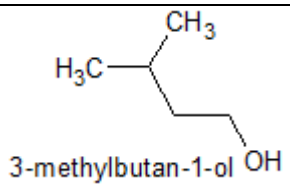
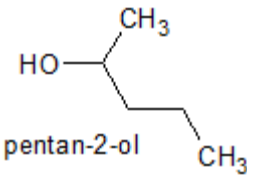
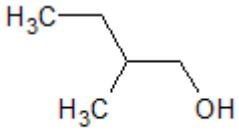
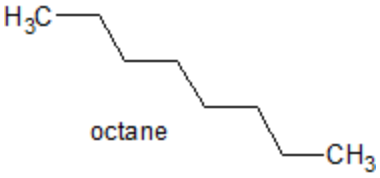
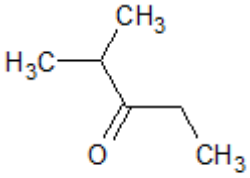
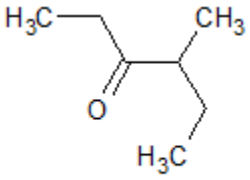
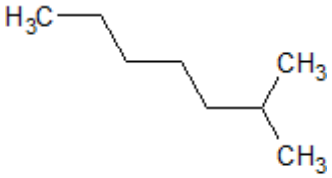
Subproblem 5:

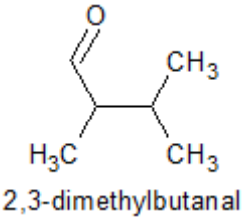
All possible combinations of higher order groups for each group subset capable of forming structurally and functionally feasible molecule are enumerated. 42 group subsets have been identified and checked if a molecule with targeted properties is possible from those subsets. The molecules identified by solving molecular design model are listed below in Table 3.7.

Table 3.7: Possible blanket wash solvents.

	Molecule	H_v (kJ/mol)	T_b (K)	T_m (K)	H_{fus} (kJ/mol)
1	 2-oxopropanal	39.298	385.643	263.653	16.586
2	 pentan-2-one	37.368	374.201	210.510	12.205
3	 pentanal	39.050	380.590	220.789	18.101
4	 butan-1-ol	50.894	381.714	213.001	11.555
5	 3-methylbutan-2-one	35.648	359.287	248.030	10.111
6	 pentan-3-one	36.469	362.051	210.369	11.979
7	 3-methylpentan-2-one	40.957	388.989	244.027	12.354

8	 4-methylpentan-2-one	40.238	391.122	226.134	11.756
9	 3-methylbutanal	37.010	368.624	227.685	15.013
10	 2-methylbutanal	36.859	363.247	201.392	14.248
11	 hexan-3-one	41.379	391.275	219.031	14.618
12	 4-methylpentanal	41.920	397.049	235.412	17.652
13	 2-methylpentanal	41.769	392.324	210.581	16.887
14	 3-methylpentanal	42.319	397.180	223.081	17.256

15	 3-methylbutan-1-ol	53.764	398.093	228.376	11.106
16	 pentan-2-ol	53.957	387.465	193.351	10.111
17	 2-methylbutan-1-ol	54.163	398.223	215.415	10.710
18	 octane	41.627	398.074	157.093	16.348
19	 2-methylpentan-3-one	39.339	379.885	226.007	11.530
20	 4-methylhexan-3-one	39.738	380.026	212.826	11.134
21	 2-methylheptane	39.587	387.035	167.584	13.260

22	 2,3-dimethylbutanal	40.261	393.333	245.513	12.033

Vapor pressure *(VP),* and solubility *(R_{ij})* of the solvent are calculated for the above solvents and listed in the following Table 3.8.

Table 3.8: Vapor pressure and solubility calculations for identified blanket wash solvents.

	Molecule	VP (mmHg)	Solubility
1	2-oxopropanal	24.822	15.047
2	pentan-2-one	40.154	15.087
3	pentanal	30.735	15.113
4	butan-1-ol	29.313	15.542
5	3-methylbutan-2-one	74.022	15.638
6	pentan-3-one	66.176	14.886
7	3-methylpentan-2-one	21.523	15.779
8	4-methylpentan-2-one	19.645	15.779
9	3-methylbutanal	50.577	15.636
10	2-methylbutanal	63.035	15.636
11	hexan-3-one	19.516	15.007
12	4-methylpentanal	15.213	15.734
13	2-methylpentanal	18.656	15.734

14	3-methylpentanal	15.127	15.734
15	3-methylbutan-1-ol	14.539	15.999
16	pentan-2-ol	22.970	15.992
17	2-methylbutan-1-ol	14.457	15.992
18	octane	14.551	17.297
19	2-methylpentan-3-one	31.660	15.564
20	4-methylhexan-3-one	31.473	15.564
21	2-methylheptane	23.394	17.946
22	2-3-dimethylbutanal	17.863	16.279

Since all the molecules found by solving the molecular design satisfy the property target limits for vapor pressure and solubility, they can be considered for further tests like experimentation or other property checks. Molecules 1, 2, 3, 4, 6, 7, 10, 21, and 22 are the ones identified by Chemmangattuvalappil et al. (2009) The other molecules identified in that work are excluded due the feasibility constraints applied in subproblem 2. For example, the 2-Hydroxypropanol molecule has OH and CHO groups on the same carbon and hence is likely to be unstable as a solvent. By applying feasibility constraints based on class and category classification of first-order groups, such infeasible molecules can be excluded. Also, the methodology developed here identifies the possible structural isomers as the possibility of nonexistence of each higher order group is considered while generating the combination of groups. Pentan-3-one identified by Sinha and Achenie (2001) is identified using current methodology, the other two namely propanol and methyl-ethyl ketone were not identified owing to the change in selection of property targets. Due

to the usage of algebraic approaches in the developed methodology, it generated the feasible structures more efficiently than the visual approach by Eljack and Eden (2008).

3.5.2 Case Study – Design of cyclic molecules

A general problem is considered to illustrate the capability of the developed molecular design methodology for generating cyclic molecules.

The property constraints for the solvents are listed in Table 3.9.

Table 3.9: Property targets for cyclic molecules.

Property Targets	Lower Limit	Upper Limit
H_v (kJ/mol)	20	60
T_b (K)	350	400
T_m (K)	150	300
H_{fus} (kJ/mol)	10	20

The property models for estimating H_v, T_b, T_m and H_{fus} are given by (Marrero & Gani, 2001):

$$f^M(H_v) = H_v - h_{v0} \ , h_{v0} = 11.733 \quad \textbf{3.34}$$

$$f^M(T_b) = exp\left(\frac{T_b}{t_{bo}}\right) \ , t_{bo} = 222.543 \quad \textbf{3.35}$$

$$f^M(T_m) = exp\left(\frac{T_m}{t_{mo}}\right) \ , t_{mo} = 147.45 \quad \textbf{3.36}$$

$$f^M\left(H_{fus}\right) = H_{fus} - h_{fus0} \ , h_{fus0} = -2.806 \quad \textbf{3.37}$$

The molecular property targets based on Equations 3.34 - 3.37 is given in Table 3.10.

Table 3.10: Molecular property targets for cyclic molecules.

Molecular Property Targets	Lower Limit	Upper Limit
H_v (kJ/mol)	8.268	48.268
T_b (K)	4.8195	6.034
T_m (K)	2.7657	7.6489
H_{fus} (kJ/mol)	12.806	22.806

The following first order groups have been considered for molecular design:

1	CH_3	5	CH-O
2	CH_2	6	CH_2 (ring)
3	CH	7	CH (ring)
4	OH		

Subproblem 1:

The maximum values of each first-order group selected are as follows:

	max		**max**
CH_3	4	CH-O	1
CH_2	4	CH_2 (ring)	3
CH	1	CH (ring)	4
OH	0		

This shows that molecules with OH functional group do not possess the required properties.

Subproblem 2:

The class and category of the groups considered above are obtained from the group classification tables by Gani et al. (1991).

Group	Class	Category
CH_3	1	1
CH_2	2	1
CH	3	1
OH	1	4
CH-O	1	4
CH_2 (ring)	2	1
CH (ring)	3	1

187 group subsets whose FBN is zero are generated. FBN here is calculated using the following equation. The number of rings is pre-specified to be 1.

$$FBN = \sum_{f=1}^{N_f} n_f \, FBN_f - 2\left[\left(\sum_{f=1}^{N_f}(n_f) - 1\right) + N_{rings}\right] = 0 \qquad \textbf{3.38}$$

For each of these subsets, the number of groups, n_X in each category X such that $X = 3, 4, 5, 3+4+5, 4+5$, is identified and checked against C^X_{T,L,n_L} such that:

$$\sum_X n_X \leq C^X_{T,L,n_L} \qquad \textbf{3.39}$$

where, L is the largest possible class; n_L is number of groups in class L; T is the total number of first order groups in the group subset and C^X_{T,L,n_L} is identified using Table A.17.All 187 group

subsets satisfy constraints pertaining to class and category of the groups and hence all of them are considered for further subproblems.

Subproblem 3:

The possible higher-order groups are identified in Table 3.11.

Table 3.11: Possible higher order groups for cyclic molecules.

1	$(CH_3)_2CH$
2	$CH(CH_3)CH(CH_3)$
3	CH-CHO
4	CH_{cyc}-CH_3
5	CH_{cyc}-CH_2
6	CH_{cyc}-CH
7	CH_{cyc}-CHO

The maximum possible number of each higher-group is computed algebraically by using the methodology explained. The method to calculate the maximum number of each of groups 1-3 in Table 3.11 was explained in subproblem 3 of case study 1. Groups 4-7 can be generally represented as CH_{cyc}-A as CH_{cyc} has 3 free bonds and two of these are involved in bonding to only cyclic first-order groups and only one free bond is available for their respective non-cyclic first order groups. Hence, the maximum of these groups can be calculated by:

$$n_h = int\left(\min\left(n_{CH_{cyc}}; n_A\right)\right)$$

Subproblem 4:

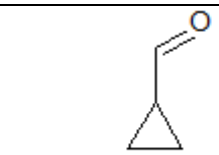
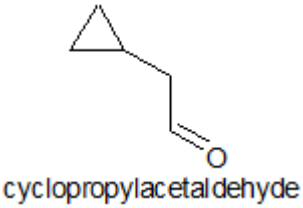
For each of group subset to be tested in this subproblem, the maximum and minimum contributions from possible higher-order groups are estimated. Based on the molecular property constraints

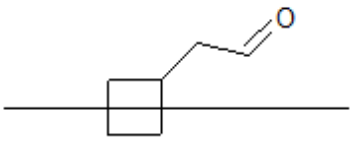
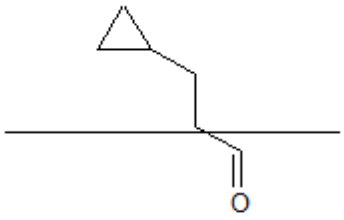
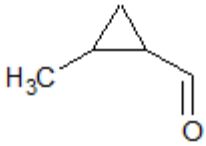
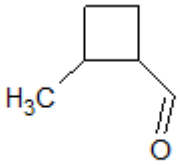
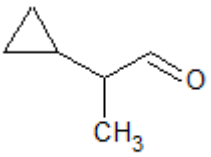
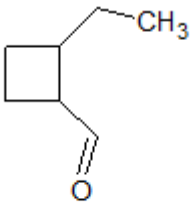
listed in Table 3.10, group subsets whose property range falls completely outside the targeted property range of molecules are excluded from being considered further. 15 group subsets obey the property constraints and are hence considered for the next subproblem.

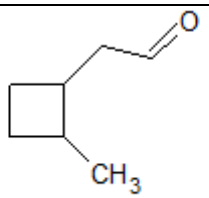
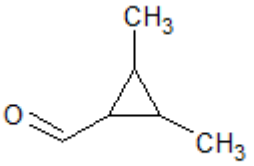
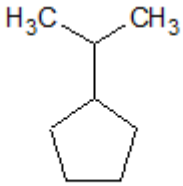
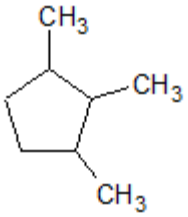
Subproblem 5:

All possible combinations of higher order groups for each group subset capable of forming structurally and functionally feasible molecule are enumerated. Additionally, for group subsets containing both cyclic and acyclic first-order groups, another constraint to check if cyclic groups (after forming a ring) can accommodate acyclic groups to form a molecule. 64 group subsets have been identified and checked if a molecule with targeted properties is possible from those subsets. The molecules identified by solving the molecular design model are listed below in Table 3.12.

Table 3.12: Possible cyclic molecules.

S. No	Molecule	H_v (kJ/mol)	T_b (K)	T_m (K)	H_{fus} (kJ/mol)
1	 cyclopropanecarbaldehyde	40.542	372.5846	245.0608	14.242
2	 cyclopropylacetaldehyde	41.683	378.5528	203.5736	14.675

~~3~~	 cyclobutylacetaldehyde	~~45.024~~	~~409.7074~~	~~223.318~~	~~15.744~~
~~4~~	 3-cyclopropylpropanal	~~46.593~~	~~405.8108~~	~~212.6318~~	~~17.314~~
5	 2-methylcyclopropanecarbaldehyde	40.589	358.6316	228.9984	16.308
6	 2-methylcyclobutanecarbaldehyde	43.93	392.4921	245.7865	17.377
7	 2-cyclopropylpropanal	44.983	396.2557	229.1574	17.029
8	 2-ethylcyclobutanecarbaldehyde	44.975	392.37	226.408	17.777

9	 (2-methylcyclobutyl)acetaldehyde	45.071	397.9558	204.5344	17.81
10	 2,3-dimethylcyclopropanecarbaldehyde	43.977	379.7679	229.8074	19.443
11	 propan-2-ylcyclopentane	39.663	399.3066	166.4824	10.262
12	 1,2,3-trimethylcyclopentane	38.602	383.9691	158.3832	11.944

Molecules 3 and 4 have properties outside of the targeted range and hence are eliminated from the list of valid molecules.

4. Property Based Process Design and its Integration with Molecular Design

4.1 Property Operators and Clustering Techniques

Standard process design techniques are chemo-centric in nature, i.e. they are based on tracking, manipulation, and allocation of individual chemical species. But, many processes are driven and governed by properties or functionalities of the streams and not by their chemical constituency. For instance, the usage of material utilities (e.g. solvents) relies on their characteristics, such as equilibrium distribution coefficients, viscosity, and volatility without the need to chemically characterize these materials. Constraints on process units that can accept recycled/reused process streams and wastes are not limited to compositions of components but are also based on the properties of the feeds to processing units (El-Halwagi, 2006). In the design of paper with a specified quality, the quality is specified in terms of the physical properties and not in terms of components as the basic component of all types of paper is cellulose (Eden et al., 2004). Since properties (or functionalities) form the basis of performance for many processes, design procedures based on key properties instead of key compounds are needed. But, unlike mass, properties are not conserved and cannot be tracked among units without undertaking component material balances. Therefore, to resolve these limitations, property-based clusters which are conserved are used (Shelley & El-Halwagi, 2000). The property clusters are formed based on property operators, which are functions of actual physical properties that obey linear additive rules (Eden et al., 2004; Shelley & El-Halwagi, 2000).

For a mixture made up of N_s streams and described by j properties, the property operator, $\Psi_j(P_{jM})$ corresponding to the property P is formulated as follows:

$$\Psi_j(P_{jM}) = \sum_{s=1}^{N_s} x_s \cdot \Psi_j(P_{js}) \quad \textbf{4.1}$$

Here, $\Psi j\ (P_{js})$ is the operator of the j^{th} property P_{js} of stream s and x_s is the fractional contribution given by

$$x_s = \frac{F}{\sum_{s=1}^{N_s} F_s} \quad \textbf{4.2}$$

The property operators can be evaluated from first principles or estimated through empirical or semi-empirical methods. Density for instance, where the resulting property of mixing two streams is given as the inverse of the summation over the reciprocal property values multiplied by their fractional contribution x_s as shown below

$$\frac{1}{\rho_M} = \sum_{s=1}^{N_s} x_s \cdot \frac{1}{\rho_s} \qquad \Psi_j(P_{jM}) = \frac{1}{\rho_M} \qquad \Psi_j(P_{js}) = \frac{1}{\rho_s} \quad \textbf{4.3}$$

The properties involved in the system are of different units and magnitudes. So, the operators are normalized into a dimensionless form by dividing with an appropriately chosen reference operator. The normalized property operator is given as:

$$\Omega_{js} = \frac{\Psi_{js}(P_{js})}{\Psi_{js}(P_j^{ref})} \quad \textbf{4.4}$$

An Augmented Property index AUP for each stream s is the sum of all the NP dimensionless property operators:

$$AUP_s = \sum_{j=1}^{NP} \Omega_{js} \quad \textbf{4.5}$$

Finally, the property cluster C_{js} for property j is defined as:

$$C_{js} = \frac{\Omega_{js}}{AUP_s} \quad \textbf{4.6}$$

These clusters, in a way, give the contribution of each property to its respective targeted value. Clusters enable the conserved tracking of properties and the derivation of visualization design tools. Also, these clusters possess two characteristics: intra- and inter-stream conservation (Eden et al., 2004). Figure 4.1 shows the ternary representation of these characteristics.

4.1.1 Intra-Stream Conservation

For any stream *s*, the sum of clusters corresponding to N_c properties is constant and adds up to unity, i.e.

$$\sum_{j=1}^{N_c} C_{js} = 1 \quad \textbf{4.7}$$

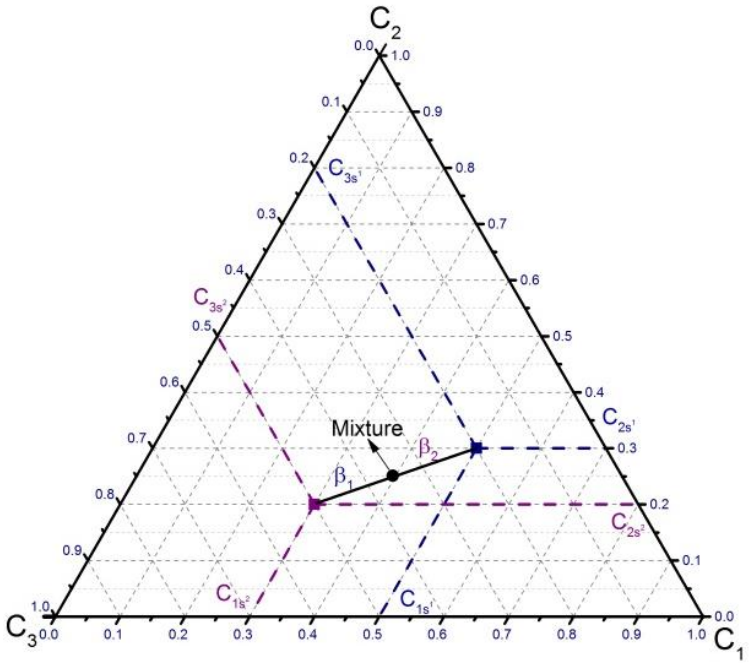


Figure 4.1: Ternary representation of clusters and their intra- and inter-stream conservation characteristics.

Two points $C_{js^1}(C_{1s^1}, C_{2s^1}, C_{3s^1})$ and $C_{js^2}(C_{1s^2}, C_{2s^2}, C_{3s^2})$ are seen to be having the sum of their respective property clusters to be unity. Hence, if (N_c-1) clusters for a stream are given, the N_c^{th} cluster can be uniquely identified.

4.1.2 Inter-Stream Conservation

For any two or more streams that are mixed, the resulting individual cluster is conserved. Additive rules in the form of lever-arm rules aid in obtaining the mean cluster property of two or more mixed streams. The lever-arm rule can be represented by (Eden et al., 2004):

$$C_{jM} = \sum_{s=1}^{N_s} \beta_s \cdot C_{js} \qquad \textbf{4.8}$$

where C_{jM} is the mean cluster of j^{th} property and β_s represents the fractional lever arm of cluster, C_{js}, of stream s. The cluster arm is given by:

$$\beta_s = \frac{x_s \cdot AUP_s}{AUP_M} \qquad \textbf{4.9}$$

and also

$$\sum_{s=1}^{N_s} \beta_s = 1 \qquad \textbf{4.10}$$

Indicating

$$AUP_M = \sum_{s=1}^{N_s} x_s . AUP_s \qquad \textbf{4.11}$$

Using AUP in the lever-arm rule enables one-to-one mapping from raw properties to property clusters and vice versa when streams are mixed (El-Halwagi, 2006). Hence, if clusters of (N_s-1) streams are given, owing to the inter-stream conservation property, the N_s^{th} clusters can be

uniquely found. Also the mixture cluster of two streams will lie on a straight line connecting those two points.

4.2 Process Design by Visualization tools

The objective here is to develop visualization tools that systematically minimize usage of fresh resources and maximize utilization of process resources into a process sink along with identification of the fresh resource's feasibility region and the corresponding property targets without committing to any components in the fresh source (until the final step).

4.2.1 Identification of feasibility region for sink

The geometric shape of the feasibility region for each sink (units capable of processing the sources) has to be identified first in order to address the above mentioned problem. Constructing the boundaries of the feasibility region (BFR) for the sink is not that straightforward and hence needs definite construction rules (El-Halwagi, Glasgow, Qin, & Eden, 2004)

Consider a sink with three targeted properties, each of them being bounded by a lower and upper limit.

$$P_{j,sink}^{min} \leq P_j \leq P_{j,sink}^{max} \quad \rightarrow \quad \Omega_{j,sink}^{min} \leq \Omega_j \leq \Omega_{j,sink}^{max} \qquad \textbf{4.12}$$

Based on Equation 4.12, the rules developed by El-Halwagi et al. (2004) for a system involving three properties are summarized below.

Rule 17: The BFR is accurately represented by six line segments.

Rule 18: The extended linear segments of the BFR constitute three convex hulls (cones) with their heads lying on the three vertices of the ternary cluster diagram.

Rule 19: The cluster boundary values defining the BFR are characterized by the following values of dimensionless operators for the sink constraints.

$$C_{1,sink}^{min} = \frac{\Omega_{1,sink}^{min}}{\Omega_{1,sink}^{min} + \Omega_{2,sink}^{max} + \Omega_{3,sink}^{max}} \quad \textbf{4.13}$$

$$C_{1,sink}^{max} = \frac{\Omega_{1,sink}^{max}}{\Omega_{1,sink}^{max} + \Omega_{2,sink}^{min} + \Omega_{3,sink}^{min}} \quad \textbf{4.14}$$

$$C_{2,sink}^{min} = \frac{\Omega_{2,sink}^{min}}{\Omega_{1,sink}^{max} + \Omega_{2,sink}^{min} + \Omega_{3,sink}^{max}} \quad \textbf{4.15}$$

$$C_{2,sink}^{max} = \frac{\Omega_{2,sink}^{max}}{\Omega_{1,sink}^{min} + \Omega_{2,sink}^{max} + \Omega_{3,sink}^{min}} \quad \textbf{4.16}$$

$$C_{3,sink}^{min} = \frac{\Omega_{3,sink}^{min}}{\Omega_{1,sink}^{max} + \Omega_{2,sink}^{max} + \Omega_{3,sink}^{min}} \quad \textbf{4.17}$$

$$C_{3,sink}^{min} = \frac{\Omega_{3,sink}^{max}}{\Omega_{1,sink}^{min} + \Omega_{2,sink}^{min} + \Omega_{3,sink}^{max}} \quad \textbf{4.18}$$

By simply bounding the region within the minimum and maximum values of the clusters identified above, the overestimation of the feasibility region is determined. While this boundary guarantees the existence of feasible points inside it, the true feasible region is obtained by connecting the six points corresponding to the above six cluster boundary values (each coordinate of whose represent a property operator value) corresponding to respective boundary clusters. This stems from the fact that all these points are part of the true feasibility region and that any mixtures of those points must also lie within the true feasibility region. This is depicted in Figure 4.2 (Eden et al., 2004).

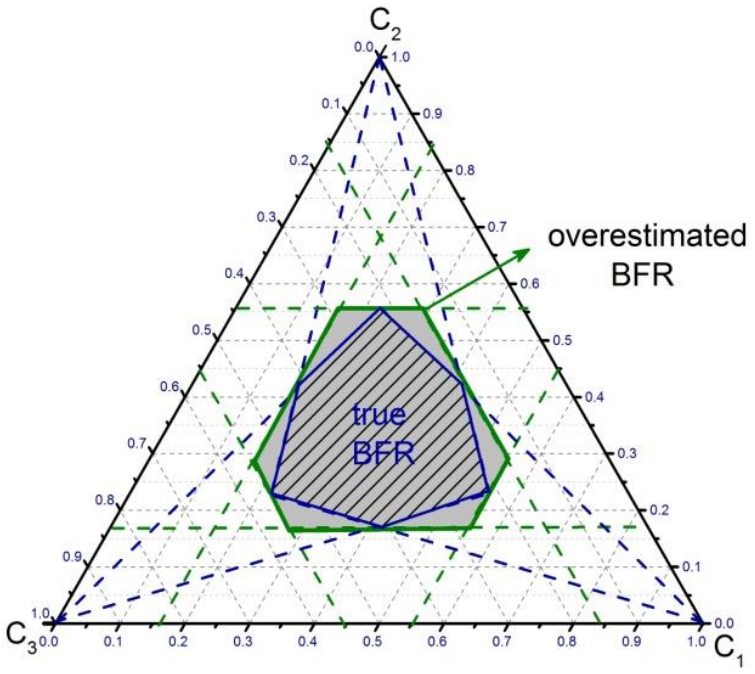


Figure 4.2: Boundary Feasible region

4.2.2 Source - Sink Mapping

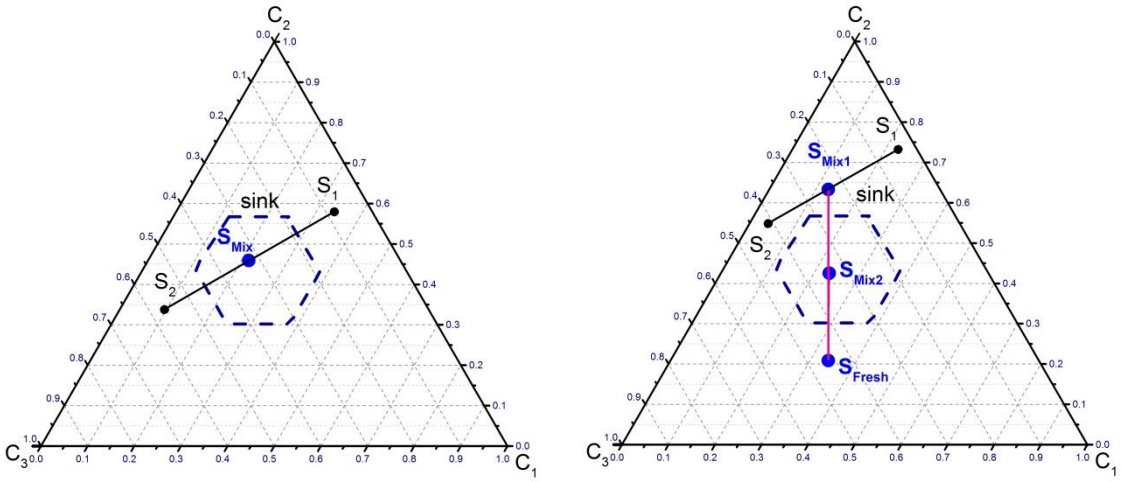
After the BFR of the sink has been found as illustrated in section 4.2.1, the next step is to map source streams and/or their mixtures into this sink. The stream or mixture stream is qualified to be processed by a sink if

- The point representing stream clusters is contained within the feasibility region of the sink on the cluster ternary diagram.
- The values of the augmented property index (AUP) for the source (or mixture of sources) and the sink must match.
- The flow rate of the source (or mixture of sources) must lie within the acceptable feed flow rate range for the sink.

Any cluster point in the ternary diagram corresponds to multiple combinations of property points due to the nonlinear mapping from the property to cluster domain (Section 4.1).

Therefore, having the cluster value inside the sink region alone will not ensure that the properties are in the correct range. In order to make sure that the properties match the sink requirements, the *AUP* values of sink and source streams must also match. Additionally the flow rate of the source, or mixture of sources, must be within the upper and lower limits of sink's capacity, otherwise only a fraction can be recycled.

As all the possible mixture points of two streams lie on a straight line connecting those two points, if the straight line passes through the sink region, the two streams can be mixed to get the required stream, S_{mix}, for the sink as shown in Figure 4.3(a).



(a) (b)

Figure 4.3: Source - Sink Mapping

Here S_1 and S_2 can be mixed to get the optimum output S_{mix}. If S_{mix1} lies outside the sink's BFR as in Figure 4.3(b), another source is added to meet the sink property targets. Sometimes the S_1-S_2 line may pass through that sink but the mixture point S_{mix} may lie outside the sink BFR. In these cases, for the mixture to be accepted by the sink, the individual flowrates of S_1 and S_2 must be changed.

4.2.3 Identification of feasibility region for fresh source

Process constraints based on the sink's BFR and available source streams in a process are utilized to design a needed fresh source. Given two available source streams S_1 and S_2 to be recycled to the process sink, the source mixture (*SM*) is identified using the lever-arm principles as shown in Figure 4.3(b). Owing to the flow rate constraint of the sink, the flow rate of fresh stream is identified by:

$$F_{fresh} = F_{sink} - F_{SM} \quad \textbf{4.19}$$

In, Figure 4.4, the first feasible region reflects the sink's original property demands as given by Equation 4.12. Lever-arm principles are utilized to identify a new feasibility region for the fresh source. This region serves to integrate the process requirements with the design of fresh source to be mixed with source mixture.

The feasibility region points *(A′, B′, C′, D′, E′, F′)* for the fresh stream are determined from point *SM* and *A, B, C, D, E, F* respectively by using Equation 4.20.

For generalization, the line segment connecting points *SM* and D' in Figure 4.4 has been magnified with *SM* and D' shown as S_1 and S_2 respectively in the magnification. D is marked as the mixture point; the cluster point for D' is given by:

$$C_{jS_2} = \frac{C_{jM} - \beta_{S_1} C_{jS_1}}{(1 - \beta_{S_1})} \quad \textbf{4.20}$$

Given C_{jS_1} and C_{jM} and calculating β_{S_1} by Equation 4.9, C_{jS_2} is determined. Similarly, *A′, B′, C′, E′, F′* can be easily determined.

Hence, the new property requirements for the fresh source are back calculated from the determined cluster values while also satisfying the AUP match condition mentioned in section 4.2.2:

$$x_{S_1} AUP_{S_1} + (1 - x_{S_1}) AUP_{S_2} = AUP_M \quad \textbf{4.21}$$

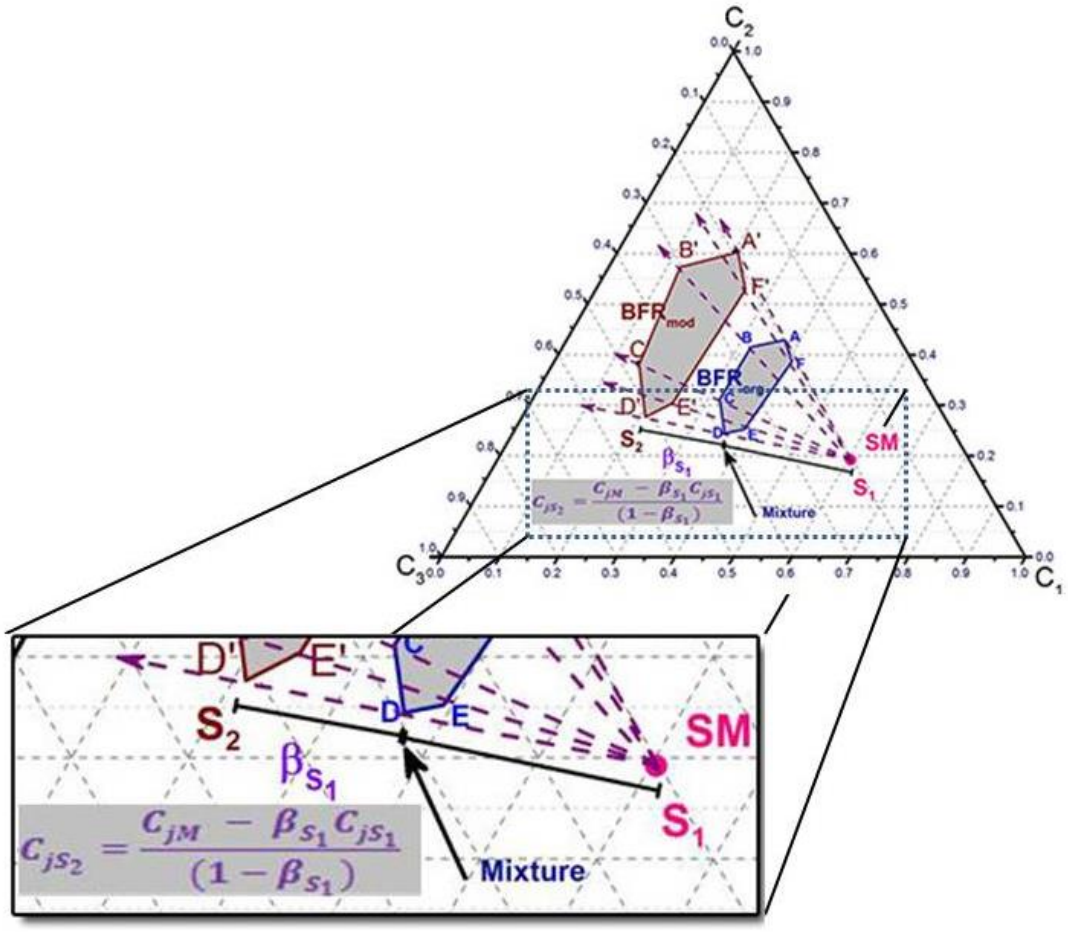


Figure 4.4: Identification of feasibility region for fresh source (Eljack, Solvason, Chemmangattuvalappil, & Eden, 2008)

4.3 Process Design by Mathematical Programming

The graphical approach has limitations pertaining to the number of properties it can handle and also its practicality is inversely proportional to the number of streams and sinks present in the process. It becomes quite hard to track the properties visually and hence, to overcome these

limitations, a mathematical programming formulation is used to address the process design problem.

Consider a process sink i, $i = 1,2,3 \ldots\ldots N_{sinks}$, The sink's BFR is defined in terms of j bounded properties, P_k, $k = 1,2,3 \ldots\ldots N_{properties}$ and a permissible flow rate, F_i^{sink}. The property bounds for each sink, i are given by:

$$P_{k,i}^{L} \leq P_{k,i} \leq P_{k,i}^{U} \qquad \textbf{4.22}$$

Since the properties may not be conserved, they are reformulated in terms of their conserved surrogates by the application of the property operators as mentioned in section 4.1:

$$[\Psi_k(P_{k,i})]^L \leq \Psi_k(P_{k,i}) \leq [\Psi_k(P_{k,i})]^U \qquad \textbf{4.23}$$

If available source streams j, $j = 1,2,3 \ldots\ldots N_{sources}$ with properties, P_k, j and flow rates, F_i^{source} are to be recycled to the process sinks along with the fresh source, *fresh*, the process design problem aims at mathematically finding the admissible property ranges $P_{k,fresh}^{lower} \leq P_{kf,resh} \leq P_{k,fresh}^{upper}$ and flow rate, F_{fresh} of fresh source into the sinks without committing to any components in the fresh source.

4.3.1 Mathematical model of the process design problem

Min $\pm\Psi_k(P_{k,fresh})$

Subject to:

Flow rate constraints:

$$F_i^{sink} = F_{i,fresh} + \sum_{j=1}^{N_{sources}} F_{i,j} \qquad \forall\, i = 1\ to\ N_{sinks} \qquad \textbf{4.24}$$

$$F_j^{source} = W_j + \sum_{i=1}^{N_{sinks}} F_{i,j} \qquad \forall\, j = 1\ to\ N_{sources} \tag{4.25}$$

$$F_{fresh} = \sum_{i=1}^{N_{sinks}} \left(F_i^{sink}\right) - \sum_{i=1}^{N_{sources}} F_i^{source} \quad ; \quad If \sum_{i=1}^{N_{sinks}} F_i^{sink} \geq \sum_{i=1}^{N_{sources}} F_i^{source} \tag{4.26}$$

If $\sum_{i=1}^{N_{sinks}} F_i^{sink} \leq \sum_{i=1}^{N_{sources}} F_i^{source}$ and the source stream mixture do not meet the property targets of the sinks, the proper bounds on the total fresh stream (Equation 4.27) are added to the mathematical formulation. Also from upper bounds on waste streams (Equation 4.28), these bounds on fresh stream can be calculated.

$$F_{fresh}^{L} \leq F_{fresh} \leq F_{fresh}^{U} \tag{4.27}$$

$$Waste = \sum_{j=1}^{N_{sources}} W_j \leq W^U \tag{4.28}$$

$$0 \leq F_{i,j} \leq F_j^{source} \qquad \forall\, j = 1\ to\ N_{sources}\ \ and\ \ \forall\, i = 1\ to\ N_{sinks} \tag{4.29}$$

Property Constraints

$\forall\, i = 1\ to\ N_{sinks}$

$$F_i^{sink}\, \Psi_k(P_{k,i}) = F_{i,fresh}\, \Psi_k(P_{k,fresh}) + \sum_{j=1}^{N_{sources}} F_{i,j}\, \Psi_k\left(P_{k,j}\right) \tag{4.30}$$

$$\Psi_k(P_{k,i})^L \leq \Psi_k\left(P_{k,i}\right) \leq \Psi_k\left(P_{k,i}\right)^U \ if\ \Psi_k(P_k) \propto P_k \tag{4.31}$$

$$\Psi_k(P_{k,i})^U \leq \Psi_k\left(P_{k,i}\right) \leq \Psi_k\left(P_{k,i}\right)^L \ if\ \Psi_k(P_k) \propto \frac{1}{P_k} \tag{4.32}$$

Where,

$F_{i,j}$ is the flow rate of stream j entering the sink i, $F_{i,\text{fresh}}$ is the flow rate of fresh stream entering sink i and W^U is the flow rate of the maximum waste that can be handled.

4.3.2 Global optimal solution

Although the nonlinearity in the process design problem is greatly reduced by using property operators, it is still a nonlinear programming model because of the presence of bilinear terms, $F_{i,fresh}\ \Psi_k\left(P_{k,fresh}\right)$, which in general is not globally solvable.

The global optimal solution is obtained using a reformulation – linearization technique (Quesada & Grossmann, 1995). This method involves reformulating the nonlinear problem by linearizing the bilinear terms and then using its solution within a spatial branch and bound enumeration.

A bilinear term, $\beta = xy$, over the domain $[x^L, x^U] \times [y^L, y^U]$ can be tightly bounded by a relaxed convex underestimator and concave overestimator (Al-Khayyal & Falk, 1983; McCormick, 1976). These linear estimators are given by:

$$\beta \geq x^L y + y^L x - x^L y^L \quad \textbf{4.33}$$

$$\beta \geq x^U y + y^U x - x^U y^U \quad \textbf{4.34}$$

$$\beta \leq x^U y + y^L x - x^U y^L \quad \textbf{4.35}$$

$$\beta \leq x^L y + y^U x - x^L y^U \quad \textbf{4.36}$$

For the current problem the binary terms involved are given by:

$$\beta_i = F_{i,fresh}\ \Psi_k\left(P_{k,fresh}\right) \quad \forall\ i = 1\ to\ N_{sinks} \quad \textbf{4.37}$$

It is clearly evident from the model in section 4.3.1 that the bounds on $F_{i,\,fresh}$ is

$$0 \leq F_{i,fresh} \leq\ F_i^{sink} \quad \forall\ i = 1\ to\ N_{sinks} \quad \textbf{4.38}$$

The bounds on $P_{k,\,fresh}$ is obtained as follows (Qin, 2007).

Taking the summation of the inequality equation 4.31 over all sinks, i, $i = 1,2,3 \ldots\ldots N_{sinks}$, and substituting $\Psi_k(P_{k,i})$ with Equation 4.30, we get

$$\sum_{i=1}^{N_{sinks}} F_i^{sink}\,\Psi_k(P_{k,i})^L \le F_{fresh}\Psi_k(P_{k,fresh}) + \sum_{i=1}^{N_{sinks}}\sum_{j=1}^{N_{sources}} F_{i,j}\Psi_k(P_{k,j}) \le \sum_{i=1}^{N_{sinks}} F_i^{sink}\,\Psi_k(P_{k,i})^U \quad \textbf{4.39}$$

Furthermore,

$$\sum_{i=1}^{N_{sinks}}\sum_{j=1}^{N_{sources}} F_{i,j}\Psi_k[min(P_{k,j})] \le \sum_{i=1}^{N_{sinks}}\sum_{j=1}^{N_{sources}} F_{i,j}\Psi_k(P_{k,j}) \le \sum_{i=1}^{N_{sinks}}\sum_{j=1}^{N_{sources}} F_{i,j}\Psi_k[max(P_{k,j})] \quad \textbf{4.40}$$

And since,

$$\sum_{i=1}^{N_{sinks}}\sum_{j=1}^{N_{sources}} F_{i,j} = \sum_{i=1}^{N_{sinks}} (F_i^{sink}) - F_{fresh} \quad \textbf{4.41}$$

Equation 4.40 is rewritten as

$$\Psi_k[min(P_{k,j})]\left(\sum_{i=1}^{N_{sinks}} (F_i^{sink}) - F_{fresh}\right) \le \sum_{i=1}^{N_{sinks}}\sum_{j=1}^{N_{sources}} F_{i,j}\Psi_k(P_{k,j}) \le \Psi_k[max(P_{k,j})]\left(\sum_{i=1}^{N_{sinks}} (F_i^{sink}) - F_{fresh}\right) \quad \textbf{4.42}$$

By adding $F_{fresh}\Psi_k(P_{k,fresh})$ to Equation 4.42 and combining it with Equation 4.39

$$\sum_{i=1}^{N_{sinks}} F_i^{sink}\,\Psi_k(P_{k,i})^L - \Psi_k[max(P_{k,j})]\left(\sum_{i=1}^{N_{sinks}} (F_i^{sink}) - F_{fresh}\right) \quad \textbf{4.43}$$

$$\leq F_{fresh}\Psi_k\left(P_{k,fresh}\right)$$

$$\leq \sum_{i=1}^{N_{sinks}} F_i^{sink}\,\Psi_k(P_{k,i})^U - \Psi_k\left[min\left(P_{k,j}\right)\right]\left(\sum_{i=1}^{N_{sinks}}\left(F_i^{sink}\right) - F_{fresh}\right)$$

Hence,

$$if\ \Psi_k(P_k) \propto P_k \quad : \quad \Psi_k\left(P_{k,fresh}\right) \in$$

$$\left\{\frac{\sum_{i=1}^{N_{sinks}} F_i^{sink}\,\Psi_k(P_{k,i})^L - \Psi_k\left[max\left(P_{k,j}\right)\right]\left(\sum_{i=1}^{N_{sinks}}\left(F_i^{sink}\right) - F_{fresh}\right)}{F_{fresh}}\right\},$$

$$\left\{\frac{\sum_{i=1}^{N_{sinks}} F_i^{sink}\,\Psi_k(P_{k,i})^U - \Psi_k\left[min\left(P_{k,j}\right)\right]\left(\sum_{i=1}^{N_{sinks}}\left(F_i^{sink}\right) - F_{fresh}\right)}{F_{fresh}}\right\} \quad \textbf{4.44}$$

$$if\ \Psi_k(P_k) \propto \frac{1}{P_k} \quad : \quad \Psi_k\left(P_{k,fresh}\right) \in$$

$$\left\{\frac{\sum_{i=1}^{N_{sinks}} F_i^{sink}\,\Psi_k(P_{k,i})^U - \Psi_k\left[max\left(P_{k,j}\right)\right]\left(\sum_{i=1}^{N_{sinks}}\left(F_i^{sink}\right) - F_{fresh}\right)}{F_{fresh}}\right\},$$

$$\left\{\frac{\sum_{i=1}^{N_{sinks}} F_i^{sink}\,\Psi_k(P_{k,i})^L - \Psi_k\left[min\left(P_{k,j}\right)\right]\left(\sum_{i=1}^{N_{sinks}}\left(F_i^{sink}\right) - F_{fresh}\right)}{F_{fresh}}\right\} \quad \textbf{4.45}$$

For cases, $\sum_{i=1}^{N_{sinks}} F_i^{sink} \leq \sum_{i=1}^{N_{sources}} F_i^{source}$, F_{fresh}^L is used to substitute F_{fresh} for calculating the upper limit and F_{fresh}^U is used to substitute F_{fresh} for calculating the lower limit. These bounds may not be exact but since overestimation of target regions does not do any harm to the model, they serve the purpose of fitting in the spatial branch and bound algorithm. With bounds on

$F_{i,fresh}$, (Equation 4.38) and $\Psi_k(P_{k,fresh})$, (Equation 4.44 and 4.45) known, the mathematical model described in section 4.3.1 is reformulated into a relaxed linear problem formulation using Equations 4.33 - 4.36. The global optimal solution of the proposed problem is obtained using the following procedure as reported by Quesada and Grossmann (1995).

