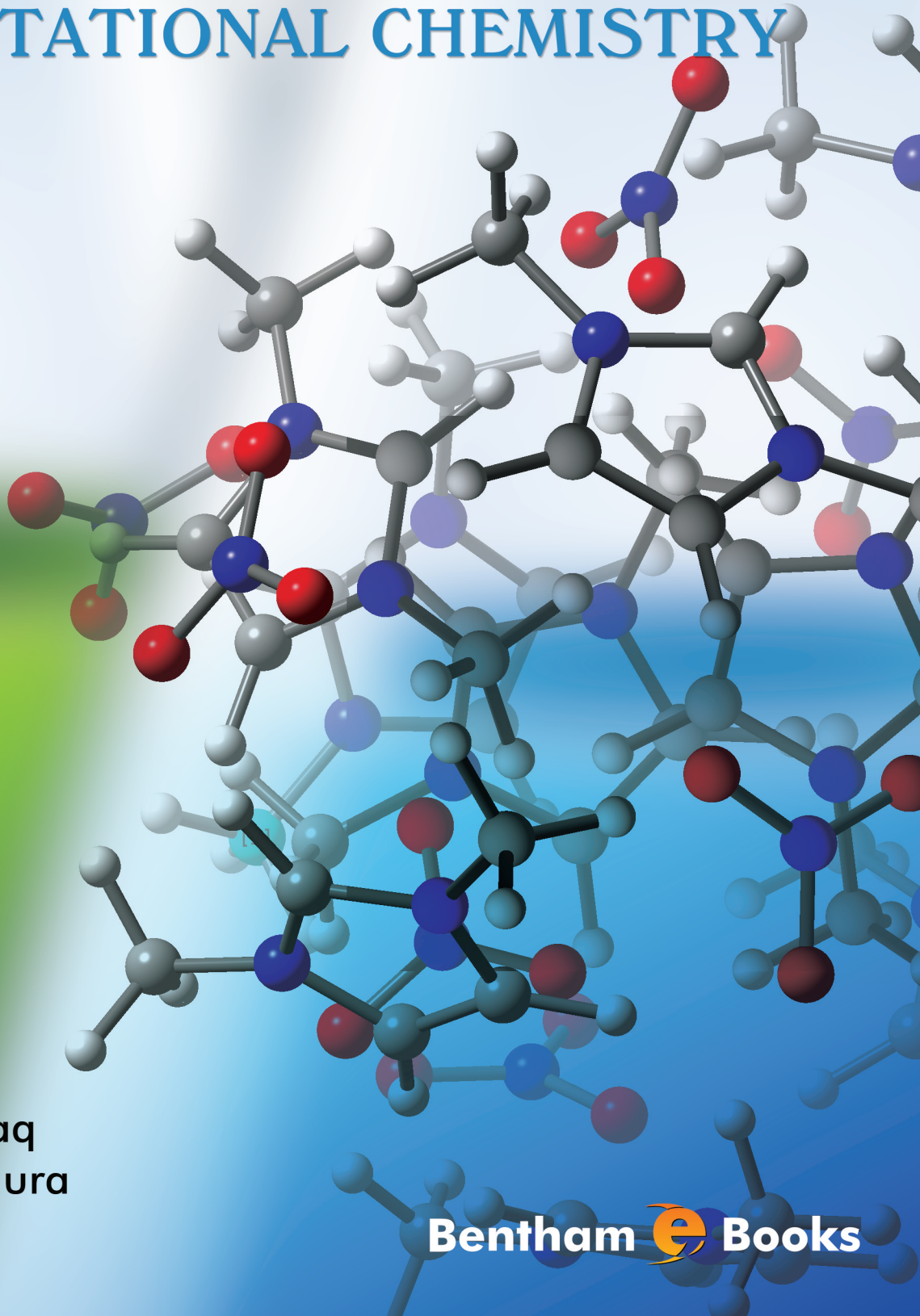
eISBN: 978-1-68108-167-0
ISBN: 978-1-68108-168-7

eISSN: 2352-9458
ISSN: 2352-944X

FRONTIERS IN COMPUTATIONAL CHEMISTRY

(VOLUME 3)



