Summary

'*Structure vs. Property: Algorithms and Models*' thesis has been written as a transdisciplinary approach joining together concepts of mathematics and informatics to solve problems in chemistry. The main problem treated is the use of the chemical structure to explain measured (physical, chemical) properties and (biological) activities of the compounds.

We use models and algorithms all the time, but the most difficult is not to use, but to design and/or recognize them. Having this in mind, the thesis has been written as training for their recognition and design, and therefore it may seem too structured.

First chapter, *From observation to experiment, and to model*, comes from a introspection about the experimental design, the one which we provide raw data, and from which we extract models, patterns and relations between our observables and may be considered the basis for the following.

The second chapter, *Representation with formulas and properties*, defines our researching material while the third chapter, *Representation levels* synthesizes the key links between the structure and the property of the chemical compounds. An original result is given at the end of its, and is about the link between the spread of the chlorophylls in living organisms and their efficiency in converting the solar energy.

The fourth chapter, *Molecular topology and geometry*, defines the informational representation of the chemical structure and is a bridge to our methods of research. It contains also a series of original results about the developing and the use of molecular topology tools to investigate the chemical structure.

The fifth chapter, *Theory of chemical reactions and equilibrium*, discuss the second using of mathematical modelling in chemistry, for study of the reaction kinetics. Software implementations easy to do are provided.

The sixth chapter, *Separation in solvent mixtures: Snyder triangle's method*, provides the link between the theory of solutions and the use of models to estimate the optimum mobile phase in chromatography. Applications of the models derived here are given in Chapter 16.

The seventh chapter, *Binary and multinomial variables analysis: binomial confidence intervals*, is a turning back to the challenge of measurement function use, discussed in the first chapter. Namely, this chapter (containing original results too), takes the first level of the measurement function (on binary and multinomial scales) and provides solutions to express the confidence for expressions containing binary ratios (when are used to measure the association between observables).

Chapter 8, *Ordinal variables analysis: ranks statistic*, takes the second level of the measurement function (on ordinal scales) and provides solutions to express the strength of the

monotonic association.

Chapter 9, *Linear regression analysis: linear models*, takes the last level of measurement (on ratio scales) and provides analysis alternatives for the most important model of association: the linear association. Two original results are given here: about the generalization of the Gauss and Laplace distributions and its use to assess the type of the experimental error, and the other is about the alternatives in linear association analysis.

Chapter 10, *Structure-activity relationships: families of descriptors method*, gives my main contribution at developing of the methods of analysis the link between the structure and the property of the chemical compounds.

Chapter 11, *Reducing computing complexity: projecting and using of genetic algorithms*, gives my main contribution at the design of genetic algorithms for heuristic finding of structure-activity relationships.

Chapter 12, *Population distribution analysis applications*, gives a series of original results obtained involving the analysis of population distribution from which a sample was drawn.

Chapter 13, *Graph theory and statistical analysis applications*, gives a series of original results in graph theory, information theory, and applied statistic, as well as their use to assess the relation between the structure and the property of the chemical compounds.

Chapter 14, *Software design applications*, briefly describes the results from development and assessment of the software for conducting researches.

Chapter 15, *Meta-heuristics applications*, points out the results obtained from the use of a contingency of evolution strategies to the study of the evolution supervised by genetic algorithms.

Chapter 16, *Analysis of factors in separating from solvent mixtures*, point out the results obtained from the analysis of intrinsic factors occurring at the chromatographic separation of chemical compounds.

Chapter 17, *Chemicals dynamics applications*, discuss four original results obtained when dynamics of chemicals was the subject of research.

Following three chapters (18, 19, and 20, named *Structure-activity relationships with families of molecular descriptors (1)*, *(2)*, and *(3)*) structures the main results obtained from the use of the families of molecular descriptors to provide relations between the structure and the property of the chemical compounds.

Chapter 21, *Subject of interest: design of algorithms and models*, explain my plans for further development on this assumed key subject: design of algorithms and models, implementation of programs and obtaining of equations for relating the structure with the property.

Chapter 22, *Action plans*, gives a series of identified actions, and thus probable to follow the assumed plan of development.