Lower bounds of the global minimum value of the objective function are computed by solving a reformulated linear relaxation model of the original non convex problem. Upper bounds on the global minimum are obtained by any feasible solution of the nonlinear model. The lower bound found above is not a feasible solution to the original nonlinear model, but it can be used as a good initial point to solve the model. The feasible region of the relaxed problem is divided into efficient subregions depending on its parent subregion's lower bound. This is done to effectively improve the quality of lower bound with each partition. Lower and upper bounds over these smaller partition regions are then computed. The upper bound for each subregion is updated whenever a feasible point with an objective function value less than the parent subregion's upper bound is found. Subregions with infeasible solutions with their lower bound close to or above the upper bound, are discarded. If no subregions are left and if the relaxation gap (the difference between upper and lower bounds) is within the specified tolerance the global solution corresponds to the best upper bound. Since the relaxed linear formulation of the problem provides valid lower bounds over the specified partition, the above procedure is guaranteed to converge to a global optimum. Hence, solving the model to find the maximum and minimum values of required properties of the fresh feed into the sink would provide the target properties for molecular design problem. Also the initial bounds on properties of fresh solvent are overestimated, but overestimation of the boundaries of the feasibility region (BFR) is allowable although, more accurate solutions can be found if the BFR is tighter.

4.4 Framework for Integrated Process and Product Design

Having in the hand, the methods to design a process visually and mathematically, this chapter concentrates on pointing out the need for a simultaneous solution to process and molecular design problems. The chapter also gives the developed framework that integrates process design with molecular design. It is evident from above sections of the chapter that solving process design problem based on properties enables it to fall into reverse problem formulation paradigm. It can thus be termed as reverse simulation problem. If the process design problem is solved chemo centrically, the nature of the problem would have been forward.

Solving process design and molecular design problems individually limits the solution space. For example, the properties of fresh material to a process depend on the existing recycle streams within the process. Solving process design problems alone would require committing to specific raw materials well in advance in order to lead to a solution. Hence, when process and product design problems are solved together each benefits from other in the method of designing molecules that meet process performance. The identification of optimal molecule(s) corresponding to optimum process performance is a challenging issue. Molecular design subject to process constraints through Property Integration and Group Contribution Methods is one possible solution to overcome the above limitation. The concept of reverse problem formulation (RPF) (Eden et al., 2003a) has helped formulate integrated process-product design problems without leading to MINLP formulations by insightful decoupling of constitutive equations from the process model. Reverse Problem Formulation enables design of novel molecules and solution of process design problems without commitment to specific components during the solution step. An outline of how the process design and molecular design problems are solved simultaneously is given by Figure 4.5.

Once the set of molecules having properties within the limits set by solving the process design problems are identified, the optimal molecule(s) and the corresponding optical process design(s) can be easily obtained as a solution to a simple linear optimization problem.

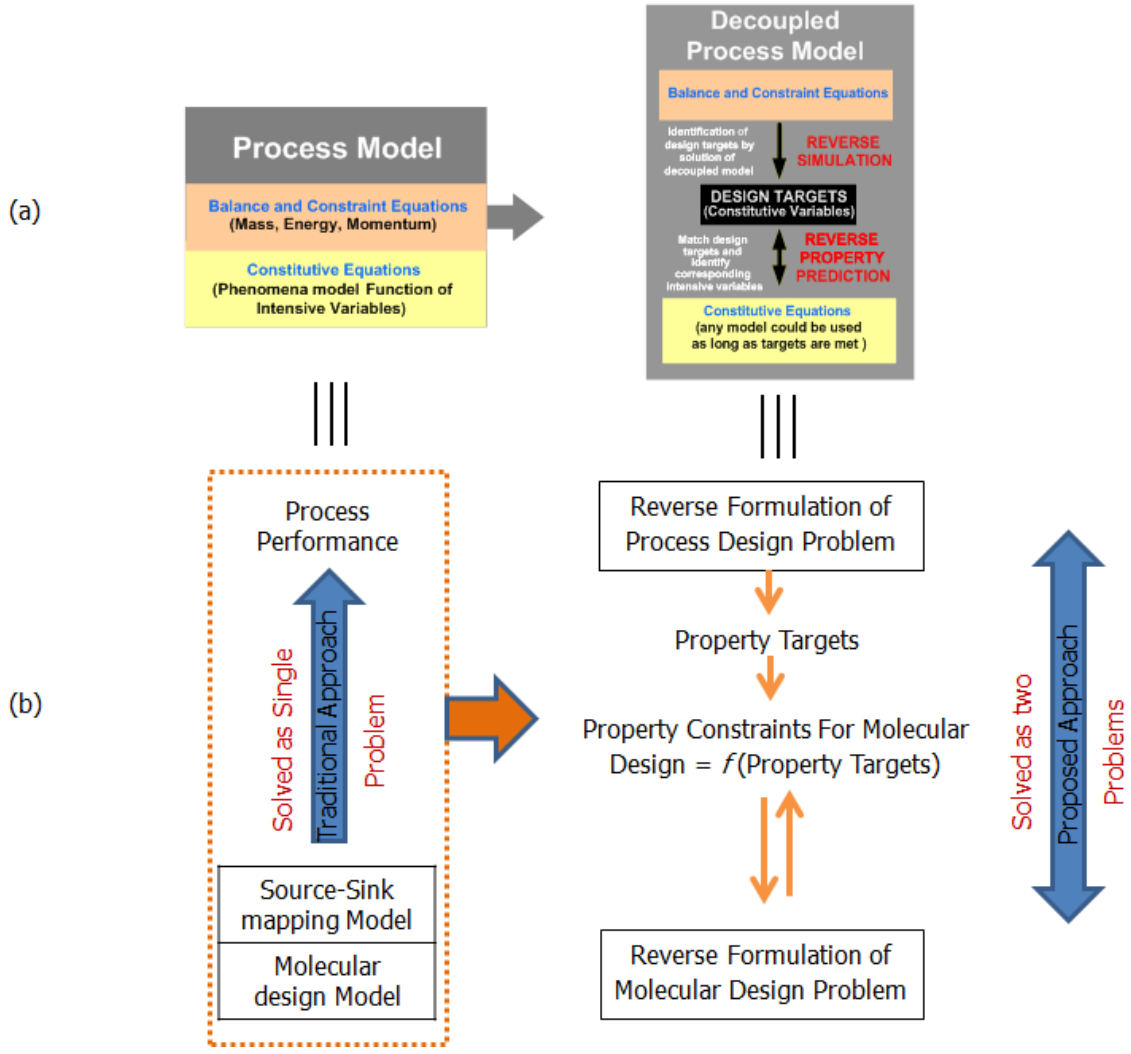


Figure 4.5: (a) Reverse Problem Formulation by Eden et al. (2003a). (b) Proposed framework for simultaneous solution to process & product design problems

Techniques developed by Eden et al. (2003a); Eden, Jorgensen, Gani, and El-Halwagi (2003b), 2003b) for the identification of property targets corresponding to the optimum process performance using a visual approach are shown in section 4.2. Techniques to mathematically

identify the same property targets are shown in section 4.3. The targets thus found from solving a reversely formulated decoupled process design model, $\left[\Psi_k(P_k^L), \Psi_k(P_k^U)\right]$ or $\Psi_k\left(P_k^{goal}\right)$ are translated into property targets for molecular design problem in terms of molecular property functions, $[f^M(P_k^L), f^M(P_k^U)]$ or $f^M{}_k\left(P_k^{goal}\right)$. This kind of translation can be made as a simple correlation as both process and molecular property functions/operators are functions of a known property value. Algorithms to identify the molecules that meet the process are discussed in chapter 3.

4.5 Optimal Solution to Integrated Process & Product Design Problem

As explained above, it is important that process and product design are solved simultaneously as a single problem. In order to achieve an optimal solution, the complexity of such design problems is shown to have been handled by insightfully decoupling the process and product design problems and solving them piecewise based on a reverse solution methodology to achieve their respective new targets as shown in Sections 4.3 and 3.1. These new targets are surrogates of the overall design performance target. On the process side, the task was to find the upper and lower limits of the fresh solvent that may be needed subject to the targeted minimum usage of fresh solvent or minimum waste discharge. On the product side, the task is to identify an optimal solvent or set of solvents subject to the property constraints on the fresh solvent identified by the process design problem. Solving the problem by dividing it into two reverse problems and then integrating them from a property platform, gives us a list of molecules that meet the process performance targets. The integrated process and product design model now is free of the constitutive equations (here the molecular property models) and also, we have solutions to the constitutive equations in the form of a set of feasible molecules and their properties. Allocation of the sources to the sink along with

the fresh solvent make the problem complete. Since the properties of the fresh solvent are now known, the source-sink allocation is easily solved using the synthesized molecules. The model for this becomes linear and thus a global minimum is obtained. The model is comprised of Equations 4.24 - 4.32.

In the case where multiple optimal solvents are identified by the CAMD framework and all the solvents need to be considered, the optimal solvent or set of solvents (different solvent for each sink) is identified based on energy or cost considerations. The integrated process and product design model can now be represented as follows:

Min *Cost*

Subject to:

$$\text{Cost} = f(\text{Amount of fresh solvent for each sink, Cost of the solvent for each sink, waste disposal costs of each source stream, piping costs etc. }) \quad \textbf{4.46}$$

and Equations 4.24 - 4.32.

Though the properties of the fresh molecule(s) are now known, the model becomes nonlinear as the selection of a fresh solvent for a given sink is not known. This problem can again be solved again by a decomposition based methodology. All combinations of identified fresh solvents to the sinks can be generated first and then by fixing those variables, the problem becomes linear which can be easily solved. The combination of fresh solvents that leads to the minimum cost solution can be verified further with rigorous simulation.

4.6 Summary

This chapter covers techniques to solve process design problems from a property platform. The concept of property operators, property clusters and their visual/mathematical treatments have

been analyzed. The chapter finally concentrates on showing the need to solve process and product design problems simultaneously and give a framework that enable to identify an optimal solution for a simultaneous process and product design problem.

4.7 Case Studies – Integrated Process & Product Design

4.7.1 Case Study – Design of solvent for a gas treatment process

Consider a gas treatment process involving five units to purify a gaseous mixture that contains acid gases. Currently four source streams, S_1, S_2, S_3 and S_4 with properties given in Table 4.1 are available in the process as feed to the acid gas removal unit. Design objectives and requirements are to find a fresh solvent source to be mixed with the process sources such that the maximum amount of source streams are utilized.

Qin (2007) solved this type of problem as a mixed-integer non-linear programming problem (MINLP) and Kazantzi, Qin, El-Halwagi, Eljack, and Eden (2007) solved it visually using molecular property clusters. Solving the molecular and process design problems together makes it highly non-linear; hence the problem is divided into two reverse problems approaching the same targets. In this case study, the process design part comprises of a reverse simulation problem, in which the property bounds for the fresh solvent are targeted using process property models. The process design problem involves a non-linear model with bi-linear terms, so the reformulation – linearization technique by Quesada and Grossmann (1995) is used to solve it. The molecular design part is a reverse property prediction problem in which the same property bounds are targeted by using molecular property models. This is solved by a GC based decomposition approach as explained in the previous chapter. This method helps in preventing the formulation of MINLP problems and thus ensures that the problem does not suffer from combinatorial explosion. Group

contribution data for the properties considered were taken from Marrero and Gani (2001, 2002). Hence, the process and molecular design problems are integrated through the same property targets.

The following three properties are considered: critical temperature (T_c), heat of vaporization (H_v) and heat of fusion (H_{fus}). Additionally, two thermal constraints are imposed on the synthesized molecules to make sure that the designed molecule(s) will remain in liquid state at the process conditions and to prevent excessive solvent losses via evaporation. Also the boiling point constraint ensures that the solvents' flammability is checked, as the parameter used in the context of flammability, the flash point, directly depends on boiling point (Affens, 1966). An additional environmental constraint is imposed to check the toxicity level of solvents used. LC_{50} is the parameter that measures the limiting concentration of material to which test organisms can be exposed. The higher the LC_{50} value, the less toxic the substance is. The octanol/water partition coefficient (K_{ow}) is defined as the ratio of a chemical's concentration in the octanol phase to its concentration in the aqueous phase of a two-phase octanol/water system. It represents the tendency of the chemical to partition itself between an organic phase (e.g., a fish, soil) and an aqueous phase. The LC_{50} value is related to the K_{ow} value (Konemann, 1981) and hence a constraint on the log K_{ow} value places a bound on the toxicity. A constraint of log $K_{ow} < 4$ is imposed based on studies of reducing the environmental impact of acid gas control technologies.

The process sinks' property bounds as well as flow rate and property data for all streams (S_1, S_2, S_3 and S_4) are summarized in Table 4.1.

Table 4.1: Property data for gas purification process.

Property	**Property bounds on process sinks**	**S_1**	**S_2**	**S_3**	**S_4**
T_c (K)	Sink 1: [667.0 ; 730.0]	678.0	670.0	715.0	699.0

	Sink 2: [672.0 ; 730.0]				
	Sink 3: [672.0 ; 735.0]				
	Sink 4: [675.0 ; 735.0]				
	Sink 5: [675.0 ; 740.0]				
	Sink 1: [90.5 ; 100.0]				
	Sink 2: [90.5 ; 104.5]				
H_v (kJ/mol)	Sink 3: [95.0 ; 104.5]	95.0	64.0	85.0	82.0
	Sink 4: [95.0 ; 110.0]				
	Sink 5: [100.5 ; 110.0]				
	Sink 1: [18.0 ; 34.0]				
	Sink 2: [18.0 ; 34.5]				
H_{fus} (kJ/mol)	Sink 3: [20.0 ; 34.5]	23.0	18.4	22.1	21.7
	Sink 4: [20.0 ; 35.0]				
	Sink 5: [21.0 ; 35.0]				
	Sink 1: 200				
	Sink 2: 210				
Flowrate, *F*	Sink 3: 230	60	90	70	60
(kmol/month)	Sink 4: 190				
	Sink 5: 170				

As explained in the previous chapter, though only three properties are handled, it becomes hard to keep track of five sinks and four process streams to identify optimal property values of the required fresh solvent. Hence, a mathematical method is used to identify the property targets for the fresh stream.

The selected process properties combine linearly and hence the property operator of each is the property itself. The process property models of the above properties are given by:

$$f^P(T_c) = T_c = \sum_{S_i} x_{S_i} T_{c,S_i} \quad \textbf{4.47}$$

$$f^P(H_v) = H_v = \sum_{S_i} x_{S_i} V_{c,S_i} \quad \textbf{4.48}$$

$$f^P(H_{fus}) = H_{fus} = \sum_{S_i} x_{S_i} V_{c,S_i} \quad \textbf{4.49}$$

$$x_{S_i} = \frac{F_{S_i}}{F} \quad \textbf{4.50}$$

where, S_i represents the streams that are mixed and F represents the total flowrate of the stream mixture. f^P indicates process operator, while Ψ_k is used for simplicity purposes.

The property bounds for the fresh solvent are determined by employing the reformulation-linearization technique and the reformulated model is solved using the commercial optimization software, LINGO. The identified property targets are listed in Table 4.2.

These values are utilized as the property constraints in the molecular design problem.

Table 4.2: Property targets for fresh solvent.

Process Property Targets	Lower Limit	Upper Limit
T_c (K)	665.4	751.2

H_v (kJ/mol)	100.5	115.6
H_{fus} (kJ/mol)	19.4	39.9

Additionally, property targets for boiling and melting points and environment related properties are given in Table 4.3.

Table 4.3: Additional property constraints.

Property	Lower Limit	Upper Limit
T_m (K)		340
T_b (K)	480	
log K_{ow}		4

Molecular Design by decomposition approach

Based on the problem data and constraints, 13 first-order groups are pre-selected from the group tables in Marrero and Gani (2001, 2002). The molecular property models for the targeted properties are given by:

$$f^M(T_c) = \sum_{g=1}^{N_g} n_g T_{cg} \quad \textbf{4.51}$$

$$f^M(H_v) = \sum_{g=1}^{N_g} n_g H_{vg} \quad \textbf{4.52}$$

$$f^M(H_{fus}) = \sum_{g=1}^{N_g} n_g H_{fusg} \quad \textbf{4.53}$$

$$f^M(T_b) = \sum_{g=1}^{N_g} n_g T_{bg} \quad \textbf{4.54}$$

$$f^M(T_m) = \sum_{g=1}^{N_g} n_g T_{mg} \quad \textbf{4.55}$$

$$f^M(logK_{ow}) = \sum_{g=1}^{N_g} n_g logK_{0wg} \quad \textbf{4.56}$$

Since the molecular property operators and process property operators target the same property, a relation between them is needed to integrate the process and molecular design problems. The functions that link molecular property operators to the process property operators are given below:

$$f^M(T_c) = exp\left(\frac{f^P(T_c)}{t_{co}}\right) \quad , t_{co} = 231.239 \quad \textbf{4.57}$$

$$f^M(H_v) = f^P(H_v) - h_{v0} \quad , h_{v0} = 11.733 \quad \textbf{4.58}$$

$$f^M\left(H_{fus}\right) = f^P\left(H_{fus}\right) - h_{fus0} \quad , h_{fus0} = -2.806 \quad \textbf{4.59}$$

$$f^M(T_b) = exp\left(\frac{T_b}{t_{bo}}\right) \quad , t_{bo} = 222.543 \quad \textbf{4.60}$$

$$f^M(T_m) = exp\left(\frac{T_m}{t_{mo}}\right) \quad , t_{mo} = 147.45 \quad \textbf{4.61}$$

$$f^M(logK_{ow}) = logK_{ow} - K_{ow0} \quad , K_{ow0} = 0.543 \quad \textbf{4.62}$$

The property targets from the process design problem are translated using Equations 4.57- 4.62 and are used as property targets in the molecular design problem.

Table 4.4 gives the property target inputs to the molecular design model.

Table 4.4: Molecular property targets.

Molecular Property Targets	Lower Limit	Upper Limit
T_c (K)	17.73983	25.62062
H_v (kJ/mol)	88.267	103.267
H_{fus} (kJ/mol)	22.806	42.806
T_m (K)		10.032
T_b (K)	8.644	
log K_{ow}		3.457

The following first order groups have been considered for molecular design:

1	CH_3	5	OH	9	C-O	13	CH-NH
2	CH_2	6	CH_3-O	10	CH_2-NH_2		
3	CH	7	CH_2-O	11	CH3-NH		
4	C	8	CH-O	12	CH_2-NH		

Subproblem 1:

Property data for these selected groups is taken from Marrero and Gani (2001, 2002), which are included in Appendix A. Now, inequality expressions for each property are formulated. The number of preselected first order molecular groups is maximized subject to the specific constraints mentioned in Table 4.4. The reason behind maximizing these groups is to ensure that no potential molecule is left out. The variations in the property values caused by inclusion of the higher order groups will be considered in the later stages. To insure water solubility and to reduce vapor

pressure, the molecule must have two or more –OH groups. To limit the extent of corrosion, only one amino group is allowed to be in the amine (N in the amino group either connects to H or C). Finally, to limit detrimental effects of direct exposure to the solvent, tertiary amines are ruled out in this case study.

The maximum values are as follows:

	max		**max**		**max**		**max**
CH_3	5	OH	3	C-O	2	CH-NH	1
CH_2	7	CH_3-O	3	CH_2-NH_2	1		
CH	4	CH_2-O	4	CH3-NH	1		
C	2	CH-O	2	CH_2-NH	1		

Subproblem 2:

The class and category of the groups considered above are obtained from the group classification tables by Gani et al. (1991).

Table 4.5: Class and Category of selected first-order groups (Gani et al., 1991).

Group	**Class**	**Category**	**Group**	**Class**	**Category**
CH_3	1	1	CH-O	3	4
CH_2	2	1	C-O	4	4
CH	3	1	CH_2-NH_2	1	4
C	4	1	CH3-NH	1	4

OH	1	4	CH_2-NH	2	4
CH_3-O	1	4	CH-NH	3	4
CH_2-O	2	4			

1500 group subsets whose FBN is zero are generated. For each of these subsets, the number of groups, n_X in each of category *X* such that *X= 3, 4, 5, 3+4+5, 4+5,* is identified and checked against C^X_{T,L,n_L} such that:

$$\sum_X n_X \leq C^X_{T,L,n_L} \quad \textbf{4.63}$$

where, *L* is the largest possible class; n_L is number of groups in class *L*; *T* is the total number of first order groups in the group subset and $C^X_{T,L,n}$ is identified using Table A.15

44 of the tested group subsets satisfy the class and category constraints and these are considered for further subproblems.

Subproblem 3:

The possible higher-order groups are identified in Table 4.6. Groups 1- 9 are second-order and 10 – 13 are third-order groups. The third-order groups are considered in cases of open-chain polyfunctional compounds with more than four carbon atoms in the main chain.

Table 4.6: Possible higher order groups.

1	$(CH_3)_2CH$
2	$(CH_3)_3C$
3	$CH(CH_3)CH(CH_3)$

4	$CH(CH_3)C(CH_3)_2$	
5	$C(CH_3)_2C(CH_3)_2$	
6	CHOH	
7	COH	
8	$CH_m(OH)CH_n(OH)$	(m,n, 0..2)
9	$CH_m(OH)CH_n(NH_p)$	(m,n,p, 0..2)
10	$NH_2(CH_n)_mOH$	(m>2, n in 0..2)
11	$HO(CH_n)_mOH$	(m>2, n in 0..2)
12	$HO(CH_p)_k$-O-$(CH_n)_mOH$	(m,k>0, p,n in 0..2)
13	$HO(CH_p)_k$-NH_x-$(CH_n)_m$-OH	(m,k>0, p,n,x in 0..2)

Where, m,n,p,k,x are integers within the bounds mentioned for them respectively for each higher-order group. The maximum possible number of each higher-group is computed algebraically as shown below.

1. $(CH_3)_2CH$

   If the group subsets comprises of only 3 CH3 and 1CH groups, $n_h = 2$;

   Else, $n_h = int\left(\min\left(\frac{n_{CH_3}}{2}; n_{CH}\right)\right)$

2. $(CH_3)_3C$

   If the group subsets comprises of only 4 CH3 and 1Cgroups, $n_h = 2$

   Else, $n_h = int\left(\min\left(\frac{n_{CH_3}}{3}; n_C\right)\right)$

3. $CH(CH_3)CH(CH_3)$

Maximum number of groups of this kind can exist when the groups $CH(CH_3)$ are positioned in a series, Hence, if *a* is the number of $CH(CH_3)$ groups i.e.,

$If\ a = int\left(\min\left(n_{CH_3}; n_{CH}\right)\right) \quad and \quad a > 1;$

$n_h = a - 1 \quad Else \quad n_h = 0$

4. $CH(CH_3)C(CH_3)_2$

Maximum number of groups of this kind can exist when the groups $CH(CH_3)C(CH_3)_2$ are positioned in series. Hence, if *a* is the number of $CH(CH_3)C(CH_3)_2$ groups i.e.,

$If\ a = int\left(\min\left(\frac{n_{CH_3}}{3}; n_{CH}; n_C\right)\right) \quad and \ \ a > 0$

$n_h = 2a - 1 \quad Else \quad n_h = 0$

Again, if a balance of CH, CH3 and C groups exists in the group subset, existence of one $CH(CH_3)$ or one $C(CH_3)_2$ group is possible. In these cases n_h is incremented by 1.

5. $C(CH_3)_2C(CH_3)_2$

Maximum number of groups of this kind can exist when the groups $C(CH_3)_2$ are positioned in a series, Hence, if *a* is the number of C $(CH_3)_2$ groups i.e.,

$If\ a = int\left(\min\left(\frac{n_{CH_3}}{2}; n_C\right)\right) \quad and \quad a > 1;$

$n_h = a - 1 \quad Else \quad n_h = 0$

6. CHOH

If *a* is the number of $CH(OH)_2$ groups, *b* is the number of CHOH groups from balance CH and OH groups, i.e.

$If\ a = int\left(\min\left(\frac{n_{OH}}{2}; n_{CH}\right)\right) \ \ and \ \ b = int(\min(n_{OH}; n_{CH}))$

$n_h = 2a + b$

7. COH

If *a* is the number of $C(OH)_3$ groups, *b* is the number of $C(OH)_2$ groups from balance CH and OH groups and *c* is the is the number of C(OH) groups from balance CH and OH groups , i.e.

$If\ a = int\left(\min\left(\frac{n_{OH}}{3}; n_C\right)\right)\ ; a = int\left(\min\left(\frac{n_{OH}}{2}; n_C\right)\right)\ and$

$c = int(\min(n_{OH}; n_C))$

$n_h = 3a + 2b + c$

8. $CH_m(OH)CH_n(OH)$; (*m*,*n*, 0..2)

   If *a* is the number of COH groups, *b* is the number of CHOH groups using balance OH groups and *c* is the number of $C(OH)_2$ groups from balance OH groups and *d* is the balance OH groups ranging from 0 to 2, i.e.

   $If\ a = int(\min(n_{OH}; n_C))\ ;\ \ b = int(\min(n_{OH}; n_{CH}))\ \ and$

   $c = \ int\left(\min\left(\frac{n_{OH}}{2}; n_C\right)\right)$

   $n_h = a + b + c - 1 + d$

9. $CH_m(OH)CH_n(NH_p)$; (*m*,*n*,*p*, 0..2)

   If *a* is the number of $CH_x(OH)CH_y(NH_p)$ with *x*,*y* ranging from 0 to 1, *b* is the number of $C(OH)_2$ groups from balance OH groups, *c* is the sum of balance OH groups and $CH_2(NH_p)$ groups ranging from 0 to 2, i.e.

   $If\ a = int\left(\min\left(n_{OH}\,;\ n_{CHNH_p} + n_{CNH_p};\ n_C + n_{CH}\right)\right)\ ;$

   $a = int\left(\min\left(\frac{n_{OH}}{2}; n_C\right)\right)\ \ and$

   $n_h = 2a - 1 + b + c$

10. $NH_2(CH_n)_mOH$; (*m*>2, *n* in 0..2)

If *a* is the number of $(NH_2)CH_n$ and OH sets, *b* is number of sets of three CH_n groups and *c* is balance OH groups ranging from 0 to min of *b* and $2n_C + n_{CH}$, i.e.

$$If\ a = int\left(\min\left(n_{CH_n(NH_2)};\ n_{OH}\right)\right) \quad and \quad b = \frac{n_{CH_n}}{3}$$

$$n_h = int(\min(a;\ b)) + c$$

11. $HO(CH_n)_mOH$; (*m*>2, *n* in 0..2)

    If *a* is the number of sets of 2 OH groups, *b* is number of sets of three CH_n groups, and *c* is balance OH groups ranging from 0 to min of *b* and $2n_C + n_{CH}$, i.e.

$$If\ a = \frac{n_{OH}}{2} \quad and \quad b = \frac{n_{CH_n}}{3}$$

$$n_h = int(\min(a;\ b)) + c$$

12. $HO(CH_p)_k$-O-$(CH_n)_mOH$; (*m*,*k*>0, *p*,*n* in 0..2)

    If *a* is the number of sets of 2 OH groups, *b* is number of sets of 2 CH_n groups and *1* CH_p-O group , *c* is balance OH groups ranging from 0 to min of *b* and $2n_C + n_{CH}$, i.e.

$$If\ a = \frac{n_{OH}}{2} \quad and \quad b = int\left(\min\left(\frac{n_{CH_n}}{2};\ n_{CH_p-O}\right)\right)$$

$$n_h = int(\min(a;\ b)) + c$$

13. $HO(CH_p)_k$-NH_x-$(CH_n)_m$-OH ; (*m*,*k*>0, *p*,*n*,*x* in 0..2)

    If *a* is the number of sets of 2 OH groups, *b* is number of sets of 2 CH_n groups and *1* $CH_p(NH_p)$ group, *c* is balance OH groups ranging from 0 to min of *b* and $2n_C + n_{CH}$, i.e.

$$If\ a = \frac{n_{OH}}{2} \quad and \quad b = int\left(\min\left(\frac{n_{CH_n}}{2};\ n_{CH_p(NH_p)}\right)\right)$$

$$n_h = int(\min(a;\ b)) + c$$

Subproblem 4:

For each of group subset to be tested in this subproblem, maximum and minimum contribution from possible higher-order groups is estimated. Based on the molecular property constraints as

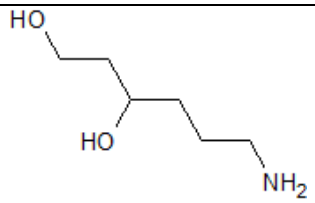
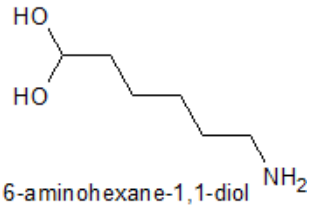
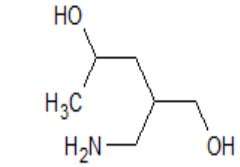
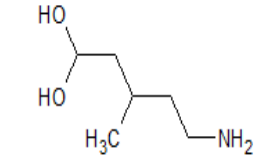
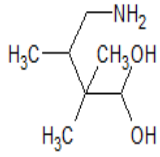
listed in Table 4.4, group subsets whose property range falls completely outside the targeted property range of molecules are excluded from being considered further. 32 group subsets obey the property constraints and are hence considered for the next subproblem.

Subproblem 5:

All possible combinations of higher order groups for each group subset capable of forming structurally and functionally feasible molecules are enumerated. This ensures the identification of structural isomers as the possibility of nonexistence of each higher order groups is considered which in turn leads to different conformations of first order groups. 260 group subsets have been enumerated and their property values estimated using Equation 2.12. Group subsets that possess the targeted properties as listed in Table 4.4 are identified. Finally 31 group subsets are considered for checking if these subsets form structurally feasible molecules i.e. having the list of first- and higher-order groups, molecules are formed by satisfying the bonding requirements of each first-order group and also by checking if the molecular conformation leads to the specified higher-order groups.

The following molecules that could be used as a fresh solvent are identified.

Table 4.7: Valid Molecules for Acid Gas Problem

	Molecule	T_c (K)	H_v (kJ/mol)	H_{fus} (kJ/mol)	T_m (K)	T_b (K)	log K_{ow}
1	 6-aminohexane-1,3-diol	733.5575	102.989	30.3352	326.2293	529.5349	3.92638
2	 6-aminohexane-1,1-diol	726.9041	102.783	29.7362	320.4894	523.6358	3.94885
3	 2-(aminomethyl)pentane-1,4-diol	730.4473	101.348	26.8512	323.2631	523.5617	3.74937
4	 5-amino-3-methylpentane-1,1-diol	723.7024	101.142	26.2522	317.4042	517.4999	3.77184
5	 4-amino-2,2,3-trimethylbutane-1,1-diol	740.9033	103.109	23.4132	327.0837	530.1933	3.95992

The generated molecular structures have properties consistent with the property limits of the sinks and also since subproblem 5 of the molecular design considers the elimination of each possible

higher order group, structural isomers due to the positioning and existence of higher-order can be identified.

Optimal Solution to Integrated Process & Product Design Problem

Once the variables of the reverse property prediction problem (molecular design) have been identified, an optimal molecule can be selected by virtue of its cost or any additional constraints. No economic data was found in the open literature for the identified solvents, therefore the final optimal solution set in terms of cost is hard to find. All the molecules above lead to a solution where the minimum amount of fresh solvent is used. If, molecule 3 is considered to obtain a source sink allocation, a simple linear optimization problem can be formulated and the source-sink allocations can be easily identified for the selected molecule.

Table 4.8: Source-Sink allocation with molecule 3 as fresh solvent.

	Sink 1	**Sink 2**	**Sink 3**	**Sink 4**	**Sink 5**
Source 1	0.00	0.00	60.00	0.00	0.00
Source 2	27.45	0.00	26.39	32.29	3.86
Source 3	70.00	0.00	0.00	0.00	0.00
Source 4	0.00	55.17	4.83	0.00	0.00
Fresh solvent	102.55	154.83	138.78	157.71	166.14

5. Integrated Process & Product Synthesis/Design

"Given the product requirements, develop a framework for solving process and product synthesis/design problems and their integration to ultimately identify the optimal process flowsheet to manufacture it"

In this chapter the details of the solution to the above question is presented. A systematic framework for integrated process and product synthesis/design from a property perspective that is also non-iterative in nature is presented here. This chapter starts with a main section that gives the outline of the integrated framework followed by three important sections that present the requirements of the framework. Finally, the software implementation of this part of the framework is given in detail in this chapter.

5.1 Framework for Integrated Process and Product Design

The interactions among process synthesis, process design and molecular design is through a common set of properties that are employed to analyze the processes as well as external agents involved in the process. Knowledge of these specific properties is needed to establish the feasibility of a unit operation in a process and the corresponding conditions of operation. The same information is needed for design of a component as an appropriate external agent. This forms the very basis of the proposed hybrid methodology for flowsheet synthesis/design integrated with

molecular design. As pointed out in Figure 5.1, the relation between properties and process/product synthesis and design has to be exploited to be carefully used in the framework.

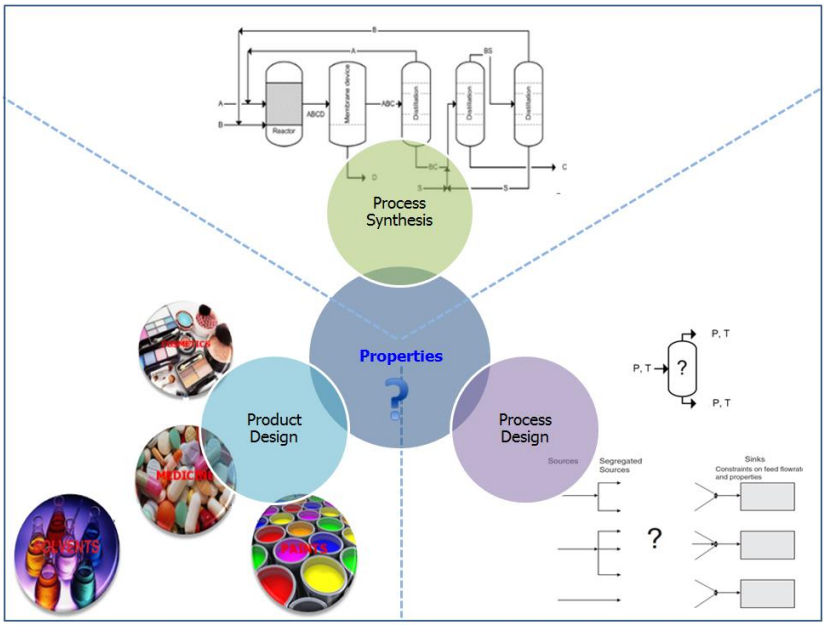


Figure 5.1: Relation between properties and process synthesis , design, product design

With product and process synthesis/design being highly integrated, when isolated from each other, it is evident that both have some inherent limitations due to the amount of information that is required prior to invoking the design algorithm. To overcome the limitations encompassed by isolating the process and molecular design problems, a hybrid method for Computer Aided Flowsheet Design (CAFD) and its effective integration with molecular design (CAMD) is developed. Using this approach, the process flowsheet is synthesized and its design variables, involved molecules, which facilitate the desired process performance target, are identified.

Given the product that needs to be synthesized, Figure 5.2 shows the outline of methodology employed for solving integrated process and product design problems.

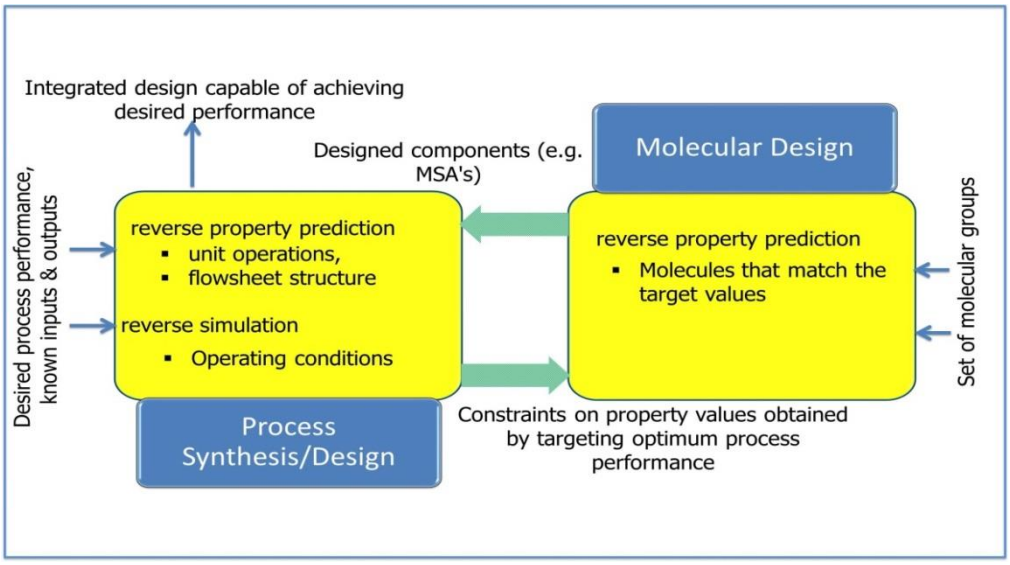


Figure 5.2: Methodology for integrated process and design.

Both the developed CAFD and CAMD frameworks are group contribution (GC) based approaches. CAFD makes use of functional process groups, characterized by the type of unit operation/process and their corresponding driving force, to generate and represent flowsheets; process group contribution based property models to predict flowsheet properties from apriori regressed contributions of process groups; a notation system for storing the flowsheet structural information; and a synthesis method to generate and identify the feasible flowsheets. The identified candidate flowsheets are ranked based on flowsheet properties (e.g. energy consumption, amount (mass) of external agents used and/or cost/profit) representing flowsheet performance in a quantitative sense. Once the promising flowsheet structures are identified, the flowsheet design parameters that describe the process will be estimated. The reverse simulation method is used to calculate the design variables of the unit operations involved in the process. This also gives a good estimate of the important design parameters. Some alternatives may involve unit operations that require an external agent. Also, the properties of any external agent needed by the process may depend on potential existing recycle streams within the process. Thus, a property based platform is utilized

for integrating process and molecular design. The property clustering technique introduced by Shelley and El-Halwagi (2000) provides the tools to track properties. Conventional agents may not always meet the property constraints set by the reverse simulation design problem of such operations. Novel agents can be identified by solving the product design problem, which includes satisfying these property constraints. This is done by integrating the flowsheet design problem with a molecular design problem. Depending on the type of unit operation in the process where an external agent is required, the CAMD problem is formulated accordingly and the effect of the solution alternatives from the CAMD problem on the process is evaluated by the process models. CAMD includes building blocks (atoms and functional groups) to generate and represent molecules; group contribution based property models to predict target properties; a standard molecular structure notation system to store and visualize the molecular structure information; and a synthesis method to generate and screen molecules that match the target (design) properties. Once a set of near optimal flowsheet alternatives have been identified, rigorous simulation is used to verify the predicted performance and select the best flowsheet. The framework also aims at maintaining a good accuracy of solutions and large application range.

The principal concepts introduced in the chapter 2 are utilized to develop the framework shown in Figure 5.2. The requirements for developing the framework are listed below:

Process Synthesis/Design

a) Systematic methods for selecting/screening unit operations involved in the process.
b) Simple string-like representations capable of systematically representing process alternatives.
c) Models to calculate each flowsheet property.
d) Methods to synthesize flowsheets that meet the targets.

e) Method to calculate design parameters of the flowsheet.

Requirements a-c and e are based on available literature and listed in the next sections in the sequence they are utilized. Requirement d is what is addressed in this dissertation and it is one of the important step in CAFD.

Integration of Process & Molecular Design

a) Translation of information from process design problem to molecular design.
b) Nature of external agent based on the process recycle streams, type of unit operation and solubility data.

These requirements are addressed in chapter 4.

Molecular Design

a) Models to calculate each desired molecular property.
b) Methods to synthesize molecules that meet target properties.

These requirements are addressed in chapter 3.

Hence, this chapter deals with developing an integrated framework for process and product synthesis/design by using the concepts from literature and those shown in chapters 3 and 4.

5.2 Process Synthesis/Design by decomposition based approach

There are many different approaches to process synthesis including expert systems, optimization or algorithmic methods, and conceptual methods based on physical insights. The objective here is, given the raw materials and the desired product specifications, to develop a methodology that systematically identifies the optimal process flowsheet transforming the raw materials into the desired products in a more efficient manner. This dissertation highlights a novel hybrid method for Computer Aided Flowsheet Design (CAFD) that combines physical insights with algorithmic

reverse design approaches to enable systematic identification of feasible flowsheets at significantly reduced computational expense. Large number of components have to be dealt with while developing a process flowsheet. Also, there are a wide range of unit operations that are available. This leads to the possibility of a large number of process flowsheet alternatives. This complexity is handled by a decomposition approach based on a targeted reverse methodology. Preliminary screening of alternatives is done based on thermodynamic insights. Secondly, more promising alternatives are screened based on performance indices. Finally the reduced list of alternatives are verified using rigorous simulations. The developed framework utilizes the group contribution (GC) approach developed by d'Anterroches and Gani (2005) in which, the flowsheets could be represented using appropriate process descriptors. Also the framework relies on the exploited relationships between properties and separation process principles by (Jaksland et al., 1995).

5.2.1 Methods for selecting/screening unit operations

For a typical process synthesis problem, the first thing to be known is if the required product is naturally available (although in most cases with some impurities) or if a reaction route is needed to synthesize it. Different reaction pathways based on the problem can be identified from literature. Based on this background information different mixture streams can be imitated and the tasks of identifying a) the respective upstream or downstream separation schemes b) the optimal among all the identified process flowsheets remains to be worked on.