Editors:
Zaheer Ul-Haq
Jeffry D. Madura

Bentham Books

Frontiers in Computational Chemistry

(Volume 3)

Edited by

Dr. Zaheer Ul-Haq
Panjwani Center for Molecular Medicine & Drug Research
International Center for Chemical & Biological Sciences
University of Karachi
Pakistan

&

Dr. Jeffry D. Madura
Department of Chemistry & Biochemistry
Center for Computational Sciences Duquesne University
Pittsburgh
USA

Frontiers in Computational Chemistry

Volume # 3

Editors: Dr. Zaheer Ul-Haq and Dr. Jeffry D. Madura

eISSN (Online): 2352-9458

ISSN (Print): 2352-944X

eISBN (Online): 978-1-68108-167-0

ISBN (Print): 978-1-68108-168-7

First published in 2017.

advertisements or ideas contained in the Work.

Limitation of Liability:

In no event will Bentham Science Publishers, its staff, editors and/or authors, be liable for any damages, including, without limitation, special, incidental and/or consequential damages and/or damages for lost data and/or profits arising out of (whether directly or indirectly) the use or inability to use the Work. The entire liability of Bentham Science Publishers shall be limited to the amount actually paid by you for the Work.

General:

1. Any dispute or claim arising out of or in connection with this License Agreement or the Work (including non-contractual disputes or claims) will be governed by and construed in accordance with the laws of the U.A.E. as applied in the Emirate of Dubai. Each party agrees that the courts of the Emirate of Dubai shall have exclusive jurisdiction to settle any dispute or claim arising out of or in connection with this License Agreement or the Work (including non-contractual disputes or claims).
2. Your rights under this License Agreement will automatically terminate without notice and without the need for a court order if at any point you breach any terms of this License Agreement. In no event will any delay or failure by Bentham Science Publishers in enforcing your compliance with this License Agreement constitute a waiver of any of its rights.
3. You acknowledge that you have read this License Agreement, and agree to be bound by its terms and conditions. To the extent that any other terms and conditions presented on any website of Bentham Science Publishers conflict with, or are inconsistent with, the terms and conditions set out in this License Agreement, you acknowledge that the terms and conditions set out in this License Agreement shall prevail.

Bentham Science Publishers Ltd.
Executive Suite Y - 2
PO Box 7917, Saif Zone
Sharjah, U.A.E.
Email: subscriptions@benthamscience.org

CONTENTS

PREFACE

The branch of chemistry that uses computers to study chemical questions is known as Computational Chemistry which is a very diverse field spanning from the development and application of linear free energy relationships (*e.g.* QSAR, QSPR), to electronic structure calculations, molecular dynamics simulations, and to solving coupled differential equations (*e.g.* drug metabolism). The focus of ***Frontiers in Computational Chemistry*** is to present material for the application of computational techniques used in biological processes. Topics falling under this umbrella include computer aided molecular design, drug discovery and development, lead generation, lead optimization, database management, computer and molecular graphics, and the development of new computational methods or efficient algorithms for the simulation of chemical phenomena including the analysis of biological activity. In this third volume, we have collected five different perspectives on the application of computational methods towards drug design.

Chapter 1 "*In Silico* Approaches for Drug Discovery and Development" reviews the main computational tools used in the drug discovery process. Joseph, *et al.* also presented the application of physics-based methods that are currently being developed and applied to the drug discovery process.

The removal of toxic metal ions from nuclear and chemical waste streams is an imperative and demanding problem. In Chapter 2 "Computational Chemistry Assisted Design and Screening of Ligand-Solvent Systems for Metal Ion Separation" Ali *et al.* review electronic structure methods to aid the design and development of new ligands that can be used to extract metal ions from the environment. The goal is to use electronic structure methods to identify a suitable ligand anchored on a solid matrix that can be used in a complex separation process.

One challenge in the biochemical field is understanding the side effects of anti-cancer drugs containing platinum. The authors of Chapter 3 "Molecular Mechanisms of Cellular Transport, Resistance and Cytotoxic Side Effects of Platinum and Adjuvant Anti-cancer Drugs — A Molecular Orbital Study" present a review of the application of electronic structure methods to understand the side effects, acquired resistance, and combination of platinum drugs with adjuvant drugs in treating cancer.

