

Computer Aided Molecular Design (CAMD): Designing Better Chemical Products

Dr. Kevin G. Joback
Molecular Knowledge Systems, Inc.

Abstract

Today's solvents, lubricants, heat transfer fluids, and coatings will all be replaced within the next 25 years. A new generation of chemical products with superior performance and low or zero environmental impact will take their place. Specialty chemicals tailored for specific applications will be produced in small quantities and have market lives of only a few years. This vision may be hard to believe but consider the past 25 year history of CFC refrigerants or chlorinated solvents. One of the factors driving this vision into reality is the continuing advances made in the computer aided design of molecules. Most of these software tools assist the chemist in designing chemicals with desired structural properties. Now new tools are available to assist chemical engineers in designing chemicals with desired physical properties.

Introduction

Today many companies are undertaking process improvements to reduce waste and prevent pollution. Zero emission valves, improved maintenance, and enhanced quality control are just some process modifications which successfully minimize pollution.

In addition to these very visible process improvements, a small but growing number of companies are investigating product modifications to minimize pollution. Environmentally compatible surfactants, ozone-safe refrigerants, and biodegradable hydraulic fluids are just some of the new products being developed. Combining process and product modifications have resulted in environmentally compatible solvents for reaction, separation, and cleaning.

However, a product's environmental compatibility is important but not sufficient to ensure pollution prevention. Life-cycle analysis teaches us that we must consider synthesis, performance, and disposal to understand the full environmental impact of a chemical product. Water is an excellent example of a solvent which is extremely environmentally compatible. However, water's high enthalpy of vaporization often makes its processing extremely energy intensive resulting in the secondary pollution associated with power generation. Water's propensity for promoting corrosion often requires the use of corrosion inhibitors which often have considerable environmental impact.

Designing a product which minimizes waste throughout its entire life cycle requires trading off between multiple objectives, understanding the relationships between molecules and properties, and a thorough understanding of the processes and applications in which the chemical products will be used. Chemical engineers have good experiences in these areas and thus are very well suited to undertake this task of designing better chemical products.

Computer-Aided Molecular Design (CAMD)

Software tools greatly assist with designing a chemical product which must satisfy the multiple specifications of high performance and environmental compatibility. When most chemical engineers think about molecular design they envision three dimensional proteins displayed in vivid computer graphics on high performance workstations. This class of CAMD software is very common in pharmaceutical research and provides great insight into the structure and activity of drugs.

A new class of molecular design, oriented more towards chemical engineering problems, has developed over the last several years [1,2]. This class of CAMD software focuses on three major design steps:

1. Identifying target physical property constraints. If the chemical must be a liquid at certain temperatures we can develop constraints on melting and boiling points. If the chemical must solvate a particular solute we can develop constraints on activity coefficients.
2. Automatically generating molecular structures. Using structural groups as building blocks CAMD software generates all feasible molecular structures. During this step we can restrict the types of chemicals designed. We could eliminate all structural groups which contain chlorine or we might require that an ether group always be included.
3. Estimating physical properties. Using structural groups as our building blocks enables us to use group contribution estimation techniques to predict the properties of all generated structures. Using group contribution estimation techniques enables CAMD software to evaluate new compounds.

These three steps of CAMD are described further using the example of extracting phenol from waste water stream.

Solvent Design Example

To demonstrate the three steps of CAMD we use an example of designing an extraction solvent. This example was taken from the open literature [3].

A number of processes, including the production of certain polymers, produce aqueous waste streams containing quantities of phenol. The traditional route for removing phenol from the water stream was extraction with toluene. Figure 1 shows a typical extraction column's conditions. Since toluene is continuing to be strongly regulated, finding an environmentally friendlier solvent is highly desirable. We will use the three steps of CAMD to search for a toluene substitute.

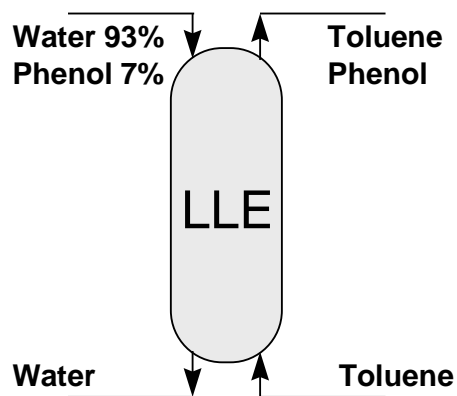


Figure 1: Solvent Design Example

An extraction solvent must satisfy numerous property constraints. The selectivity and capacity for the solute must be high, the density should be significantly different from the parent liquor to facilitate phase separation, and the vapor-liquid equilibrium with the solute should promote easy solvent recovery.