Separation of a mixture is caused by the addition of a separating agent, which takes the form of another stream of matter or of energy (King, 1980). Separation processes in general can be classified as three types – Mechanical separation processes, Equilibrium separation processes (energy separating agent based and mass- separating agent based) and Rate – governed separation processes. For any separation process classified as above, having an insight on their underlying

physical or chemical principles and phenomena, one or more specific pure component property can be associated with it. Based on the evaluation of appropriateness of those characteristic property differences between the key components of a mixture to effect the separations, respective separation process may be selected as an alternative for separating/splitting the mixture across the key components. One such physical insights based methodology is by (Jaksland et al., 1995). Here, for a separation between two components and a given separation process, the separation process can be checked for its feasibility based on the ratios of its corresponding pure component properties. If the ratio is above a threshold value corresponding to this operation task and if some additional property constraints are satisfied the operation task is considered as feasible. For example, if the melting point ratio is larger or equal to 1.20 and the minimum value of melting point is greater than a T_{min} of 250K, crystallization is considered as very feasible separation process for the binary mixture. The characteristic properties and their respective threshold limits are given in the appendix. In some cases, where the very feasible separation processes cannot be identified, just feasible separation processes can be considered as an alternative with a warning. In identifying these, the trapezoidal representation (Qian & Lien, 1995) of threshold limits for a separation process is applied. This representation helps in predicting the extent of feasibility of a separation process for binary mixture.

5.2.2 Process Descriptors

The process groups by (d'Anterroches & Gani, 2005) are used in development of the framework. The process flowsheet can be easily generated and represented by functional process-groups (PGs). These PGs represent different unit operations such as reactor, distillation, extractive distillation etc. and hence characterized by the type of unit operation. The bonds among the process groups

represent the streams connecting the unit operations. Also the separation PGs are characterized by driving force and reactor PGs are characterized by their attainable region.

Driving Force

The driving force (DF) of component *i* with respect to component *k* in a binary pair for property (or separation process) *j*, D_{ij}, is given by

$$D_{ij} = y_i - x_i = \frac{x_i \beta_{ij}}{1 + x_i(\beta_{ij} - 1)} \qquad \textbf{5.1}$$

Where, x_i and y_i are compositions of component *i* in two co-existing phases and β_{ij} is an adjustable parameter with respect to separation process, *j*, that may be a constant or a function of other variables (Gani & Bek-Pedersen, 2000). Also, it is evident that all the component indices are in principle with respect to the second component *k* (in the binary pair). This difference in composition of *i* in coexisting phases may be due to thermodynamic equilibrium in case of equilibrium separation processes like distillation. Nevertheless, transport mechanisms other than thermodynamic equilibrium, for instance diffusion or convection, can also promote driving forces, and enable the separation to take place. For different compositions of *i*, we can hence obtain corresponding DF by knowing β_{ij} for a given separation process or phase composition data. Figure 5.3 gives the graphical representation of the phase composition data (that is, plot of the DF as a function of liquid (or vapor) composition). It can be seen that DF in general is a concave function with respect to composition with a well-defined maximum. Since, using the maximum driving force for a given operation enables it to cause the change easily, the energy needed to be added to or removed from the system to create and maintain two coexisting vapor liquid phases is indirectly related to the DF. i.e. if DF is large, less energy is involved, while if DF is small more energy is involved. Hence having this information readily available for a PG, we can have an

estimate of minimum energy needed for utilizing and hence the optimal conditions of operation of the separation technique without trial and error calculations.

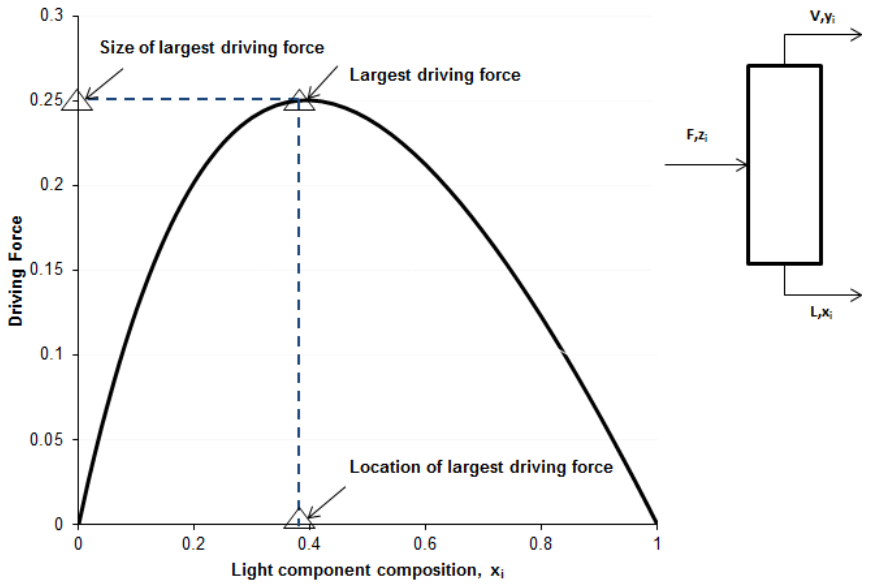


Figure 5.3: Driving force as a function of composition (Bek-Pedersen, 2003)

Attainable region

The attainable region (AR) is defined as the set of all possible outcomes, for the considered system, that can be achieved using the fundamental processes operating within the system, and that satisfies all constraints placed on the system. Fundamental processes that may be considered are physical and chemical phenomena such as mixing, reaction, and heat or mass transfer. Based on the available kinetics of the reaction scheme and having defined the constraints of the system, the boundary of the AR can be constructed. Once the AR has been found the resulting boundary is interpreted as series of various processing units and the optimal network can be effortlessly obtained. Hence, having the reactor process group characterized by its AR, we can ensure that the superstructure of all reactor systems is included in the analysis, removing the reliance on the user's imagination to create reactor structures and the optimal network can be identified avoiding trial

and error calculations. For example, for the trambouze reaction scheme, defined by the following reaction set and kinetics, the AR in concentrations of A&C space is shown in Figure 5.4. It shows that the boundary of AR is defined by a "bypassed" CSTR reactor between point O and A, followed by plug flow reactor.

Reactions of trambouze reaction scheme:

$$A \rightarrow B \quad (1)$$
$$A \rightarrow C \quad (2) \qquad \textbf{5.2}$$
$$A \rightarrow D \quad (3)$$

Reaction kinetics of trambouze reaction scheme:

$$r_1 = \frac{k_1 c_A}{1 + k_4 c_A}$$
$$r_2 \rightarrow k_2 c_A \quad (2) \qquad \textbf{5.3}$$
$$r_3 = \frac{k_3 {c^2}_A}{1 + k_5 {c^2}_A}$$

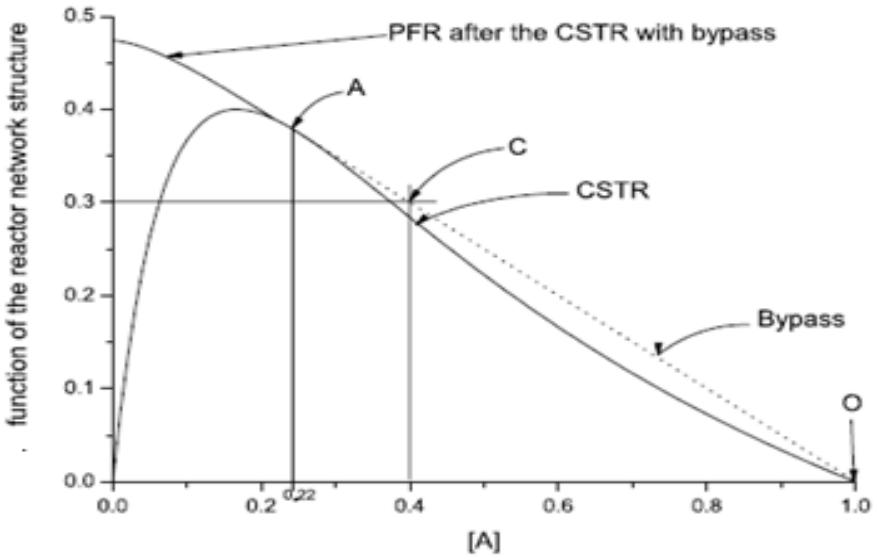


Figure 5.4: Example of attainable region for the trambouze reaction scheme.

Again, it is also evident that since flowsheet synthesis is done based on the premise that each separation process operates utilizing maximum driving force, the purity of outlet process groups can be ensured to be at least 99.5% pure. This enables the simple mass balance around process groups at the final stage.

Figure 5.5 shows a flowsheet with reaction process group, distillation, membrane separation process group and a extractive distillation process group along. It can be seen that the process groups represent either a unit operation (such as reactor, distillation, flash), or a set of unit operations (such as, two distillation columns in extractive distillation, pressure swing distillation). Consider the process flowsheet in Figure 5.5; the feed mixtures are represented by two inlet process groups; (*iA*), (*iB*). The end products are represented by four outlet process groups; (*oA*), (*oB*), (*oC*), (*oD*). The four process groups representing a reactor *(rAB/pABCD)*, distillation (*AB*/*CD*), a solvent based separation (*cycB/C*) and a molecular sieve separation *(msA/B)* have at least one inlet and one outlet streams.

It is evident that, the same process groups can be used to represent different components having similar properties. Note, however, that the inlet and outlet streams (bonds) of process groups must maintain the list of components present in them and that the path of a component through a process group establishes the flowsheet structure. That is, process groups (*AB/CD*) and (ms*A/B*) can be connected to form [(*AB/CD*)(ms*A/B*)] without knowing the identities of the components *A*, *B*, and *C*. Only when the properties of the flowsheet need to be calculated, the identities of the chemicals (components) are needed.

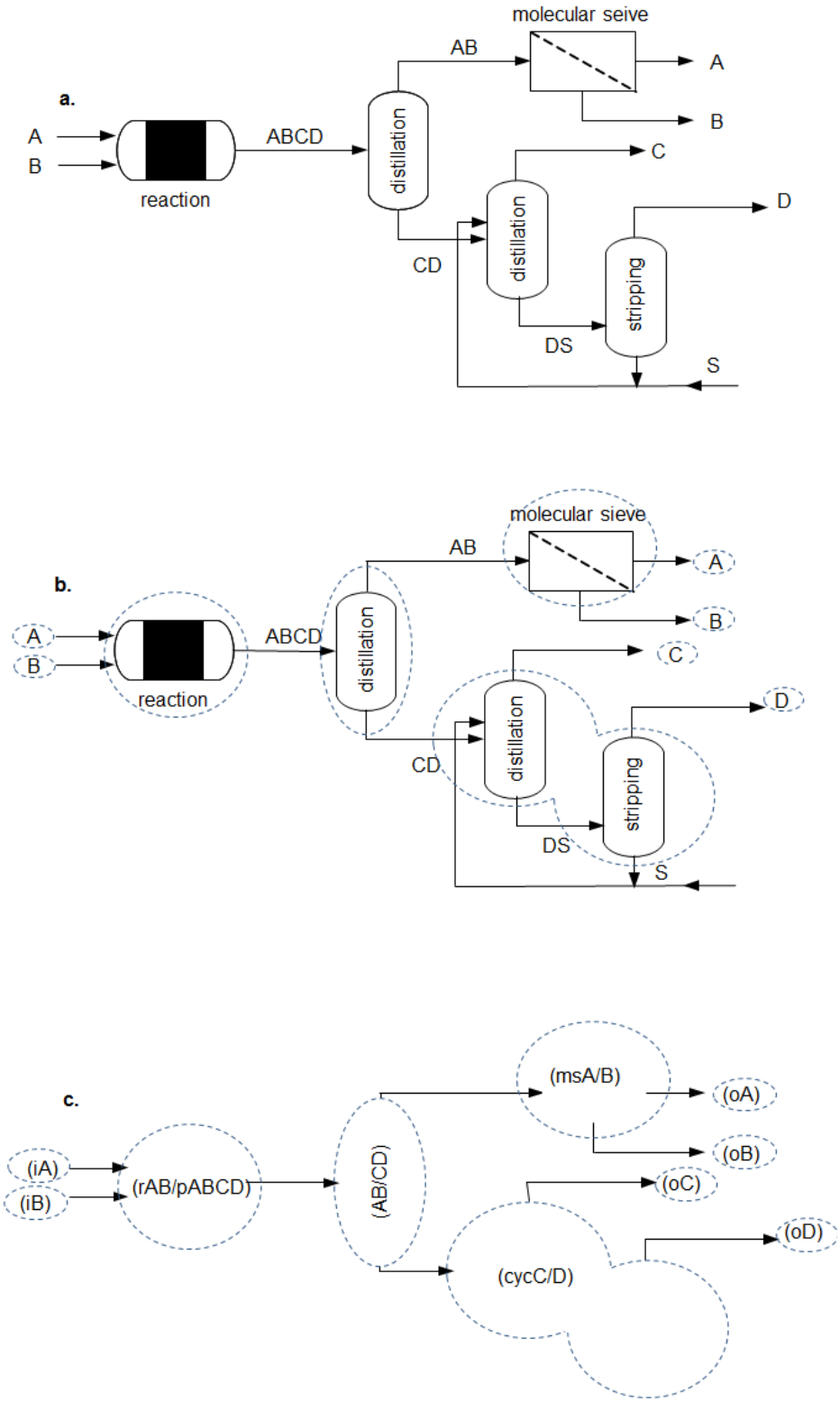


Figure 5.5: Representation of flowsheet (a). with process groups (b, c) process groups

Currently, seventeen process groups are available (Alvarado, 2010). These are listed in Table 5.1.

Table 5.1: Available Process groups (Alvarado, 2010).

Unit Operation	Process-group example
Simple Distillation Column	(AB/C)
Solvent Based Azeotropic Separation	(cycA/B)
Flash Separation	(fAB/CD)
Kinetic Model Based Reactor	(rABC/nE/pABCD)
Fixed Conversion Reactor	(rABC/nE/pABCD)
Pressure Swing Distillation	(swA/B)
Polar Molecular Sieve Based Separation	(pmsABC/D)
Molecular Sieve Based Separation	(msABC/D)
Liquid Membrane Based Separation	(lmemABC/D)
Liquid Adsorption Based Separation	(ladsABC/D)
Gas Membrane Based Separation	(gmemABC/D)
Pervaporation Based Separation	(pervABC/D)
Crystallization Based Separation	(crsABC/D)
Liquid Liquid Extraction Based Separation	(lleABC/S/SC/AB)
Simple Solid Liquid Separation	(slAB/CD)
Absorption	(abEAB/eF/EABF/EF)
Ion Exchange Separation	(ieABCD/ABC)

This kind of representation of the involved streams and unit operations in a process also gives an additional benefit of being able to represent the flowsheets by a simple line entry called SFILES (Simplified Flowsheet Input Line Entry System) (d'Anterroches, 2006). This enables easy representation of the synthesized flowsheets.

Simplified Flowsheet Input Line Entry System

According to d'Anterroches (2006), any flowsheet comprising of the above described PGs can be represented by a unique SFILES. This concept in turn can be traced back for its roots to SMILES (Simplified Molecular Input Line Entry System) (Weininger, 1988). If in a flowsheet, the analogy between flowsheet and a graph; its PGs and graph nodes; connection between PGs and graph edges respectively is established, a unique representation of each flowsheet by a SFILES representation is maintained by canonicalization of a graph using the product of corresponding primes function on the ranks of each node based on the ranks of their invariant set. In Figure 5.6, flowsheets with and without recycle streams are shown to be represented by their respective SFILES. Branches in the flowsheet are represented within square brackets and recycle stream are represented by numbers. Also, owing to the fact that PGs can be connected in only one direction, the orientation of the group when not following left to right direction is clarified by a smaller than symbol (<).

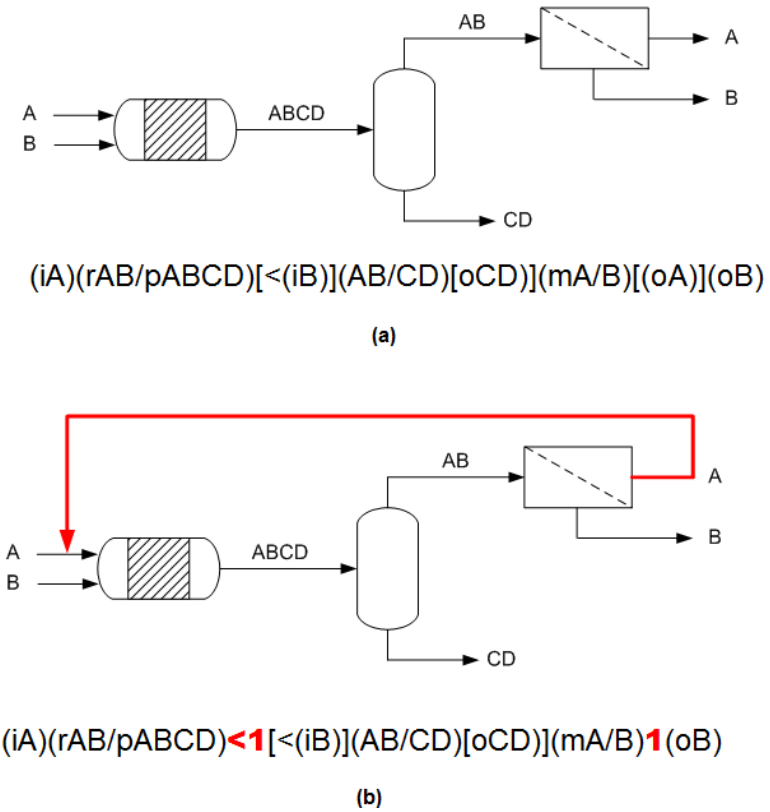


Figure 5.6: SFILES notation of a simple flowsheet (a) without recycle (b) with recycle.

Determining unique SFILES for a flowsheet prevents ambiguity and provides its own share of computational benefits during flowsheet synthesis. For example, when a number of flowsheet alternatives are generated for a problem, checking if an alternative is already found becomes easy with this SFILES representation. Also, by virtue of how these are made to describe a flowsheet uniquely, they can help in efficient transfer/storage of PG information for a flowsheet.

Finally having the flowsheets described by well characterized PGs and represented by SFILES, the property of a flowsheet can be easily identified using a GC based property model. This surfaces from the fact that this representation of flowsheets by PGs is part of utilizing GC techniques to flowsheet synthesis. d'Anterroches and Gani (2005) pointed out that energy consumption of a process flowsheet is one flowsheet property which can be predicted by a GC based property model.

Hence, the separation PGs along with being characterized by type of unit process/operation and driving force should be also characterized by their contributions towards the flowsheet property of interest.

Flowsheet property model

d'Anterroches (2006) shows that a flowsheet property – energy consumption of a flowsheet can be represented by a sum of process group contributions towards this property.

$$f(P) = E_x = \sum_{k=1}^{NG} pos_k \cdot a_k \qquad \textbf{5.4}$$

where, E_x is the flowsheet property – energy consumption; *NG* is the number of process groups; a_k is the regressed contribution of PG, *k* and pos_k is the topology factor.

As stated above, it is clear that any PG is component independent if it is based on the driving force. Hence, a general property model irrespective of its components can be derived. The contributions a_k of the process groups can be regressed by means of fitting experimental data.

5.3 The CAFD framework integrated with CAMD framework.

In any GC based method, as long as the structure can be fully described with the groups, the properties of the structure are immediately available. So, these methods can be used to synthesize new structures easily as the evaluation of the properties of a structure is straightforward given the models and the group contributions. In CAFD, based on a GC method, flowsheets are described by process groups and the goal of generating flowsheet structures matching the target properties within the structural constraints is achieved easily without the need of rigorous models. Also from the information the process groups carry, the minimum set of design parameters to fully describe the flowsheet can be identified.

The CAFD framework, based on the GC-concept is handled by the following steps: a) Problem Definition and Analysis (analyze the synthesis/design problem to define the performance criteria in terms of a set of target properties and to establish the desired target property values; define the initial search space in terms of a superstructure of alternatives); b) Process Synthesis (generate and screen process flowsheet alternatives); c) Process Design (determine the equipment/operation designs for the selected feasible process flowsheets); and d) Process Verification (verify the design through rigorous simulation and/or experiments). The outline of the CAFD framework and necessary information at each stage is shown in Figure 5.7.

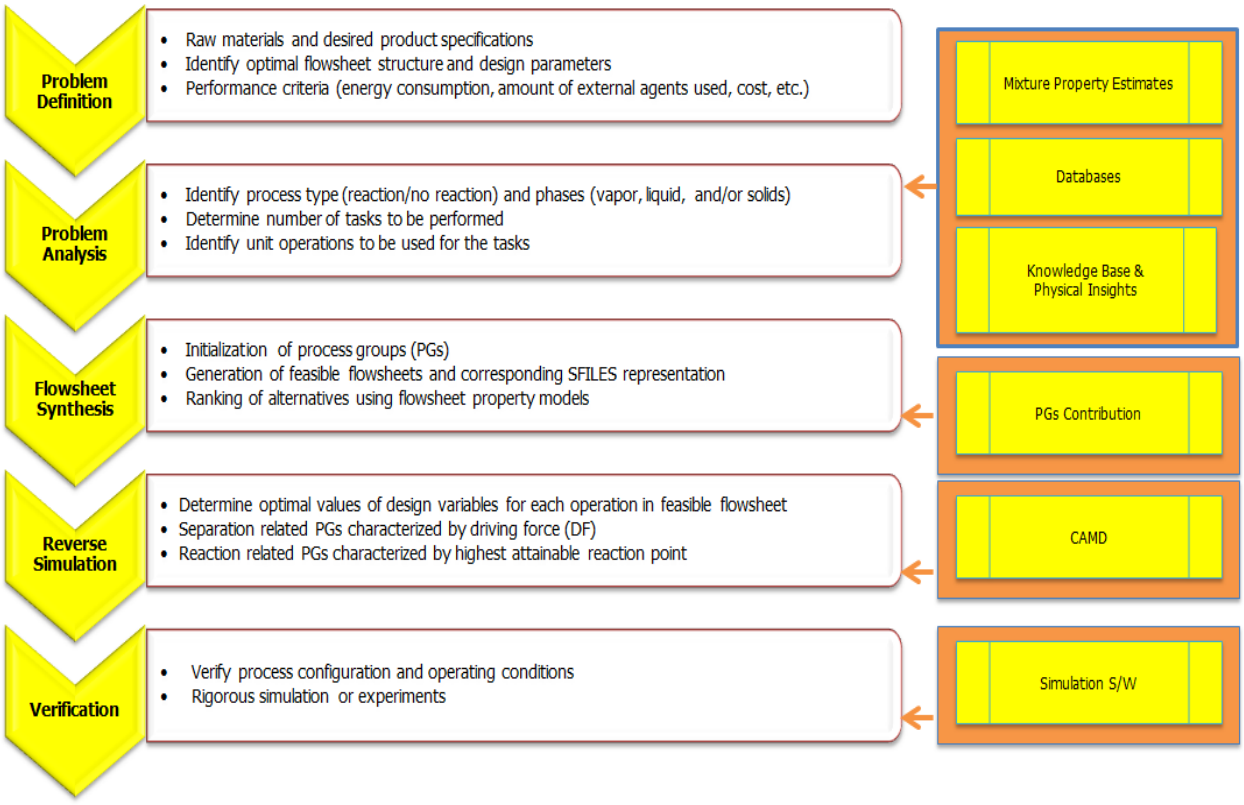


Figure 5.7: CAFD framework.

5.3.1 Problem Definition & Analysis

Given the raw materials and the desired product specifications, the problem is to identify the optimal process flowsheet that transforms the raw materials into the desired products in the most efficient manner. The problem is later analyzed to define the type of process that needs to be synthesized, e.g. processes with or without reaction based on the raw materials and product specifications; to identify the phases (vapor, liquid, and/or solids) that may be involved; to determine the number of operations/tasks that need to be performed; to select the types of unit process/operations that may be involved in the flowsheet alternatives based on rules proposed by Jaksland et al. (1995); to select the performance criteria by which to evaluate the process flowsheet alternatives.

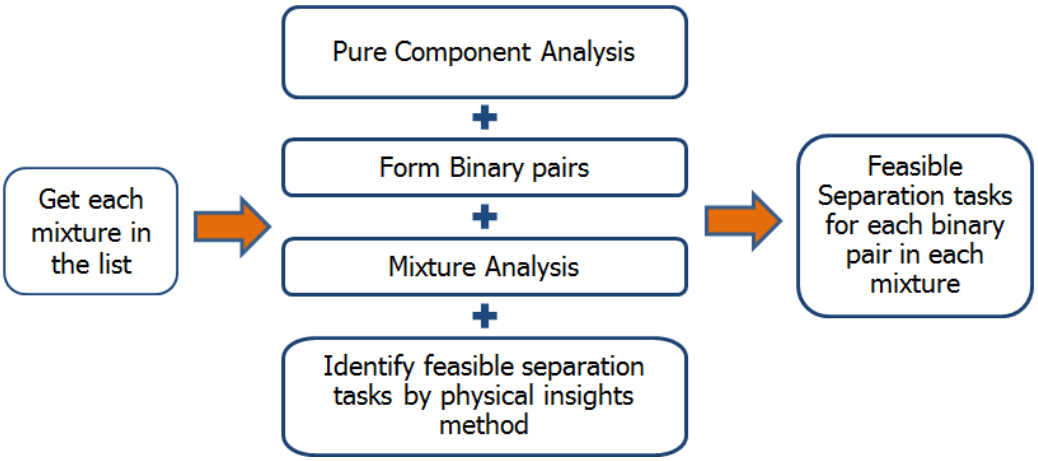


Figure 5.8: Problem analysis steps (Jaksland et al., 1995).

Reaction Analysis

If in the definition of the problem, desired product components are different from raw materials, it is clearly evident that chemical conversion of raw materials to products is needed. In such cases, a database search can be performed to identify the reaction routes making the chemical transformation possible. Depending on the knowledge of reaction rates or conversion rates of the

involved reaction routes, we can initialize the kinetic reactor or fixed conversion reactor PGs, respectively along with the inlet PGs for raw materials. If more than one reaction are involved and have to occur sequentially in the selected route, corresponding reaction PGs are initialized for each of them which may have to be connected by separation PGs or recycle streams. Another interesting evaluation would be to check for the possibility of simultaneous reaction+separation processes. This kind of evaluation about the possibility of simultaneous reaction+separation process can be made by having an insight of the nature of side reactions, involved inerts, as well as Damkohler numbers (Kulprathipanja, 2001) and is beyond the scope of this dissertation. But once the possibility of such groups is established, the reaction+separation PGs can be easily used within the scope of the developed framework. Finally with the reaction process groups in hand, resulting mixtures from each of them can be initialized and the separation schemes for the mixtures depending on the problem definition have to be synthesized.

Pure Component and Mixture Analysis

The nature of each of the mixtures obtained is determined at this stage. This kind of preliminary assessment enables the selection or rejection of separation operations. If for example, two of the components in the mixture tend to form azeotropes, distillation operation may be ruled out for that pair. Here preliminarily, the following information about the mixture is identified at the given reference conditions.

a. State of components (vapor, liquid, and/or solids).
b. Pure component properties.
c. Bulk/dilute components (if compositions of components are known).
d. Polar/non polar components.
e. Azeotropes [hetero/homo] and eutectic points.

f. Temperature sensitive components.

g. Toxic components.

Also, all possible preliminarily information is gathered at this stage. This information helps in deciding on the thermodynamic models for predicting phase equilibrium when necessary. The 22 pure component properties listed in Table 5.2 are identified using available chemical databases or using any property prediction models.

Table 5.2: List of considered pure component properties

	Property	**Secondary/Primary**
MW	Molecular Weight (g/mol)	Primary
ω	Acentric Factor	Primary
T_c	Critical Temperature (K)	Primary
P_c	Critical Pressure (bar)	Primary
Z_c	Critical Compressibility Factor	Primary
V_c	Critical Volume (m^3/kmol)	Primary
T_b	Normal Boiling Point (K)	Secondary
dm	Dipole Moment $\times 1 \cdot 10^{-30}$ (C · m)	Primary
rg	Radius of Gyration (nm)	Primary
T_m	Melting Point (K)	Secondary
T_{tp}	Triple Point Temperature (K)	Primary
P_{tp}	Triple Point Pressure (Pa)	Primary
M_v	Molar Volume (m^3/kmol)	Secondary
H_f	Ideal Gas Heat of Formation (kJ/kmol)	Secondary

G_f	Ideal Gas Gibbs Energy of Formation (kJ/kmol)	Secondary
$S_{I\,G}$	Ideal Gas Absolute Entropy (kJ/(kmol · K))	Secondary
H_{fus}	Heat of Fusion at T_m (kJ/kmol)	Secondary
H_{comb}	Standard Net Heat of Combustion (MJ/kmol)	Secondary
δ	Solubility parameter (kJ/m^3)	Secondary
V_{vw}	Van der Waals Volume (m^3/kmol)	Primary
A_{vw}	Van der Waals Area (m^2/kmol)	Primary
P_{nvap}	Normal Vapor Pressure (Pa)	Secondary

Having this data in hand, mixture analysis can be made by enumerating all possible binary pairs and calculating their respective ratio of all above 22 properties.

If *NC* is the number of components in the mixture, the number of binary pairs, n_{bin} is given by:

$$n_{bin} = \frac{NC(NC-1)}{2} \quad \textbf{5.5}$$

And the number of separation tasks, n_{sep} if sharp split between components is aimed at is given by:

$$n_{sep} = NC - 1 \quad \textbf{5.6}$$

This enables the identification of feasible separation techniques/operations for each binary separation task.

For all $i = 1\ to\ n_{bin}$ binary pairs, the property ratio, r_{ijk} corresponding to j^{th} property and k^{th} separation technique is given by

$$r_{ij} = \frac{P_{j,A}}{P_{j,B}} \quad ; \quad P_{j,A} \geq P_{j,B} \tag{5.7}$$

For the properties marked as secondary in Table 5.2, the binary values may be calculated over a range of pressure or/and temperature. Doing this gives the initial estimates of conditions of operation of a separation technique.

Selection of Candidate Separation operations

The rules proposed by Jaksland et al. (1995) as explained in section 5.2.1. are utilized here. With the limits on the property ratios for a separation operations/techniques to be used for a binary separation task set a-priori from literature, by comparing the above tabulated binary property ratios against these limits, candidate separation tasks can be identified. The limits on property ratios for different separation operations are given in Appendix B.

This can be mathematically explained as follows:

For a binary pair i = 1 to n_{bin} and separation operation/technique, k and and property j, the upper and lower limits of corresponding separation technique's property, r_{ijk} are given by a trapezoidal form of limits (Qian & Lien, 1995) as below:

$$\begin{aligned} Lower\ limits: \left[m_{k,j}, M_{k,j}\right] \\ Upper\ limits: \left[N_{k,j}, n_{k,j}\right] \end{aligned} \tag{5.8}$$

The feasibility of a separation technique is then decided by:

$$\begin{aligned} r_{ijk} \leq m_{k,j} \ &\longrightarrow \mu(r_{ijk}) = 0 \ \& \ k\ is\ infeasible\ for\ i \\ m_{k,j} \leq r_{ijk} \leq M_{k,j} \ &\longrightarrow \mu(r_{ijk}) = \frac{(r_{ijk} - m_{k,j})}{(M_{k,j} - m_{k,j})} \ \& \ k\ is\ feasible\ for\ i \\ M_{k,j} \leq r_{ijk} \leq N_{k,j} \ &\longrightarrow \mu(r_{ijk}) = 1 \ \& \ k\ is\ very\ feasible\ for\ i \end{aligned} \tag{5.9}$$

$$N_{k,j} \leq r_{ijk} \leq n_{k,j} \;\; \longrightarrow \;\; \mu(r_{ijk}) = \frac{(n_{k,j} - r_{ijk})}{(n_{k,j} - N_{k,j})} \;\; \& \; k \; is \; infeasible \; for \; i$$

$$r_{ijk} \geq n_{k,j} \;\; \longrightarrow \;\; \mu(r_{ijk}) = 0 \;\; and \; k \; is \; infeasible \; for \; i$$

Additionally

$$0 \leq \mu(r_{ijk}) \leq 1 \;\; \longrightarrow \;\; A \; warning \; on \; k \; for \; i \; is \; issued \qquad \textbf{5.10}$$

Finally, separation techniques which might be feasible at extreme T, P ranges are considered as infeasible at the preliminary stage. Only when no flowsheet can be synthesized, these separation techniques are reconsidered.

5.3.2 Flowsheet Synthesis

Once the problem has been analyzed, the objective here is to generate and evaluate feasible flowsheet alternatives based on the selected process groups; the minimum number of processing steps and the product specifications. A number of methods and tools are needed here. Flowsheet synthesis was earlier done by d'Anterroches and Gani (2005) and Alvarado (2010) but owing to the high combinatorial nature of the problem, a systematic method for processing of data from problem analysis is needed. In earlier works, all the possible PGs are initialized irrespective of constraints related to connections and are then connected. In such cases, the list would be large and may contain many useless PGs leading to structurally infeasible flowsheets. This tedious task can be avoided by a much more sophisticated way of generating the flowsheets as explained below. Hence, initialization and connecting the PGs is done hand in hand in this work.

Owing to the PG's characteristics, the PGs by themselves are property dependent and component independent while connections between the PGs are component dependent. As PGs are component independent, for any PG of a particular separation technique type, one for each set of components

can be initialized provided its characteristic property ratio between key components falls into feasible limits of the separation technique.

Generation of Superstructure with Initialized Process Groups

Initialization of Process Groups

Here, instead of initializing all PGs for different separation techniques at once, for each product mixture when encountered, the separation PGs of different separation technique type are initialized with respect to the number of chemicals and their identities. This is done through the following three step procedure and each step is illustrated through an imaginary mixture list.

1. Order mixture components wrt. characteristic property for each feasible separation task.

   For example:

Table 5.3: Illustration of component order table.

Components in the mixture	A,B and C
Feasible separation techniques	S_1, S_2, and S_3
Characteristic Property of	
S_1	P_1 : $P_{1A} > P_{1B} > P_{1C}$
S_2	P_2 : $P_{2A} > P_{2C} > P_{2B}$
S_3	P_3 : $P_{3C} > P_{3B} > P_{3A}$
Component order table	S_1: A, B, C S_2: A, C, B S_3: C, B, A

2. Order feasible binary splits based on property ratio (ease of separation)

Table 5.4 : Illustration of binary split order table.

Components in the mixture	A,B and C
Feasible separation techniques	S_1, S_2, and S_3
Possible splits by	
S_1	B/C, A/C : $r_{S_1P_1A/C} > r_{S_1P_1B/C}$
S_2	B/C
S_3	B/C , A/B : $r_{S_3P_3A/B} > r_{S_3P_3B/C}$
Binary Splits order table	S_1: A/C , B/C S_2: B/C S_3: A/B , B/C

3. Initialize the PGs involving all components of the mixture: For each separation technique, only the PGs with the highest property ratios are initialized. This constraint reduces the number of initialized PGs that could form practically feasible flowsheets. If the binary pair with the highest property ratio for any of the separation technique in not of adjacent components from the component order table, the PGs with next highest property ratio are initialized. From Table 5.4, the S_1A/C PG falls into such category.

Table 5.5: Initialized PGs of a ABC mixture.

Components in the mixture	A,B and C
Feasible separation techniques	S_1, S_2, and S_3
Initialized PGs	(S_1AB/C) (S_2AC/B) (S_3CB/A)

Again, if separation barriers affect the separation task further analysis is done to keep the PG in and also the effect of non-key components on the properties enabling separation are checked before initializing the process group. For example, If S_3 is a distillation based PG, and AC forms an azeotropic mixture and only sharp splits are considered, S_3(CB/A) is eliminated from the set of initialized PGs as azeotropic AC becomes a separation barrier

to S_3. If azeotropes or eutectic mixtures exist in the component, they have to be considered as a single entity and ordered accordingly in the component order table before initializing any PG involving them.

The algorithm for initializing PGs for a mixture is shown in Figure 5.9.

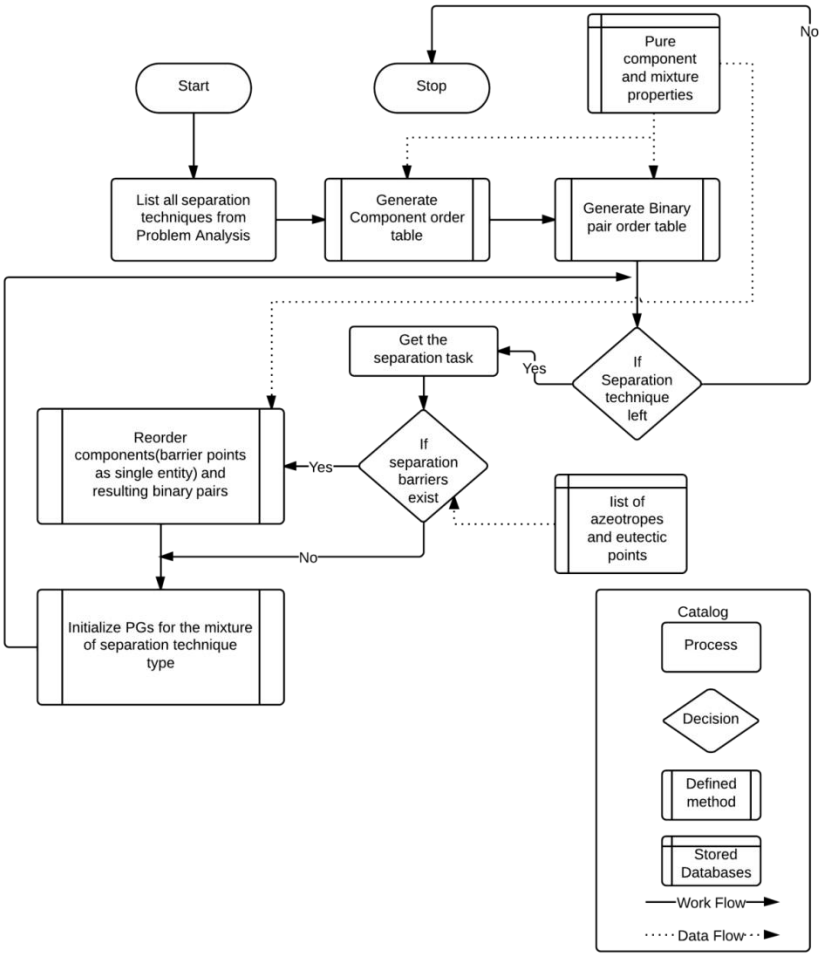


Figure 5.9: Initialization of PGs for an mixture encountered in runtime.

Generation of Superstructure

The superstructure for each of the initial mixtures (defined by the problem or from initialized reaction PGs) is generated using the following methodology. Once the PGs for each of these mixtures are initialized, the consequent mixture lists resulting from these PGs are listed and if they comprise of more than one component, their matching PGs are initialized using the algorithm in Figure 5.9 else an outlet PG is initialized. This is continued until no mixture is left. As each PG carries the information of the number and identity of the components along with the separation task, this information can be utilized to impose structural constraints on the consequent PGs by limiting the number of PGs that could be linked to the parent PG in the super structure. Hence, here basically, the tree lists of PGs resulting from its parent PGs are made. Figure 5.10 gives an example of one such superstructure involving imaginary component lists and unit operations/processes.

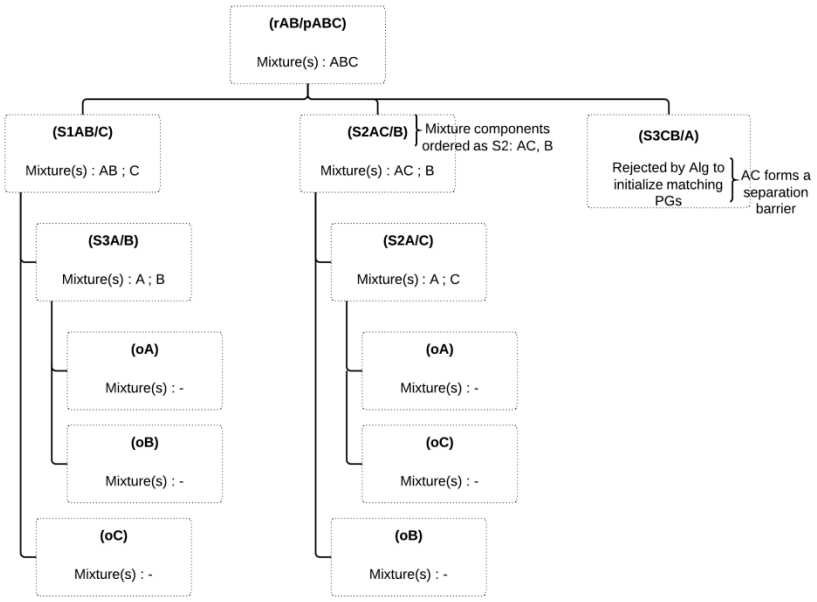


Figure 5.10: Illustration of PGs superstructure.

If for all the components from the initial mixture (A,B,C), outlet PGs exist in the superstructure, the header PG can be termed as a feasible first order group and the superstructure remains in the list for further processing else it can be discarded or the problem can be reanalyzed relieving some constraints on separation techniques. The algorithm for generating a superstructure is shown in Figure 5.11.

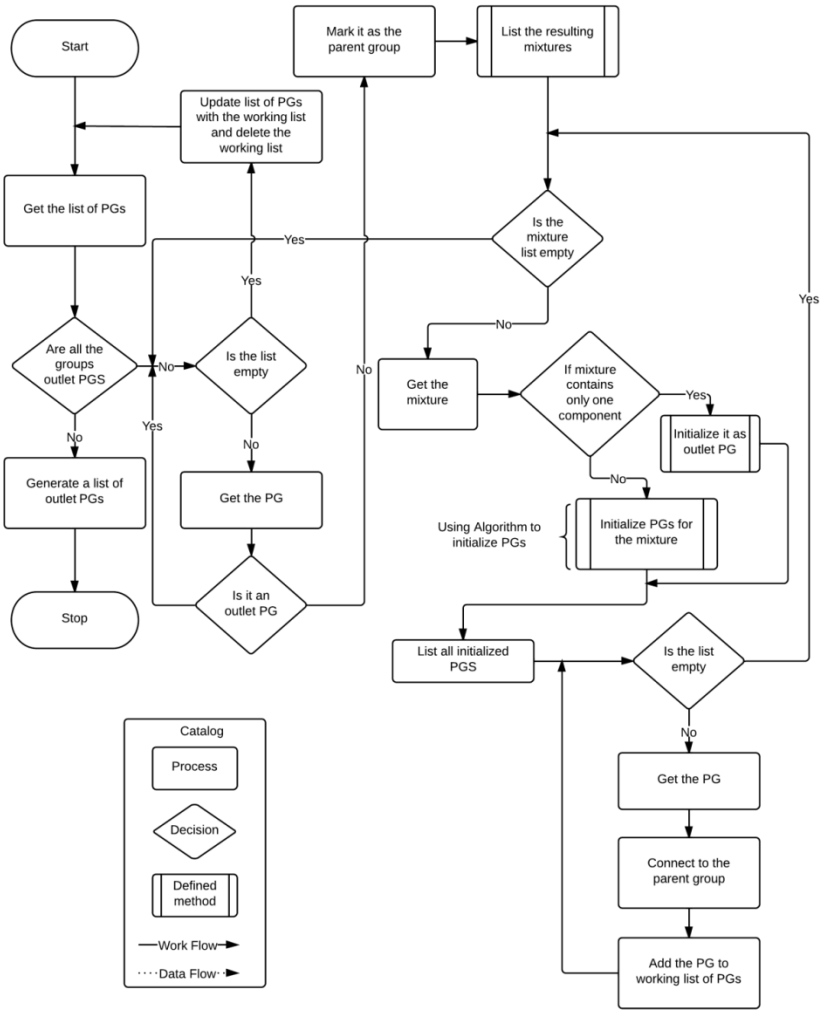
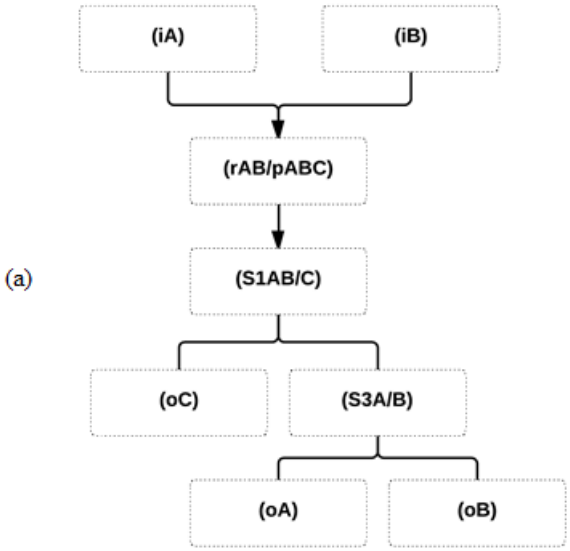


Figure 5.11: Generation of superstructure of PGs.