In Chapter 4 "Elucidating Allosteric Communications in Proteins *Via* Computational Methods", the authors present a review of the application of different normal mode analyses based on molecular dynamics methods to understanding allosteric communication in proteins. Alakent and Ince also present the application of graph theory, perturbation methods, and

statistical methods to investigate allosteric mechanisms.

The authors of Chapter 5 “Information-theoretic chemical space for many electron systems: from atoms to biological and pharmacological molecules” review the utility of an information-theoretic three-dimensional (IT-3D) space to unveil the unique physical, chemical and biological aspects of a great diversity of many electron systems. These multiple electrons systems range from simple atomic systems to more complex systems such as amino acids. Esquivel *et al.* claim that “All chemical families recognized by the existing energy-based classifications are embraced by this entropic scheme”.

Zaheer Ul Haq
Panjwani Center for Molecular Medicine & Drug Research
International Center for Chemical & Biological Sciences
University of Karachi
Pakistan

&

Jeffry D. Madura
Department of Chemistry & Biochemistry
Center for Computational Sciences Duquesne University
Pittsburgh
USA

List of Contributors

Anil Boda	Chemical Engineering Division, Bhabha Atomic Research Centre, Mumbai, India
Ashish Kumar Singha Deb	Chemical Engineering Division, Bhabha Atomic Research Centre, Mumbai, India
Burak Alakent	Department of Chemical Engineering, Bogazici University, Istanbul, Turkey
Clifford W. Fong	Eigenenergy, Adelaide, South Australia, Australia
C. Soriano-Correa	Qu´ımica Computacional, FES-Zaragoza, Universidad Nacional Aut´onoma de M´exico, 09230-Iztapalapa, M´exico, D.F, Mexico
J.C. Angulo	Instituto Carlos I de F´ısica Te´orica y Computacional, Universidad de Granada, 18071-Granada, Spain Departamento de F´ısica At´omica, Molecular y Nuclear, Universidad de Granada, 18071-Granada, Spain
J.S. Dehesa	Instituto Carlos I de F´ısica Te´orica y Computacional, Universidad de Granada, 18071-Granada, Spain Departamento de F´ısica At´omica, Molecular y Nuclear, Universidad de Granada, 18071-Granada, Spain
KalsankaTrivikram Shenoy	Chemical Engineering Division, Bhabha Atomic Research Centre, Mumbai, India
M. Molina-Esp´ıritu	Departamento de Qu´ımica, Universidad Aut´onoma Metropolitana, 09340-M´exico, D.F., M´exico
Pooja Sahu	Chemical Engineering Division, Bhabha Atomic Research Centre, Mumbai, India
R.O. Esquivel	Departamento de Qu´ımica, Universidad Aut´onoma Metropolitana, 09340-M´exico, D.F., M´exico Instituto Carlos I de F´ısica Te´orica y Computacional, Universidad de Granada, 18071-Granada, Spain
S. L´opez-Rosa	Instituto Carlos I de F´ısica Te´orica y Computacional, Universidad de Granada, 18071-Granada, Spain Departamento de F´ısica Aplicada II, Universidad de Sevilla, 41012-Sevilla, Spain
Srinivasaraghavan Kannan	Bioinformatics Institute, A STAR, Singapore 138671,
Sk. Musharaf Ali	Chemical Engineering Division, Bhabha Atomic Research Centre, Mumbai, India
Thomas Leonard Joseph	Bioinformatics Institute, A STAR, Singapore 138671,

Vasanthanathan Poongavanam	Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, DK-5230, Odense M, Denmark
Vigneshwaran Namasivayam	Department of Life Science Informatics, B-IT, Rheinische Friedrich-Wilhelms-Universitaet, Dahlmannstr, 2, 53113 Bonn, Germany
Z. Nevin Gerek Ince	Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, USA

CHAPTER 1

In Silico Approaches for Drug Discovery and Development

Thomas Leonard Joseph[1], Vigneshwaran Namasivayam[2], Vasanthanathan Poongavanam[3] and **Srinivasaraghavan Kannan[1,*]**

