A New Optimization Model for Computer-Aided Molecular Design Problems

Computer-Aided Molecular Design (CAMD) is a method to design molecules with desired properties. That is, through CAMD, it is possible to generate molecules that match a specified set of target properties. CAMD has attracted much attention in recent years due to its ability to design novel as well as known molecules with desired properties. The attention is in particular targeted at the design of chemical based products, such as solvents, refrigerants, active pharmaceutical ingredients, polymers, surfactants, lubricants, and more [1]. Property prediction methods are needed in molecular design, as they enable the prediction of the target properties of the generated candidate molecules from there structural information. Here, CAMD methods can be regarded as the reverse engineering approach to property prediction, as the target properties are known while the molecules that match them need to be determined. In this way, CAMD problems can be formulated as a Mixed Integer Linear/Non-Linear Program (MILP/MINLP). With the advent of connectivity-based prediction methods, several researchers have developed new strategies for embedding it with CAMD method. Constantinou et al. [2] proposed a systematic strategy for generating isomers from a set of groups. Harper et al. [3] proposed a framework for CAMD method, where the pre-design phase defines the basic needs, the design phase determines the feasible candidates (generates molecules and tests for desired properties) and the post-design phase performs higher level analysis of the molecular structure and the final selection of the product. Samudra and Sahinidis [4] proposed a new optimization model using relaxed property targets and refined property targets with structural corrections. It is usually difficult to model and solve the MILP/MINLP problem with structure information considered due to the increased size of the mathematical problem and number of alternatives. Thus, decomposition-based approach is proposed to solve the problem. In this approach, only first-order groups are considered in the first step to obtain the building block of the designed molecule, then the property model is refined with second-order groups based on the results of the first step. However, this may result in the possibility of an optimal solution being excluded. Samudra and Sahinidis [4] used property relaxation method in the first step to avoid this situation, but it is not always easy for the users to find the appropriate relaxations. On the other hand, the feasible region of the optimization problem will become larger when relaxations are applied, which makes the solution of the problem harder. In this paper, a new model for CAMD problems is proposed. The model has been developed for the consideration of higher order groups in the molecular generation step of CAMD through mathematical optimization [5]. The model can consider both first and second order groups simultaneously in the MILP/MINLP formulation through a set of mathematical constraints. Structural constraints are defined through a set of linear mathematical equations for the feasible generation of molecules and the connectivity of molecular groups through the adjacency matrix. Property constraints are defined from a set of linear constraints based on the group contribution method [2]. The structural information of the molecule is obtained from the solution of the adjacency matrix. The adjacency matrix provides the adjacent connectivity of first order molecular groups. From this, the second order group description is found, which increases the structural information and property prediction accuracy. This will avoid the possible situation in which the optimal point is excluded from the feasible region due to inaccurate property prediction and ensures the obtainability of a global optimal solution. The model is implemented into a GAMS-based environment for the efficient optimization of a given problem. The model applicability will be demonstrated through the solution of a range of product design problems from literature, such as design of simple molecules (solvents and refrigerants) to design of complex molecules (polymers, lipids and surfactants).