As each PG carries the information of the number and identity of the components along with the separation task, this information is utilized to impose structural constraints on the consequent PGs. Finally from the list of superstructures combinations, flowsheets leading to targeted product components are identified and represented by their unique SFILES.

SFILES Representation of Flowsheets

All feasible flowsheets generated in this step are represented by their corresponding SFILES notation according to the method developed by d'Anterroches (2006) and stored in terms of this notation. In order to generate the feasible SFILES, all paths from the generated superstructures are populated and combined to form a binary tree structure such that the combination has *NC* components (the number of components in the initial mixture/product of the initial reactor PG of a superstructure), sharp splits and *NC*-1 separation PGs involved. Once these trees are formed, each PG is denoted with an invariant as described in d'Anterroches (2006). The root node is first selected each time and branching decisions are taken based on the invariant of PG, the one with lowest is selected first. The rules to denote a PG with an invariant is shown in appendix. Figure 5.12 gives an example of a tree representation and its SFILES representation. The algorithm to generate SFILES is given by Figure 5.13.



(a)

(b) (iA)(rAB/pABC)[<(iB)](S$_1$AB/C)[(oC)](S$_3$A/B)[(oA)](oB)

Figure 5.12: (a) Tree representation of a combination, (b) SFILES representation of the feasible flowsheet

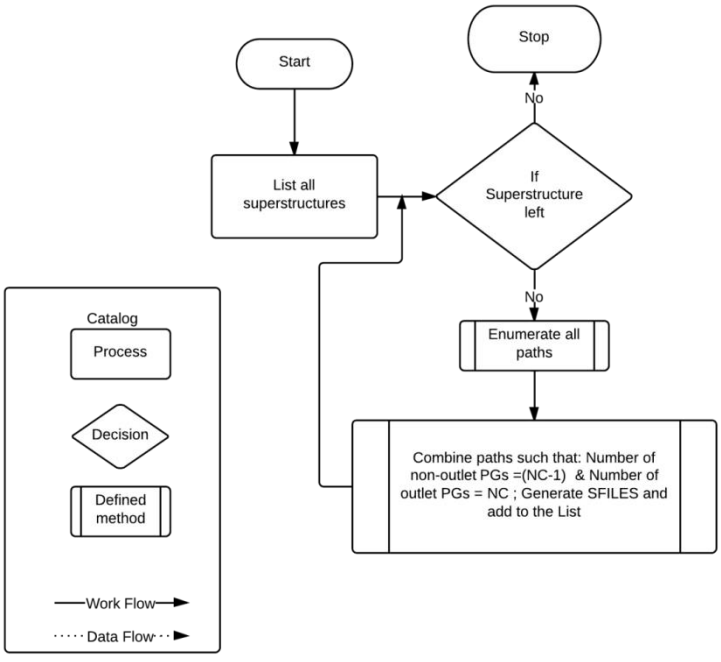


Figure 5.13: Algorithm for SFILES generation.

Ranking of Flowsheet Alternatives

The target properties for each generated feasible flowsheet alternative are calculated using the flowsheet property models combined with a-priori calculated property contributions of each PG. The target property models may be only structure dependent (primary properties), or, they may be dependent on multiple phenomena (effect of energy and mass separating agents). Based on these target properties, the flowsheet alternatives that are structurally feasible and that satisfy the property targets are identified. Currently, there are property models to assess the performance in terms of energy consumption for distillation, flash, solvent based distillation process groups and pressure swing distillation PGs. The flowsheet property model is given by:

$$E = \sum_{k=1}^{NG} \frac{1 + p_k}{d_{ij}^k} \times a_k \qquad \textbf{5.11}$$

where, E is the energy consumption performance of the flowsheet, *NG* is the number of PGs, d_{ij}^k is the maximum driving force of PG, *k*, a_k is the contribution of PG, *k* and p_k is the topology factor given by

$$p_k = \sum_{i=1}^{nt} \bar{D}_\iota \qquad \textbf{5.12}$$

where, *nt* is the number of separation tasks that should be performed before the task, *k* and $\bar{D}_\iota$ is the maximum driving force of task, *i*. The PG contribution are taken from Alvarado (2010); d'Anterroches (2006) and are listed in appendix.

5.3.3 Process Design and integration with Molecular Design

The objective here is to determine the optimal values of the minimum number of design variables corresponding to each process equipment and/or operation in the selected feasible flowsheet. For

counter-current staged separation processes, this includes variables such as number of stages, feed location, product specifications and reflux ratio. For reactors, examples of the design variables are reactor volume, residence time, reactor effluent composition, and temperature. With the reverse approach, separation related PGs such as distillation, extractive distillation, and flash drums are characterized in terms of their driving force (Bek-Pedersen, 2003) while reaction related PGs are characterized in terms of their highest attainable reaction point (Horn, 1964). The design method back-calculates the design variables from the highest driving force or highest attainable reaction point. The detailed methodology for each of this can be understood through methodologies shown in appendix as well as part of the case studies.

Some of the unit operations may be in need of an external agent and as stated, the performance of the flowsheet varies depending on the external agent. Process and product design are closely related and failing to correctly achieve the simultaneous design can lead to unexpected results like increase in cost of production etc. Having process models that can rapidly evaluate the impact on a process of a change of an involved component, based on the process type a CAMD problem is set up to identify the feasible external agent corresponding to the molecular design targets translated from the targets set by the process design problem. Also as the PGs carry all information about the properties of its inlet and outlet components, the process design methodology given in section 4.3 can be utilized to identify property targets on the external agent (like a solvent) depending on the other streams in the process. Finally, the best option among the solutions to CAMD problem can be chosen based on its effect on the process.

CAMD

To design the external agents, the CAMD methodology developed by Bommareddy et al. (2010a) is used. Depending on the type of unit operation the process targets are translated into targets for

the CAMD problem. In the developed CAMD methodology which is given in chapter 3, the molecular structures that match the targets are identified by a reverse property prediction problem. The molecules designed are further screened based on the mixture property targets. With each feasible molecule found, different flowsheets with corresponding properties are identified and ranked.

5.3.4 Final Verification

The objective here is to verify the synthesis/design results from the previous steps as well as to determining the remaining unknown variables of the process. This is done through rigorous simulation using design variables identified by reverse simulation provided the necessary models are available. As the design corresponds to maximum driving force and/or highest attainable region, the candidate design should correspond to the minimum energy requirement design. If necessary, the design can be verified experimentally with the established optimal operating conditions and equipment parameters.

5.3.5 Software Implementation

The CAFD and CAMD methods involve screening a lot of data and it is practically impossible to apply them without an efficient software implementation. A beta version of the tool for the CAFD has been developed based on the framework using VC++ - Microsoft Foundation Class Libraries. The data flow of the CAFD framework along with the solution steps is given by Figure 5.14.

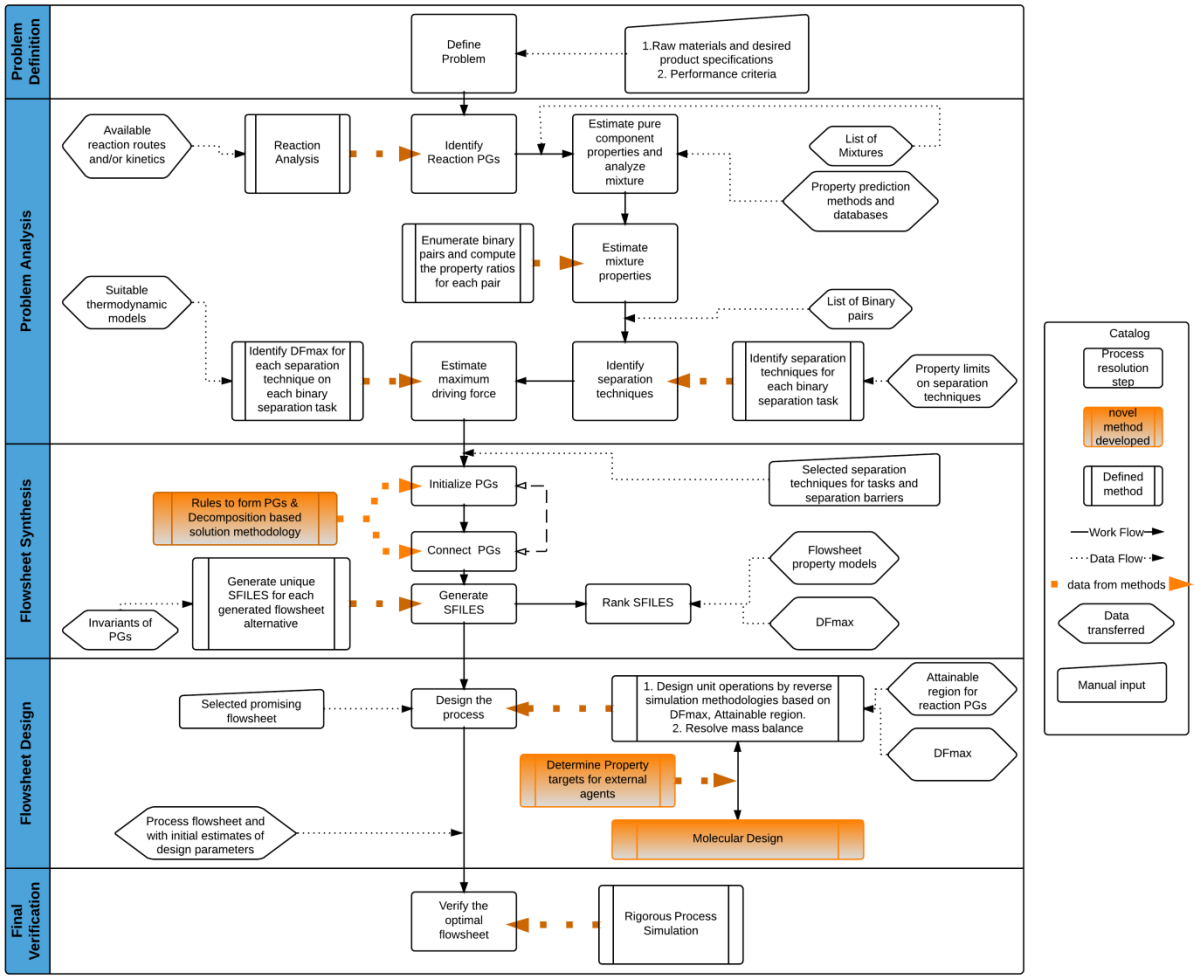


Figure 5.14: Data flow in the CAFD framework.

The beta version of the tool considers the problem analysis being done and given the unit operations that could perform the binary separation tasks, the tools works to initialize the PGs and generate the list of SFILES with their respective property. Once the tool generates the list the optimal flowsheet could be selected, designed and verified. The steps as seen by the user of the tool are shown in the Figure 5.15.

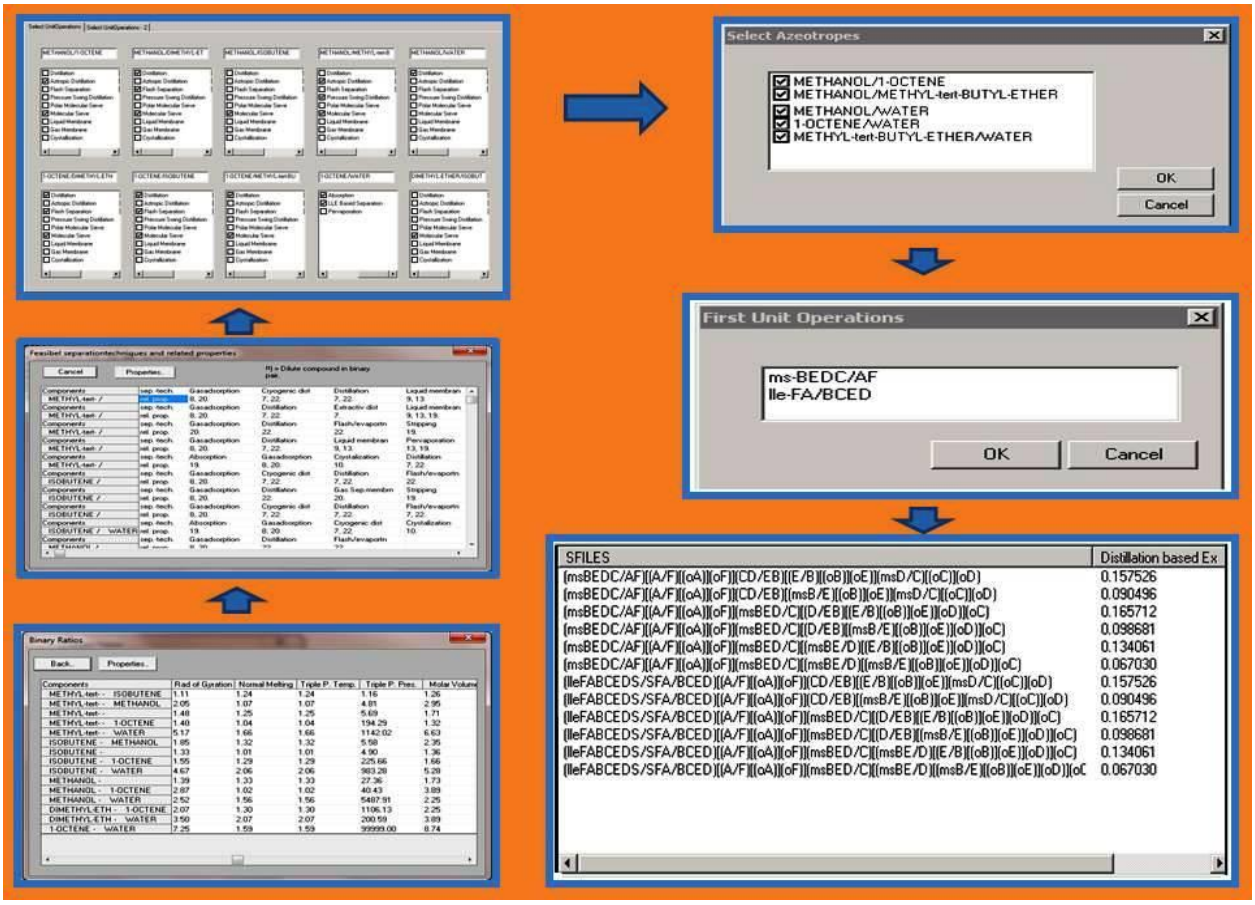


Figure 5.15: View of the developed CAFD tool

5.4 Summary

This chapter presented a novel systematic framework for CAFD integrated with CAMD. The CAFD methodology is based on GC concepts, and when integrated using the reverse problem formulation technique with CAMD leads to very efficient simultaneous process and product synthesis/design, while keeping the high level of accuracy associated with the group contribution methods. The architecture of the developed prototype software has been presented and its utilization is shown by developing the following interesting application examples.

5.5 Case Studies – Computer Aided Flowsheet Design

5.5.1 Case Study – Production of Isobutene

Problem statement

Isobutene is a potential starting material for production of butyl rubber, poly isobutylene etc. In this case study, the production of isobutene by decomposition of Methyl tert Butyl Ether at about 493K is investigated. The objective is to identify the optimal process configuration by minimizing the energy requirements. The potential reaction pathways to form Isobutene are identified as shown in Equations 5.13, 5.14 and 5.15.

$$\text{Methyl tert Butyl Ether} \rightarrow \text{Methanol} + \text{Isobutene} \quad \textbf{5.13}$$

$$\text{Methanol} \rightarrow \text{Dimethyl Ether} + \text{Water} \quad \textbf{5.14}$$

$$\text{Isobutene} \rightarrow \text{Isobutene Dimer} \quad \textbf{5.15}$$

For the sake of initializing PGs and easy representation of components in a mixture, the components involved in the above reactions are denoted by

A	Methyl tert Butyl Ether
B	Methanol
C	Isobutene
D	Dimethyl Ether
E	Water
F	Isobutene Dimer

Since the conversion is incomplete, the product (Isobutene) needs to be recovered and then purified from the reactor effluent while the reactants recycled to the reactor. The reactor affluent stream consists of all A-F components. Each of the components is to be recovered at > 99% purity.

Problem Analysis

Since the reaction route is given, the reactor PG - (rA/pABCDEF) is initialized. The total number of components in the resulting mixture from reaction is six. Hence minimum number of separation processing steps is five. Pure component properties and mixture properties are analyzed at reference conditions of 298K and 1 atm. The mixture properties are in terms of the ratios of properties of all the possible fifteen binary pairs among the components in the mixture. The initial mixture analysis identifies the azeotropes among the components listed in Table 5.6.

Table 5.6: Azeotropes at 1 atm pressure.

Azeotrope	T(K) & x(mole %)	Azeotrope type
B/C	266.2K – 0.2% B	Min Boiling
B/F	336K – 86.9% B	Min Boiling
F/E	346.1K – 67.3% E	Min Boiling
A/E	325.6K – 96.5% A	Min Boiling
B/A	325.54K – 68.2% A	Pressure sensitive

The feasible separation techniques identified using rules from Jaksland (1996) are shown in Table 5.7.

Table 5.7: Separation tasks and potential techniques for Isobutene production.

Separation Technique	Property Ratio Threshold Values	Separation Tasks
Azeotropic Distillation	Azeotrope	B/F, B/A, F/E, A/E
Liquid-Liquid Extraction	Azeotrope	B/F, B/A, F/E, A/E
Molecular Sieve	Kinetic Diameter > 1.05 van der Walls Volume > 1.07 Polarisability > 1.0 Dipolemoment>1.05	B/F, B/D, B/C, B/A, F/D, D/C, D/E
Pressure Swing Distillation	Pressure sensitive azeotrope	B/A
Distillation	Boiling point > 1.01 Vapor Pressure>1.05	B/D, B/E, F/D, F/C, F/A, D/A, D/E,C/E
Flash	Boiling point > 1.23 Vapor Pressure>10	B/D, F/D, F/C, D/A, D/E, C/E

Flowsheet Synthesis

At this stage the developed software is initiated with all the information and all the SFILES with their respective flowsheet properties (based on only distillation columns at this stage) are obtained.

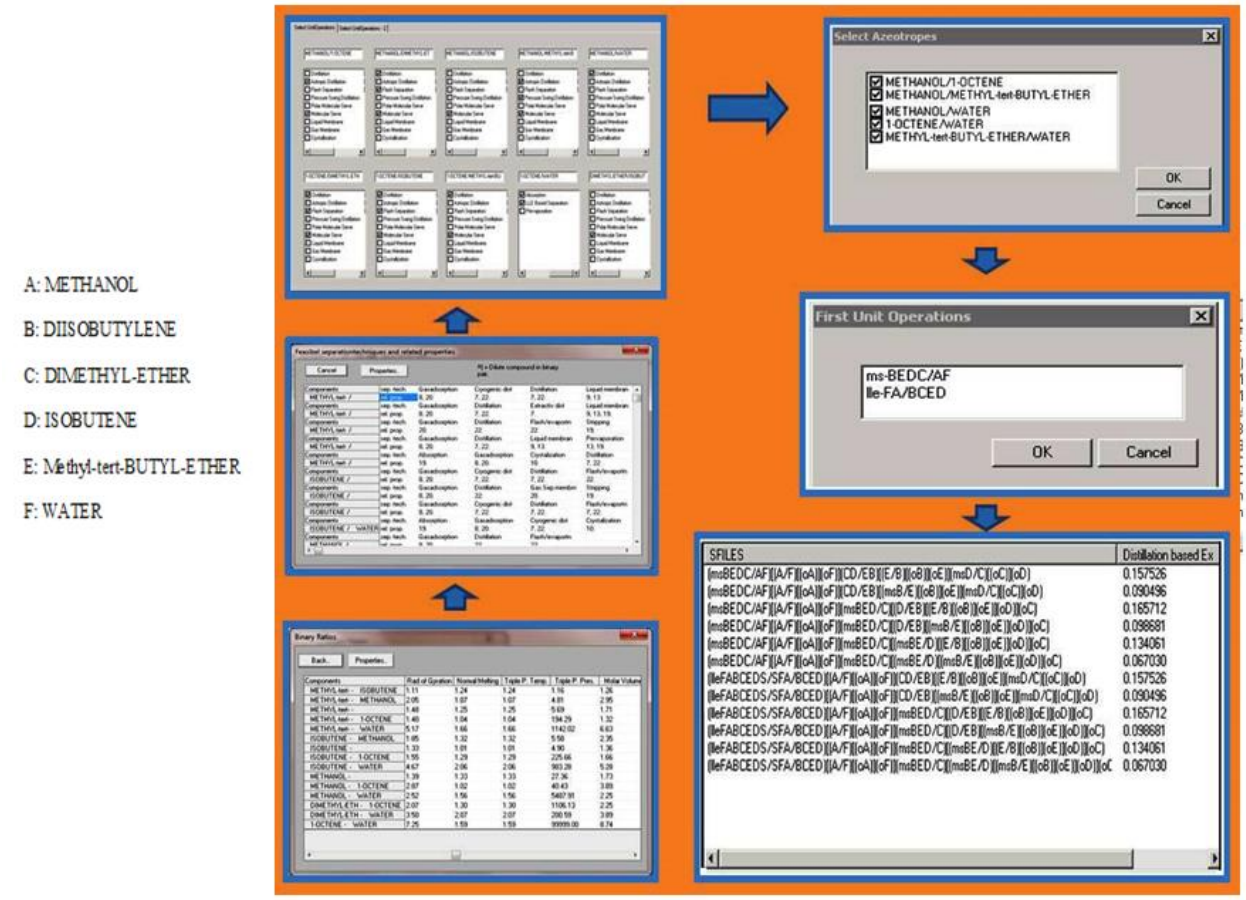


Figure 5.16: Generation of SFILES using the CAFD tool.

This information hence was provided to the developed CAFD tool and 12 feasible flowsheets were identified as shown in Figure 5.17 from the candidate process groups represented by the corresponding SFILES notation.

SFILES	Distillation based Ex
(msBEDC/AF[(A/F[(oA)](oF)]CD/EB[(E/B[(oB)](oE)]msD/C[(oC)](oD)	0.157526
(msBEDC/AF[(A/F[(oA)](oF)]CD/EB[(msB/E[(oB)](oE)]msD/C[(oC)](oD)	0.090496
(msBEDC/AF[(A/F[(oA)](oF)]msBED/C[(D/EB[(E/B[(oB)](oE)](oD)](oC)	0.165712
(msBEDC/AF[(A/F[(oA)](oF)]msBED/C[(D/EB[(msB/E[(oB)](oE)](oD)](oC)	0.098681
(msBEDC/AF[(A/F[(oA)](oF)]msBED/C[(msBE/D[(E/B[(oB)](oE)](oD)](oC)	0.134061
(msBEDC/AF[(A/F[(oA)](oF)]msBED/C[(msBE/D[(msB/E[(oB)](oE)](oD)](oC)	0.067030
(lleFABCEDS/SFA/BCED[(A/F[(oA)](oF)]CD/EB[(E/B[(oB)](oE)]msD/C[(oC)](oD)	0.157526
(lleFABCEDS/SFA/BCED[(A/F[(oA)](oF)]CD/EB[(msB/E[(oB)](oE)]msD/C[(oC)](oD)	0.090496
(lleFABCEDS/SFA/BCED[(A/F[(oA)](oF)]msBED/C[(D/EB[(E/B[(oB)](oE)](oD)](oC)	0.165712
(lleFABCEDS/SFA/BCED[(A/F[(oA)](oF)]msBED/C[(D/EB[(msB/E[(oB)](oE)](oD)](oC)	0.098681
(lleFABCEDS/SFA/BCED[(A/F[(oA)](oF)]msBED/C[(msBE/D[(E/B[(oB)](oE)](oD)](oC)	0.134061
(lleFABCEDS/SFA/BCED[(A/F[(oA)](oF)]msBED/C[(msBE/D[(msB/E[(oB)](oE)](oD)](oC)	0.067030

Figure 5.17: SFILES identified by the CAFD tool for Isobutene production problem.

The energy index flowsheet property was calculated for all candidate configurations and the SFILES string with the energy index (0.1575) is selected for design. It is combined with the inlet PGs and reaction PGs. The obtained configuration consists of a reactor and five separation units: molecular sieve, extraction and three distillation columns.

(rA/pABCDEF)(lleEBFDAC/S/SEB/FDAC)[(B/SE)(oSE)[(oB)]][(DC/AF)[(A/F)[(oF)](oA)](msC/D)[(oD)](oC)

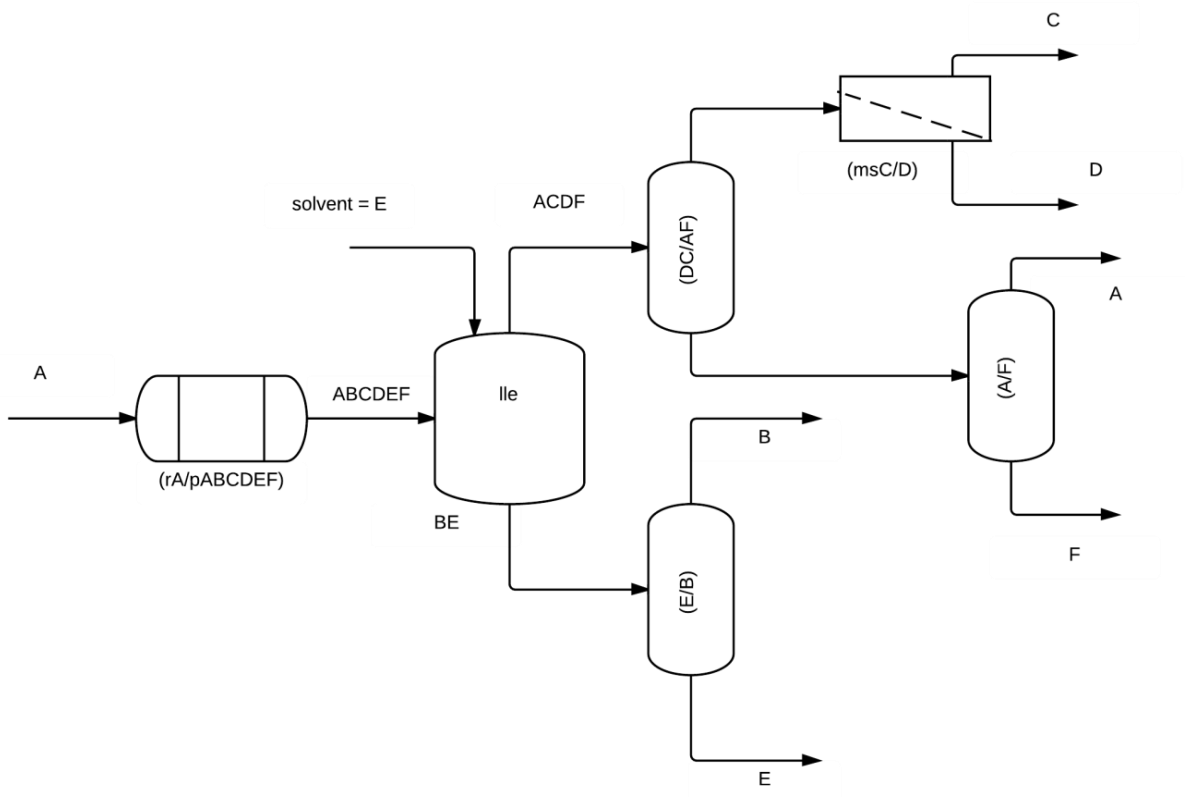


Figure 5.18: Selected optimal flowsheet.

Before invoking the molecular design problem to identify the solvent for LLE PG, it is evident that methanol is highly soluble in water (one of the components in the mixture) while other

components in the mixture are not soluble in water. Hence, this information confirms water as the candidate solvent for the extraction unit. But in many cases it's not simple to find a solvent like water, hence a thorough emphasis has to be put on setting the targets for the molecular design problem and the solvents may be found using any CAMD tool (like the one by Bommareddy et al. (2010a).

The contributions towards the flowsheet property model are currently available for only distillation process groups. The flowsheets generated using the software have different number of distillation type PGs and hence the flowsheet having less number of distillation PGs in them might have the least property value. The contribution of PGs other than distillation type may be high or low. Therefore in the current case study an old heuristic: "prefer distillation columns first" is used to select the flowsheet shown in Figure 5.18. Once the contributions from PGs of other than distillation type are found as cited in the future work shown in this dissertation, the optimal flowsheet could be selected without the help of heuristics. Also this case study can be compared with the flowsheet for separation of the mixture from isobutylene production (Yamase & Suzuki, 2005) – as shown in Figure 5.19. Sharp splits have been considered at each stage. The unit operations used and their order looks to be similar to the published one.

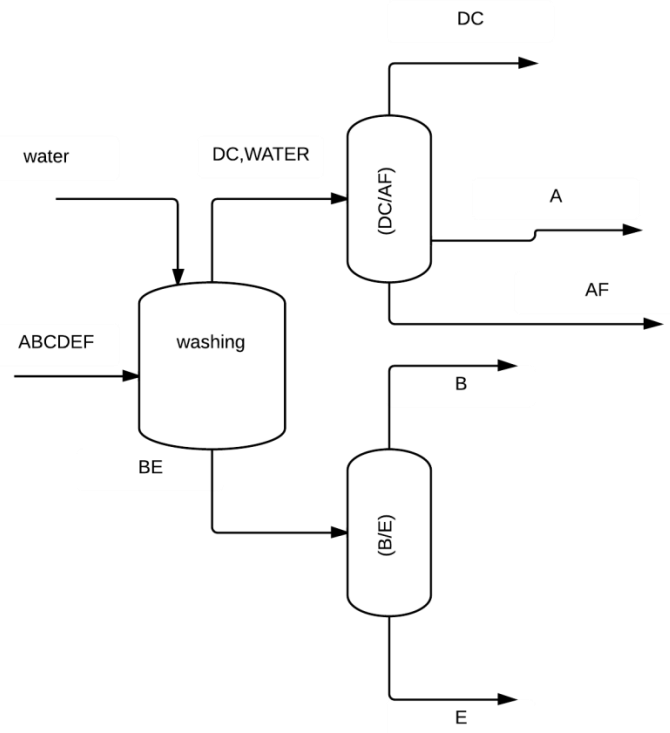


Figure 5.19: Flowsheet from literature (Yamase & Suzuki, 2005).

Flowsheet Design

The LLE PG does not involve energy consumption; the maximum solvent free driving force is used as the performance criterion. Knowing the solvent free driving force of the unit, the amount of solvent can be estimated. The higher the maximum driving force, the less solvent is required. Here for this process group the LLE data between water, methanol, and Isobutene dimer are to be available for further calculations. This information could not be readily retrieved from literature. If more than two solvents are feasible, their respective maximum driving force corresponding to given solvent amount or solvent amount corresponding to selected driving force can be used in a comparative sense. The solvent with larger maximum driving force value or the one that when used in lesser quantity does the job is the optimal solvent. The reverse simulation of the distillation

columns using the driving force approach (Gani & Bek-Pedersen, 2000) yielded a design operating at the maximum driving force. The design parameters are shown in Table 5.8. The VLE data for the key components in each distillation column is taken from literature and the maximum driving force (difference between vapor phase composition and liquid phase composition) is noted.

Table 5.8: Design parameters of distillation columns

Parameter	(DC/AF)	(E/B)	(A/F)
Max driving force (from VLE data):	0.5078	0.3916	0.4919
Light key composition at max driving force, D_x (from VLE data):	0.22	0.2	0.2
Recovery of light key (from definition of PG):	>0.995	>0.995	>0.995
Ideal no. of stages (from (Gani & Bek-Pedersen, 2000) or knowledge) :	20	15	20
Feed location (= Ideal No of stages *(1- D_x)):	15	12	16
Min Reflux Ratio (from (Gani & Bek-Pedersen, 2000)) :	0.4105	1.095	0.515

These conditions can be used as initial conditions for performing the rigorous simulation to validate the obtained optimal flowsheet further.

5.5.2 Case Study – Production of Diethyl Succinate

Succinic acid is a potential co-product from bioethanol manufacture, which can be further reacted with ethanol to produce diethyl succinate (DES), a useful solvent for cleaning metal surfaces and paint stripping. In this case study, the production of diethyl succinate from ethanol and succinic acid is investigated. The objective is to identify the optimal process configuration by minimizing

the energy requirements. The potential reaction pathways to form diethyl succinate are identified as shown in Equations 5.16 and 5.17

$$\text{Succinic Acid} + \text{Ethanol} \rightarrow \text{Monoethyl Succinate} + \text{Water} \quad \textbf{5.16}$$

$$\text{Monoethyl Succinate} + \text{Ethanol} \rightarrow \text{Diethyl Succinate} + \text{Water} \quad \textbf{5.17}$$

For the sake of initializing PGs and easy representation of components in a mixture, the components involved in the above reactions are denoted by

A	Ethanol
B	Water
C	Diethyl Succinate
D	Monoethyl Succinate
E	Succinic Acid

Since the conversion of succinic acid and ethanol is incomplete, the product (diethyl succinate) needs to be recovered and then purified from the reactor effluent while the reactants recycled to the reactor. Each of components is to be recovered at > 99% purity.

Problem Analysis

Since the reaction route is given, the reactor PG - (rAE/pABCDE) is initialized. From literature Kolah, Asthana, Vu, Lira, and Miller (2008), a pervaporation assisted reactor is also considered as one of the options for this system. Hence the reaction+separation PG – (rpervAE/pB/ACDE) is also initialized. The total number of components in the resulting mixtures from the reaction PGs are 5 and 4, respectively. Hence the minimum number of separation processing steps is four and three, respectively. Pure component properties and mixture properties are analyzed at reference conditions of 298K and 1 atm. The mixture properties are in terms of the ratios of properties of all the possible ten binary pairs among the components in the mixture.

The initial problem analysis identifies the feasible separation techniques shown in Table 5.9.

Table 5.9: Separation tasks and potential techniques for DES production

Separation Technique	Property Ratio Threshold Values	Separation Tasks
Crystallization	Melting Point > 1.27	A/C, C/E, D/E
Liquid Membrane	Radius of Gyration > 1.03 Molar Volume > 1.08 Solubility Parameter > 1.28	A/B, A/D, A/E, B/C, B/D, C/D, C/E, D/E
Pervaporation	Molar Volume > 1.08	A/B, A/D, A/E, B/C, B/D, C/D, C/E, D/E
Distillation	Boiling point > 1.01 Vapor Pressure>1.05	A/C, B/C
Flash	Boiling point > 1.23 Vapor Pressure>10	A/C, B/C

The initialized PGs while generating the flowsheets is listed in Table 5.10.

Flowsheet Synthesis

Table 5.10: Initialized process groups for DES production.

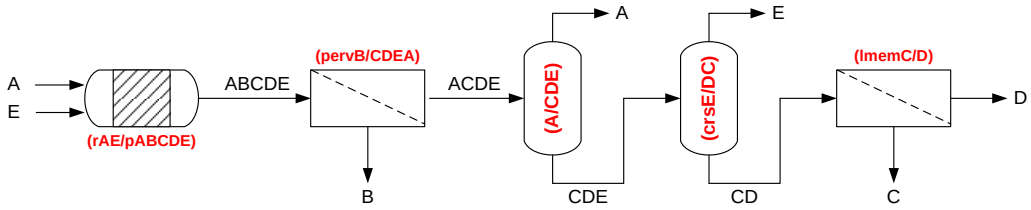
Unit Operation	Process Group
Reactor	rAE/pABCDE, rpervAE/pB/ACDE
Crystallization	crsE/DBCA, crsE/DCA, crsDBC/A, crsE/DC, crsDC/A

Liquid Membrane	lmemCDEA/B, lmemCDE/A, lmemCDA/B, lmemCD/E, lmemCD/A, lmemCD/B, lmemC/D, lmemA/B
Pervaporation	pervCDEA/B, pervCDE/A, pervCDA/B, pervCD/E, pervCD/A, pervCD/B, pervC/D, pervA/B
Distillation	AB/CDE, AB/CD, A/CDE, A/CD, B/CD
Flash	fAB/CDE, fAB/CD, fA/CDE,fA/CD, fB/CD

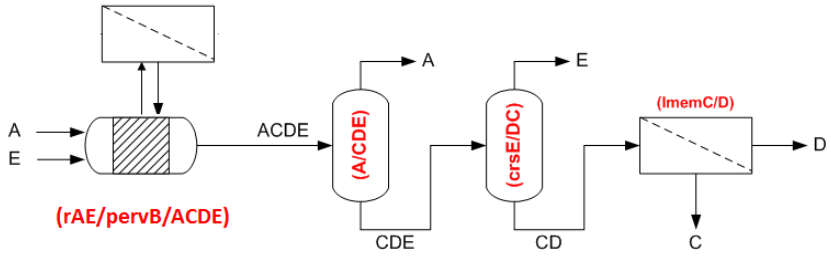
The mixture analysis also reveals the existence of two binary azeotropes (water/diethyl succinate and water/ethanol). Therefore, azeotropic distillation, extractive distillation, and liquid-liquid extraction might also be potential separation techniques to be considered in the synthesis problem. However, pervaporation and the liquid membrane selectively remove water from the mixture thus alleviating the need for azeotropic separation. To enable the user taking such decisions the developed tool interacts with the user to select the separation techniques for the tasks. A pervaporation assisted reactor is found to be an efficient configuration for esterification reactions. The PGs highlighted in blue in Table 5.10 are the potential first separation operations for the reactor mixture, ABCDE. Similarly they could be identified for mixture ACDE. A total of 176 feasible flowsheets were identified from the candidate process groups and represented by the corresponding SFILES notation. The energy index flowsheet property was calculated for all candidate configurations and the two SFILES strings with the lowest value of the energy index (0.051) are shown below. The first configuration consists of a reactor and four separation units: pervaporation, distillation, crystallization and liquid membrane. The second configuration

involves a pervaporation reactor and three separations: distillation, crystallization and liquid membrane.

1. (iAE)(rAE/ABCDE)(pervCDEA/B)[(A/CDE)[(crsE/DC)[(oE)](lmemC/D)[(oC)] (oD)](oA)](oB)



2. (iAE)(rAE/pervB/ACDE)[(A/CDE)[(crsE/DC)[(oE)](lmemC/D)[(oC)] (oD)](oA)](oB)



Flowsheet Design

It is assumed that the membranes exhibit very high selectivity thus leading to a near perfect separation and recovery. The reverse simulation of the distillation column using the driving force approach yielded a design operating at a maximum driving force of 0.85 corresponding to a column with 15 stages (feed location 13.5) and a reflux ratio of 0.552 (minimum 0.368). For the two feasible flowsheets selected for final verification, one has already been shown to be the same as found by Alvarado (2010) while the other with the pervaporation assisted reactor is found by the tool showing that new process groups generated could fit into the framework. Reaction kinetics from Kolah et al. (2008) for macro porous Amberlyst-15 ion-exchange resin are used to design the reactor PG - (rAE/pABCDE).

$$\text{Succinic Acid} + \text{Ethanol} \underset{k_1/K_{eqx,1}}{\overset{k_1}{\rightleftharpoons}} \text{Monoethyl Succinate} + \text{Water} \quad \textbf{5.18}$$

$$\text{Monoethyl Succinate} + \text{Ethanol} \underset{k_2/K_{eqx,2}}{\overset{k_2}{\rightleftharpoons}} \text{Diethyl Succinate} + \text{Water} \quad \textbf{5.19}$$

Parameter	Units	value
$k_1^{\,0}$	$kg_{soln}/kg_{cat}S$	$5.3*10^7$
$k_2^{\,0}$	$kg_{soln}/kg_{cat}S$	$8.0*10^7$
$E_{A,1}$	kJ/kmol	66000
$E_{A,2}$	kJ/kmol	70000
$K_{eqx,1}$		5.3
$K_{eqx,2}$		1.2

Using 85% conversion, attainable region analysis gives the optimal configuration as CSTR with bypass (Alvarado, 2010). For a given residence time, reactor volume and outlet concentrations can easily be calculated as shown in appendix. These conditions can be used as initial conditions for performing the rigorous simulation to validate the obtained optimal flowsheet further.

6. Conclusions

6.1 Achievements

In this dissertation, the main achievement is the development of a systematic methodology that enables solution of process and product synthesis/design problems. The dissertation clearly articulates the need for solving the process and product design problems in an integrated fashion. The methodologies for three different problems viz, process synthesis, process design, product design and their integration by linking the targets for each of the respective problems have been shown via Chapters 4, 5 and 6. All the decoupled problems are clearly defined and their interlinked targets irrespective of their respective solutions are carefully set and methods to identify an optimal solution with respect to overall targeted performance are developed.

In this work, the main achievement on the product design side has been the development of a molecular design framework for its effective integration with process synthesis and design. Algorithms to identify the molecules that meet the process targets based on group contribution approaches have been developed in the past, but, the CAMD framework developed here differs from the prior works by considering the efficient inclusion of the property contribution from higher-order groups at an early stage of molecular design. The earlier methods either did not efficiently incorporate higher groups within the algorithm or did not incorporate the contribution of higher order groups during the initial stage thus increasing the number of group subsets which need to be checked in order to assess the molecules' structural stability. Incorporating higher order

groups at a later stage in the algorithm may also lead to a situation where some potential group subsets are omitted without being considered in further stages of the algorithm. The accuracy of property prediction is enhanced by using these improved techniques to enumerate higher order groups. Incorporation of these higher order enumeration techniques increases the efficiency of property prediction and thus the range of applicability of group contribution methods to molecular design problems. The developed methodology also enables the identification of structural isomers as it puts a check on the possibility of nonexistence of each higher order group in each group subset. An algebraic approach is an efficient extension to earlier visual (geometric) methods, particularly for large problems that normally are prone to combinatorial explosion. Unlike the geometric approach, the developed algebraic approach automatically generates a complete solution set which will ensure that no potential solution is missing.

