Mathematical Models for Combinatorial Chemistry and Molecular Structure Elucidation

Abstract

Mathematical Models are indispensable in nearly all subjects of science and technology. More and more problems from organic chemistry are simulated and solved using mathematical modeling.
The motivation is the search for new drugs and materials with certain desired biological–pharmaceutical or physico–chemical properties. If such a compound is synthesized and found, another important task is the determination of the *molecular structure* of the usually unknown substance. For this purpose you will use certain physico–chemical properties obtained from *spectroscopic* methods.
When searching for new drugs and materials techniques from *combinatorial chemistry* are increasingly applied. Starting from several sets of chemical building blocks all combinations are synthesized and then tested for biological–pharmaceutical activity or examined for desired physico–chemical properties. The enormous advantage of this method is the ability to automate the synthesis and screening. Using these methods high throughput rates can be realized, but the prohibitive costs involved makes the automative planning and evaluation of such experiments necessary.
Optimization of experiments in combinatorial chemistry and automated molecular structure elucidation gives rise to various problems, which can be solved using mathematical modeling followed by algebraic combinatorial construction algorithms, graph theoretical invariants and statistical learning techniques. Most important goals are

- Structure generation:
  In combinatorial chemistry methods for the construction of virtual combinatorial libraries are needed. Often such structural spaces are defined by *reactants* and *reactions*. In this work algorithms for *reaction based* structure generation are described.
  For molecular structure elucidation, methods are used that are able to generate possible structural formulae of the analyte starting from its *brutto formula* considering *structural restrictions*. Strategies for brutto formula based structure generation are introduced.
  Important methods for the aforementioned are *canonical labeling* and *orderly generation*.

- Relating structure and property:
  In order to predict properties of structures in virtual combinatorial li-

braries, *quantitative structure property relationships (QSPR)* are used. These are calculated based on a precompiled and screened smaller real library.
Computer aided structure elucidation (CASE) seeks to determine the molecular structure of an unknown chemical compound starting from its spectroscopically measured properties.
Mathematical tools for these tasks are *molecular* and *spectral descriptors* respectively along with *statistical learning*.

In the first part of this thesis the required models and methods are described: The representation of chemical compounds, molecular *substructures* and chemical *reactions*, *orderly* and *constrained generation* for molecular formula based structure generation, *kernel ligand attachments* and construction by *network principle* for reaction based structure generation, as well as different techniques for *supervised statistical learning*, including *classification* and *regression* using *linear models*, *artificial neural networks*, *support vector machines*, *decision trees* and *k nearest neighbors*.
The second part contains concrete applications for problems in chemistry: Combinatorial libraries are generated, for instance by *Ugi's* seven component reaction. Using different molecular descriptors (*arithmetical*, *topological* and *geometrical indices*, *substructure counts*) and statistical learning QSPR are determined, compared and applied for prediction. Problems considered are *boiling points* of *decanes*, *physical densities* of *propylacrylates* and the *antibacterial activity* of *quinolones*. In this work the purpose of CASE is the *interpretation* and *verification* of *mass spectra*. Ranking functions for molecular and structural formulae are defined and tested. Different classification techniques for the determination of MS classifiers are compared. Advanced studies examine among others possibilities of *high resolution* MS.

Keywords: structure generation, combinatorial chemistry, molecular descriptor, quantitative structure property relationships, molecular structure elucidation, mass spectra classification.