To satisfy these property constraints it is often easy to simply specify that the substitute should have the same properties as the original solvent. Using this approach and focusing only on a single constraint for this example our CAMD task is to find a new chemical which has the same selectivity for extracting phenol from water as does toluene.

To quantify selectivity we can use activity coefficients, infinite dilution activity coefficients, or solubility parameters. Although they are less accurate, solubility parameters are very simple to use and present a very simple example. Using solubility parameters, our CAMD target is

$$SP_d = 16.4$$

$$SP_p = 8.0$$

$$SP_h = 1.6$$

where SP_d , SP_p , and SP_h are respectively the dispersive, polar, and hydrogen-bonding solubility parameters in units of $\text{MPa}^{1/2}$. Since equality constraints will never be matched exactly we add a small tolerance to each target value.

The second step in our CAMD procedure is to generate molecular structures. Instead of starting with individual atoms we start with structural groups. Groups are more expressive than atoms enabling us to restrict chemical families such as acids, alcohols, and unsaturated compounds. Groups also enable us to use group contribution estimation techniques to estimate properties.

Table 1 lists the groups used to generate molecular structures. Halogenated groups were not considered because of environmental concerns. Acidic groups were not considered because of corrosion concerns. Our CAMD software examines all molecular structures which can be generated from the resulting table of 16 groups. The software begins by selecting all combinations of two groups, then all combinations of three groups, and continues up to selecting all combination of ten groups. This limit of ten groups is an upper limit we set.

Table 1: Groups Used to Generate Structures

-CH3	-CH2-	>CH-	>C<
H2C=CH-	-CH=CH-	H2C=C<	>C=C<
-HC=C<	-O-	>C=O	O=CH-
-COO-	>NH	-CN	-NO2

In total 5,311,718 molecular structures can be generated. However, many of these group combinations can not form feasible molecular structures. For example, the three groups



can not be connected into a feasible structure. To address this problem CAMD software uses a number of constraints from the mathematical field of graph theory. Molecules can be represented as simple connected graphs. Such graphs must satisfy the following constraint:

$$b/2 = n + r - 1$$

where b is the number of bonds, n is the number of groups, and r is the number of rings in the resulting molecule.

For our set of three carbon groups shown above we have b equal to 6, n equal to 3, and r equal to 0. CAMD software can test these values and find that they do not satisfy the above equality and therefore can not form a feasible molecule.

Using a group basis and graph theoretic constraints, CAMD software can quickly generate many feasible molecular structures. The third step in our CAMD procedure is to evaluate the properties of these structures and compare them to our property constraints developed in step 1.

The group basis of CAMD software enables us to use group contribution estimation techniques to predict the properties of new structures. Table 2 shows some of the contributions toward the solubility parameters for some of the groups used in our structure generation.

Table 2: Sample Group Contributions

Group	Contributions		
	SPd	SPp	SPh
-CH3	0.344	-0.591	-0.847
-CH2-	0.267	-0.377	-0.595
-CH=CH-	-0.566	-0.034	-0.775
-O-	-0.637	2.315	1.804
>C=O	-1.145	4.670	4.846
O=CH-	-1.114	5.922	5.256
-COO-	-0.861	4.729	4.012
>NH	-1.074	3.875	2.772

Group contribution techniques assume that each group contributes a certain value to the molecule's property. Totaling these contributions leads to a property estimate. For example, to estimate the solubility parameters of ethyl acetate we total the contributions for each group:

Group	Contributions		
	SPd	SPp	SPh
-CH3	0.344	-0.591	-0.847
-CH2-	0.267	-0.377	-0.595
-COO-	-0.861	4.729	4.012
-CH3	0.344	-0.591	-0.847
Intercept	13.290	5.067	7.229
Total	13.384	8.236	8.950

An intercept term is added and the estimate is generated. These estimates agree very well with the literature values of 13.4, 8.6, and 8.9 MPa^{1/2}[4].