Traditionally process design and molecular design problems have been treated as two separate problems, with little or no feedback between them. But, solving process design and molecular design problems individually limits the solution space. The properties of fresh material to a process depend on the existing recycle streams within the process and solving process design problems alone would require committing to specific raw materials well in advance in order to lead to a solution. Hence, when process and product design problems are solved together each benefits from other in the method of designing molecules that meet process performance. The property clustering techniques are used to track properties in both process and molecular design problems. The property integration framework has allowed for simultaneous representation of processes and products from a property perspective and hence established a link between molecular and process design. The simultaneous approach involves solving two reverse problems. The first reverse problem identifies the input molecules' property targets corresponding to the desired process

performance. The second reverse problem is the reverse of a property prediction problem, which identifies the molecular structures that match the targets identified in the first problem. The developed CAMD framework helps in solving this second reverse problem.

The main target on the process synthesis side of this work was to develop a systematic process synthesis and design framework to be integrated with molecular design. A systematic framework for computer aided flowsheet design and its completely automated tool are developed. Overall, analogous to the CAMD framework based on group contribution methods, a systematic group contribution based CAFD framework is developed for synthesis of process flowsheets from a given set of input and output specifications. Feasible flowsheet configurations are generated using efficient generate and test algorithms and the performance of each candidate flowsheet is evaluated using a set of flowsheet properties. A systematic notation system called SFILES is used to store the structural information of each flowsheet. The design variables for the selected flowsheet(s) are identified through a reverse simulation approach and are used as initial estimates for rigorous simulation to verify the feasibility and performance of the design

In the developed methodology for CAFD, it can been seen that, having initialized PGs from their base PGs (defined in section 5.2.2) after analyzing the problem and connecting them based on the connectivity rules to synthesize flowsheet structures with required process performance before committing to any design aspects makes the methodology fall into the reverse problem formulation paradigm. In contrast to a forward problem where the flowsheets are synthesized on a trial and error based methodology here in the reverse property prediction problem having the targets set, we could use the PGs and their systematic processing methodology as described above to find the optimal solution without trial and error. This was possible as the initialized PGs are also characterized by a flowsheet property contribution as they are derived from base PGs whose

contribution towards a flowsheet property is available apriori; the connection between these PGs determine the topology factor (similar to higher order group correction in molecular groups) and process flowsheet models, by just using the contributions of each PG and not committing to any design aspects, easily calculate the targeted property. The optimal flowsheet synthesized can be designed by the reverse simulation techniques which may sometimes involve invoking a molecular design (again a reverse property prediction) problem when external agents are needed. The target for the reverse simulation problem is the flowsheet property and for molecular design problem, the targets are a feasible property range for the molecule. This dissertation hence presented novel methodologies for CAFD, CAMD and their integration. The two methods based on GC concepts, when integrated using reverse problem formulation techniques lead to very efficient simultaneous process and product synthesis/design, while keeping the high level of accuracy associated to the group contribution methods.

A beta version of the prototype software for solving CAFD problems has been developed. As pointed out in the dissertation, the size of the process synthesis problem being too large to be solved without the help of computer aided tools, development of one such tool helped in solving interesting application examples.

Also, the developed framework as well as its software implementation are seen to be modular in nature thus allowing space for updating of the methodologies within the framework as well as new methodologies to be developed and integrated with the methodologies of the framework.

6.2 Remaining challenges for CAFD and CAMD framework

Mixture design problems are very complex in nature since it involves identifying the new compounds as well as their proportion in the mixture. Mathematical formulation of mixture design problems may be highly non-linear in nature, thus challenging the usage of generate and test methods used in this work. However, modifying the current framework may help in selection of mixture components whose various combinations can be generated and tested if they have the required property targets. Research of this kind is being actively done using past molecular design techniques. Also, some molecules may not be described by using predefined molecular descriptors. Instead of using these pre-described groups for a given problem, methods to generate descriptors and their respective property models within the developed framework can efficiently increase the applicability of the framework.

Thermodynamic insights based identification of candidate tasks and respective unit operations in the problem analysis part of CAFD should be extended to include a wider range of unit operations that may be utilized in the process. The process synthesis methodology depends on the availability of process groups and their contribution towards a flowsheet property. Efforts need to be focused on extending the scope of the methodology by developing more process groups (PGs). The current CAFD methodology employs only one property related to process performance. Additional property models to cater for environmental effects have to be developed for extending the scope of the developed methodology. Also, in order to use the available property model, apriori regressed values of the contributions of each involved PGs have to be tabulated, these are available for only certain PG's and hence effort has to be put in obtaining such data by regression of available experimental data. On the integration of process synthesis and molecular design front, real time

case studies with available equilibrium data, needs to be done to show the evaluation of flowsheets generated by the developed CAFD methodology based on the designed molecules.

Again, sufficient work needs to be put for developing the beta version of the prototype tool. It should be improved in terms of writing the code for other parts of the CAFD framework like - programming the thermodynamic insights based unit operation identification method and integrating it with the current CAFD tool.

7. References

Affens, W. A. (1966). Flammability Properties of Hydrocarbon Fuels. Interrelations of Flammability Properties of n-Alkanes in Air. *Journal of Chemical & Engineering Data, 11*(2), 197-202.

Al-Khayyal, F. A., & Falk, J. E. (1983). Jointly constrained biconvex programming. *Mathematics of Operations Research, 8*(2), 273-286.

Alvarado, M. (2010). *Process-product synthesis, design and analysis through the Group Contribution (GC) approach.* (Doctoral dissertation), Technical University of Denmark Lyngby, Denmark. Retrieved from http://orbit.dtu.dk/fedora/objects/orbit:83238/datastreams/file_5231342/content

Barnicki, S D., & Fair, J R. (1990). Separation system synthesis: a knowledge-based approach. 1. Liquid mixture separations. *Industrial & Engineering Chemistry Research, 29*(3), 421-432.

Barnicki, S D., & Fair, J R. (1992). Separation system synthesis: a knowledge-based approach. 2. Gas/vapor mixtures. *Industrial & Engineering Chemistry Research, 31*(7), 1679-1694.

Barton, A. F. M. (1985). *CRC handbook of solubility parameters and other cohesion parameters*. Boca Raton, Fla.: CRC press.

Bek-Pedersen, E. (2003). *Synthesis and Design of Distillation based Separation Schemes.* (Doctoral dissertation), Technical University of Denmark, Lyngby, Denmark. Retrieved from http://orbit.dtu.dk/fedora/objects/orbit:78003/datastreams/file_2658142/content

Bommareddy, S., Chemmangattuvalappil, N. G., & Eden, M. R. (2012). An Integrated Framework for Flowsheet Synthesis and Molecular Design. *Computer Aided Chemical Engineering, 30A*, 662-666.

Bommareddy, S., Chemmangattuvalappil, N. G., Solvason, C. C., & Eden, M. R. (2009a). Simultaneous Consideration of Process and Product Design Problems using an Algebraic Property Based Approach *Design for Energy & the Environment* (pp. 851-860): Taylor and Francis.

Bommareddy, S., Chemmangattuvalappil, N. G., Solvason, C. C., & Eden, M. R. (2009b). Simultaneous Solution of Process and Molecular Design Problems Using an Algebraic Approach *Computer Aided Chemical Engineering* (Vol. Volume 27, pp. 1095-1100): Elsevier.

Bommareddy, S., Chemmangattuvalappil, N. G., Solvason, C. C., & Eden, M. R. (2010a). An algebraic approach for simultaneous solution of process and molecular design problems. *Brazilian Journal of Chemical Engineering, 27*, 441-450.

Bommareddy, S., Chemmangattuvalappil, N. G., Solvason, C. C., & Eden, M. R. (2010b). Simultaneous solution of process and molecular design problems using an algebraic approach. *Computers & Chemical Engineering, 34*(9), 1481-1486.

Bommareddy, S., Eden, M. R., & Gani, R. (2011). Computer Aided Flowsheet Design using Group Contribution Methods. *Computer Aided Chemical Engineering, 29A*, 321-325.

Bondy, A., & Murty, U.S.R. (2008). *Graph Theory*: Springer.

Brignole, E. A., & Cismondi, M. (2002). Molecular design-generation & test methods. In Rafiqul Gani Luke E.K. Achenie & Venkatasubramanian Venkat (Eds.), *Computer Aided Chemical Engineering* (Vol. 12, pp. 23-41): Elsevier

Cerda, J., Westerberg, A. W., Mason, D., & Linnhoff, B. (1983). Minimum utility usage in heat exchanger network synthesis A transportation problem. *Chemical Engineering Science, 38*(3), 373-387.

Chemmangattuvalappil, Nishanth G., Eljack, Fadwa T., Solvason, Charles C., & Eden, Mario R. (2009). A novel algorithm for molecular synthesis using enhanced property operators. *Computers & Chemical Engineering, 33*(3), 636-643.

Chen, Yi-Ming, & Fan, L. T. (1993). Synthesis of complex separation schemes with stream splitting. *Chemical Engineering Science, 48*(7), 1251-1264.

Constantinou, L., Bagherpour, K., Gani, R., Klein, J. A., & Wu, D. T. (1996). Computer aided product design: problem formulations, methodology and applications. *Computers & Chemical Engineering, 20*(6–7), 685-702.

Constantinou, L., & Gani, R. (1994). New Group-Contribution Method for Estimating Properties of Pure Compounds. *Aiche Journal, 40*(10), 1697-1710.

Constantinou, L., Prickett, S E., & Mavrovouniotis, M L. (1993). Estimation of thermodynamic and physical properties of acyclic hydrocarbons using the ABC approach and conjugation operators. *Industrial & Engineering Chemistry Research, 32*(8), 1734-1746.

d'Anterroches, L. (2006). *Process Flow Sheet Generation & Design through a Group Contribution Approach.* (Doctoral dissertation), Technical University of Denmark Lyngby, Denmark. Retrieved from http://orbit.dtu.dk/fedora/objects/orbit:82699/datastreams/file_5067161/content

d'Anterroches, L., & Gani, R. (2005). Group contribution based process flowsheet synthesis, design and modelling. *Fluid Phase Equilibria, 228–229*(0), 141-146.

Douglas, J. M. (1985). A hierarchical decision procedure for process synthesis. *AIChE Journal, 31*(3), 353-362.

Dunn, R F., & Bush, G E. (2001). Using process integration technology for cleaner production. *Journal of Cleaner Production, 9*(1), 1-23.

Duvedi, Amit P., & Achenie, Luke E. K. (1996). Designing environmentally safe refrigerants using mathematical programming. *Chemical Engineering Science, 51*(15), 3727-3739.

Eden, M. R. (2003). *Property Based Process and Product Synthesis and Design.* (Doctoral dissertation), Technical University of Denmark, Lyngby, Denmark. Retrieved from http://orbit.dtu.dk/en/publications/id(de983d05-148a-4c01-a9e3-6d670e788c92).html

Eden, M. R., Jorgensen, S. B., Gani, R., & El-Halwagi, M. M. (2003a). Reverse problem formulation based techniques for process and product synthesis and design. *Computer Aided Chemical Engineering, 15,* 451-456.

Eden, M. R., Jorgensen, S. B., Gani, R., & El-Halwagi, M. M. (2003b). Property cluster based visual technique for synthesis and design of formulations. In Chen Bingzhen & W. Westerberg Arthur (Eds.), *Computer Aided Chemical Engineering* (Vol. 15, pp. 1175-1180): Elsevier.

Eden, M. R., Jørgensen, S. B., Gani, R., & El-Halwagi, M. M. (2004). A novel framework for simultaneous separation process and product design. *Chemical Engineering and Processing: Process Intensification, 43*(5), 595-608.

El-Halwagi, M. M., Glasgow, I. M., Qin, X. Y., & Eden, M. R. (2004). Property integration: Componentless design techniques and visualization tools. *AIChE Journal, 50*(8), 1854-1869.

El-Halwagi, M.M. (1997). *Pollution Prevention through Process Integration: Systematic Design Tools*: Elsevier Science.

El-Halwagi, M.M. (2006). *Process Integration*: Elsevier

Eljack, F T., & Eden, M R. (2008). A systematic visual approach to molecular design via property clusters and group contribution methods. *Computers & Chemical Engineering, 32*(12), 3002-3010.

Eljack, F. T., Solvason, C.C., Chemmangattuvalappil, N. G., & Eden, M. R. (2008). A Property Based Approach for Simultaneous Process and Molecular Design. *Chinese Journal of Chemical Engineering, 16*(3), 424-434.

Fedors, R. F. (1982). A relationship between chemical structure and the critical temperature. *Chemical Engineering Communications, 16*(1-6), 149-151.

Friedler, F., Fan, L. T., Kalotai, L., & Dallos, A. (1998). A combinatorial approach for generating candidate molecules with desired properties based on group contribution. *Computers & Chemical Engineering, 22*(6), 809-817.

Friedler, F., Tarjan, K., Huang, Y. W., & Fan, L. T. (1993). Graph-theoretic approach to process synthesis: Polynomial algorithm for maximal structure generation. *Computers & Chemical Engineering, 17*(9), 929-942.

Gani, R. (2004). Chemical product design: challenges and opportunities. *Computers & Chemical Engineering, 28*(12), 2441-2457.

Gani, R., & Bek-Pedersen, E. (2000). Simple new algorithm for distillation column design. *AIChE Journal, 46*(6), 1271-1274.

Gani, R., Harper, P. M., & Hostrup, M. (2005). Automatic Creation of Missing Groups through Connectivity Index for Pure-Component Property Prediction. *Industrial & Engineering Chemistry Research, 44*(18), 7262-7269.

Gani, R., Nielsen, B., & Fredenslund, A. (1991). A Group Contribution Approach to Computer-Aided Molecular Design. *AIChE Journal, 37*(9), 1318-1332.

Gani, R., & O'Connell, J. P. (2001). Properties and CAPE: from present uses to future challenges. *Computers & Chemical Engineering, 25*(1), 3-14.

Glasser, D., Crowe, C., & Hildebrandt, D. (1987). A geometric approach to steady flow reactors: the attainable region and optimization in concentration space. *Industrial & Engineering Chemistry Research, 26*(9), 1803-1810.

Grossmann, I. E., Aguirre, P. A., & Barttfeld, M. (2005). Optimal synthesis of complex distillation columns using rigorous models. *Computers & Chemical Engineering, 29*(6), 1203-1215.

Gundersen, T., & Naess, L. (1988). The synthesis of cost optimal heat exchanger networks: An industrial review of the state of the art. *Computers & Chemical Engineering, 12*(6), 503-530.

Harper, P. M., & Gani, R. (2000). A multi-step and multi-level approach for computer aided molecular design. *Computers & Chemical Engineering, 24*(2–7), 677-683.

Harper, P.M. (2000). *A Multi-Phase, Multi-Level Framework for Computer Aided Molecular Design.* (Doctoral dissertation), Technical University of Denmark,Lyngby, Denmark.

Hildebrandt, D., & Biegler, L.T. (1995). Synthesis of Reactor Networks. *AIChE Symposium Series, 91*, 52-67.

Horn, F. (1964). Attainable and Non-Attainable Regions in Chemical Reaction Technique. *Third European Symposium on Chemical Reaction Engineering*, 1-10.

Hostrup, M. (2002). *Integrated Approach to Computer Aided Process Synthesis.* (Doctoral dissertation), Technical University of Denmark, Lyngby, Denmark. Retrieved from http://orbit.dtu.dk/en/publications/integrated-approach-to-computer-aided-process-synthesis(d477b3c5-64f8-450e-9ab9-a4d258c73945).html

Jaksland, C. (1996). *Separation Process Design and Synthesis Based on Thermodynamic Insights.* (Doctoral dissertation), Technical University of Denmark, Lyngby, Denmark.

Jaksland, C., Gani, R., & Lien, K M. (1995). Separation process design and synthesis based on thermodynamic insights. *Chemical Engineering Science, 50*(3), 511-530.

Joback, K. G., Stephanopoulos, G. (1995). Searching Spaces of Discrete Solutions: The Design of Molecules Possessing Desired Physical Properties, *Advanced Chemical Engineering*, 21, 257-311.

Kazantzi, V., Qin, X., El-Halwagi, M., Eljack, F., & Eden, M. R. (2007). Simultaneous Process and Molecular Design through Property Clustering Techniques: A Visualization Tool. *Industrial & Engineering Chemistry Research, 46*(10), 3400-3409.

Kehiaian, H. V. (1983). Group contribution methods for liquid mixtures: A critical review. *Fluid Phase Equilibria, 13*(0), 243-252.

Kier, L.B., & Hall, L.H. (1986). *Molecular connectivity in structure-activity analysis*. New York: John Wiley & Sons.

King, C. J. (1980). Separation Processes, Second Edition.

Kolah, Aspi K., Asthana, Navinchandra S., Vu, Dung T., Lira, Carl T., & Miller, Dennis J. (2008). Reaction Kinetics for the Heterogeneously Catalyzed Esterification of Succinic Acid with

Ethanol. *Industrial & Engineering Chemistry Research, 47*(15), 5313-5317. doi: 10.1021/ie0706616

Konemann, H. (1981). Quantitative structure-activity relationships in fish toxicity studies. Part 1: relationship for 50 industrial pollutants. *Toxicology, 19*(3), 209-221.

Kulprathipanja, S. (2001). *Reactive Separation Processes*: Taylor & Francis.

Linnhoff, B., & Hindmarsh, E. (1983). The pinch design method for heat exchanger networks. *Chemical Engineering Science, 38*(5), 745-763.

Marcoulaki, E. C., & Kokossis, A. C. (1998). Molecular design synthesis using stochastic optimisation as a tool for scoping and screening. *Computers & Chemical Engineering, 22, Supplement 1*(0), S11-S18.

Marrero, J., & Gani, R. (2001). Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria, 183–184*(0), 183-208.

Marrero, J., & Gani, R. (2002). Group-Contribution-Based Estimation of Octanol/Water Partition Coefficient and Aqueous Solubility. *Industrial & Engineering Chemistry Research, 41*(25), 6623-6633.

McCarthy, E., Fraga, E S., & Ponton, J W. (1998). An automated procedure for multicomponent product separation synthesis. *Computers & Chemical Engineering, 22, Supplement 1*(0), S77-S84.

McCormick, G. P. (1976). Computability of global solutions to factorable nonconvex programs: Part I — Convex underestimating problems. *Mathematical Programming, 10*(1), 147-175.

Metzger, M., Glasser, B.J., Glasser, D., Hausberger, B., & Hildebrandt, D. (2007). Teaching Reaction Engineering Using the Attainable Region. *Chemical Engineering Education, 41:4*, 258-264.

Moggridge, G. D., & Cussler, E. L. (2000). An Introduction to Chemical Product Design. *Chemical Engineering Research and Design*, *78*(1), 5-11.

Pistikopoulos, E. N., & Stefanis, S. K. (1998). Optimal solvent design for environmental impact minimization. *Computers & Chemical Engineering, 22*(6), 717-733.

Qian, Y., & Lien, K. M. (1995). Rule based synthesis of separation systems by predictive best first search with rules represented as trapezoidal numbers. *Computers & Chemical Engineering, 19*(11), 1185-1205. doi: http://dx.doi.org/10.1016/0098-1354(94)00116-2

Qin, X.Y. (2007). *Simultaneous process and molecular design/selection through property integration.* (Doctoral dissertation), Texas A&M University, College Station, Texas. Retrieved from http://hdl.handle.net/1969.1/4918

Quesada, I., & Grossmann, I. E. (1995). Global optimization of bilinear process networks with multicomponent flows. *Computers & Chemical Engineering, 19*(12), 1219-1242.

Russel, Boris M., Henriksen, Jens P., Jørgensen, Sten B., & Gani, Rafiqul. (2002). Integration of design and control through model analysis. *Computers & Chemical Engineering, 26*(2), 213-225.

Shah, P B., & Kokossis, A C. (2002). New synthesis framework for the optimization of complex distillation systems. *AIChE Journal, 48*(3), 527-550.

Shelley, Mark D., & El-Halwagi, Mahmoud M. (2000). Component-less design of recovery and allocation systems: a functionality-based clustering approach. *Computers & Chemical Engineering, 24*(9–10), 2081-2091.

Shenoy, U.V. (1995). *Heat Exchanger Network Synthesis: Process Optimization by Energy and Resource Analysis*: Gulf Publishing Company.

Siirola, J J. (1996). Strategic process synthesis: Advances in the hierarchical approach. *Computers & Chemical Engineering, 20, Supplement 2*(0), S1637-S1643.

Sinha, M., & Achenie, L E. K. (2001). Systematic design of blanket wash solvents with recovery considerations. *Advances in Environmental Research, 5*(3), 239-249.

Stefanis, E., & Panayiotou, C. (2008). Prediction of Hansen Solubility Parameters with a New Group-Contribution Method. *International Journal of Thermophysics, 29*(2), 568-585. doi: 10.1007/s10765-008-0415-z

Vaidyanathan, R., & El-Halwagi, M. (1994). Global optimization of nonconvex nonlinear programs via interval analysis. *Computers & Chemical Engineering, 18*(10), 889-897.

Venkatasubramanian, V., Chan, K., & Caruthers, J. M. (1994). Computer-aided molecular design using genetic algorithms. *Computers & Chemical Engineering, 18*(9), 833-844.

Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences, 28*(1), 31-36. doi: 10.1021/ci00057a005

Westerberg, Arthur W. (2004). A retrospective on design and process synthesis. *Computers & Chemical Engineering, 28*(4), 447-458.

Wu, H S., & Sandler, S I. (1989). Proximity effects on the predictions of the UNIFAC model: I. Ethers. *AIChE Journal, 35*(1), 168-172.

Wu, H S., & Sandler, S I. (1991). Use of ab initio quantum mechanics calculations in group contribution methods. 1. Theory and the basis for group identifications. *Industrial & Engineering Chemistry Research, 30*(5), 881-889.

Yamase, M., & Suzuki, Y. (2005). *U.S. Patent No*. 10/173,415. Washington DC: U.S. Patent and Trademark Office.

Appendix A.

The property contributions used in the case studies are given in Table A.1 through Table A.12. In the first three tables, three levels of property contributions given in the property model used by Marrero and Gani (2001) are given. The following properties are predicted using these models:

- Normal boiling and melting temperatures
- Critical pressure, critical volumes and critical temperature
- Standard enthalpy of vaporization and standard Gibbs energy, and standard enthalpy of formation

Table A.6 through Table A.8 provide the group contribution values estimated to be used in the property models used by Marrero and Gani (2002) for the prediction of K_{ow} value.

Table A.10 and Table A.11 provide the group contribution values estimated to be used in the property models used by Stefanis and Panayiotou (2008) for the prediction of Hansen solubility parameters. Table A.12 provides the property models for predicting Hansen solubility parameters. Table A.13 provides the regressed parameters in the equation for the connectivity index method. These expressions are used for the prediction of properties of groups by Marrero and Gani (2001) whose contributions are missing.

Table A.14 through Table A.17 gives the classification and rules to generate molecules.

Table A.1: First order group contribution data (Marrero & Gani, 2001)

	Group	Example	T_{mi}	T_{bi}	T_{ci}	P_{ci}	V_{ci}	G_{fi}	H_{fi}	H_{vi}	H_{msi}
1	CH_3	*n*-Tetracontane (2)	0.6953	0.8491	1.7506	0.018615	68.35	2.878	−42.479	0.217	1.660
2	CH_2	*n*-Tetracontane (38)	0.2515	0.7141	1.3327	0.013547	56.28	8.064	−20.829	4.910	2.639
3	CH	2-Methylpentane (1)	−0.3730	0.2925	0.5960	0.007259	37.50	8.254	−7.122	7.962	0.134
4	C	2,2-Dimethylbutane (1)	0.0256	−0.0671	0.0306	0.001219	16.01	16.413	8.928	10.730	−1.232
5	CH_2=CH	1-Hexene (1)	1.1728	1.5596	3.2295	0.025745	111.43	95.738	57.509	4.031	1.268
6	CH=CH	2-Hexene (1)	0.9460	1.5597	3.0741	0.023003	98.43	92.656	69.664	9.456	4.441
7	CH_2=C	2-Methyl-1-butene (1)	0.7662	1.3621	2.7717	0.021137	91.40	85.107	61.625	8.602	2.451
8	CH=C	2-Methyl-2-butene (1)	0.1732	1.2971	2.5666	0.019609	83.89	88.691	81.835	14.095	3.032
9	C=C	2,3-Dimethyl-2-butene (1)	0.3928	1.2739	2.6391	0.014114	90.66	93.119	95.710	19.910	2.616
10	CH_2=C=CH	1,2-Butadiene (1)	1.7036	2.6840	5.4330	0.035483	143.57	229.906	198.840	11.310	7.076
11	CH_2=C=C	3-Methyl-1,2-butadiene (1)	1.5453	2.4014	4.8219	0.029678	146.36	226.710	208.490	*****	7.435
12	CH=C=CH	2,3-Pentadiene (1)	1.2850	2.5400	*****	*****	*****	*****	*****	*****	6.000
13	CH≡C	1-Pentyne (1)	2.2276	1.7618	3.7897	0.014010	84.60	230.029	224.902	6.144	−1.548
14	C≡C	3-Decyne (1)	2.0516	1.6767	4.5870	0.010888	74.66	216.013	228.282	12.540	6.128
15	aCH	Benzene (6)	0.5860	0.8365	2.0337	0.007260	42.39	26.732	12.861	3.683	1.948
16	aC fused with aromatic ring	Naphthalene (2)	1.8955	1.7324	5.4979	0.003564	35.71	20.379	20.187	6.631	0.845
17	aC fused with non-aromatic subring	Indane (2)	1.2065	1.1995	3.1058	0.006512	34.65	33.912	30.768	6.152	1.095
18	aC except as above	Benzophenone (1)	0.9176	1.5468	4.5344	0.012859	26.47	23.331	24.701	6.824	−0.531
19	aN in aromatic ring	Pyridine (1)	2.0438	1.3977	4.0954	−0.003339	36.47	89.902	70.862	9.420	2.555
20	aC–CH_3	Toluene (1)	1.0068	1.5653	3.4611	0.020907	97.33	24.919	−19.258	8.279	2.969
21	aC–CH_2	Ethylbenzene (1)	0.1065	1.4925	2.9003	0.018082	87.19	31.663	4.380	11.981	0.948
22	aC–CH	Cumene (1)	−0.5197	0.8665	1.9512	0.011795	73.51	30.393	18.440	13.519	−1.037
23	aC–C	*tert*-Butylbenzene (1)	−0.1041	0.5229	0.8576	0.011298	67.20	40.127	35.297	16.912	−2.856
24	aC–CH=CH_2	Styrene (1)	1.2832	2.4308	5.7861	0.030637	134.69	114.531	77.863	*****	4.013
25	aC–CH=CH	1-Propenylbenzene (1)	1.7744	2.9262	6.5062	0.026282	128.84	111.216	88.084	*****	8.274
26	aC–C=CH_2	α-Methylstyrene (1)	1.2612	2.1472	4.9967	0.026371	110.74	115.728	90.927	*****	3.324
27	aC–C=CH	Phenylacetylene (1)	1.7495	2.3057	6.4572	0.019507	112.08	263.205	257.448	*****	2.514
28	aC–C=C	1-Phenyl-1-propyne (1)	*****	2.7341	*****	*****	*****	*****	*****	*****	*****
29	OH	1,4-Butanediol (2)	2.7888	2.5670	5.2188	−0.005401	30.61	−144.051	−178.360	24.214	4.786
30	aC–OH	Phenol (1)	5.1473	3.3205	9.3472	−0.008788	50.77	−131.327	−164.191	34.099	8.427
31	COOH	1,5-Pentanedioic acid (2)	7.4042	5.1108	14.6038	0.009885	90.66	−337.090	−389.931	17.002	10.692
32	aC–COOH	Benzoic acid (1)	12.4296	6.0677	15.4515	0.017100	119.10	−312.422	−361.249	*****	14.649
33	CH_3CO	2-Butanone (1)	2.9588	3.1178	7.0058	0.025227	127.99	−120.667	−180.604	15.195	8.062
34	CH_2CO	3-Pentanone (1)	2.5232	2.6761	5.7157	0.019619	112.79	−120.425	−163.090	19.392	8.826
35	CHCO	2,4-Dimethyl-3-pentanone (1)	1.1565	2.1748	4.4743	0.012487	97.16	−116.799	−139.909	20.350	7.205
36	CCO	2,2,4,4-Tetramethyl-3-pentanone (1)	1.0638	1.7287	*****	*****	*****	*****	*****	*****	*****
37	aC–CO	Acetophenone (1)	2.9157	3.4650	9.4806	0.011007	90.69	−91.812	−106.965	25.036	4.852
38	CHO	1-Hexanal (1)	3.0186	2.5388	5.8013	0.010204	71.08	−100.882	−130.816	12.370	11.325
39	aC–CHO	Benzaldehyde (1)	2.4744	3.5172	9.4795	0.019633	122.91	−80.222	−107.159	*****	7.273
40	CH_3COO	Butyl acetate (1)	2.1657	3.1228	6.3179	0.033812	148.91	−306.733	−387.458	19.342	7.910
41	CH_2COO	Methyl Butyrate (1)	1.6329	2.9850	5.9619	0.026983	132.89	−298.332	−364.204	21.100	9.479
42	CHCOO	Ethyl isobutyrate (1)	1.0668	2.2869	4.7558	0.021990	125.52	−301.414	−352.057	24.937	9.317
43	CCOO	Ethyl 2,2-dimethylpropionate (1)	0.3983	1.6918	*****	*****	*****	*****	*****	23.739	*****
44	HCOO	Propyl formate (1)	2.0223	2.5972	5.6064	0.015249	93.29	−276.878	−327.678	15.422	8.115

Table A.1 (contd.)

45	aC–COO	Methyl benzoate (1)	1.3348	3.1952	6.7311	0.018948	105.53	−291.662	−307.727	25.206	8.149
46	aC–OOCH	Phenyl formate (1)	*****	0.4621	*****	*****	*****	*****	*****	*****	*****
47	aC–OOC	Phenyl acetate (1)	4.8044	3.0854	*****	*****	*****	*****	*****	*****	5.875
48	COO except as above	Ethyl acrylate (1)	1.5038	2.1903	4.7346	0.013087	81.17	−299.803	−331.397	*****	10.573
49	CH_3O	Methyl butyl ether (1)	1.3643	1.7703	3.4393	0.020084	88.20	−90.329	−156.062	5.783	5.089
50	CH_2O	Di-*n*-butyl ether (1)	0.8733	1.3368	2.4217	0.017954	74.03	−105.579	−152.239	9.997	4.891
51	CH–O	sec-Butyl ether (1)	0.2461	0.8924	0.7889	0.014487	60.06	−101.207	−147.709	14.620	4.766
52	C–O	*tert*-Butylether (1)	−0.4446	0.4983	0.2511	0.005613	52.96	−92.804	−121.608	13.850	2.458
53	aC–O	Methyl phenyl ether (1)	1.3045	1.8522	3.6588	0.005115	47.27	−83.354	−101.783	16.151	−0.118
54	CH_2NH_2	Ethylamine (1)	3.2742	2.7987	8.1745	0.011413	117.62	68.812	−10.703	15.432	13.482
55	$CHNH_2$	sec-Butylamine (1)	30.8394	2.0948	4.2847	0.013049	76.36	61.452	0.730	16.048	6.283
56	CNH_2	*tert*-Butylamine (1)	11.7400	1.6525	2.8546	0.010790	80.01	55.202	2.019	17.257	*****
57	CH_3NH	Dimethylamine (1)	2.4034	2.2514	4.5529	0.015863	77.04	88.512	24.740	11.831	4.490
58	CH_2NH	Dipropylamine (1)	1.7746	1.8750	3.2422	0.020482	95.15	88.874	23.610	13.067	7.711
59	CHNH	Diisopropylamine (1)	1.7577	1.2317	2.0057	0.005329	99.16	73.101	21.491	14.048	2.561
60	CH_3N	Methyldiethylamine (1)	0.9607	1.3841	3.0106	0.021186	94.94	125.906	55.024	9.493	6.008
61	CH_2N	Triethylamine (1)	0.0442	1.1222	2.1673	0.027454	74.05	121.247	65.331	12.636	1.756
62	aC–NH_2	Aniline (1)	3.9889	3.8298	10.2155	0.005335	81.40	66.470	17.501	23.335	6.542
63	aC–NH	*N*-methyl aniline (1)	1.4837	2.9230	8.4081	−0.005596	86.37	98.195	53.274	23.026	0.624
64	aC–N	*N*,*N*-dimethyl aniline (1)	1.7618	2.1918	5.8536	−0.000838	108.39	143.280	115.606	22.249	−2.576
65	NH_2 except as above	Cyclobutylamine	3.3478	2.0315	4.7420	0.000571	63.39	42.687	−8.556	13.425	6.158
66	CH=N	Acetaldazine (2)	8.8492	1.5332	*****	*****	*****	*****	*****	*****	*****
67	C=N	Ketazine (2)	1.4621	1.4291	*****	*****	*****	*****	*****	*****	*****
68	CH_2CN	Propionitrile (1)	2.5760	4.5871	12.9827	0.036523	133.62	134.997	99.245	21.923	7.303
69	CHCN	Isobutyronitrile (1)	2.1393	3.9774	8.4309	0.029034	134.73	142.475	151.390	24.963	9.464
70	CCN	2,2-Dimethylpropionitrile (1)	3.3807	2.8870	5.8829	0.024654	120.74	142.295	124.770	24.967	4.166
71	aC–CN	Benzonitrile (1)	5.1346	4.1424	10.4124	0.020978	119.08	162.175	148.968	*****	6.788
72	CN except as above	Acrylonitrile (1)	3.2747	3.0972	8.1381	0.024346	94.91	130.986	124.917	16.639	6.867
73	CH_2NCO	Ethyl isocyanate (1)	4.2256	3.4891	*****	*****	*****	*****	*****	*****	*****
74	CHNCO	Isopropyl isocyanate (1)	*****	3.1220	*****	*****	*****	*****	*****	*****	*****
75	CNCO	*tert*-Butyl isocyanate (1)	9.1492	*****	*****	*****	*****	*****	*****	*****	*****
76	aC–NCO	Phenyl isocyanate (1)	2.2327	3.1853	6.5884	0.025065	141.24	*****	*****	*****	*****
77	CH_2NO_2	1-Nitropropane (1)	3.2131	4.5311	10.9507	0.021056	157.57	25.783	−65.620	29.640	10.989
78	$CHNO_2$	2-Nitropropane (1)	0.7812	3.8069	9.5487	0.014899	143.36	16.407	−60.750	29.173	*****
79	CNO_2	2-Methyl-2-nitropropane (1)	5.6280	3.3059	*****	*****	*****	*****	*****	*****	−4.187
80	aC–NO_2	Nitrobenzene (1)	4.3531	4.5750	12.1243	0.018311	133.06	57.352	−22.931	24.863	7.572
81	NO_2 except as above	Nitrocyclohexane (1)	3.0376	3.2069	*****	*****	*****	*****	*****	*****	6.302
82	ONO	Butyl nitrite (1)	*****	1.8896	*****	*****	*****	*****	*****	*****	*****
83	ONO_2	*n*-Butyl nitrate (1)	2.5974	3.2656	*****	*****	*****	*****	*****	*****	9.353
84	$HCON(CH_2)_2$	Diethylformamide (1)	*****	5.8779	*****	*****	*****	*****	*****	*****	*****
85	$HCONHCH_2$	Ethylformamide (1)	*****	7.4566	*****	*****	*****	*****	*****	46.490	*****
86	$CONH_2$	Butyramide (1)	13.2124	6.5652	25.1184	0.001467	138.71	−127.512	−201.369	44.240	16.840
87	$CONHCH_3$	Methylacetamide (1)	5.4720	5.0724	20.5590	0.023455	190.71	−102.912	−203.069	*****	17.429
88	$CONHCH_2$	Ethylacetamide (1)	5.8825	6.6810	*****	*****	*****	*****	−183.613	52.723	*****
89	$CON(CH_3)_2$	Dimethylacetamide (1)	4.1720	6.0070	15.4603	0.043090	244.71	−56.412	−188.069	38.290	11.553
90	$CONCH_3CH_2$	Methylethylacetamide (1)	*****	*****	*****	*****	*****	*****	−48.210	*****	*****
91	$CON(CH_2)_2$	Diethylacetamide (1)	*****	5.0664	*****	*****	*****	*****	*****	*****	*****
92	CONHCO	Diacetamide (1)	9.1763	7.6172	*****	*****	*****	*****	*****	*****	*****
93	CONCO	Methyldiacetamide	3.2657	5.6487	*****	*****	*****	*****	*****	*****	*****

Table A.1 (contd.)

	Group	Example	T_{m1i}	T_{b1i}	T_{c1i}	P_{c1i}	V_{c1i}	G_{f1i}	H_{f1i}	H_{v1i}	H_{fus1i}
94	aC–$CONH_2$	Benzamide	12.8071	8.3775	*****	*****	*****	*****	*****	*****	16.811
95	aC–NH(CO)H	*N*-phenylformamide (1)	5.6631	7.3497	19.8979	0.023447	162.08	−44.595	−125.052	*****	8.658
96	aC–N(CO)H	*N*-methyl-*N*-phenylmethanamide (1)	3.3602	5.1373	*****	*****	*****	*****	*****	*****	*****
97	aC–CONH	*N*-methylbenzamide (1)	6.5160	7.5850	*****	*****	*****	*****	*****	*****	10.959
98	aC–NHCO	*N*-(2-methylphenyl)acetamide (1)	9.8204	7.4955	*****	*****	*****	*****	*****	*****	4.370
99	aC–NCO	Phenylmethylacetamide (1)	7.2552	*****	*****	*****	*****	*****	*****	*****	*****
100	NHCONH	*N*,*N*′-dimethylurea (1)	9.3110	8.9406	*****	*****	*****	*****	*****	*****	9.862
101	NH_2CONH	Methylurea (1)	14.2020	16.3539	*****	*****	*****	*****	*****	*****	12.845
102	NH_2CON	*N*,*N*-dimethylurea (1)	13.0856	2.0796	*****	*****	*****	*****	*****	*****	10.958
103	NHCON	Trimethylurea (1)	8.4447	7.1529	*****	*****	*****	*****	*****	*****	12.098
104	NCON	Tetramethylurea (1)	3.5041	4.1459	*****	*****	*****	*****	*****	*****	9.557
105	aC–$NHCONH_2$	Phenylurea (1)	13.4695	5.7604	*****	*****	*****	*****	*****	*****	16.703
106	aC–NHCONH	*N*,*N*′-diphenylurea	23.2570	1.1633	*****	*****	*****	*****	*****	*****	18.460
107	NHCO except as above	*N*-chloroacetamide (1)	3.0882	*****	*****	*****	*****	*****	*****	*****	*****
108	CH_2Cl	1-Chlorobutane (1)	1.9253	2.6364	6.2561	0.021419	112.12	−19.484	−65.056	11.754	6.353
109	CHCl	2-Chloropropane (1)	1.0224	2.0246	4.3756	0.015640	100.78	−31.933	−65.127	12.048	*****
110	CCl	2-Chloro-2-methylpropane (1)	1.8424	1.7049	3.7063	0.009187	87.01	−37.848	−62.881	16.597	−0.082
111	$CHCl_2$	1,1-Dichloroethane (1)	2.5196	3.3420	7.8956	0.028236	159.79	−24.214	−80.812	17.251	6.781
112	CCl_2	2,2-Dichloropropane (1)	3.6491	2.9609	*****	*****	*****	*****	*****	20.473	1.823
113	CCl_3	1,1,1-Trichloroethane	4.4493	3.9093	8.8073	0.036746	204.71	−44.122	−105.369	20.550	3.492
114	CH_2F	1-Fluorobutane (1)	1.5597	1.5022	3.3179	0.023315	87.71	−180.212	−227.469	8.238	7.139
115	CHF	2-Fluorobutane (1)	1.1289	1.3738	2.6702	0.020040	78.08	−228.239	−261.901	*****	3.917
116	CF	2-Fluoro-2-methylpropane (1)	2.5398	1.0084	2.1633	−0.010120	*****	*****	*****	6.739	*****
117	CHF_2	1,1-Difluoroethane (1)	2.1689	2.2238	3.5702	0.031524	102.71	−411.239	−463.901	*****	7.011
118	CF_2	Perfluorohexane (4)	0.1312	0.5142	0.8543	0.018572	95.09	*****	*****	1.621	*****
119	CF_3	Hexafluoroethane (2)	1.4828	1.1916	1.7737	0.048565	108.85	−615.333	−673.875	7.352	2.526
120	CCl_2F	Tetrachloro-1,2-difluoroethane (2)	3.2035	2.5053	5.1653	0.037948	171.04	−249.020	−306.765	8.630	3.114
121	HCClF	1-Chloro-1,2,2,2-tetrafluoroethane (1)	*****	2.0542	*****	*****	*****	*****	*****	*****	*****
122	$CClF_2$	1,2-Dichlorotetrafluoroethane (2)	1.7510	1.7227	3.0593	0.041641	146.01	−396.814	−458.074	8.086	2.156
123	aC–Cl	Chlorobenzene (1)	1.7134	2.0669	5.7046	0.016033	92.67	1.985	−17.002	11.224	4.435
124	aC–F	Hexafluorobenzene (6)	0.9782	0.7945	1.5491	0.014037	54.36	−141.306	−160.965	3.965	2.003
125	aC–I	Iodobenzene (1)	2.1905	3.7739	12.4470	0.014403	131.08	91.505	95.048	*****	2.814
126	aC–Br	Bromobenzene (1)	2.4741	2.8414	8.4199	0.010199	104.12	42.977	38.917	14.393	5.734
127	I^- except as above	Iodoethane (1)	1.9444	3.1778	8.5775	−0.004637	104.28	43.910	47.632	14.171	6.103
128	Br^- except as above	Bromoethane (1)	1.7641	2.4231	4.5036	−0.001460	77.99	5.528	−1.703	9.888	4.826
129	F^- except as above	Benzyl fluoride (1)	1.2308	0.8504	0.8976	0.012034	24.62	−182.973	−201.968	*****	3.096
130	Cl^- except as above	Ethyl chloroacetate (1)	1.5454	1.5147	4.0947	0.007923	57.77	−29.876	−46.963	*****	5.181
131	CHNOH	Propionaldehyde oxime (1)	3.9813	4.5721	*****	*****	*****	*****	*****	*****	*****
132	CNOH	Diethyl ketoxime (1)	3.5484	4.0142	*****	*****	*****	*****	*****	*****	*****
133	aC–CHNOH	Phenyl oxime (1)	10.5579	*****	*****	*****	*****	*****	*****	*****	*****
134	OCH_2CH_2OH	2-Ethoxyethanol (1)	2.3651	4.8721	10.4579	0.025986	159.33	−233.335	−343.903	31.493	8.454
135	$OCHCH_2OH$	2-Ethoxy-1-propanol (1)	*****	4.2329	*****	*****	*****	*****	*****	*****	*****
136	OCH_2CHOH	1-Methoxy-2-propanol (1)	1.5791	3.6653	*****	0.018783	147.66	−239.423	−333.385	*****	12.594
137	–O–OH	*tert*-Butylhydroperoxide (1)	4.8181	3.1669	5.8307	−0.002815	58.01	−75.568	−125.111	*****	*****
138	CH_2SH	Ethanethiol (1)	2.2992	3.1974	7.7300	0.017299	105.68	27.469	−8.021	16.815	10.068
139	CHSH	2-Propanethiol (1)	0.9704	2.5910	5.8527	0.008968	109.36	27.030	3.510	17.098	4.266
140	CSH	2-Methyl-2-propanethiol (1)	4.2329	2.0902	4.6431	0.005118	94.01	27.338	12.589	18.397	−0.623

Table A.1 (contd.)

