

## Chapter 58

# Graph Applications in Chemoinformatics and Structural Bioinformatics

**Eleanor Joyce Gardiner**  
*University of Sheffield, UK*

### ABSTRACT

*The focus of this chapter will be the uses of graph theory in chemoinformatics and in structural bioinformatics. There is a long history of chemical graph theory dating back to the 1860's and Kekule's structural theory. It is natural to regard the atoms of a molecule as nodes and the bonds as edges (2D representations) of a labeled graph (a molecular graph). This chapter will concentrate on the algorithms developed to exploit the computer representation of such graphs and their extensions in both two and three dimensions (where an edge represents the distance in 3D space between a pair of atoms), together with the algorithms developed to exploit them. The algorithms will generally be summarized rather than detailed. The methods were later extended to larger macromolecules (such as proteins); these will be covered in less detail.*

### INTRODUCTION

Chemistry space is exceedingly large. Recent estimates put the number of potentially 'drug-like' molecules at anything between  $10^{12}$  and  $10^{180}$  (Gorse, 2006). The overwhelming majority of these molecules never has been, and never will be, synthesized but methods are nevertheless required to determine which of these potential compounds should be made. Some large pharmaceutical/

agrochemical companies maintain corporate databases of millions of molecules. The discovery of New Chemical Entities (NCEs) which may become drugs depends on the successful mining of the information stored in these databases. Such information may be explicit (e.g. the molecules may be annotated with chemical reaction or activity data) or may be implicit in the structure of a molecule. For many years isomorphism algorithms have formed the basis of structural comparison between pairs of molecules in these databases, designed to extract this implicit information.

DOI: 10.4018/978-1-4666-3604-0.ch058

The main purpose of this chapter is to introduce the concept of chemoinformatics to practitioners from the field of graph theory and to demonstrate the widespread application of graph-theoretic techniques to the solving of chemoinformatics problems. However, many graph-theoretic algorithms from chemoinformatics have subsequently been adapted for the structural comparison of macromolecules in the field known as structural bioinformatics. A secondary aim of the chapter is to provide a brief overview of these applications.

The layout of the chapter will now be described. In the Background section some definitions of chemoinformatics are given and the topic is placed within the context of the drug discovery process. Structural bioinformatics is also defined and the necessary graph theoretic notation is introduced. The two main sections deal with the use of algorithms from graph theory in chemoinformatics and structural bioinformatics respectively. In the Chemoinformatics section the reader is introduced to the concept of a molecule as a molecular graph; this concept informs the rest of the section. Then the use of algorithms for graph labeling in order to register molecules is discussed, followed by the use of graph invariants as part of a molecular description. The use of subgraph isomorphism algorithms for molecular substructure searching is considered and the importance of the Ullmann algorithm is emphasized. Next the important concept of molecular similarity is considered along with the reformulation of this problem as a maximum clique problem. Two key clique-detection algorithms in chemoinformatics, Bron-Kerbosch and RASCAL, are considered in more detail. These algorithms are then shown to be important in protein-ligand docking, pharmacophore elucidation and molecular clustering. Reduced graphs for molecular representation and searching are then introduced and example graph reductions are considered. In the Structural Bioinformatics section the necessary elements of protein structure are first presented. Many methods from chemoinformatics have been adapted for use in structural bioinformatics: some

of these are described here, in particular uses of the Ullmann algorithm for substructural searching and the Bron-Kerbosch algorithm for similarity searching and protein-protein docking. The final sections are the Future Directions section, where the need for graph-theoreticians to present algorithms in a way which is accessible to chemoinformaticians is discussed, and the Conclusions section where the chapter is summarized.

## **BACKGROUND**

The first question to be answered is “What is chemoinformatics?” Chemoinformatics (also known as cheminformatics) is an interface science (N. Brown, 2009) which includes expertise from chemistry, biology, physics, mathematics and computer science. No definitive definition of the term “chemoinformatics” has ever been given but, as chemoinformaticians, we know what we mean when we say it. The term chemoinformatics was first used by Frank Brown:

*The use of information technology and management has become a critical part of the drug discovery process. Chemoinformatics is the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and optimization. (F. K. Brown, 1998).*

Other definitions from leading practitioners in the field include:

*Chemoinformatics is the application of informatics methods to solve chemical problems. (Gasteiger, 2006).*

*...the computer manipulation of two- or three-dimensional chemical structures and excludes textual information. This distinguishes the term from chemical information, largely a discipline of*

30 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/graph-applications-chemoinformatics-structural-bioinformatics/76111](http://www.igi-global.com/chapter/graph-applications-chemoinformatics-structural-bioinformatics/76111)

## Related Content

---

### Improving Prediction Accuracy via Subspace Modeling in a Statistical Geometry Based Computational Protein Mutagenesis

Majid Masso (2010). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 54-68).  
[www.irma-international.org/article/improving-prediction-accuracy-via-subspace/49549](http://www.irma-international.org/article/improving-prediction-accuracy-via-subspace/49549)

### In Silico Analysis of the CST6 Tumor Suppressor Gene

Athanasia Pavlopoulou and Georgios Tsaramiris (2013). *International Journal of Systems Biology and Biomedical Technologies* (pp. 42-58).  
[www.irma-international.org/article/in-silico-analysis-of-the-cst6-tumor-suppressor-gene/97741](http://www.irma-international.org/article/in-silico-analysis-of-the-cst6-tumor-suppressor-gene/97741)

### Impact of the COVID-19 Lockdown on the Quality of Higher Education

Danis Amirov (2022). *International Journal of Applied Research in Bioinformatics* (pp. 1-8).  
[www.irma-international.org/article/impact-of-the-covid-19-lockdown-on-the-quality-of-higher-education/290340](http://www.irma-international.org/article/impact-of-the-covid-19-lockdown-on-the-quality-of-higher-education/290340)

### Parameterless Clustering Techniques for Gene Expression Analysis

Vincent S. Tseng and Ching-Pin Kao (2006). *Advanced Data Mining Technologies in Bioinformatics* (pp. 155-173).  
[www.irma-international.org/chapter/parameterless-clustering-techniques-gene-expression/4251](http://www.irma-international.org/chapter/parameterless-clustering-techniques-gene-expression/4251)

### Using Case Costing Data and Case Mix for Funding and Benchmarking in Rehabilitation Hospitals

Grace Liu (2014). *Research Perspectives on the Role of Informatics in Health Policy and Management* (pp. 62-78).  
[www.irma-international.org/chapter/using-case-costing-data-and-case-mix-for-funding-and-benchmarking-in-rehabilitation-hospitals/78689](http://www.irma-international.org/chapter/using-case-costing-data-and-case-mix-for-funding-and-benchmarking-in-rehabilitation-hospitals/78689)