

Cheminformatics B.S

1. General description and characteristics of the program.

Cheminformatics integrates a comprehensive knowledge of chemistry with an extensive understanding of information technology. The intersection of chemistry and information technology embraces an expanding territory; it includes computational modeling of individual molecules, thermodynamic methods of estimating chemical properties, methods of predicting biological activity of hypothetical compounds, and organization and classification of chemical information. Students in this program will complete upper division coursework in both chemistry and computer science. Graduates will be prepared to work with chemical databases, computational chemistry, and modeling of physicochemical and biochemical activities of chemical compounds.

2. Rationale

Chemistry is no longer exclusively a laboratory science. With over 24 million chemical substances known (CAS registry) computer methods are increasingly essential to anyone who works with chemistry at any level from students to specialized researchers. The basic curriculum in chemistry includes an introduction to the use of chemical databases; it does not cover the mathematical and computation methods underlying the generation and organization of chemical information.

Advances in theoretical and computational chemistry now allow chemists to model chemical compounds “in silico” with ever-increasing accuracy. Molecular properties now becoming accessible through computation include molecular shape, electronic structure, physical properties, chemical reactivity, protein folding, structures of materials and surfaces, catalytic activity, and biochemical activities.

In the realm of detailed quantum-level modeling of individual molecules, reaction pathways, and solvent interactions computational chemists are still making progress on the problem posed by Dirac in 1929:

“The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved.”

People working in this field must thoroughly understand computer programming as well as fundamental physical chemistry.

An important aspect of cheminformatics is the development of (quantitative) structure activity relationships, QSARs or SARs, which are used to predict chemical behavior from molecular structure. Behaviors that are predicted in this way include physical properties (e.g. boiling point, vapor pressure, aqueous solubility, hydrophobicity, dipole moment), reactivity (e.g. hydrolysis, oxidation rates), and bioactivity (e.g. inhibition of an enzyme, antibiotic activity, toxicity). These methods are typically based in chemical thermodynamics coupled with statistical methods.

Biological and environmental systems are complex mixtures of interacting chemical species. In modeling the natural world at the level of a cell or a planet vast numbers of compounds must be accounted for, along with their equilibrium and kinetic behavior in multiphase environments. Because equilibrium and kinetic equations are non-linear, numerical methods are normally required to handle these problems.

Additional computational challenges lie in indexing and classifying the infinite population of chemical compounds that could be synthesized or are already known. Specific indexing and search problems include how to find a compound that might block a specific biological target; how to predict the most efficient synthetic strategy for a desired compound from available precursors; how to employ results of bioactivity tests on a family of molecules to design improved versions; how to name new classes of compounds (e.g. fullerenes and nanotubes).

Currently combinatorial chemists are developing new methods of synthesizing libraries of related compounds on an unprecedented scale. Such libraries can be used to produce huge arrays of materials for investigation of biochemical, catalytic, or material properties. Systems are required to design, catalog, and search these libraries, assess test results in a meaningful way, and integrate new information with existing chemical databases.

Finally, investigations into information storage at the molecular level are underway, bringing to full circle the link between chemistry and information technology.