141	aC–SH	Benzenethiol (1)	2.8464	3.2675	9.5115	0.010086	95.08	48.905	41.648	17.413	4.513
142	–SH (except as above)	Cyclohexanethiol (1)	0.9600	2.3323	7.7987	0.006399	57.89	15.818	11.339	9.813	5.829
143	CH_3S	Dimethylsulfide (1)	1.7150	2.9892	6.9733	0.018013	122.03	35.845	−3.337	14.296	7.497
144	CH_2S	Diethylsulfide (1)	1.0063	2.6524	6.4871	0.015254	106.60	42.684	21.492	16.965	4.096
145	CHS	Diisopropylsulfide (1)	0.7892	2.0965	*****	*****	*****	*****	*****	19.038	*****
146	CS	di-*tert*-Butylsulfide (1)	1.1170	1.6412	*****	*****	*****	*****	*****	19.996	*****
147	aC–S–	Phenyl methyl sulfide (1)	0.9646	2.9731	*****	*****	*****	*****	*****	*****	*****
148	SO	Dimethyl sulfoxide (1)	5.3663	6.2796	19.8953	−0.005534	82.36	−52.231	−71.050	*****	13.403
149	SO_2	Dimethyl sulfone (1)	7.0778	7.0976	17.2586	−0.000784	89.95	−257.608	−305.498	*****	17.748
150	SO_3 (sulfite)	Dimethyl sulfite (1)	*****	3.9199	8.6910	0.004240	115.80	*****	−430.833	*****	*****
151	SO_3 (sulfonate)	Dimethyl sulfonate (1)	5.8426	6.7785	*****	*****	*****	*****	*****	*****	*****
152	SO_4 (sulfate)	Dimethyl sulfate (1)	3.6976	5.5627	18.9366	−0.027208	144.58	−519.853	−621.412	*****	*****
153	aC–SO	Phenyl methyl sulfoxide (1)	3.9911	6.1185	*****	*****	*****	*****	*****	*****	*****
154	aC–SO_2	Diphenyl sulfone (1)	5.2948	8.4333	*****	*****	135.47	−314.643	−370.493	*****	3.281
155	PH (phosphine)	Dimethylphosphine (1)	*****	2.0536	*****	*****	*****	*****	*****	*****	*****
156	P (phosphine)	Trimethylphosphine (1)	*****	1.0984	*****	*****	*****	*****	*****	*****	*****
157	PO_3 (phosphite)	Triethylphosphite (1)	1.0306	2.7900	*****	*****	*****	*****	*****	*****	*****
158	PHO_3 (phosphonate)	Dimethylphosphonate (1)	*****	5.6433	*****	*****	*****	*****	*****	*****	*****
159	PO_3 (phosphonate)	Trimethylphosphonate (1)	*****	4.5468	*****	*****	*****	*****	*****	*****	*****
160	PHO_4 (phosphate)	Diethylphosphate (1)	2.7461	5.1567	*****	*****	*****	*****	*****	*****	*****
161	PO_4 (phosphate)	Trimethylphosphate (1)	2.0330	3.7657	16.9914	−0.029036	85.59	*****	−1060.325	*****	*****
162	aC–PO_4	Triphenylphosphate (1)	−1.7840	2.3522	*****	*****	*****	*****	−1005.161	*****	4.256
163	aC–P	Triphenylphosphine (1)	0.2337	2.9272	38.6148	−0.126108	−142.79	*****	72.339	*****	−5.654
164	CO_3 (carbonate)	Diethylcarbonate (1)	3.6593	2.8847	6.6804	0.007235	93.56	−447.186	−516.282	21.613	8.363
165	C_2H_3O	Ethyl oxirane (1)	1.3135	2.8451	6.6418	0.021238	125.43	11.149	−52.241	*****	*****
166	C_2H_2O	2,2-Dimethyl oxirane (1)	*****	2.6124	6.0159	0.010678	194.36	1.890	−51.390	*****	*****
167	C_2O	Trimethyl oxirane (1)	*****	2.2036	*****	*****	*****	*****	*****	*****	*****
168	CH_2 (cyclic)	Cyclopentane (5)	0.5699	0.8234	1.8815	0.009884	49.24	13.287	−18.575	3.341	1.069
169	CH (cyclic)	Methylcyclopentane (1)	0.0335	0.5946	1.1020	0.007596	44.95	6.107	−12.464	6.416	2.511
170	C (cyclic)	1,1-Dimethylcyclohexane (1)	0.1695	0.0386	−0.2399	0.003268	33.32	−0.193	−2.098	7.017	−0.921
171	CH=CH (cyclic)	Cyclobutene (1)	1.1936	1.5985	3.6426	0.013815	83.91	86.493	59.841	7.767	1.185
172	CH=C (cyclic)	1-Methylcyclopentene (1)	0.4344	1.2529	3.5475	0.010576	70.98	67.056	64.295	7.171	2.559
173	C=C (cyclic)	1,2-Dimethylcyclopentene (1)	0.3048	1.1975	*****	*****	*****	*****	*****	*****	*****
174	CH_2=C (cyclic)	Methylene cyclohexane (1)	0.2220	1.5109	4.4913	0.019101	83.96	*****	*****	*****	5.351
175	NH (cyclic)	Cyclopentimine (1)	3.4814	2.1634	5.9726	−0.003678	51.80	72.540	23.138	13.700	8.655
176	N (cyclic)	*N*-methylpyrrolidine (1)	0.6040	1.6541	4.3905	−0.001179	31.41	83.779	65.622	*****	0.269
177	CH=N (cyclic)	Imidazole (1)	5.5779	6.5230	*****	*****	*****	*****	*****	*****	3.993
178	C=N (cyclic)	2-Methyl-1H-imidazole (1)	6.6382	6.6710	*****	*****	*****	*****	*****	*****	*****
179	O (cyclic)	Tetrahydropyran (1)	1.3828	1.0245	2.7409	−0.000387	17.69	−114.062	−137.353	6.877	3.806
180	CO (cyclic)	Cyclobutanone (1)	3.2119	2.8793	12.6396	−0.000207	57.38	−156.672	−180.166	17.124	6.137
181	S (cyclic)	2-Methyl-thiophene (1)	1.6023	2.3256	5.5523	0.001540	45.45	12.020	15.453	12.262	5.170
182	SO_2 (cyclic)	Cyclobutadiene sulfone (1)	6.1006	*****	24.3995	0.002487	96.66	−241.601	−283.839	*****	9.934

Table A.2: Second-order group contribution data (Marrero & Gani, 2001).

	Group	Example	T_{m2j}	T_{b2j}	T_{c2j}	P_{c2j}	V_{c2j}	G_{f2j}	H_{f2j}	H_{v2j}	H_{fus2j}
1	$(CH_3)_2CH$	2-Methylpentane (1)	0.1175	−0.0035	−0.0471	0.000473	1.71	−0.418	−0.419	−0.399	0.396
2	$(CH_3)_3C$	2,2,4,4-Tetramethylpentane (2)	−0.1214	0.0072	−0.1778	0.000340	3.14	−2.776	−1.967	−0.417	0.554
3	$CH(CH_3)CH(CH_3)$	2,3,4-Trimethylpentane (2)	0.2390	0.3160	0.5602	−0.003207	−3.75	6.996	6.065	0.532	−1.766
4	$CH(CH_3)C(CH_3)_2$	2,2,3,4,4-Pentamethylpentane (2)	−0.3276	0.3976	0.8994	−0.008733	−10.06	8.938	8.078	0.623	0.351
5	$C(CH_3)_2C(CH_3)_2$	2,2,3,3,4,4-Hexamethylpentane (2)	3.3297	0.4487	1.5535	−0.016852	−8.70	10.735	10.535	5.086	−1.089
6	$CH_n{=}CH_m{-}CH_p{=}CH_k$ (k, m, n, p in 0..2)	1,3-Butadiene (1)	0.7451	0.1097	0.4214	0.000792	−7.88	−6.562	−11.786	1.632	1.408
7	$CH_3{-}CH_m{=}CH_n$ (m, n in 0..2)	2-Methyl-2-butene (3)	0.0524	0.0369	−0.0172	−0.000101	0.50	−0.120	−0.048	0.064	0.070
8	$CH_2{-}CH_m{=}CH_n$ (m, n in 0..2)	1,4-Pentadiene (2)	−0.1077	−0.0537	0.0262	0.000815	0.14	1.006	1.449	−0.060	−0.632
9	$CH_p{-}CH_m{=}CH_n$ (m, n in 0..2; p in 0..1)	3-Methyl-1-butene (1)	−0.2485	−0.0093	−0.1526	−0.000163	−2.67	3.857	3.964	0.004	−0.368
10	CHCHO or CCHO	2-Methylbutyraldehyde (1)	0.5715	−0.1286	−1.0434	0.005789	10.36	−0.525	1.514	−0.550	−0.369
11	CH_3COCH_2	2-Pentanone (1)	−0.0968	−0.0215	−0.0338	−0.000111	−4.08	−1.543	0.033	−0.403	0.105
12	CH_3COCH or CH_3COC	3-Methyl-2-pentanone (1)	−0.6024	−0.0803	−0.3658	−0.001892	3.02	2.202	4.994	0.723	1.005
13	CHCOOH or CCOOH	2-Methyl butanoic acid (1)	−3.1734	−0.3203	−4.7275	0.006916	10.56	3.920	1.121	7.422	5.475
14	CH_3COOCH or CH_3COOC	Isopropyl acetate (1)	0.2114	−0.2066	−0.5537	−0.000569	4.28	−11.779	−12.295	−1.871	1.208
15	CO–O–CO	Propanoic anhydride (1)	−1.2441	−0.0500	−0.3576	0.001812	2.98	−16.075	−14.140	*****	−2.666
16	CHOH	2-Butanol (1)	−0.3489	−0.2825	−0.6768	0.000246	−3.04	−5.614	−4.422	−0.206	−0.599
17	COH	2-Methyl-2-butanol (1)	0.3695	−0.5325	−1.5224	0.003224	13.98	−25.382	−25.929	−1.579	−0.459
18	CH_3COCH_nOH (n in 0..2)	3-Hydroxy-2-butanone (1)	0.9886	−0.2987	−0.3940	−0.002912	5.17	6.621	8.244	*****	*****
19	NCCHOH or NCCOH	2-Hydroxypropionitrile (1)	−1.1810	0.2981	0.3414	−0.000516	0.68	4.833	0.000	*****	−0.149
20	OH–CH_n–COO (n in 0..2)	Ethyl lactate (1)	−0.1526	−0.2310	*****	*****	*****	*****	*****	*****	*****
21	$CH_m(OH)CH_n(OH)$ (m, n in 0..2)	Ethylene glycol (1)	−0.0414	0.8854	1.9395	−0.004712	7.54	−1.051	−0.592	−6.611	−0.306
22	$CH_m(OH)CH_n(-)$ (m, n, p in 0..2)	2-Amino-1-butanol (1)	−0.5941	0.5082	1.2342	0.002581	5.58	−1.506	−0.959	*****	−0.041
23	$CH_m(NH_2)CH_n(NH_2)$ (m, n in 0..2)	Ethylenediamine (1)	0.3258	−0.0064	−3.3555	0.000726	20.82	0.344	−1.443	2.384	−1.575
24	$CH_m(NH)CH_n(NH_2)$ (m, n in 1..2)	Diethylenetriamine (1)	−1.8403	0.2318	−1.1598	0.000157	−26.31	3.848	3.608	*****	*****
25	$H_2NCOCH_nCH_mCONH_2$ (m, n in 1..2)	Butanediamide (1)	11.5351	*****	*****	*****	*****	*****	*****	*****	*****
26	$CH_m(NH_2)$–COOH (m, n in 0..2)	L-Alanine (1)	12.3481	*****	62.4740	−0.002696	17.78	3.145	6.598	*****	7.032
27	HOOC–CH_n–COOH (n in 1..2)	Malonic acid (1)	0.9327	−0.1222	1.9595	−0.001479	12.46	−5.217	−6.058	*****	4.264
28	HOOC–CH_n–CH_m–COOH (n, m in 1..2)	Succinic acid (1)	7.5057	*****	0.7686	0.000090	15.17	−4.281	−6.929	*****	29.245
29	HO–CH_n–COOH (n in 1..2)	2-Hydroxyisobutyric acid (1)	−0.4531	−0.4625	*****	*****	*****	*****	*****	*****	*****
30	$NH_2{-}CH_n{-}CH_m{-}COOH$ (n, m in 1..2)	β-Alanine (1)	14.1593	*****	*****	*****	*****	*****	*****	*****	*****
31	CH_3–O–CH_n–COOH (n in 1..2)	Methoxyacetic acid (1)	−2.3026	0.9198	0.4750	−0.001445	7.91	−2.678	−1.727	*****	*****
32	HS–CH_n–COOH	2-Mercaptopropionic acid (1)	−2.1535	*****	*****	*****	*****	*****	*****	*****	*****
33	HS–CH_n–CH_m–COOH (n, m in 1..2)	β-Thiolactic acid (1)	−2.7514	*****	−0.2697	0.000655	20.43	−7.376	7.292	*****	−3.623
34	NC–CH_n–CH_m–CN (n, m in 1..2)	1,2-Dicyanoethane (1)	4.0747	1.8957	1.9699	0.002330	24.82	18.974	5.661	*****	−8.038
35	OH–CH_n–CH_m–CN (n, m in 1..2)	3-Hydroxypropanenitrile (1)	−0.9493	1.3434	0.2311	−0.001022	14.54	0.558	−3.906	*****	−4.371
36	HS–CH_n–CH_m–SH (n, m in 1..2)	1,2-Ethanedithiol (1)	0.2232	0.1815	2.1272	0.001321	−10.31	6.728	0.794	−0.683	−0.931
37	COO–CH_n–CH_m–OOC (n, m in 1..2)	Ethylene glycol diacetate (1)	−0.5946	0.3401	1.5418	−0.003385	−2.33	1.306	4.025	1.203	*****
38	OOC–CH_m–CH_m–COO (n, m in 1..2)	Dimethylsuccinate (1)	2.5962	0.5794	*****	*****	*****	*****	*****	*****	2.303
39	NC–CH_n–COO (n in 1..2)	Methylcyanoacetate (1)	−0.2509	1.2171	2.7051	−0.001999	−0.73	*****	*****	*****	1.100
40	$COCH_nCOO$ (n in 1..2)	Methylacetoacetate (1)	0.6304	0.2427	0.7502	−0.000231	1.69	10.556	−7.261	*****	*****
41	CH_m–O–$CH_n{=}CH_p$ (m, n, p in 0..3)	Ethyl viny ether (1)	−0.0811	0.1399	0.2900	−0.000432	−4.54	−10.098	−9.411	0.372	3.169
42	$CH_m{=}CH_n$–F (m, n in 0..2)	1-Fluoro-1-propene (1)	−0.2568	0.0591	*****	*****	*****	*****	*****	*****	2.823
43	$CH_m{=}CH_n$–Br (m, n in 0..2)	1-Bromo-1-propene (1)	−0.4329	−0.3192	*****	−0.010021	2.63	14.470	17.014	*****	2.212
44	$CH_m{=}CH_n$–I (m, n in 0..2)	1-Iodo-1-propene (1)	*****	−0.3486	*****	*****	*****	*****	*****	*****	*****
45	$CH_m{=}CH_n$–Cl (m, n in 0..2)	1-Chloro-2-methylpropene (1)	0.0446	−0.0268	−0.0188	0.000152	2.80	8.207	9.715	*****	−0.480
46	$CH_m{=}CH_n$–CN (m, n in 0..2)	Acrylonitrile (1)	0.1027	0.0653	−1.1249	0.000893	3.82	−8.304	−16.903	*****	−0.405

Table A.2 (contd.)

47	$CH_n{=}CH_m{-}COO{-}CH_p$ (m, n, p in 0..3)	Ethyl Acrylate (1)	0.2117	−0.0430	−0.0880	0.000044	0.21	−12.085	−12.509	*****	−0.014
48	$CH_m{=}CH_n{-}CHO$ (m, n in 0..2)	Propenaldehyde (1)	−0.7191	0.1102	*****	*****	*****	*****	*****	*****	*****
49	$CH_m{=}CH_n{-}COOH$ (m, n in 0..2)	Acrylic Acid (1)	2.4103	0.0667	−1.7762	−0.000763	4.36	10.194	9.090	*****	1.291
50	$aC{-}CH_n{-}X$ (n in 1..2) X: Halogen	Benzyl bromide (1)	0.8092	0.4537	2.2630	0.002464	−4.88	−8.081	−8.570	*****	*****
51	$aC{-}CH_n{-}NH_m$ (n in 1..2; m in 0..2)	Benzyl amine (1)	−1.0802	0.2590	1.4069	−0.000034	2.50	−2.044	−3.447	4.608	−0.639
52	$aC{-}CH_n{-}O{-}$ (n in 1..2)	Benzyl ethyl ether (1)	0.8607	−0.0425	0.2698	−0.000417	−7.49	6.043	5.486	*****	0.969
53	$aC{-}CH_n{-}OH$ (n in 1..2)	Benzyl alcohol (1)	0.8981	0.1005	−1.0107	0.002944	−0.25	*****	*****	*****	−2.754
54	$aC{-}CH_n{-}CN$ (n in 1..2)	Benzyl cyanide (1)	0.1088	1.0587	2.4950	−0.000796	−11.01	25.157	16.950	*****	*****
55	$aC{-}CH_n{-}CHO$ (n in 1..2)	Phenyl acetaldehyde (1)	1.9470	−0.0177	*****	*****	*****	*****	*****	*****	*****
56	$aC{-}CH_n{-}SH$ (n in 1..2)	Phenyl methanethiol (1)	1.2057	0.1702	0.8705	0.000183	2.00	16.725	7.568	*****	0.890
57	$aC{-}CH_n{-}COOH$ (n in 1..2)	Phenyl acetic acid (1)	0.3666	0.1584	*****	*****	*****	*****	*****	*****	−4.086
58	$aC{-}CH_n{-}CO{-}$ (n in 1..2)	Phenyl acetone (1)	−0.2363	0.3094	*****	*****	*****	*****	*****	*****	*****
59	$aC{-}CH_n{-}S{-}$ (n in 1..2)	Benzyl methyl sulfide (1)	0.4506	0.1030	*****	*****	*****	*****	*****	*****	*****
60	$aC{-}CH_n{-}OOC{-}H$ (n in 1..2)	Benzyl formate (1)	*****	0.2238	1.7860	0.004195	−3.40	3.020	4.145	*****	*****
61	$aC{-}CH_m{-}NO_2$ (n in 1..2)	Phenyl nitromethane (1)	*****	0.5390	*****	*****	*****	*****	*****	*****	*****
62	$aC{-}CH_n{-}CONH_2$ (n in 1..2)	Phenyl ethanamide (1)	2.2421	−0.2197	*****	*****	*****	*****	*****	*****	*****
63	$aC{-}CH_n{-}OOC$ (n in 1..2)	Benzyl acetate (1)	−0.6997	0.0886	1.1629	−0.000384	−7.02	1.556	4.066	*****	*****
64	$aC{-}CH_n{-}COO$ (n in 1..2)	Methyl phenyl acetate (1)	−0.2636	0.0352	*****	*****	*****	*****	*****	*****	*****
65	$aC{-}SO_2{-}OH$	Benzenesulfonic acid (1)	−1.1057	*****	*****	*****	*****	*****	*****	*****	*****
66	$aC{-}CH(CH_3)_2$	Cumene (1)	0.0642	0.0196	0.1565	−0.001446	−2.04	1.238	−0.751	1.030	−0.270
67	$aC{-}C(CH_3)_3$	*tert*-Butylbenzene (1)	0.0790	0.0494	0.8016	−0.006495	−5.70	0.354	−0.192	*****	−0.878
68	$aC{-}CF_3$	Perfluorotoluene (1)	−10.8058	−1.5974	*****	*****	*****	*****	*****	*****	*****
69	$(CH_n{=}C)$(cyclic)–CHO (n in 0..2)	Furfural (1)	−1.0516	0.4267	2.4070	−0.002650	0.39	−6.438	−12.517	*****	−1.670
70	$(CH_n{=}C)_{cyc}{-}COO{-}CH_m$ (n, m in 0..3)	Methyl furanyrate (1)	−6.9427	0.0879	*****	*****	*****	*****	*****	*****	*****
71	$(CH_n{=}C)_{cyc}{-}CO{-}$ (n in 0..2)	2-Acetylfuran (1)	0.6572	0.6115	*****	*****	*****	*****	*****	*****	*****
72	$(CH_n{=}C)_{cyc}{-}CH_3$ (n in 0..2)	1,2-Dimethylcyclopentene (2)	0.0416	0.0173	−0.2509	−0.000624	0.03	28.972	24.560	*****	2.235
73	$(CH_n{=}C)_{cyc}{-}CH_2$ (n in 0..2)	2-Ethylfuran (1)	−0.3151	−0.0504	−1.1019	0.003921	−4.43	−22.533	−12.044	*****	0.961
74	$(CH_n{=}C)_{cyc}{-}CN$ (n in 0..2)	3-Cyanofuran (1)	1.5819	−0.2474	*****	*****	*****	*****	*****	*****	*****
75	$(CH_n{=}C)_{cyc}{-}Cl$ (n in 0..2)	2-Chlorofuran (1)	−0.8604	−0.5736	*****	*****	*****	*****	*****	*****	*****
76	$CH_{cyc}{-}CH_3$	Methylcyclopentane (1)	−0.1326	−0.1210	−0.1233	0.000779	2.79	4.178	4.452	0.096	0.033
77	$CH_{cyc}{-}CH_2$	Ethylcyclohexane (1)	−0.4669	−0.0148	0.3816	0.001694	−2.95	5.332	4.428	−0.428	−1.137
78	$Ch_{cyc}{-}CH$	Isopropylcyclopentane (1)	−0.3548	0.1395	0.1093	0.000124	6.19	6.084	−4.128	0.153	2.421
79	$Ch_{cyc}{-}C$	*tert*-Butylcyclohexane (1)	−0.1727	0.1829	*****	*****	*****	*****	*****	*****	*****
80	$Ch_{cyc}{-}CH{=}CH_n$ (n in 1..2)	Vinylcyclopentane (1)	0.6817	−0.1192	*****	*****	*****	*****	*****	*****	*****
81	$Ch_{cyc}{-}C{=}CH_n$ (n in 1..2)	Limonene (1)	−1.0631	−0.0455	−0.2832	0.002114	−16.97	6.768	10.390	*****	*****
82	$Ch_{cyc}{-}Cl$	Chloro cyclopentane (1)	0.5124	0.2667	*****	*****	*****	*****	*****	*****	*****
83	$Ch_{cyc}{-}F$	Fluoro cyclohexane (1)	2.8497	−0.1899	*****	*****	*****	*****	*****	*****	*****
84	$Ch_{cyc}{-}OH$	Cyclohexanol (1)	1.3691	−0.3179	0.8973	0.004640	−7.73	−3.024	−8.050	2.134	*****
85	$Ch_{cyc}{-}NH_2$	Cyclohexylamine (1)	1.5069	−0.3576	−0.9610	0.000039	−2.50	2.046	3.446	−4.607	0.328
86	$Ch_{cyc}{-}NH{-}CH_n$ (n in 0..3)	*N*-methylcyclohexylamine (1)	0.0370	−0.7458	−2.0833	−0.014535	−51.50	−11.965	14.531	*****	0.402
87	$Ch_{cyc}{-}N{-}CH_n$ (n in 0..3)	*N*,*N*-dimethylcyclohexanamine (1)	*****	0.1218	*****	*****	*****	*****	*****	*****	*****
88	$Ch_{cyc}{-}SH$	Cyclohexanethiol (1)	−0.3312	−0.0569	−0.6447	−0.000199	−2.00	−16.723	−7.569	*****	−0.878
89	$Ch_{cyc}{-}CN$	Cyanocyclopentane (1)	*****	0.4649	*****	*****	*****	*****	*****	*****	*****
90	$Ch_{cyc}{-}COOH$	Cyclopropanecarboxylic acid (1)	−2.0822	0.1506	*****	*****	*****	*****	*****	*****	*****
91	$Ch_{cyc}{-}CO$	Methyl cyclohexyl ketone (1)	0.7743	0.1300	*****	*****	*****	*****	*****	−0.616	*****
92	$Ch_{cyc}{-}NO_2$	Nitrocyclohexane (1)	−0.8578	0.6540	*****	*****	*****	*****	*****	*****	*****
93	$Ch_{cyc}{-}S{-}$	Methyl cyclopentyl sulfide (1)	−0.8638	0.0043	*****	*****	*****	*****	*****	*****	*****
94	$Ch_{cyc}{-}CHO$	Cyclohexanecarboxaldehyde (1)	0.5076	−0.2692	*****	*****	*****	*****	*****	*****	*****
95	$Ch_{cyc}{-}O{-}$	Methoxycyclohexane (1)	−0.3978	−0.2787	*****	*****	*****	*****	*****	*****	*****

Table A.2 (contd.)

	Group	Example	T_{m2j}	T_{b2j}	T_{c2j}	P_{c2j}	V_{c2j}	G_{f2j}	H_{f2j}	H_{v2j}	H_{fus2j}
96	Ch_{cyc}–OOCH	Cyclohexyl ester formic acid (1)	*****	−0.2107	*****	*****	*****	*****	*****	*****	*****
97	Ch_{cyc}–COO	Ethyl cyclobutyrate (1)	*****	0.0926	*****	*****	*****	*****	*****	*****	*****
98	Ch_{cyc}–OOC	Cyclohexyl acetate (1)	−0.4666	−0.4495	−0.3450	−0.000692	−12.03	4.358	−15.751	*****	*****
99	C_{cyc}–CH_3	1,1-Dimethyl-cyclohexane (2)	0.1737	0.0722	0.1607	0.001235	1.95	0.107	0.238	0.808	−1.237
100	C_{cyc}–CH_2	1-Ethyl-1-methyl-cyclopentane (1)	−1.9233	0.0319	0.1090	−0.000610	−5.17	18.755	21.498	0.585	*****
101	C_{cyc}–OH	1-Methylcyclopentanol (1)	0.7334	−0.6775	−2.1303	−0.004683	−14.40	−18.970	−21.975	*****	0.235
102	>N_{cyc}–CH_3	*N*-methyl-2-pyrrolidone (1)	−0.0383	0.0604	−0.0003	0.000058	*****	*****	*****	*****	*****
103	>N_{cyc}–CH_2	*N*-ethylpyrrole (1)	1.0497	−0.3080	*****	*****	*****	*****	*****	*****	*****
104	AROMRINGs^{1}s^2	2-Methyl-phenol (1), 2-Et-toluene (1)	−0.6388	−0.1590	−0.3161	0.000522	2.86	1.577	1.486	1.164	−1.470
105	AROMRINGs^{1}s^3	3-Methyl-phenol (1), 3-Et-toluene (1)	−0.6218	0.0217	−0.0693	0.001790	6.54	−1.037	0.294	−1.910	−1.059
106	AROMRINGs^{1}s^4	4-Methyl-phenol (1), 4-Et-toluene (1)	0.9840	0.1007	0.0803	0.000467	3.70	−0.709	0.384	0.331	1.244
107	AROMRINGs^{1}s^{2}s^3	1,2,3-Trimethylbenzene (1)	−0.2762	−0.1647	1.0088	−0.005598	−9.58	7.731	5.743	1.433	0.473
108	AROMRINGs^{1}s^{2}s^4	1,2,4-Trihydroxybenzene (1)	−0.3689	−0.1387	0.0908	0.000255	−2.05	−2.767	−0.449	0.313	−0.302
109	AROMRINGs^{1}s^{3}s^5	3,5-Diethyltoluene (1)	−0.3841	−0.1314	−0.6412	0.004090	−7.67	−2.148	−7.538	−0.117	−2.530
110	AROMRINGs^{1}s^{2}s^{3}s^4	3-Ethyl-1,2,4-trimethylbenzene (1)	1.7722	0.2745	2.1116	−0.007612	−7.04	14.226	12.710	*****	−1.736
111	AROMRINGs^{1}s^{2}s^{3}s^5	1,2,3,5-Tetramethylbenzene (1)	0.4553	0.1645	0.9353	−0.001811	−0.04	4.926	5.220	*****	−2.246
112	AROMRINGs^{1}s^{2}s^{4}s^5	1,2,4,5-Tetramethylbenzene (1)	2.0561	0.0754	0.6241	−0.000500	−0.04	−0.474	−1.340	*****	8.034
113	PYRIDINEs^2	2-Methylpyridine (1)	−0.5769	−0.1196	−1.0256	0.007006	8.68	−9.713	−9.644	−1.683	−0.786
114	PYRIDINEs^3	3-Methylpyridine (1)	−0.2556	0.0494	0.5784	0.007006	8.68	−2.523	−2.446	0.277	3.671
115	PYRIDINEs^4	4-Methylpyridine (1)	1.6282	0.1344	0.6595	0.001283	14.28	−4.703	−6.466	0.397	5.975
116	PYRIDINEs^{2}s^3	2,3-Dimethylpyridine (1)	−0.1341	0.0032	*****	*****	*****	*****	*****	−0.939	*****
117	PYRIDINEs^{2}s^4	2,4-Dimethylpyridine (1)	−1.6848	−0.0817	*****	*****	*****	*****	*****	−1.269	*****
118	PYRIDINEs^{2}s^5	2,5-Dimethylpyridine (1)	−0.9802	−0.1564	*****	*****	*****	*****	*****	−1.719	*****
119	PYRIDINEs^{2}s^6	2,6-Dimethylpyridine (1)	0.3018	−0.5176	−2.2773	0.008029	−50.26	−16.570	−17.778	−3.419	−1.487
120	PYRIDINEs^{3}s^4	3,4-Dimethylpyridine (1)	0.1018	0.5477	*****	*****	*****	*****	*****	1.742	*****
121	PYRIDINEs^{3}s^5	3,5-Dimethylpyridine (1)	0.2811	0.3533	*****	*****	*****	*****	*****	0.572	*****
122	PYRIDINEs^{2}s^{3}s^6	2,3,6-Trimethylpyridine (1)	−0.3189	−0.3888	*****	*****	*****	*****	*****	−2.744	*****

Table A.3: Third-order group contribution data (Marrero & Gani, 2001).

	Group	Example	T_{m3k}	T_{b3k}	T_{c3k}	P_{c3k}	V_{c3k}	G_{f3k}	H_{f3k}	H_{tas3k}
1	HOOC–$(CH_n)_m$–COOH ($m > 2$, n in 0..2)	1,5-Pentanedioic acid (1)	−1.5257	1.6498	−1.6986	0.001544	−3.72	−4.708	−6.572	−7.583
2	NH_n–$(CH_n)_m$–COOH ($m > 2$, n in 0..2)	4-Aminobutyric acid (1)	11.2271	*****	*****	*****	*****	*****	*****	*****
3	NH_2–$(CH_n)_m$–OH ($m > 2$, n in 0..2)	4-Aminobutanol (1)	0.7732	1.0750	0.4950	0.000728	−23.74	3.079	4.171	−4.840
4	OH–$(CH_n)_m$–OH ($m > 2$, n in 0..2)	1,9-Nonanediol (1)	0.6674	0.7193	0.1725	−0.000327	−0.84	7.536	5.411	−0.272
5	OH–$(CH_p)_k$–O–$(CH_n)_m$–OH ($m, k > 0$; p, n in 0..2)	Dipropylene glycol (1)	−0.1073	1.1867	6.6872	0.001937	1.44	−8.397	−8.651	1.661
6	OH–$(CH_p)_k$–S–$(CH_n)_m$–OH ($m, k > 0$; p, n in 0..2)	2,2′-Diethyl-dihydroxy sulfide (1)	−1.3891	*****	2.6769	0.003792	−1.62	10.194	8.164	−3.479
7	OH–$(CH_p)_k$–NHx–$(CH_n)_m$–OH ($m, k > 0$; p, n, x in 0..2)	Diethanolamine (1)	−0.0781	0.2991	*****	0.003254	−0.69	1.662	1.753	0.301
8	CH_p–O–$(CH_n)_m$–OH ($m > 2$; n, p in 0..2)	Butoxypropanol (1)	*****	−0.4605	*****	*****	*****	*****	*****	*****
9	NH_2–$(CH_n)_m$–NH_2 ($m > 2$; n in 0..2)	1,5-Diaminopentane (1)	−0.0604	0.0060	−4.3195	0.006734	6.69	4.100	0.371	5.666
10	NH_k–$(CH_n)_m$–NH_2 ($m > 2$; k in 0..1; n in 0..2)	*N,N*-dimethylpropylenediamine (1)	−1.1888	−0.1819	*****	*****	*****	*****	*****	*****
11	SH–$(CH_n)_m$–SH ($m > 2$; n in 0..2)	1,5-Pentanedithiol (1)	0.6669	0.4516	*****	*****	*****	*****	*****	*****
12	NC–$(CH_n)_m$–CN ($m > 2$)	Glutaronitrile (1)	−0.3798	1.3440	0.0834	−0.011090	−36.89	−7.035	7.782	−0.607
13	COO–$(CH_n)_m$–OOC ($m > 2$; n in 0..2)	Glyceryl tridodecanoate (1)	−2.6542	*****	*****	*****	*****	*****	*****	*****
14	aC–$(CH_n{=}CH_m)_{cyc}$ (fused rings) (n, m in 0..1)	Indene (1), Acenaphtylene (2)	0.2479	−0.3741	−0.0185	0.000851	−8.87	−1.601	2.689	−2.703
15	aC–aC (different rings)	Biphenylene (2), Biphenyl (1)	1.1395	−0.4961	6.1894	−0.040100	−26.26	−4.459	−4.558	−0.385
16	aC–$CH_{n,cyc}$ (different rings) (n in 0..1)	Cyclohexylbenzene (1)	0.0570	−0.4574	−0.2474	−0.005826	−8.55	−5.267	−5.914	−0.442
17	aC–$CH_{n,cyc}$ (fused rings) (n in 0..1)	Tetralin (2), Indane (2)	−0.5640	−0.1736	0.5060	−0.003746	−11.56	−4.203	−4.863	−0.143
18	aC–$(CH_n)_m$–aC (different rings) ($m > 1$; n in 0..2)	Bibenzyl (1)	1.9902	0.3138	3.0321	0.003007	9.73	1.318	0.084	5.377
19	aC–$(CH_n)_m$–CH_{cyc} (different rings) ($m > 0$; n in 0..2)	1-Cyclopentyl-3-phenylpropane (1)	*****	0.5928	*****	*****	*****	*****	*****	*****
20	CH_{cyc}–Ch_{cyc} (different rings)	Cyclohexylcyclohexane (1)	0.5460	0.4387	2.1761	0.002745	7.72	−67.517	−66.870	*****
21	CH_{cyc}–$(CH_n)_m$–CH_{cyc} (different rings) ($m > 0$; n in 0..2)	1,2-Dicyclohexylethane (1)	0.4497	0.5632	*****	*****	*****	*****	*****	*****
22	CH multiring	Hexahydroindan (2), Decalin (2)	0.6647	0.1415	0.4963	−0.000985	−3.33	*****	*****	0.223
23	C multiring	Spiropentane (1)	0.0792	*****	*****	*****	*****	*****	*****	*****
24	aC–CH_m–aC (different rings) (m in 0..2)	Diphenylmethane (1)	0.6457	0.2391	0.1174	−0.002673	−4.67	−0.729	0.866	−0.958
25	aC–$(CH_m{=}CH_n)$–aC (different rings) (m, n in 0..2)	1,2-Diphenylethylene (1)	0.9608	0.7192	0.7039	−0.004661	14.31	−0.702	−2.291	3.275
26	$(CH_m{=}C)_{cyc}$–CH=CH–$(C{=}CH_n)_{cyc}$ (different rings)	1,2-Furanyl ethene (1)	16.2235	*****	*****	*****	*****	*****	*****	*****
27	$(CH_m{=}C)_{cyc}$–CH_p–$(C{=}CH_n)_{cyc}$ (different rings)	Difuranyl methane (1)	16.8558	*****	*****	*****	*****	*****	*****	*****
28	aC–CO–aC (different rings)	Benzophenone (1)	−1.0394	1.0171	−0.2678	−0.001837	−7.05	11.125	7.108	−4.091
29	aC–CH_m–CO–aC (different rings) (m in 0..2)	Benzyl phenone (1)	−0.4486	0.9674	*****	*****	*****	*****	*****	*****
30	aC–CO–$(C{=}CH_n)_{cyc}$ (different rings) (n in 0..1)	Phenyl-2-furanyl-methanone (1)	−0.1376	0.1126	*****	*****	*****	*****	*****	*****
31	aC–CO–CO–aC (different rings)	Diphenylethanedione (1)	0.4361	0.9317	*****	*****	*****	*****	*****	−3.687
32	aC–CO_{cyc} (fused rings)	Phenolphthalein (1)	3.6847	0.5031	*****	*****	*****	*****	*****	2.047
33	aC–CO–$(CH_n)_m$–CO–aC (different rings) ($m > 0$; n in 0..2)	1,4-Diphenyl-1,4-butanedione (1)	4.9038	*****	*****	*****	*****	*****	*****	7.327
34	aC–CO–$CH_{n,cyc}$ (different rings) (n in 0..1)	Cyclohexyl phenyl methanone (1)	−7.0038	*****	*****	*****	*****	*****	*****	*****
35	aC–CO–NH_n–aC (different rings) (n in 0..1)	*N*-phenyl benzamide (1)	5.9653	*****	*****	*****	*****	*****	*****	2.510
36	aC–NH_n $CONH_m$–aC (different rings) (n, m in 0..1)	*N,N′*-diphenylurea (1)	1.5629	*****	*****	*****	*****	*****	*****	0.018
37	aC–CO–N_{cyc} (different rings)	*N*-phenonyl-piperidine (1)	−9.1856	*****	*****	*****	*****	*****	*****	*****
38	aC–S_{cyc} (fused rings)	Dibenzothiophene (2)	0.2612	0.2242	3.5541	0.004600	12.60	8.333	9.212	−0.784
39	aC–S–aC (different rings)	Diphenyl sulfide (1)	−1.8403	0.0185	*****	*****	*****	*****	*****	*****
40	aC–PO_n–aC (different rings) (n in 0..4)	Triphenylphosphate (3)	0.0393	*****	*****	*****	*****	*****	*****	*****
41	aC–SO_n–aC (different rings) (n in 1..4)	Diphenyl sulfone (1)	0.9514	−0.0850	*****	*****	*****	*****	*****	−2.485
42	aC–$NH_{n,cyc}$ (fused rings) (n in 0..1)	Carbazole (2)	3.4983	1.1457	3.5541	0.017201	0.44	−2.221	−16.080	0.196
43	aC–NH–aC (different rings)	Diphenylamine (1)	−0.3048	0.5768	0.9519	0.008484	1.42	−0.596	−1.994	1.934
44	aC–(C=N)$_{cyc}$ (different rings)	Phenyl-3-pyrazole (1)	−1.3060	−0.5335	*****	*****	*****	*****	*****	*****
45	aC–(N=CH$_n$)$_{cyc}$ (fused rings) (n in 0..1)	Benzoxazole (1)	−4.9289	−5.2736	*****	*****	*****	*****	*****	−0.599
46	aC–$(CH_n{=}N)_{cyc}$ (fused rings) (n in 0..1)	Benzoisoxazole (1)	−10.1007	*****	*****	*****	*****	*****	*****	*****

Table A.3 (contd.)