Our CAMD software can now estimate the properties of each of the feasible structures generated in step 2. Those structures which satisfy the property constraints identified in step 1 are the toluene substitutes we are searching for.

For this particular example one of the CAMD generated solvents, butyl acetate, matched the solvent chosen as the toluene substitute in the plant.

Synapse CAMD Software

Synapse is a CAMD software package capable of rapidly generating and testing billions of candidate molecular structures. Synapse allows the user to enter their own groups and estimation techniques thus tailoring the program to the specific application being investigated.

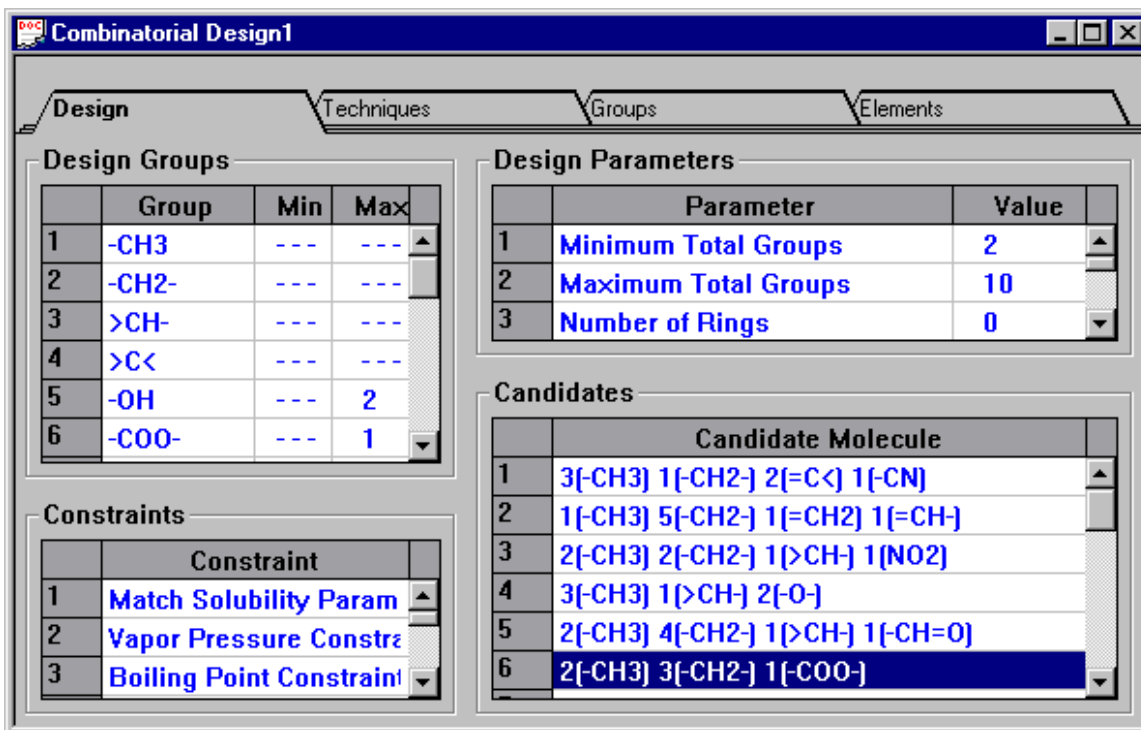


Figure 2: Synapse Computer-Aided Molecular Design Software

Contact Information

For discussions about this paper you may contact Dr. Kevin G. Joback at Molecular Knowledge Systems, Inc., PO Box 10755, Bedford, NH 03110-0755, USA, Phone: 1-603-472-5315, FAX: 1-603-472-5359, eMail: kevin@molknow.com. Additional information on Synapse can also be found at Molecular Knowledge Systems' web site: <http://www.molknow.com>.

References

1. K.G. Joback and G. Stephanopoulos. Searching Spaces of Discrete Solutions: The Design of Molecules Possessing Desired Physical Properties. In *Advances in Chemical Engineering*, Volume 21. 1995.
2. R. Gani and E.A. Brignole. Molecular design of solvents for liquid extraction based on UNIFAC. *Fluid Phase Equilibrium*. Volume 82, 1993.

3. A.E. Hodel. Butyl Acetate Replaces Toluene to Remove Phenol from Water. Chemical Processing. March 1993.
4. A.F.M. Barton. Handbook of Solubility Parameters. CRC Press. 1983.