	Group	Example	T_{m3k}	T_{h3k}	T_{c3k}	P_{c3k}	V_{c3k}	G_{f3k}	H_{f3k}	H_{fus3k}
47	aC–O–CH_n–aC (different rings) (n in 0..2)	Benzyl phenyl ether (1)	1.0834	0.6571	*****	*****	*****	*****	*****	*****
48	aC–O–aC (different rings)	Diphenyl ether (1)	−0.4803	−0.8252	−0.9785	0.001162	−2.63	2.668	−5.074	1.193
49	aC–CH_n–O–CH_m–aC (different rings) (n, m in 0..2)	Benzyl ether (1)	−3.2676	0.2790	−1.4002	−0.004716	28.42	−4.229	−2.303	−3.971
50	aC–O_{cyc} (fused rings)	Benzoxazole (1)	−0.3545	−0.6848	*****	*****	*****	*****	*****	−1.153
51	AROMFUSED[2]	Naphthalene (2)	0.2825	0.0441	−1.0095	−0.001332	−6.88	1.993	1.904	0.694
52	AROMFUSED[2]s^1	1-Methylnaphtalene (1)	−1.2836	−0.1666	0.1605	−0.002030	−3.17	−2.940	−2.274	−3.699
53	AROMFUSED[2]s^2	2,7-Dimethylnaphtalene (2)	0.3378	−0.2692	−0.6765	−0.002436	−3.85	−1.873	−1.316	2.037
54	AROMFUSED[2]s^2s^3	2,3-Dimethylnaphtalene (1)	1.8941	−0.2807	*****	*****	*****	*****	*****	2.150
55	AROMFUSED[2]s^1s^4	1,4-Dimethylnaphtalene (1)	−2.7585	−0.3294	*****	*****	*****	*****	*****	*****
56	AROMFUSED[2]s^1s^2	1,2-Dimethylnaphtalene (1)	−3.0362	−0.2931	*****	*****	*****	*****	*****	*****
57	AROMFUSED[2]s^1s^3	1,3-Dimethylnaphtalene (1)	−3.2228	−0.3360	*****	*****	*****	*****	*****	*****
58	AROMFUSED[3]	Phenalene (3), Pyrene (2)	1.6600	0.0402	−1.0430	0.004695	35.21	3.896	5.819	1.176
59	AROMFUSED[4a]	Anthracene (1)	7.0402	1.0466	3.3011	0.015244	−6.96	13.843	11.387	5.027
60	AROMFUSED[4a]s^1	9-Methylanthracene (1)	−3.3463	−7.8521	*****	*****	*****	*****	*****	*****
61	AROMFUSED[4a]s^1s^4	9,10-Dimethylanthracene (1)	6.8373	*****	*****	*****	*****	*****	*****	*****
62	AROMFUSED[4p]	Phenanthrene (1), Pyrene (2)	−1.5856	0.9126	2.8885	0.007280	−24.02	−16.040	−19.089	−3.417
63	AROMFUSED[4p]s^3s^4	9,10-Dimethylphenanthrene (1)	2.0821	*****	*****	*****	*****	*****	*****	*****
64	PYRIDINE.FUSED[2]	Quinoline (1)	−4.4725	−0.9432	1.1251	−0.005369	63.29	8.688	13.586	−4.967
65	PYRIDINE.FUSED[2–iso]	Isoquinoline (1)	−2.5898	−0.5844	3.9241	−0.011207	−2.71	−5.112	−0.314	−2.587
66	PYRIDINE.FUSED[4]	Acridine (1)	1.0358	0.1733	7.7134	−0.001275	−12.04	20.073	15.786	−1.365

Table A.4: Property model for each property (Marrero & Gani, 2001)

Property (X)	Left-hand side of Eq. (1) (function $f(X)$)	Right-hand side of Eq. (1) (group-contribution terms)
Normal melting point (T_m)	$\exp(T_m/T_{m0})$	$\sum_i N_i T_{m1i} + \sum_j M_j T_{m2j} + \sum_k O_k T_{m3k}$
Normal boiling point (T_b)	$\exp(T_b/T_{b0})$	$\sum_i N_i T_{b1i} + \sum_j M_j T_{b2j} + \sum_k O_k T_{b3k}$
Critical temperature (T_c)	$\exp(T_c/T_{c0})$	$\sum_i N_i T_{c1i} + \sum_j M_j T_{c2j} + \sum_k O_k T_{c3k}$
Critical pressure (P_c)	$(P_c - P_{c1})^{-0.5} - P_{c2}$	$\sum_i N_i P_{c1i} + \sum_j M_j P_{c2j} + \sum_k O_k P_{c3k}$
Critical volume (V_c)	$V_c - V_{c0}$	$\sum_i N_i V_{c1i} + \sum_j M_j V_{c2j} + \sum_k O_k V_{c3k}$
Standard Gibbs energy at 298 K (G_f)	$G_f - G_{f0}$	$\sum_i N_i G_{f1i} + \sum_j M_j G_{f2j} + \sum_k O_k G_{f3k}$
Standard enthalpy of formation at 298 K (H_f)	$H_f - H_{f0}$	$\sum_i N_i H_{f1i} + \sum_j M_j H_{f2j} + \sum_k O_k H_{f3k}$
Standard enthalpy of vaporization at 298 K (H_v)	$H_v - H_{v0}$	$\sum_i N_i H_{v1i} + \sum_j M_j H_{v2j}$
Standard enthalpy of fusion (H_{fus})	$H_{fus} - H_{fus0}$	$\sum_i N_i H_{fus1i} + \sum_j M_j H_{fus2j} + \sum_k O_k H_{fus3k}$

Table A.5: Value of Adjustable Parameters (Marrero & Gani, 2001)

Adjustable parameter (universal constants)	Value
T_{m0}	147.450 K
T_{b0}	222.543 K
T_c	231.239 K
P_{c1}	5.9827 bar
P_{c2}	0.108998 $bar^{-0.5}$
V_{c0}	7.95 cm^3/mol
G_{f0}	−34.967 kJ/mol
H_{f0}	5.549 kJ/mol
H_{v0}	11.733 kJ/mol
H_{fus0}	−2.806 kJ/mol

Table A.6: First-order group contribution data (Marrero & Gani, 2002)

	group	example	log K_{ow}	SE[a,b]	log W_s	SE[b,c]
1	CH_3	*n*-tetracontane (2)	0.257 78	0.117 47	−5.944 17	1.635 13
2	CH_2	*n*-tetracontane (38)	0.450 05	0.056 58	−5.779 18	1.579 37
3	CH	2-methylpentane (1)	0.465 31	0.131 16	−5.549 66	1.521 57
4	C	2,2-dimethylbutane (1)	0.748 06	0.194 44	−5.162 26	1.461 49
5	CH_2═CH	1-hexene (1)	0.511 34	0.172 70	−10.859 55	2.193 09
6	CH═CH	2-hexene (1)	0.758 03	0.178 04	−10.788 85	2.151 80
7	CH_2═C	2-methyl-1-butene (1)	0.733 37	0.233 79	−10.499 65	2.151 80
8	CH═C	2-methyl-2-butene (1)	0.701 01	0.188 91	−10.624 20	2.109 79
9	C═C	2,3-dimethyl-2-butene (1)	0.796 44	0.339 70	−9.600 40	2.066 75
10	CH≡C	1-pentyne (1)	−0.320 01	0.195 46	−10.281 21	2.109 81
11	C≡C	3-decyne (1)	0.430 81	0.239 51	−12.354 65	2.066 90
12	aCH	benzene (6)	0.216 94	0.070 64	−5.190 40	1.521 57
13	aC fused with aromatic ring	naphthalene (2)	0.364 01	0.089 33	−5.230 00	1.461 47
14	aC fused with non-aromatic subring	indane (2)	0.339 82	0.084 12	−5.229 94	1.461 47
15	aC except as above	benzophenone (1)	0.331 52	0.117 74	−5.198 05	1.461 44
16	aN in aromatic ring	pyridine (1)	−0.498 33	0.098 41	−5.070 59	1.578 23
17	aC−CH_3	toluene (1)	0.642 29	0.107 05	−11.011 06	2.193 07
18	aC−CH_2	ethylbenzene (1)	0.565 09	0.123 07	−10.631 42	2.151 80
19	aC−CH	cumene (1)	0.754 38	0.171 86	−10.360 32	2.109 78
20	aC−C	*t*-butbenzene (1)	0.963 77	0.205 85	−10.111 12	2.066 83
21	aC−CH═CH_2	styrene (1)	1.275 75	0.440 53	−15.795 01	2.635 36
22	aC−CH═CH	1-propenylbenzene (1)	0.936 95	0.187 66	−16.026 30	2.601 18
23	aC−C═CH_2	α-methylstyrene (1)	0.939 88	0.389 09	−15.966 77	2.601 16
24	aC−C≡CH	phenylacetylene (1)	0.902 34	0.579 30	−15.467 68	2.566 61
25	aC−C≡C	1-phenyl-1-propyne (1)	1.254 31	0.394 01	−15.535 57	2.531 35
26	OH	1,4-butanediol (2)	−1.096 58	0.120 49	−5.801 15	1.739 09
27	aC−OH	phenol (1)	−0.025 44	0.111 60	−11.010 26	2.271 35
28	COOH	pentanoic acid (2)	−0.883 14	0.148 47	−16.828 40	2.829 39
29	aC−COOH	benzoic acid (1)	0.140 76	0.162 31	−22.220 05	3.184 52
30	CH_3CO	2-butanone (1)	−0.480 75	0.172 92	−16.588 23	2.766 70
31	CH_2CO	3-pentanone (1)	−0.174 07	0.181 41	−17.117 42	2.734 11
32	CHCO	2,4-dimethyl-3-pentone (1)	0.204 53	0.248 97	−16.072 62	2.701 17
33	CCO	2,2,4,4-tetramethyl-3-pentanone (1)	0.256 51	0.284 97	−16.257 10	2.667 84
34	aC−CO	acetophenone (1)	−0.175 31	0.160 73	−16.029 85	2.667 72
35	CHO	1-hexanal (1)	−0.633 06	0.241 64	−11.007 55	2.271 39
36	aC−CHO	benzaldehyde (1)	−0.117 82	0.221 42	−16.121 27	2.701 16
37	CH_3COO	butyl acetate (1)	−0.490 06	0.179 31	−22.694 05	3.240 35
38	CH_2COO	methyl butyrate (1)	−0.317 51	0.155 32	−22.394 91	3.212 55
39	CHCOO	ethyl isobutyrate (1)	−0.522 03	0.213 19	−22.052 00	3.184 54
40	CCOO	ethyl 2,2-dimethyl propionate (1)	0.154 21	0.257 40	−22.388 09	3.156 26
41	HCOO	propyl formate (1)	−0.887 47	0.440 97	−17.013 00	2.829 40
42	aC−COO	methyl benzoate (1)	−0.036 86	0.148 80	−21.808 00	3.156 28
43	aC−OOCH	phenyl formate (1)	−0.367 66	0.579 30	–	–
44	aC−OOC	phenyl acetate (1)	−0.266 05	0.170 48	−21.647 34	3.156 29
45	COO except as above	ethyl acrylate (1)	−0.474 98	0.134 65	−16.929 62	2.797 53
46	CH_3O	methyl butyl ether (1)	−0.350 70	0.158 79	−11.374 75	2.349 22
47	CH_2O	di-*n*-butyl ether (1)	−0.123 97	0.144 88	−11.159 87	2.310 76
48	CH−O	*sec*-butyl ether (1)	−0.033 33	0.253 13	−11.380 73	2.271 69
49	C−O	*tert*-butyl ether (1)	0.923 07	0.392 67	−10.541 61	2.231 87
50	aC−O	methyl phenyl ether (1)	0.018 73	0.111 12	−10.941 66	2.231 83
51	CH_2NH_2	ethylamine (1)	−1.413 22	0.202 42	−10.732 52	2.311 65
52	$CHNH_2$	*sec*-butylamine (1)	−1.922 14	0.235 61	−11.885 81	2.272 54
53	CNH_2	*tert*-butylamine (1)	−1.139 49	0.443 77	−9.421 10	2.232 83
54	CH_3NH	dimethylamine (1)	−0.741 95	0.198 26	−10.616 30	2.311 59
55	CH_2NH	dipropylamine (1)	−0.983 92	0.182 90	−10.016 56	2.272 58
56	CHNH	diisopropylamine (1)	−0.377 16	0.239 33	−11.041 53	2.232 69
57	CH_3N	methyldiethylamine (1)	−0.460 04	0.157 60	−10.632 51	2.272 54
58	CH_2N	triethylamine (1)	−0.655 52	0.196 59	−10.075 54	2.232 72
59	aC−NH_2	aniline (1)	−0.292 10	0.117 44	−10.902 99	2.232 74
60	aC−NH	*N*-methyl aniline (1)	0.360 82	0.147 91	−11.056 61	2.192 24
61	aC−N	*N,N*-dimethyl aniline (1)	0.188 04	0.201 39	−10.547 09	2.150 95
62	NH_2 except as above	cyclobutylamine	−0.815 10	0.148 95	−5.799 13	1.688 00
63	CH═N	acetaldazine (2)	0.280 70	0.200 76	−11.164 41	2.192 26
64	C═N	ketazine (2)	0.804 52	0.210 80	−10.172 86	2.150 96
65	CH_2CN	propionitrile (1)	−0.573 04	0.221 52	−15.266 75	2.668 54
66	CHCN	isobutyronitrile (1)	−0.571 77	0.410 04	–	–
67	CCN	2,2-dimethylpropionitrile (1)	0.242 21	0.411 77	−14.998 22	2.600 60
68	aC−CN	benzonitrile (1)	−0.080 22	0.169 60	−15.295 57	2.600 44
69	CN except as above	acrylonitrile (1)	−0.101 22	0.175 30	−10.759 93	2.150 94
70	CH_2NO_2	1-nitropropane (1)	−0.233 57	0.327 86	−23.287 17	3.267 34
71	$CHNO_2$	2-nitropropane (1)	−0.172 17	0.371 04	−23.290 88	3.239 76
72	CNO_2	2-methyl-2-nitropropane (1)	−0.548 63	0.445 80	−22.782 89	3.212 14
73	aC−NO_2	nitrobenzene (1)	0.144 07	0.117 49	−23.073 88	3.212 00
74	NO_2 except as above	nitrocyclohexane (1)	−0.425 43	0.164 17	−19.017 16	2.860 26
75	ONO_2	*n*-butyl nitrate (1)	−0.096 41	0.270 78	−24.054 06	3.320 57
76	$HCONHCH_2$	ethylformamide (1)	−1.733 07	0.492 48	–	–
77	$CONH_2$	butyramide (1)	−1.444 27	0.166 11	−16.555 13	2.798 30
78	$CONHCH_3$	methylacetamide (1)	−0.456 69	0.184 98	−22.061 05	3.213 19
79	$CONHCH_2$	ethylacetamide (1)	−0.835 43	0.164 26	−22.698 44	3.185 21
80	$CON(CH_3)_2$	dimethylacetamide (1)	−0.365 98	0.209 32	−26.947 67	3.580 37
81	$CONCH_3CH_2$	methylethylacetamide (1)	−0.824 94	0.260 12	−26.936 05	3.555 28
82	$CON(CH_2)_2$	diethylacetamide (1)	−0.633 88	0.235 47	−26.457 30	3.529 93
83	CONHCO	diacetamide (1)	−2.110 29	0.410 43	−27.568 03	3.554 05
84	CONCO	methyldiacetamide	−0.976 42	0.579 86	−27.390 46	3.528 82
85	aC−$CONH_2$	benzamide	−1.045 12	0.186 62	−21.938 52	3.156 94
86	aC−NH(CO)H	*N*-phenylformamide (1)	−0.405 50	0.305 96	−21.179 41	3.156 94

Table A.6 (contd.)

	group	example	log K_{ow}	SE[a,b]	log W_s	SE[b,c]
87	aC−N(CO)H	*N*-methyl-*N*-phenylmethanamide (1)	−0.630 84	0.410 63	–	–
88	aC−CONH	*N*-methylbenzamide (1)	−0.609 90	0.163 52	−21.057 13	3.128 39
89	aC−NHCO	*N*-(2-methylphenyl)acetamide (1)	−0.340 63	0.147 21	−21.596 12	3.128 39
90	aC−NCO	phenylmethylacetamide (1)	−0.980 99	0.211 34	−20.466 46	3.099 64
91	NHCONH	*N,N'*-dimethylurea (1)	−1.122 44	0.244 59	−22.432 38	3.212 64
92	NH_2CONH	methylurea (1)	−1.275 46	0.262 43	−22.333 73	3.240 44
93	NH_2CON	*N,N*-dimethylurea (1)	−0.209 20	0.487 40	–	–
94	NHCON	trimethylurea (1)	−0.638 20	0.328 80	−21.260 28	3.184 66
95	NCON	tetramethylurea (1)	−1.626 44	0.402 40	−20.104 75	3.156 27
96	$aC-NHCONH_2$	phenylurea (1)	−0.580 55	0.255 50	−26.396 89	3.554 74
97	aC−NHCONH	*N,N'*-diphenylurea	−0.197 45	0.204 09	−27.798 91	3.529 44
98	NHCO except as above	*N*-chloroacetamide (1)	−1.115 98	0.127 63	−16.321 82	2.766 03
99	CH_2Cl	1-chlorobutane (1)	0.516 13	0.170 28	−19.582 79	2.966 29
100	CHCl	2-chloropropane (1)	0.845 87	0.312 32	−19.195 05	2.935 89
101	CCl	2-chloro-2-methylpropane (1)	0.753 63	0.579 60	−19.666 88	2.905 36
102	$CHCl_2$	1,1-dichloroethane (1)	0.914 76	0.198 25	−32.937 20	3.863 20
103	CCl_2	2,2-dichloropropane (1)	0.993 98	0.492 47	−34.850 37	3.839 99
104	CCl_3	1,1,1-trichloroethane	1.829 67	0.260 16	−46.917 41	4.587 95
105	CH_2F	1-fluorobutane (1)	−0.022 95	0.274 33	−12.336 36	2.423 43
106	CHF	2-fluorobutane (1)	−0.085 13	0.389 96	−11.900 28	2.386 28
107	CHF_2	1,1-difluoroethane (1)	1.449 61	0.281 25	−20.438 12	3.011 93
108	CF_2	perfluorohexane (4)	0.671 10	0.253 55	−19.649 58	2.982 04
109	CF_3	hexafluoroethane (2)	0.874 69	0.183 87	−27.581 64	3.503 00
110	CCl_2F	tetrachloro-1,2-di.F.ethane (2)	1.599 79	0.447 19	−40.787 52	4.257 17
111	HCClF	1-Cl-1,2,2,2-tetrafluoroethane (1)	0.683 61	0.443 77	−26.712 01	3.463 78
112	$CClF_2$	1,2-dichlorotetrafluoroethane (2)	1.105 07	0.376 37	−34.467 72	3.898 37
113	aC−Cl	chlorobenzene (1)	0.896 34	0.080 76	−19.261 90	2.905 24
114	aC−F	hexafluorobenzene (6)	0.360 42	0.125 79	−12.408 22	2.348 29
115	aC−I	iodobenzene (1)	0.820 10	0.196 45	−54.387 25	4.970 21
116	aC−Br	bromobenzene (1)	1.019 63	0.136 28	−36.692 13	4.042 90
117	I− except as above	iodoethane (1)	0.692 62	0.217 88	−49.757 26	4.750 48
118	Br− except as above	bromoethane (1)	0.399 37	0.157 68	−31.182 31	3.769 50
119	F− except as above	benzyl fluoride (1)	−0.247 92	0.127 43	−7.325 22	1.838 06
120	Cl− except as above	ethyl chloroacetate (1)	0.268 73	0.120 42	−14.059 08	2.510 88
121	CHNOH	propionaldehyde oxime (1)	−0.025 78	0.294 42	−16.641 97	2.798 32
122	CNOH	diethyl ketoxime (1)	−0.495 55	0.313 64	−15.874 41	2.766 00
123	aC−CHNOH	phenyl oxime (1)	−0.018 55	0.306 73	–	–
124	OCH_2CH_2OH	2-ethoxyethanol (1)	−1.135 97	0.256 62	−22.335 94	3.295 24
125	$OCHCH_2OH$	2-ethoxy-1-propanol (1)	−0.665 51	0.413 53	−22.578 29	3.267 98
126	OCH_2CHOH	1-methoxy-2-propanol (1)	−0.706 10	0.200 23	−22.135 91	3.267 90
127	CH_2SH	ethanethiol (1)	0.379 63	0.358 56	−18.650 83	2.894 09
128	CHSH	2-propanethiol (1)	−0.087 48	0.481 73	–	–
129	CSH	2-methyl-2-propanethiol (1)	−0.033 25	0.583 42	−17.687 75	2.831 37
130	aC−SH	benzenethiol (1)	0.546 92	0.357 04	−17.397 26	2.831 48
131	−SH (except as above)	cyclohexanethiol (1)	0.640 78	0.359 32	−12.757 72	2.425 00
132	CH_3S	dimethyl sulfide (1)	0.511 21	0.219 44	−18.670 32	2.894 06
133	CH_2S	diethyl sulfide (1)	0.484 69	0.190 93	−18.745 61	2.862 93
134	CHS	diisopropyl sulfide (1)	0.584 17	0.488 36	–	–
135	CS	di-*tert*-butyl sulfide (1)	1.710 29	0.393 43	−19.277 35	2.799 53
136	aC−S−	phenyl methyl sulfide (1)	0.608 25	0.178 74	−17.689 79	2.799 59
137	SO	dimethyl sulfoxide (1)	−0.940 94	0.271 06	−17.208 89	2.923 52
138	SO_2	dimethyl sulfone (1)	−0.660 57	0.191 99	−24.891 05	3.375 26
139	SO_3 (sulfite)	dimethyl sulfite (1)	0.353 73	0.581 86	−31.686 90	3.773 18
140	SO_3 (sulfonate)	dimethyl sulfonate (1)	−0.960 34	0.357 00	−31.277 66	3.773 30
141	SO_4 (sulfate)	dimethyl sulfate (1)	−0.818 64	0.579 32	−36.953 97	4.133 18
142	aC−SO	phenyl methyl sulfoxide (1)	−0.794 49	0.265 21	–	–
143	$aC-SO_2$	diphenyl sulfone (1)	−0.474 35	0.147 65	−29.520 62	3.678 04
144	PO_3 (phosphite)	triethyl phosphite (1)	−1.756 64	0.579 91	–	–
145	PHO_3 (phosphonate)	dimethylphosphonate (1)	−2.054 70	0.579 91	–	–
146	PO_3 (phosphonate)	trimethylphosphonate (1)	−0.781 62	0.259 37	−29.150 96	3.747 48
147	PHO_4 (phosphate)	diethyl phosphate (1)	−2.094 20	0.579 41	–	–
148	PO_4 (phosphate)	trimethyl phosphate (1)	−1.785 77	0.300 43	−34.766 78	4.109 57
149	$aC-PO_4$	triphenyl phosphate (1)	−0.894 43	0.282 76	−40.755 23	4.361 70
150	aC−P	triphenylphosphine (1)	0.828 34	0.488 57	–	–
151	CO_3 (carbonate)	diethyl carbonate (1)	−0.969 86	0.359 48	−22.754 73	3.266 61
152	$C2H_3O$	ethyl oxirane (1)	−0.547 63	0.265 05	−16.190 26	2.766 72
153	C2O	trimethyl oxirane (1)	−0.201 87	0.580 95	−15.669 66	2.701 02
154	CH_2 (cyc)	cyclopentane (5)	0.173 89	0.079 11	−5.536 22	1.579 37
155	CH (cyc)	methylcyclopentane (1)	0.327 05	0.090 87	−5.309 56	1.521 55
156	C (cyc)	1,1-dimethylcyclohexane (1)	0.305 81	0.136 60	−4.780 54	1.461 47
157	CH═CH (cyc)	cyclobutene (1)	0.371 03	0.130 12	−10.347 65	2.151 82
158	CH═C (cyc)	1-methylcyclopentene (1)	0.558 72	0.111 18	−10.160 67	2.109 75
159	C═C (cyc)	1,2-dimethylcyclopentene (1)	0.733 41	0.146 16	−10.336 87	2.066 86
160	CH_2═C (cyc)	methylene cyclohexane (1)	−0.216 19	0.293 85	−9.977 48	2.151 73
161	NH (cyc)	cyclopentimine (1)	−0.450 91	0.119 62	−5.648 25	1.634 03
162	N (cyc)	*N*-methylpyrrolidine (1)	−0.623 15	0.120 76	−4.495 82	1.578 22
163	CH═N (cyc)	imidazole (1)	−0.191 66	0.140 63	−10.755 71	2.192 24
164	C═N (cyc)	2-methyl-1H-imidazole (1)	0.032 64	0.115 91	−10.566 35	2.151 00
165	O (cyc)	tetrahydropyran (1)	−0.385 24	0.105 23	−6.005 72	1.686 77
166	CO (cyc)	cyclobutanone (1)	−0.353 43	0.102 97	−11.068 62	2.231 82
167	S (cyc)	2-methylthiophene (1)	0.377 46	0.141 60	−12.497 82	2.387 84
168	SO_2 (cyc)	cyclobutadiene sulfone (1)	−1.227 31	0.174 88	−24.280 83	3.375 23

[a] SE = standard errors of log K_{ow} first-order group-contribution values. [b] Standard errors calculated through eqs 9−11. [c] SE = standard errors of log W_s first-order group-contribution values.

Table A.7: Second-order group contribution data (Marrero & Gani, 2002).

	group	example	log K_{ow}	SE[b,c]	log W_s	SE[c,d]
1	$(CH_3)_2CH$	2-methylpentane (1)	0.026 07	0.002 88	−0.040 21	0.016 01
2	$(CH_3)_3C$	2,2,4,4-tetramethylpentane (2)	0.018 06	0.005 08	−0.349 44	0.156 32
3	$CH(CH_3)CH(CH_3)$	2,3,4-trimethylpentane (2)	0.140 13	0.038 25	−0.115 51	0.035 06
4	$CH(CH_3)C(CH_3)_2$	2,2,3,4,4-pentamethylpentane (2)	−0.175 44	0.026 88	0.221 91	0.077 03
5	$CH_n{=}CH_m{-}CH_p{=}CH_k$ ($k, m, n, p = 0, 1, 2$)	1,3-butadiene (1)	0.089 58	0.016 78	−0.201 33	0.035 09
6	$CH_3{-}CH_m{=}CH_n$ ($m, n = 0, 1, 2$)	2-methyl-2-butene (3)	0.073 81	0.006 26	−0.233 72	0.033 44
7	$CH_2{-}CH_m{=}CH_n$ ($m, n = 0, 1, 2$)	1,4-pentadiene (2)	−0.032 17	0.003 70	0.000 75	0.002 36
8	$CH_p{-}CH_m{=}CH_n$ ($m, n = 0, 1, 2; p = 0, 1$)	3-methyl-1-butene (1)	0.076 22	0.017 97	0.299 39	0.041 89
9	CHCHO or CCHO	2-methylbutyl aldehyde (1)	−0.125 00	0.032 29	0.347 48	0.140 72
10	CH_3COCH_2	2-pentanone (1)	−0.155 97	0.011 32	0.425 50	0.150 05
11	CH_3COCH or CH_3COC	3-methyl-2-pentanone (1)	−0.025 09	0.023 36	0.720 01	0.317 48
12	CHCOOH or CCOOH	2-methyl butanoic acid (1)	0.099 50	0.011 81	0.715 08	0.202 54
13	CH_3COOCH or CH_3COOC	isopropyl acetate (1)	−0.028 52	0.016 93	0.149 45	0.048 22
14	CO−O−CO	propanoic anhydride (1)	−0.240 00	0.064 57	−1.124 02	0.208 46
15	CHOH	2-butanol (1)	0.022 47	0.006 25	0.555 13	0.109 34
16	COH	2-methyl-2-butanol (1)	−0.154 78	0.016 30	0.738 42	0.257 09
17	NCCHOH or NCCOH	2-hydroxypropionitrile (1)	−0.070 00	0.064 57	–	–
18	$OH{-}CH_n{-}COO$ ($n = 0, 1, 2$)	ethyl lactate (1)	0.299 07	0.031 52	−2.220 11	0.228 35
19	$CH_m(OH)CH_n(OH)$ ($m, n = 0, 1, 2$)	ethylene glycol (1)	−0.030 97	0.005 80	−0.340 51	0.023 77
20	$CH_m(OH)CH_n(NH_p)$ ($m, n, p = 0, 1, 2$)	2-amino-1-butanol (1)	0.012 39	0.003 83	0.129 35	0.032 00
21	$CH_m(NH_2)CH_n(NH_2)$ ($m, n = 0, 1, 2$)	ethylenediamine (1)	0.240 00	0.064 57	−1.345 99	0.572 25
22	$CH_m(NH)CH_n(NH_2)$ ($m, n = 1, 2$)	diethylenetriamine (1)	−0.120 78	0.029 10	–	–
23	$H_2NCOCH_nCH_mCONH_2$ ($m, n = 1, 2$)	butanediamide (1)	0.054 64	0.037 29	–	–
24	$CH_m(NH_n){-}COOH$ ($m, n = 0, 1, 2$)	L-alanine (1)	−0.240 93	0.011 39	−0.061 24	0.013 83
25	$HOOC{-}CH_n{-}COOH$ ($n = 1, 2$)	malonic acid (1)	0.155 00	0.045 66	**−0.159 23**	**0.154 85**
26	$HOOC{-}CH_n{-}CH_m{-}COOH$ ($n, m = 1, 2$)	succinic acid (1)	−0.098 69	0.028 23	−0.147 42	0.022 88
27	$HO{-}CH_n{-}COOH$ ($n = 1, 2$)	2-hydroxyisobutyric acid (1)	0.094 04	0.020 20	0.044 45	0.021 76
28	$CH_3{-}O{-}CH_n{-}COOH$ ($n = 1, 2$)	methoxyacetic acid (1)	0.097 45	0.065 15	–	–
29	HS−CH−COOH	2-mercaptopropionic acid (1)	0.050 00	0.064 57	0.876 25	0.171 49
30	$HS{-}CH_n{-}CH_m{-}COOH$ ($n, m = 1, 2$)	β-thiolactic acid (1)	−0.062 72	0.038 05	**0.582 26**	**0.572 35**
31	$NC{-}CH_n{-}CH_m{-}CN$ ($n, m = 1, 2$)	1,2-dicyanoethane (1)	−0.390 00	0.064 57	−0.485 56	0.193 45
32	$OH{-}CH_n{-}CH_m{-}CN$ ($n, m = 1, 2$)	3-hydroxypropanenitrile (1)	−0.260 00	0.064 57	0.201 48	0.047 33
33	$COO{-}CH_n{-}CH_m{-}OOC$ ($n, m = 1, 2$)	ethylene glycol diacetate (1)	−0.126 14	0.027 69	**0.041 98**	**0.572 30**
34	$OOC{-}CH_m{-}CH_m{-}COO$ ($n, m = 1, 2$)	dimethylsuccinate (1)	−0.147 50	0.032 29	−0.202 25	0.031 42
35	$NC{-}CH_n{-}COO$ ($n = 1, 2$)	methylcyanoacetate (1)	−0.220 00	0.064 57	–	–
36	$COCH_nCOO$ ($n = 1, 2$)	methylacetoacetate (1)	0.206 83	0.040 75	0.151 05	0.031 78
37	$CH_m{-}O{-}CH_n{=}CH_p$ ($m, n, p = 0, 1, 2, 3$)	ethyl vinyl ether (1)	−0.207 01	0.026 50	**−0.409 74**	**0.221 25**
38	$CH_m{=}CH_n{-}F$ ($m, n = 0, 1, 2$)	1-fluoro-1-propene (1)	−0.036 00	0.028 88	−0.118 92	0.024 82
39	$CH_m{=}CH_n{-}Br$ ($m, n = 0, 1, 2$)	1-bromo-1-propene (1)	0.042 27	0.016 02	−0.310 43	0.043 77
40	$CH_m{=}CH_n{-}Cl$ ($m, n = 0, 1, 2$)	1-chloro-2-methyl propene (1)	0.125 98	0.008 43	−0.082 32	0.033 91
41	$CH_m{=}CH_n{-}CN$ ($m, n = 0, 1, 2$)	acrylonitrile (1)	−0.027 41	0.008 84	1.173 25	0.018 94
42	$CH_n{=}CH_m{-}COO{-}CH_p$ ($m, n, p = 0, 1, 2, 3$)	ethyl acrylate (1)	−0.051 54	0.004 70	**0.415 69**	**1.067 79**
43	$CH_m{=}CH_n{-}CHO$ ($m, n = 0, 1, 2$)	propenaldehyde (1)	−0.159 64	0.026 37	0.457 84	0.108 10
44	$CH_m{=}CH_n{-}COOH$ ($m, n = 0, 1, 2$)	acrylic acid (1)	0.085 90	0.013 84	−0.355 69	0.069 67
45	$aC{-}CH_n{-}X$ ($n = 1, 2$; X = halogen)	benzyl bromide (1)	0.008 30	0.003 60	−0.207 26	0.016 74
46	$aC{-}CH_n{-}NH_m$ ($n = 1, 2; m = 0, 1, 2$)	benzylamine (1)	−0.104 55	0.006 66	0.373 39	0.007 35
47	$aC{-}CH_n{-}O{-}$ ($n = 1, 2$)	benzyl ethyl ether (1)	−0.130 08	0.006 70	−0.253 22	0.009 76
48	$aC{-}CH_n{-}OH$ ($n = 1, 2$)	benzyl alcohol (1)	−0.074 59	0.004 04	0.274 24	0.004 99
49	$aC{-}CH_n{-}CN$ ($n = 1, 2$)	benzyl cyanide (1)	−0.256 28	0.015 10	−0.239 00	0.037 70
50	$aC{-}CH_n{-}CHO$ ($n = 1, 2$)	phenyl acetaldehyde (1)	0.220 00	0.064 57	–	–

Table A.7 (contd.)

	group	example	log K_{ow}	SE[b,c]	log W_s	SE[c,d]
51	aC−CH_n−COOH (n = 1, 2)	phenyl acetic acid (1)	0.052 63	0.005 77	0.145 58	0.006 05
52	aC−CH_n−CO− (n = 1, 2)	phenyl acetone (1)	−0.014 85	0.008 55	−0.436 70	0.056 16
53	aC−CH_n−S− (n = 1, 2)	benzyl methyl sulfide (1)	−0.089 07	0.020 90	0.318 42	0.063 07
54	aC−CH_m−NO_2 (n = 1, 2)	phenyl nitromethane (1)	0.066 11	0.045 67	–	–
55	aC−CH_n−$CONH_2$ (n = 1, 2)	phenyl ethanamide (1)	−0.114 56	0.016 19	–	–
56	aC−CH_n−OOC (n = 1, 2)	benzyl acetate (1)	0.086 40	0.007 86	−0.357 28	0.014 09
57	aC−CH_n−COO (n = 1, 2)	methyl phenyl acetate (1)	−0.174 34	0.009 63	1.070 18	0.068 52
58	aC−SO_2−OH	benzenesulfonic acid (1)	−0.590 00	0.064 57	0.630 79	0.118 42
59	aC−$CH(CH_3)_2$	cumene (1)	0.130 90	0.008 56	**−0.507 38**	**0.369 08**
60	aC−$C(CH_3)_3$	*tert*-butylbenzene (1)	0.069 30	0.008 41	−0.210 83	0.033 48
61	aC−CF_3	perfluorotoluene (1)	−0.012 89	0.011 96	0.752 33	0.016 03
62	(CH_n═C)(cyc)−CHO (n = 0, 1, 2)	furfural (1)	−0.002 04	0.022 84	0.339 09	0.016 29
63	(CH_n═C)cyc−COO−CH_m (n, m = 0, 1, 2, 3)	methyl furanyrate (1)	0.025 27	0.008 78	1.660 01	0.190 32
64	(CH_n═C)cyc−CO− (n = 0, 1, 2)	2-acetylfuran (1)	0.085 21	0.006 92	**−0.012 01**	**0.057 18**
65	(CH_n═C)cyc−CH_3 (n = 0, 1, 2)	1,2-dimethylcyclopentene (2)	−0.049 24	0.003 59	0.403 69	0.029 47
66	(CH_n═C)cyc−CH_2 (n = 0, 1, 2)	2-ethylfuran (1)	−0.073 77	0.005 06	0.475 69	0.027 17
67	(CH_n═C)cyc−CN (n = 0, 1, 2)	3-cyanofuran (1)	0.025 32	0.010 02	0.260 95	0.037 78
68	(CH_n═C)cyc−Cl (n = 0, 1, 2)	2-chlorofuran (1)	0.113 50	0.005 43	−0.126 13	0.003 64
69	CHcyc−CH_3	methylcyclopentane (1)	−0.035 23	0.004 73	−0.132 21	0.001 76
70	CHcyc−CH_2	ethylcyclohexane (1)	0.018 32	0.004 22	−0.240 73	0.003 75
71	CHcyc−CH	isopropylcyclopentane (1)	−0.001 59	0.010 21	0.090 01	0.017 05
72	CHcyc−C	*tert*-butylcyclohexane (1)	0.098 50	0.019 72	−0.357 59	0.020 79
73	CHcyc−CH═CH_n (n = 1, 2)	vinylcyclopentane (1)	0.245 13	0.016 78	−0.237 34	0.019 79
74	CHcyc−C═CH_n (n = 1, 2)	limonene (1)	−0.110 26	0.046 28	**−0.022 43**	**0.042 40**
75	CHcyc−Cl	chlorocyclopentane (1)	−0.045 05	0.002 04	0.028 89	0.011 74
76	CHcyc−F	fluorocyclohexane (1)	−0.092 41	0.009 73	−1.472 65	0.001 66
77	CHcyc−OH	cyclohexanol (1)	0.059 22	0.002 85	−0.178 38	0.024 64
78	CHcyc−NH_2	cyclohexylamine (1)	−0.294 03	0.013 94	1.554 87	0.015 32
79	CHcyc−NH−CH_n (n = 0, 1, 2, 3)	*N*-methylcyclohexylamine (1)	−0.002 94	0.009 33	**1.053 17**	**5.723 63**
80	CHcyc−N−CH_n (n = 0, 1, 2, 3)	dimethylcyclohexanamine (1)	0.052 21	0.018 81	0.422 50	0.099 88
81	CHcyc−CN	cyanocyclopentane (1)	−0.440 00	0.064 57	–	–
82	CHcyc−COOH	cyclopropanecarboxylic acid (1)	0.201 62	0.013 28	**0.032 09**	**0.083 92**
83	CHcyc−CO	methyl cyclohexyl ketone (1)	0.032 76	0.007 83	0.292 32	0.030 00
84	CHcyc−NO_2	nitrocyclohexane (1)	0.059 70	0.016 71	–	–
85	CHcyc−S−	methyl cyclopentyl sulfide (1)	−0.148 07	0.013 99	–	–
86	CHcyc−O−	methoxycyclohexane (1)	0.002 36	0.006 99	−0.076 61	0.019 39
87	CHcyc−COO	ethyl cyclobutyrate (1)	−0.251 68	0.017 18	0.393 81	0.008 63
88	CHcyc−OOC	cyclohexyl acetate (1)	−0.116 73	0.009 02	0.207 81	0.007 71
89	Ccyc−CH_3	1,1-dimethylcyclohexane (2)	0.007 35	0.003 66	0.074 61	0.026 72
90	Ccyc−CH_2	1-ethyl-1-methyl-cyclopentane (1)	0.049 00	0.005 09	0.392 20	0.014 94
91	Ccyc−OH	1-methylcyclopentanol (1)	0.000 79	0.007 37	−0.280 54	0.022 51
92	>Ncyc−CH_3	*N*-methyl-2-pyrrolidone (1)	0.015 57	0.004 78	−0.212 09	0.012 79
93	>Ncyc−CH_2	*N*-ethylpyrrole (1)	−0.017 61	0.002 79	0.311 75	0.098 81
94	AROMRINGs1s2	2-methylphenol (1), 2-ethyltoluene (1)	0.005 61	0.002 83	0.091 52	0.010 28
95	AROMRINGs1s3	3-methylphenol (1), 3-ethyltoluene (1)	0.054 00	0.004 01	−0.386 00	0.012 86
96	AROMRINGs1s4	4-methylphenol (1), 4-ethyltoluene (1)	0.017 40	0.001 78	−0.164 59	0.014 32
97	AROMRINGs1s2s3	1,2,3-trimethylbenzene (1)	0.006 29	0.004 84	0.081 39	0.005 33
98	AROMRINGs1s2s4	1,2,4-trihydroxybenzene (1)	0.007 78	0.002 38	0.124 87	0.001 95
99	AROMRINGs1s3s5	3,5-diethyltoluene (1)	0.151 43	0.008 10	−0.178 49	0.006 46
100	AROMRINGs1s2s3s4	3-ethyl-1,2,4-trimethylbenzene (1)	−0.058 50	0.006 38	0.023 06	0.002 08
101	AROMRINGs1s2s3s5	1,2,3,5-tetramethylbenzene (1)	−0.040 15	0.004 19	−0.086 88	0.002 05
102	AROMRINGs1s2s4s5	1,2,4,5-tetramethylbenzene (1)	−0.036 76	0.006 01	**−0.001 06**	**0.004 82**
103	PYRIDINEs2	2-methylpyridine (1)	−0.156 05	0.008 16	0.806 72	0.001 52
104	PYRIDINEs3	3-methylpyridine (1)	−0.139 29	0.009 80	0.318 51	0.023 25
105	PYRIDINEs4	4-methylpyridine (1)	−0.161 59	0.007 93	0.154 34	0.026 52
106	PYRIDINEs2s3	2,3-dimethylpyridine (1)	−0.068 78	0.019 61	**0.392 69**	**0.246 76**
107	PYRIDINEs2s4	2,4-dimethylpyridine (1)	0.091 00	0.011 09	0.645 04	0.031 75
108	PYRIDINEs2s5	2,5-dimethylpyridine (1)	−0.068 03	0.014 05	0.092 56	0.007 40
109	PYRIDINEs2s6	2,6-dimethylpyridine (1)	−0.129 50	0.017 41	1.532 35	0.020 37
110	PYRIDINEs3s4	3,4-dimethylpyridine (1)	−0.300 00	0.064 57	–	–
111	PYRIDINEs3s5	3,5-dimethylpyridine (1)	−0.065 00	0.045 66	–	–
112	PYRIDINEs2s3s6	2,3,6-trimethylpyridine (1)	−0.168 41	0.065 06	1.399 59	0.211 55

[a] Values in bold correspond to group-contribution values that failed the *t*-Student test. [b] SE = standard errors of log K_{ow} second-order group-contribution values. [c] Standard errors are calculated through eqs 9−11. [d] SE = standard errors of log W_s second-order group-contribution values.

Table A.8: Third-order group contribution data (Marrero & Gani, 2002).

	group	example	log K_{ow}	SE[b,c]	log W_s	SE[c,d]
1	HOOC−$(CH_n)_m$−COOH ($m > 2$, $n = 0, 1, 2$)	1,5-pentanedioic acid (1)	−0.425 00	0.016 47	0.464 81	0.042 16
2	NH_2−$(CH_n)_m$−OH ($m > 2$, $n = 0, 1, 2$)	4-aminobutanol (1)	−0.050 00	0.032 94	–	–
3	NH_2−$(CH_n)_m$−NH_2 ($m > 2$; $n = 0, 1, 2$)	1,5-diaminopentane (1)	0.400 00	0.032 94	–	–
4	aC−(CH_n═CH_m)cyc (fused rings) (n, $m = 0, 1$)	indene (1), acenaphthylene (2)	0.004 40	0.002 81	0.613 27	0.048 68
5	aC−aC (different rings)	biphenylene (2), biphenyl (1)	0.036 67	0.001 73	−0.271 03	0.084 32
6	aC−CH_ncyc (different rings) ($n = 0, 1$)	cyclohexylbenzene (1)	−0.048 84	0.002 31	0.356 01	0.059 63
7	aC−CH_ncyc (fused rings) ($n = 0, 1$)	tetralin (2), indane (2)	−0.030 08	0.001 55	0.470 61	0.017 57
8	aC−$(CH_n)_m$−aC (different rings) ($m > 1$; $n = 0, 1, 2$)	bibenzyl (1)	0.130 00	0.023 29	−1.172 27	0.007 23
9	aC−$(CH_n)_m$−CHcyc (different rings) ($m > 0$; $n = 0, 1, 2$)	1-cyclopentyl-3-phenylpropane (1)	0.256 63	0.009 16	−0.092 49	0.013 12
10	CHcyc−CHcyc (different rings)	cyclohexyl cyclohexane (1)	0.027 08	0.011 80	0.031 00	0.007 86
11	CH multiring	hexahydroindan (2), decalin (2)	−0.005 70	0.001 10	**0.059 88**	**0.084 32**
12	C multiring	spiropentane (1)	0.013 67	0.001 45	**−0.050 54**	**0.049 10**
13	aC−CH_m−aC (different rings) ($m = 0, 1, 2$)	diphenylmethane (1)	−0.028 00	0.002 22	0.333 90	0.027 08
14	aC−(CH_m═CH_n)−aC (different rings) (m, $n = 0, 1, 2$)	1,2-diphenylethylene (1)	−0.156 67	0.019 02	−0.617 98	0.004 61
15	(CH_m═C)cyc−CH_p−(C═CH_n)cyc (different rings)	difuranyl methane (1)	0.203 29	0.019 13	–	–
16	aC−CO−aC (different rings)	benzophenone (1)	0.025 00	0.007 02	0.700 24	0.006 46
17	aC−CH_m−CO−aC (different rings) ($m = 0, 1, 2$)	benzyl phenone (1)	0.055 00	0.023 29	0.247 00	0.011 00
18	aC−CO−(C═CH_n)cyc (different rings) ($n = 0, 1$)	phenyl-2-furanyl-methanone (1)	−0.098 75	0.011 65	−0.312 70	0.059 63
19	aC−COcyc (fused rings)	phenolphthalein (1)	−0.007 50	0.001 61	**0.027 07**	**0.037 93**
20	aC−CO−$(CH_n)_m$−CO−aC (different rings) ($m > 0$; $n = 0, 1, 2$)	1,4-diphenyl-1,4-butanedione (1)	−0.061 59	0.008 27	0.287 36	0.084 32
21	aC−CO−CH_ncyc (different rings) ($n = 0, 1$)	cyclohexyl phenyl methanone (1)	0.608 29	0.033 02	−1.485 82	0.034 43
22	aC−CO−NH_n−aC (different rings) ($n = 0, 1$)	*N*-phenyl benzamide (1)	0.090 95	0.006 12	−1.011 01	0.006 41
23	aC−NH_nCONH$_m$−aC (different rings) (n, $m = 0, 1$)	*N*,*N*-diphenylurea (1)	0.235 00	0.016 47	0.158 97	0.028 54
24	aC−CO−Ncyc (different rings)	*N*-phenonyl piperidine (1)	0.474 08	0.010 99	−0.782 55	0.059 91
25	aC−Scyc (fused rings)	dibenzothiophene (2)	0.026 38	0.002 58	**−0.019 49**	**0.029 87**
26	aC−S−aC (different rings)	diphenyl sulfide (1)	−0.060 00	0.014 73	–	–
27	aC−SOn−aC (different rings) ($n = 1, 2, 3, 4$)	diphenyl sulfone (1)	−0.253 12	0.006 12	0.175 95	0.059 63
28	aC−NH_ncyc (fused rings) ($n = 0, 1$)	carbazole (2)	−0.013 74	0.001 73	**0.092 39**	**0.086 91**
29	aC−NH−aC (different rings)	diphenylamine (1)	0.123 31	0.006 12	0.150 97	0.013 44
30	aC−(C═N)cyc (different rings)	phenyl-3-pyrazole (1)	−0.021 89	0.002 76	1.495 06	0.042 16
31	aC−(N═CH_n)cyc (fused rings) ($n = 0, 1$)	benzoxazole (1)	0.023 15	0.002 20	0.101 46	0.010 29
32	aC−(CH_n═N)cyc (fused rings) ($n = 0, 1$)	benzoisoxazole (1)	−0.177 45	0.005 28	−0.648 33	0.025 47
33	aC−O−CH_n−aC (different rings) ($n = 0, 1, 2$)	benzyl phenyl ether (1)	−0.003 04	0.005 30	−1.405 10	0.026 29
34	aC−O−aC (different rings)	diphenyl ether (1)	0.041 50	0.002 57	−0.151 24	0.016 69
35	aC−CH_n−O−CH_m−aC (different rings) (n, $m = 0, 1, 2$)	benzyl ether (1)	−0.260 00	0.032 94	0.483 54	0.024 03
36	aC−Ocyc (fused rings)	benzoxazole (1)	0.037 73	0.001 66	−0.243 30	0.084 32
37	AROM.FUSED[2	naphthalene (2)	0.022 94	0.001 51	−0.080 61	0.009 66
38	AROM.FUSED[2]s1	1-methylnaphthalene (1)	−0.022 92	0.002 64	−0.194 82	0.084 32
39	AROM.FUSED[2]s2	2,7-dimethylnaphthalene (2)	−0.012 88	0.001 59	−0.167 46	0.005 19
40	AROM.FUSED[2]s2s3	2,3-dimethylnaphthalene (1)	0.058 84	0.003 98	−0.043 70	0.006 56
41	AROM.FUSED[2]s1s4	1,4-dimethylnaphthalene (1)	0.033 77	0.006 45	−0.260 29	0.009 90
42	AROM.FUSED[2]s1s2	1,2-dimethylnaphthalene (1)	0.044 81	0.007 49	0.547 12	0.010 52
43	AROM.FUSED[2]s1s3	1,3-dimethylnaphthalene (1)	−0.035 72	0.004 24	−0.111 65	0.017 11
44	AROM.FUSED[3]	phenalene (3), pyrene (2)	0.025 72	0.004 06	−0.138 80	0.028 85
45	AROM.FUSED[4a]	anthracene (1)	0.078 82	0.006 81	−0.653 27	0.038 40
46	AROM.FUSED[4a]s1	9-methylanthracene (1)	−0.052 67	0.011 26	**0.032 50**	**0.022 92**
47	AROM.FUSED[4a]s1s4	9,10-dimethylanthracene (1)	−0.053 95	0.023 88	0.028 51	0.012 69
48	AROM.FUSED[4p]	phenanthrene (1), pyrene (2)	0.058 47	0.003 16	0.074 96	0.022 69
49	PYRIDINE.FUSED[2]	quinoline (1)	−0.126 59	0.003 87	0.331 56	0.043 99
50	PYRIDINE.FUSED[2-iso]	isoquinoline (1)	−0.088 60	0.008 90	1.089 46	0.060 94
51	PYRIDINE.FUSED[4]	acridine (1)	−0.271 49	0.010 26	1.421 32	0.008 66

[a] Values in bold correspond to group-contribution values that failed the *t*-Student test. [b] SE = standard errors of log K_{ow} second-order group-contribution values. [c] Standard errors are calculated through eqs 9−11. [d] SE = standard errors of log W_s second-order group-contribution values.

Table A.9: Property Models for properties (Marrero & Gani, 2002).

$$\log K_{ow} = 0.543 + \Sigma_i N_i \log K_{ow}(\mathrm{I})_i + \Sigma_j M_j \log K_{ow}(\mathrm{II})_j + \Sigma_k O_k \log K_{ow}(\mathrm{III})_k$$

$$\log W_s = 4.856 + 0.385 M_w + \Sigma_i N_i \log W_s(\mathrm{I})_i + \Sigma_j M_j \log W_s(\mathrm{II})_j + \Sigma_k O_k \log W_s(\mathrm{III})_k$$

Table A.10: First-order group contributions to the dispersion partial solubility parameter, δ_d , the polar partial solubility parameter, δ_p, and the hydrogen-bonding partial solubility parameter, δ_{hb} (Stefanis & Panayiotou, 2008).

First-order groups	δ_d	δ_p	δ_{hb}	Examples (Occurrences)
$-CH_3$	−0.9714	−1.6448	−0.7813	Propane (2)
$-CH_2$	−0.0269	−0.3045	−0.4119	Butane (2)
−CH<	0.6450	0.6491	−0.2018	Isobutane (1)
>C<	1.2686	2.0838	0.0866	Neopentane (1)
CH_2=CH–	−1.0585	−2.0035	−1.2985	Propylene (1)
–CH=CH–	0.0048	−0.2984	−0.0400	*cis*-2-Butene (1)
CH_2=C<	−0.4829	−0.7794	−0.8260	Isobutene (1)
–CH=C<	0.5372	−0.9024	−1.8872	2-Methyl-2-butene (1)
>C=C<	0.3592	1.0526	−15.4659	2,3-Dimethyl-2-butene (1)
CH_2=C=CH–	−1.6518	***	−0.9980	1,2-Butadiene (1)
CH≡C–	0.2320	−1.3294	1.0736	Propyne (1)
C≡C	−0.2028	−0.7598	−1.1083	2-Butyne (1)
ACH	0.1105	−0.5303	−0.4305	Benzene (6)
AC	0.8446	0.6187	0.0084	Naphthalene (2)
$ACCH_3$	0.2174	−0.5705	−1.1473	Toluene (1)
$ACCH_2$–	0.6933	0.6517	−0.1375	*m*-Ethyltoluene (1)
CH_3CO	−0.3551	2.3192	−1.3078	Methyl ethyl ketone (1)
CH_2CO	0.6527	3.7328	−0.5344	Cyclopentanone (1)
CHO (aldehydes)	−0.4030	3.4734	0.1687	1-Butanal (1)
COOH	−0.2910	0.9042	3.7391	Vinyl acid (1)
CH_3COO	−0.5401	−0.3970	1.5826	Ethyl acetate (1)
CH_2COO	0.2913	3.6462	1.2523	Methyl propionate (1)
HCOO	***	1.9308	2.1202	*n*-Propyl formate (1)
COO	0.2039	3.4637	1.1389	Ethyl acrylate (1)
OH	−0.3462	1.1404	7.1908	Isopropanol (1)
ACOH	0.5288	1.1010	6.9580	Phenol (1)
CH_3O	−0.5828	0.1764	0.1460	Methyl ethyl ether (1)
CH_2O	0.0310	0.8826	−0.1528	Ethyl vinyl ether (1)
CHO (ethers)	0.8833	1.6853	0.4470	Diisopropyl ether (1)
$C_2H_5O_2$	−0.1249	3.6422	8.3579	2-Methoxy-ethanol (1)
CH_2O (cyclic)	0.2753	0.1994	−0.1610	1,4-Dioxane (2)
CH_2NH_2	−0.5828	1.4084	2.5920	1-Amino-2-propanol (1)
$CHNH_2$	0.0112	−1.1989	0.3818	Isopropylamine (1)
CH_3NH	***	0.6777	5.6646	*n*-Methylaniline (1)
CH_2NH	0.8116	0.9412	1.3400	di-*n*-Propylamine (1)
CH_3N	0.8769	1.2046	1.6062	Trimethylamine (1)
CH_2N	1.4681	2.8345	1.2505	Triethylamine (1)
$ACNH_2$	1.6987	1.6761	4.5274	Aniline (1)
$CONH_2$	−0.0689	6.0694	5.2280	2-Methacrylamide (1)
$CON(CH_3)_2$	0.4482	5.7899	3.0020	*n,n*-Dimethylacetamide (1)
CH_2SH	1.2797	−0.8223	4.4646	*n*-Butyl mercaptan (1)
CH_3S	***	0.4944	−1.4861	Methyl ethyl sulfide (1)
CH_2S	1.0595	0.7530	−0.2287	Diethyl sulfide (1)
I	0.7797	0.6777	0.2646	Isopropyl iodide (1)
Br	0.5717	0.6997	−1.0722	2-Bromopropane (1)
CH_2Cl	0.2623	0.5970	−0.5364	*n*-Butyl chloride (1)
CHCl	0.4462	2.8060	−1.4125	Isopropyl chloride (1)
CCl	2.7576	2.0406	0.1101	t-Butyl chloride (1)
$CHCl_2$	1.1797	1.8361	−3.2861	1,1-Dichloropropane (1)
CCl_2	0.3653	0.1696	−1.4334	Pentachlorocyclopropane (2)
CCl_3	***	1.2777	−2.6354	Benzotrichloride (1)
ACCl	0.8475	−0.0339	−0.7840	*m*-Dichlorobenzene (2)
ACF	0.1170	0.1856	−0.7182	Fluorobenzene (1)

Table A.10(contd.)

First-order groups	δ_d	δ_p	δ_{hb}	Examples (Occurrences)
Cl–(C=C)	0.2289	2.3444	3.8893	2,3-Dichloropropene (1)
CF_3	−0.2293	−1.9735	−1.4665	Perfluorohexane (2)
CH_2NO_2	***	6.8944	−1.2861	1-Nitropropane (1)
$CHNO_2$	***	8.0347	−2.3167	2-Nitropropane (1)
$ACNO_2$	1.4195	4.4838	−0.7167	Nitrobenzene (1)
CH_2CN	−0.3392	6.5341	−0.8892	*n*-Butyronitrile (1)
CF_2	−0.9729	***	***	Perfluoromethylcyclohexane (5)
CF	0.1707	***	***	Perfluoromethylcyclohexane (1)
F (except as above)	−0.7069	***	***	2-Fluoropropane (1)
CH_2=C=C<	−0.2804	***	−1.9167	3-Methyl-1,2-butadiene (1)
O (except as above)	0.0472	3.3432	0.0256	Divinyl ether (1)
Cl (except as above)	0.2256	1.8711	−0.3295	Hexachlorocyclopentadiene (2)
>C=N–	−0.3074	−0.0012	−5.3956	2,4,6-Trimethylpyridine (1)
–CH=N–	0.9672	1.9728	0.7668	Isoquinoline (1)
NH (except as above)	***	0.0103	2.2086	Dibenzopyrrole (1)
CN (except as above)	0.0861	6.5331	−0.6849	*cis*-Crotonitrile (1)
O=C=N–	−0.1306	1.6102	4.0461	*n*-Butyl isocyanate (1)
SH (except as above)	1.0427	1.9813	4.8181	2-Mercaptobenzothiazole (1)
S (except as above)	1.4899	9.2072	−0.6250	Thiophene (1)
SO_2	1.5502	11.1758	0.1055	Sulfolene (1)
>C=S	0.7747	0.0683	3.4080	*n*-Methylthiopyrrolidone (1)
>*C*=0 (except as above)	−0.4343	0.7905	1.8147	Anthraquinone (2)
N (except as above)	1.5438	2.5780	1.1189	Triphenylamine (1)

*** The specific group contributions to this delta parameter are not available

Table A.11: Second-order group contributions to the dispersion partial solubility parameter, δ_d , the polar partial solubility parameter, δ_p, and the hydrogen-bonding partial solubility parameter, δ_{hb}. (Stefanis & Panayiotou, 2008)

Second-order groups	δ_d	δ_p	δ_{hb}	Examples (Occurrences)
$(CH_3)_2$–CH–	0.0460	0.0019	0.3149	Isobutane (1)
$(CH_3)_3$–C–	−0.0738	1.1881	−0.2966	Neopentane (1)
ring of 5 carbons	−0.6681	−2.3430	−0.3079	Cyclopentane (1)
ring of 6 carbons	−0.3874	−3.6432	***	Cyclohexane (1)
–C=C–C=C–	−0.1355	−3.5085	−1.0795	1,3-Butadiene (1)
CH_3-C=	−0.0785	0.3316	0.3875	Isobutene (2)
$-CH_2-C=$	−0.3236	−2.3179	−0.5836	1-Butene (1)
>C{H or C}–C=	−0.2798	***	−1.1164	3-Methyl-1-butene (1)
string in cyclic	−0.1945	***	***	Ethylcyclohexane (1)
$CH_3(CO)CH_2$–	−0.0451	−0.3383	−0.4083	Methyl ethyl ketone (1)
C_{cyclic}=O	−0.2981	0.4497	−0.4794	Cyclopentanone (1)
ACCOOH	−0.2293	−0.6349	−0.9030	Benzoic acid (1)
>C{H or C}–COOH	***	−0.2187	1.1460	Isobutyric acid (1)
CH3(CO)OC{H or C}<	−0.5220	−0.0652	0.3085	Isopropyl acetate (1)
$(CO)C\{H_2\}COO$	***	−2.3792	0.8412	Ethyl acetoacetate (1)
(CO)O(CO)	−0.2707	−1.0562	1.6335	Acetic anhydride (1)
ACHO	0.3772	−1.8110	−1.0096	Benzaldehyde (1)
>CHOH	0.1123	0.2564	−0.1928	2-Propanol (1)
>C<OH	−0.0680	0.1075	1.2931	Tert-Butanol (1)
–C(OH)C(OH)–	***	0.6419	0.3870	1,2-Propanediol (1)
–C(OH)C(N)	−0.0809	0.5683	−0.6326	1-Amino-2-propanol (1)
C_{cyclic}–OH	−0.0876	−3.5220	0.5914	Cyclohexanol (1)
C–O–C=C	0.2063	0.6080	1.1344	Ethyl vinyl ether (1)
AC–O–C	0.2568	0.8153	0.6092	Methyl phenyl ether (1)
>N{H or C}(in cyclic)	0.2218	−2.2018	−0.0452	Cyclopentimine (1)
–S-(in cyclic)	0.4892	0.3040	0.2297	Tetrahydrothiophene (1)
ACBr	0.1234	−0.4495	0.3397	Bromobenzene (1)
(C=C)–Br	−0.4059	−0.0024	−1.1304	2-Bromopropene (1)
ring of 3 carbons	0.0200	1.8288	−0.8073	Cyclopropane (1)
ACCOO	−0.1847	0.4059	−0.1921	Methyl benzoate (1)
$AC(ACH_m)_2AC(ACH_n)_2$	−0.3751	−1.2980	0.6844	Naphthalene (1)
O_{cyclic}–C_{cyclic}=O	0.2468	2.7501	0.1220	Diketene (1)
AC–O–AC	−0.5646	−3.4329	2.0830	Diphenyl ether (1)
$C_{cyclic}H_m$=N_{cyclic}–$C_{cyclic}H_n$=$C_{cyclic}H_p$	0.7002	0.0691	−2.7661	2,6-Dimethylpyridine (1)
$N_{cyclic}H_m$–C_{cyclic}=O	0.2956	2.8958	1.3125	2-Pyrrolidone (1)
$-O-CH_m-O-CH_n-$	0.0839	0.3451	0.3767	Methylal (1)
C(=O)–C–C(=O)	−0.4862	−0.4888	1.2482	2,4-Pentanedione (1)

*** The specific group contributions to this delta parameter are not available

Table A.12: Property Models for estimation of Hansen solubility parameters.

$$\delta_d = \left(\sum_i N_i C_i + W \sum_j M_j D_j + 17.3231 \right) \text{MPa}^{(1/2)}$$

$$\delta_p = \left(\sum_i N_i C_i + W \sum_j M_j D_j + 7.3548 \right) \text{MPa}^{(1/2)}$$

$$\delta_{hb} = \left(\sum_i N_i C_i + W \sum_j M_j D_j + 7.9793 \right) \text{MPa}^{(1/2)}$$

Table A.13: Regressed Parameters for the CI Method (Gani et al., 2005).

	properties									
parameter type	T_m (10^{-1})	T_b (10^{-1})	T_c (10^{-2})	P_c (10^{-3})	V_c	H_f	G_f	H_{fus} (10^{-1})	H_V (10^{-1})	log K_{ow} (10^{-1})
a(H)	−1.951 16	−1.194 61	−44.252 84	2.022 97	7.119 75	−34.777 51	−15.256 65	−0.141 97	1.102 26	−1.175 34
a(Cl)	17.742 44	14.001 77	448.531 72	5.370 86	43.403 16	−66.442 25	−47.252 86	3.800 52	101.044 45	1.543 08
a(Br)	44.965 78	24.031 95	739.233 87	−6.189 42	51.377 39	−40.041 62	−30.932 67	5.007 44	152.640 74	−0.573 23
a(F)	−8.182 62	−0.791 56	−9.098 94	7.725 04	20.807 61	−238.125 24	−222.598 77	0.521 16	9.473 31	0.209 02
a(I)	43.472 78	35.274 55	1 312.383 03	−10.784 26	68.446 31	10.071 69	11.310 87	5.686 94	202.487 15	−5.010 69
a(N)	28.882 43	16.237 96	584.222 14	4.093 67	39.903 16	92.740 20	67.399 56	4.409 31	130.470 19	−4.673 89
a(O)	19.879 42	9.283 53	372.271 66	−1.389 01	18.047 65	−176.070 06	−168.720 71	3.621 08	100.135 61	−5.536 05
a(P)	3.304 41	2.481 35	1 795.371 60	N/A	−82.464 85	−243.529 50	N/A	−11.318 61	N/A	−2.831 10
a(S)	26.658 39	17.769 81	780.776 34	−8.430 30	32.232 63	9.576 91	−1.173 46	4.600 86	125.616 16	0.223 41
a(C)	10.864 15	11.312 90	324.582 68	5.499 90	31.797 84	40.155 90	35.307 60	2.214 78	55.237 68	2.433 30
a(Si)	−1.340 33	3.731 42	N/A	N/A	N/A	N/A	N/A	N/A	N/A	7.160 47
b	2.631 05	−9.382 97	−327.749 36	2.129 80	3.096 75	−7.395 81	−21.519 51	−3.663 88	−52.334 20	2.696 38
c	−10.868 99	4.604 18	125.059 14	6.521 88	7.874 95	11.717 23	15.348 11	3.113 80	25.389 32	0.423 65
d	0.000 00	18.371 91	388.551 35	−18.972 22	8.673 18	61.926 11	97.288 21	1.099 72	−23.39402	0.940 70
E	1474.5	2225.4	23123.9	5982.7	7.95	5.549	−34.967	−28.06	117.33	5.429
G				108.998						

Table A.14: Classification of Groups (Gani et al., 1991).

	Category				
Class	1	2	3	4	5
0	CH3OH CH3SH (CH2OH)2 NMP Diethyl glycol 2-Propanol	CH3NO2 CH3CN CH2CL2 CH3NH2 CCL3F C4H4S	H2O Furfural CHCL3 TCE Pyridine CHCL2F Morpholine	CH3NH2 HCOOH ACRY MFA 1-propanol CHCLF2	DMSO DMF TMS CS2 CCL2F2 CCLF3 CF3
	CH3	CH2CN CH2NO2 CH2NH2 I BR CL F	CH3CO CONH2 CONHCH3 CON(CH3)2 CH3NH CH2NH2	OH CHO COOH CH2CL CH3COO CH3O C2H5O2 CH3S CCL2F CH3NH CHCL2 SH CH2SH COO CCL3 CF3 CCL2F CHCLF HCOO	C4H3S C≡CH SiH3 CH2=CH
2	CH2	CHNO2	CH2CO CH2COO CH2O CONCH3CH2	CHNH2 CH2NH CHCL CONHCH2 C2H4O2 CH2S CH3N CCL2	CH=CH CH2=C C4H2S C≡C SiH2
3	CH		CON(CH2)2	CHNH CH2N CCL CH-O CHS	CH=C SiH SiH2O SiHO
4	C				C=C SiO, Si
5	ACH	ACCH3	ACOH ACNH2 ACCL ACNO2		
6		ACCH2			
7		ACCH			
8		AC			

The "A" denotes aromatic groups.

Table A.15: Rules for generation of acyclic molecules (Gani et al., 1991).

Total number of groups	Largest class of group	Number of groups from largest class	Maximum number of groups allowed from sum of categories				
			3	4	5	3+4+5	4+5
2	1	2	2	1	1	2	1
3	2	1	2	1	1	2	1
4	3	1	2	1	1	2	1
4	2	2	2	2	1	2	2
5	4	1	2	1	1	2	1
5	3	1	2	2	1	2	2
5	2	3	3	2	1	3	2
6	4	1	3	2	1	3	2
6	3	2	3	2	1	3	2
6	3	1	3	2	1	3	2
6	2	4	3	3	1	3	3
7	4	1	3	2	1	3	2
7	3	2	3	3	1	3	3
7	3	1	3	3	1	3	3
7	2	4	4	3	1	4	3
8	4	2	3	2	1	3	2
8	4	1	3	3	1	3	3
8	3	3	3	2	1	3	2
8	3	2	3	3	1	3	3
8	3	1	3	3	1	3	3
8	2	6	4	3	1	4	3
9	4	2	3	2	1	3	2
9	4	1	3	3	1	3	3
9	3	3	3	2	1	3	2
9	3	2	3	3	1	3	3
9	3	1	3	3	1	3	3
9	2	7	4	3	1	4	3
10	4	2	3	2	1	3	2
10	4	1	3	3	1	3	3
10	3	4	3	3	1	3	3
10	3	3	3	3	1	3	3
10	3	2	3	3	1	3	3
10	3	1	3	3	1	3	3
10	2	8	4	3	1	4	3
11	4	3	4	3	1	4	3
11	4	2	4	4	1	4	4
11	4	1	4	4	1	4	4
11	3	4	4	3	1	4	3
11	3	3	4	4	1	4	4
11	3	2	4	4	1	4	4
11	3	1	4	4	1	4	4
11	2	9	5	4	1	5	4
12	4	3	4	3	1	4	3
12	4	2	4	4	1	4	4
12	4	1	4	4	1	4	4
12	3	5	4	3	1	4	3
12	3	4	4	4	1	4	4
12	3	3	4	4	1	4	4
12	3	2	4	4	1	4	4
12	3	1	4	4	1	4	4
12	2	10	5	4	1	5	4

Table A.16: Rules for generation of aromatic molecules (Gani et al., 1991).

Total number of groups	Largest class of group	Number of groups from largest class	Maximum number of groups allowed from sum of categories				
			3	4	5	3+4+5	4+5
6	5	6	3	3	1	3	3
7	6	1	3	3	1	3	3
8	7	1	3	3	1	3	3
8	6	2	3	3	1	3	3
8	6	1	3	3	1	3	3
9	7	1	3	3	1	3	3
9	6	3	3	3	1	3	3
9	6	2	3	3	1	3	3
9	6	1	3	3	1	3	3
10	8	2	3	3	1	3	3
10	7	2	3	3	1	3	3
10	7	1	3	3	1	3	3
10	6	3	3	3	1	3	3
10	6	2	3	3	1	3	3
10	6	1	3	3	1	3	3
11	8	2	3	3	1	3	3
11	7	2	3	3	1	3	3
11	7	1	3	3	1	3	3
11	6	3	3	3	1	3	3
11	6	2	3	3	1	3	3
11	6	1	3	3	1	3	3
12	8	2	3	3	1	3	3
12	7	3	3	3	1	3	3
12	7	2	3	3	1	3	3
12	7	1	3	3	1	3	3
12	6	3	3	3	1	3	3
12	6	2	3	3	1	3	3
12	6	1	3	3	1	3	3
13	8	4	3	3	1	3	3
13	8	2	3	3	1	3	3
14	8	4	3	3	1	3	3
14	8	2	3	3	1	3	3
15	8	4	3	3	1	3	3
15	8	2	3	3	1	3	3
16	8	6	3	3	1	3	3
16	8	4	3	3	1	3	3
16	8	2	3	3	1	3	3
17	8	6	3	3	1	3	3
17	8	4	3	3	1	3	3
17	8	2	3	3	1	3	3
18	8	6	3	3	1	3	3
18	8	4	3	3	1	3	3
18	8	2	3	3	1	3	3

Table A.17: Rules for generation of cyclic molecules (Gani et al., 1991).

Total number of groups	Largest class of group	Number of groups from largest class	Maximum number of groups allowed from sum of categories				
			3	4	5	3+4+5	4+5
3	2	3	0	1	1	1	1
4	3	1	1	1	1	1	1
4	2	4	1	1	1	1	1
5	4	1	2	1	1	2	1
5	3	2	2	2	1	2	1
5	3	1	2	2	1	2	1
5	2	5	1	2	1	2	1
6	4	1	2	1	1	2	1
6	3	3	2	2	1	2	1
6	3	2	2	2	1	2	1
6	3	1	2	2	1	2	1
6	2	6	2	3	1	3	1
7	4	2	3	2	1	2	1
7	4	1	3	2	1	2	1
7	3	3	3	3	1	3	1
7	3	2	3	3	1	3	1
7	3	1	3	3	1	3	1
7	2	7	2	3	1	3	1
8	4	2	3	2	1	2	1
8	4	1	3	2	1	2	1
8	3	4	3	3	1	3	1
8	3	3	3	3	1	3	1
8	3	2	3	3	1	3	1
8	3	1	3	3	1	3	1
8	2	8	2	3	1	3	1
9	4	3	3	2	1	2	1
9	4	2	3	2	1	2	1
9	4	1	3	2	1	2	1
9	3	4	3	3	1	3	1
9	3	3	3	3	1	3	1
9	3	2	3	3	1	3	1
9	3	1	3	3	1	3	1
9	2	9	2	3	1	3	1
10	4	3	3	2	1	2	1
10	4	2	3	2	1	2	1
10	4	1	3	2	1	2	1
10	3	5	3	3	1	3	1
10	3	4	3	3	1	3	1
10	3	3	3	3	1	3	1
10	3	2	3	3	1	3	1
10	3	1	3	3	1	3	1
10	2	10	3	3	1	3	1
11	4	3	3	2	1	2	1
11	4	2	3	2	1	2	1
11	4	1	3	2	1	2	1
11	3	5	4	3	1	3	1
11	3	4	4	3	1	3	1
11	3	3	4	3	1	3	1
11	3	2	4	3	1	3	1
11	3	1	4	3	1	3	1
11	2	11	3	3	1	3	1
12	4	4	3	3	1	3	1
12	4	3	3	3	1	3	1
12	4	2	3	3	1	3	1
12	4	1	3	3	1	3	1
12	3	6	4	3	1	3	1
12	3	5	4	3	1	3	1
12	3	4	4	3	1	3	1
12	3	3	4	3	1	3	1
12	3	2	4	3	1	3	1
12	3	1	4	3	1	3	1
12	2	12	3	3	1	3	1

Appendix B.

Appendix B gives the required literature data for computer aided flowsheet design.

Table B.1: Recommended limits on properties for separation techniques (Jaksland et al., 1995).

Separation Process		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18
Absorption[1]	m							X	X									1.11	
	M																	2.18	
Azeotropic Distillation[1]								X	X									X	
Cryogenic Distillation[2]	m														1.12	1.8			
	M														1.17	2.4			
Crystallization[3]	m												1.2						
	M												1.27						
Desublimation	m													30	1.1				
	M													40	1.2				
Distillation[2]	m														1.01	1.05			
	M														1.02	1.5			
Extractive Distillation[1,2,4]	m							X	X						0	0			
	M														1	1			
	n														1.12	1.5			
	N														1.15	2			
Flash operation[2]	m														1.23	10			
	M														1.4	15			
Gas separation membranes	m		1.07							1.1									
	M		1.1							1.2									
Ion exchange	m																1		
	M																2		
Leaching[1]								X	X									X	
Liquid-Liquid Extraction[1]								X	X									X	
Liquid membranes	m						1.01	1.02											1.2
	M						1.03	1.08											1.28
Microfiltration	m				1.9														
	M				2.4														
Molecular seive adsorption	m		1.05	1.07					1.08	1.05									
	M		1.1	1.1					1.1	1.1									
Partial condensation[2]	m														1.9				
	M														2.15				
Pervaporation	m							1.02	X									1	
	M							1.08										1.01	
Stripping[3,4]	m							X	X									0	
	M																	1	
	n																	1.63	
	N																	3.9	
Sumblimation	m													30	1.1				
	M													40	1.2				
Supercritical extraction[1]								X	X	X	X							X	X
Ultrafiltration	m				2	1.7													
	M				3	2.3													

P1	Kinetic Diameter
P2	van der Walls volume
P3	Molecular diameter
P4	Molecular weight
P5	Radius of gyration
P6	Molar Volume
P7	Polarisability
P8	Dipole momemnt
P9	Critical Temperature
P10	Critical Pressure
P11	Melting Point
P12	Heat of fusion at Tm
P13	Tripple point pressure
P14	Boiling point
P15	Vapor pressure
P16	Ion charge
P17	Solubility parameter
P18	Refractive Index

Separation Techniques for barriers		
Barrier	**Separation Techniques**	**Principle**
Relative Volatility close to unity	Extractive/Azeotropic Distillation	exploit VLE/VLLE with MSA
	Supercritical Extraction	exploit LLE with MSA
Temperature sensitive components	Liquid Liquid Extraction	exploit LLE with MSA
	Stripping	exploit GLE with MSA
	Absorption	exploit GLE with MSA

[1] The methods require mass separating agent which is identified through CAMD.
[2] For these separation techniques heat of vaporization is also considered.
[3] Heat of fusion at melting point is also used.
[4] If relative volatility is greater than 2, Distillation becomes the better alternative

Table B.2: Available PGs (Alvarado, 2010; d'Anterroches, 2006).

Unit Operation	Process-group example
Simple Distillation Column	(AB/C)
Solvent Based Azeotropic Separation	(cycA/B)
Flash Separation	(fAB/CD)
Kinetic Model Based Reactor	(rABC/nE/pABCD)
Fixed Conversion Reactor	(rABC/nE/pABCD)
Pressure Swing Distillation	(swA/B)
Polar Molecular Sieve Based Separation	(pmsABC/D)
Molecular Sieve Based Separation	(msABC/D)
Liquid Membrane Based Separation	(lmemABC/D)
Liquid Adsorption Based Separation	(ladsABC/D)
Gas Membrane Based Separation	(gmemABC/D)
Pervaporation Based Separation	(pervABC/D)
Crystallization Based Separation	(crsABC/D)
Liquid Liquid Extraction Based Separation	(lleABC/S/SC/AB)
Simple Solid Liquid Separation	(slAB/CD)
Absorption	(abEAB/eF/EABF/EF)
Ion Exchange Separation	(ieABCD/ABC)

Table B.3: Rules to denote PGs by invariants through example (d'Anterroches, 2006).

The initial variant of PGs is obtained as through the following table:

Rule	Number calculated from the rule
1	1(inlet), 2(outlet), 3(intermediate)
2	Number of connections (01-99)
3	Number of non inlet/outlet connection (00-99)
4	Length of PG name (04-99)

Examples of invariants of PGs

PG	Number from rules	Initial Invariants	Letter	Equivalent Graph invariant (ascending order of initial invariants)
(iA)	1,01,01,04	1010104	A	1
(iB)	1,01,01,04	1010104	B	1
(rAB/pABCD)	3,03,01,11	3030111	C	5
(mABC/D)	3,03,02,08	3030208	D	6
(oD)	2,01,01,04	2010104	E	2
(A/BC)	3,03,01,06	3030106	F	4
(oA)	2,01,01,04	2010104	G	2
(oBC)	2,01,01,05	2010105	H	3

Discrimination of same ranked groups by their prime equivalent – Lower the prime equivalent of the letter denoting components in the PG, lesser the invariant

1,1,5, 6, 2, 4, 2, 3 → 1, 2, 7, 8, 4, 6, 3, 5

If the PGs are still tied, depending on the rank of the connected PG, its unique invariant is desiced, the lower the rank of the connected PG, lesser the invariant of the tied group

Table B.4: Contributions of the simple distillation process groups (d'Anterroches, 2006).

Group id	NC_{total}	NC_{top}	$F_{Di\|max}$	a_k	$\min \alpha_{i_{ref}}$	$\max \alpha_{i_{ref}}$
dist-1	2	1	0.1570	0.01173	13.625	67.783
dist-2	2	1	0.0664	0.04027	9.2514	13.625
dist-3	2	1	0.1092	0.01845	10.313	9.2514
dist-4	2	1	0.0632	0.09744	7.5854	10.313
dist-5	2	1	0.0168	6.33727	7.0563	7.5854
dist-6	3	1	0.1570	0.01016	9.2514	67.783
dist-7	3	2	0.0664	0.06629	9.2514	67.783
dist-8	3	1	0.0664	0.08074	10.313	13.625
dist-9	3	2	0.1092	0.01310	10.313	13.625
dist-10	3	1	0.1092	0.01476	7.5854	9.2514
dist-11	3	2	0.0632	0.52620	7.5854	9.2514
dist-12	3	1	0.0632	23.71149	7.0563	10.313
dist-13	3	2	0.0168	5.98355	7.0563	10.313
dist-14	3	1	0.0636	0.86774	4.0472	7.0563
dist-15	4	1	0.1570	0.00808	10.313	67.783
dist-16	4	2	0.0664	0.05101	10.313	67.783
dist-17	4	3	0.1092	0.00911	10.313	67.783
dist-18	4	1	0.0664	0.06426	7.5854	13.625
dist-19	4	2	0.1092	0.01113	7.5854	13.625
dist-20	4	3	0.0632	0.30529	7.5854	13.625
dist-21	4	1	0.1092	0.01278	7.0563	9.2514
dist-22	4	2	0.0632	1.03656	7.0563	9.2514
dist-23	4	3	0.0168	5.41363	7.0563	9.2514
dist-24	4	1	0.0168	5.97814	4.0472	7.5854
dist-25	4	2	0.0636	0.07139	4.0472	7.5854
dist-26	5	1	0.1570	0.01406	7.5854	67.783
dist-27	5	2	0.0664	0.04336	7.5854	67.783
dist-28	5	3	0.1092	0.00809	7.5854	67.783
dist-29	5	4	0.0632	0.19277	7.5854	67.783
dist-30	5	1	0.0664	0.02434	7.0563	13.625
dist-31	5	2	0.1092	0.00992	7.0563	13.625
dist-32	5	3	0.0632	0.52071	7.0563	13.625
dist-33	5	4	0.0168	5.11638	7.0563	13.625
dist-34	5	1	0.0632	0.12088	4.0472	10.313
dist-35	5	2	0.0168	5.68497	4.0472	10.313
dist-36	5	3	0.0636	0.05739	4.0472	10.313
dist-37	6	1	0.1570	0.01262	7.0563	67.783
dist-38	6	2	0.0664	0.01754	7.0563	67.783
dist-39	6	3	0.1092	0.00830	7.0563	67.783
dist-40	6	4	0.0632	0.02942	7.0563	67.783
dist-41	6	5	0.0168	4.53667	7.0563	67.783
dist-42	6	1	0.1092	0.00896	4.0472	9.2514
dist-43	6	2	0.0632	0.07976	4.0472	9.2514
dist-44	6	3	0.0168	5.16771	4.0472	9.2514
dist-45	6	4	0.0636	0.04244	4.0472	9.2514
dist-46	7	1	0.0664	0.01506	4.0472	13.625
dist-47	7	2	0.1092	0.00823	4.0472	13.625
dist-48	7	3	0.0632	0.06012	4.0472	13.625
dist-49	7	4	0.0168	4.89430	4.0472	13.625
dist-50	7	5	0.0636	0.03375	4.0472	13.625
dist-51	8	1	0.1570	0.01022	4.0472	67.783
dist-52	8	2	0.0664	0.01144	4.0472	67.783
dist-53	8	3	0.1092	0.00714	4.0472	67.783
dist-54	8	4	0.0632	0.01527	4.0472	67.783
dist-55	8	5	0.0168	4.33034	4.0472	67.783
dist-56	8	6	0.0636	0.02561	4.0472	67.783
dist-57	2	1	0.2199	0.01474	1.5479	4.0472
dist-58	3	2	0.2199	0.00965	1.5479	4.993e+015
dist-59	3	1	0.2199	0.00696	0.028517	4.0472
dist-60	4	3	0.2199	0.00712	1.5479	1e+100
dist-61	4	2	0.2199	0.00516	0.028517	4.993e+015
dist-62	5	3	0.2199	0.00409	0.028517	1e+100
dist-63	5	4	0.7065	0.02241	0.028517	1e+100

Table B.5: Contributions of the extractive process groups (Alvarado, 2010).

Ionic Liquids		
x_S	df_{ij}^k	a_k
0.28	0.3409	0.02150426
0.30	0.3433	0.01483252
0.33	0.3483	0.01355022
0.375	0.3542	0.01249599
0.40	0.3583	0.01197764
0.45	0.3682	0.01102273
0.50	0.3785	0.01070385
0.60	0.3996	0.01090200
0.70	0.4239	0.01181313

Organic solvents			Hyperbranched polymers		
x_S	df_{ij}^k	a_k	x_S	df_{ij}^k	a_k
0.279204	0.41837	0.0180	0.035	0.36	0.0234
0.29883	0.42339	0.0170	0.070	0.38	0.02014
0.332557	0.43182	0.0170			
0.374181	0.44189	0.0160			
0.399162	0.44811	0.0160			
0.449136	0.45997	0.0160			
0.472813	0.46529	0.0150			
0.499127	0.47096	0.0150			
0.522938	0.47568	0.0150			

Table B.6: Pre-calculated values based on driving force approach to design simple distillation columns (Bek-Pedersen, 2003)

$F_{Di\,Max}$	$X_{LK,\,Dist}$	$X_{LK,\,Bot}$	RR_{min}	$RR_{min}C$	N_{ideal}
0.045	0.995	0.005	9.89	14.83	96
	0.98	0.02	9.56	14.36	71
	0.95	0.05	8.90	13.35	54
	0.90	0.10	8.22	12.33	41
0.065	0.995	0.005	7.33	11.00	67
	0.98	0.02	7.10	10.65	50
	0.95	0.05	6.64	9.96	38
	0.90	0.10	6.64	8.58	29
0.101	0.995	0.005	4.50	6.74	44
	0.98	0.02	4.35	6.52	33
	0.95	0.05	4.05	6.08	25
	0.90	0.10	3.56	5.33	19
0.146	0.995	0.005	2.92	4.41	31
	0.98	0.02	2.84	4.26	23
	0.95	0.05	2.63	3.95	18
	0.90	0.10	2.29	3.44	14
0.172	0.995	0.005	2.35	3.53	27
	0.98	0.02	2.26	3.40	20
	0.95	0.05	2.09	3.13	15
	0.90	0.10	1.80	2.70	12
0.195	0.995	0.005	2.06	3.09	24
	0.98	0.02	1.89	2.97	18
	0.95	0.05	1.82	2.74	14
	0.90	0.10	1.57	2.35	11
0.225	0.995	0.005	1.73	2.60	21
	0.98	0.02	1.67	2.50	16
	0.95	0.05	1.53	2.30	12
	0.90	0.10	1.37	1.97	9
0.268	0.995	0.005	1.37	2.06	18
	0.98	0.02	1.31	1.97	13
	0.95	0.05	1.20	1.80	10
	0.90	0.10	1.02	1.52	8
0.382	0.995	0.005	0.82	1.23	13
	0.98	0.02	0.78	1.17	10
	0.95	0.05	0.70	1.05	8
	0.90	0.10	0.57	0.86	6
0.478	0.995	0.005	0.54	0.81	10
	0.98	0.02	0.51	0.76	8
	0.95	0.05	0.44	0.67	6
	0.90	0.10	0.34	0.51	5

Appendix C

Reverse simulation procedure for distillation PGs based on the driving force concept (Bek-Pedersen, 2003)

The procedure to determine the design parameters of the distillation column in the distillation process group is as follows:

1. Given a *NC* component process group.

2. Order the components by relative volatility and identify the key components.

3. Retrieve the maximum *DF* between the key components, F_{Dimax} and the composition of the light key D_x at its maximum, either from experimental data or *VLE* calculations.

4. Select the values of product purities or recoveries for the key components, if they are not given by default 99.5%.

5. If the input composition is between the requested purities for the bottom and top products, then get the ideal number of stages N_{ideal} for the column from the table of pre calculated values Table B.6 in Appendix.

6. Set the feed plate location of the column to be $N_F = (1\text{-}D_x)N_{ideal}$ (plate one is the top plate of the column).

Reverse simulation procedure for kinetic reactor PGs based on attainable region concept (Glasser et al., 1987) – excerpted from Metzger, Glasser, Glasser, Hausberger, and Hildebrandt (2007)

Determining the candidate attainable region for a given reaction scheme involves the following steps: (a) selecting the fundamental processes, (b) choosing the state variables, (c) defining and drawing the process vectors, (d) constructing the region, (e) interpreting the boundary as the process flow sheet, and (f) finding the optimum.

(a) Choosing the fundamental process: In this step, the possible physical and chemical phenomenon possible such as reaction mixing, separation and mass transfer etc. are selected.

(b) Choosing the state variables: In this step, the state variable that characterize the state of the system are selected. These variable must be sufficient to describe the dynamics of all the fundamental process occurring within the system and must include all the variables in the objective function

(c) Defining and drawing the process vectors: A process vector is one that gives the instantaneous change in system due to the occurrence of a fundamental process. For mixing, this vector gives the divergence from the current state, **c**, based upon the added state, **c***, or **v**(**c**, **c***) = **c*** – **c**. For reactions, vector, **r**[CA,CB], will give the instantaneous direction and magnitude of change from the current concentration position. The plug flow reactor has a defining equation given by $\frac{dC}{d\tau} = r(c)$ and $c(t=0) = c^0$. For a continuous stirred tank reactor, the same is given by $c - c^0 = r(c)t$.

(d) Construction of attainable region: According to Glasser et al. (1987), in order for some region to be a candidate for the attainable region, it is required that the following necessary conditions hold true for the Attainable region candidate.

1. The region includes all feed points.
2. No process vectors on the boundary of the region point out of the region
3. No stationary point with mixing as a process exists in compliment to the region.

With these conditions in hand, the general steps to construct an attainable region are as follows. The attainable region candidate is constructed using a growth approach and a trial and error method. At each stage the necessary conditions are checked to look for possible extensions to the region (excerpted from d'Anterroches (2006)).

1. The first step in finding the AR will be to find a CSTR curve PFR trajectory using the system feed.
2. As the region is not convex, allow mixing between the points. that can be achieved by the PFR(s). This process is finding the convex hull of the curves
3. Check whether any reaction vectors point out of the surface of the convex hull. If the reaction vector points out wards over certain regions, then find the CSTR(s) with feed points in the convex hull that extend the *AR* the most. If no reaction vectors point outwards, then check whether the third necessary condition is met. If they are not met, extend the region using the appropriate reactor
4. Find the new, enlarged convex hull. If a CSTR lies in the boundary at this stage, the reaction vectors must point out of the region, and the PFR with feed points

on the CSTR will extend the region. Extend the region by finding the convex hull with these PFR´s included.

5. Repeat the last two steps, alternating between PFR's and CSTR's, until no reaction vectors point out over the region, and necessary third condition is met.

(e) Interpretation of the boundary: Once we have attainable region in hand, to identify the process flowsheet required to obtain a particular product we trace the path to get to the particular point of concern. This allows us to determine the correct sequence of fundamental processes required to achieve the specified product. There is usually only one path to a particular point on the boundary.

(f) Finding the optimum: With the attainable region fully determined, the optimal value for any objective function may be determined in a straight forward manner along with the best process design.

A final point of note is the AR analysis does not guarantee the determination of the complete attainable region. The analysis is composed of guidelines for the creation of a candidate attainable region, as no mathematically derived sufficiency conditions exist