[1] *Bioinformatics Institute, A*STAR, Singapore 138671, Singapore*

[2] *Department of Life Science Informatics, B-IT, Rheinische Friedrich-Wilhelms-Universitaet, Dahlmannstr 53113 Bonn, Germany*

[3] *Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, DK-5230, Odense M, Denmark*

Abstract: Discovery of new therapeutics is a very challenging, expensive and time-consuming process. With the number of approved drugs declining steadily combined with increasing costs, a rational approach is needed to facilitate, expedite and streamline the drug discovery process. In this regard computational methods are playing increasingly important roles, largely assisted by developments in algorithms and greatly increased computer power. With *in silico* methods playing key roles in the discovery of growing numbers of marketed drugs, nowadays use of computational tools has become an integral part of most drug discovery programs. Computational tools can be applied at different stages: from target selection through identification of hits to optimization. In this chapter we aim to provide an overview of major tools that have been developed and are routinely being used in the search of novel drug candidates. In addition, we present recent advances, especially in the application of physics-based simulation methodologies, in the drug discovery process for the development of improved therapeutics.

1. INTRODUCTION

Drug discovery is the process of creating or finding a molecule which has a specific activity on a biological organism. The aim of the discovery process is

[*] **Corresponding author Srinivasaraghavan Kannan:** Bioinformatics Institute, A*STAR, Singapore 138671; E-mail: raghavk@bii.a-star.edu.sg

Zaheer Ul-Haq and Jeffry D. Madura (Eds.)

to identify compounds with pharmacological interest that can be used in the treatment of diseases. As several factors decide the activity of a drug molecule, undoubtedly the development of a new drug is a complex and difficult process. It is estimated that a drug discovery process can cost several hundred million dollars and a typical discovery cycle can take as many as 15 years from the first compound identified in the laboratory until the drug is brought to market [1 - 6]. Traditionally drug discovery starts with an experimental screening of compound libraries of molecule that bind to biomolecular targets and modulate their activity. This is followed by subsequent rounds of iterative chemical modifications to enhance their potency, with further optimization for increased selectivity and pharmacological properties [5, 6]. The emergence of combinatorial chemistry combined with rapid developments in high throughput screening (HTS) technologies have speeded up the discovery process by enabling huge libraries of compounds to be screened in short periods of time [7 - 10]. However the hit rates for high throughput screens are often extremely low and most identified hits do not proceed to actual leads [7 - 10].

The sequencing of human genome has revealed unknown proteins that might serve as new drug targets. However the therapeutic importance of most of these proteins is either unknown or poorly characterized. The routine set of experiments (blind expression, purification and *in vitro* assays) that are typically used, cannot be applied for thousands of proteins against libraries of several hundreds of thousands of compounds. Therefore new approaches are needed to speed up and streamline drug discovery and development process to save time, money and resources. In this regard computational approaches have a major role to play.

A variety of computational approaches can be applied at different stages of the drug-design process; right from target identification and validations, identification of initial hits, hit-to-lead selection, and optimization of leads to avoid safety issues.

In this chapter we aim to provide an overview of major *in silico* tools and approaches that have been developed and are routinely being used to search for novel drug candidates. In addition we will also present recent advances (enhancements), especially the application of physics-based simulation

methodologies that lead to a dynamic view of receptor drug interaction, replacing the traditional dogma of single structure-based drug design with the concept of ensemble–based drug design, where conformational flexibility of a receptor molecule plays key roles.

In the first section, we introduce two major Computer Aided Drug Design (CADD) strategies namely ligand based and structure based methods that are widely used in the drug discovery process. Next we briefly introduce several computational techniques that are routinely used. In the third section we will introduce Molecular Dynamics (MD) simulations and applications at various steps of the drug discovery process. We then discuss computational methods for predictions and optimization of drug metabolism and pharmacokinetics. Finally we will discuss targeting protein-protein interactions and briefly introduce peptide based inhibitor design for inhibitions of protein-protein interactions. The goal here is to offer an overview of highly promising themes and tools in this interdisciplinary field.

2. COMPUTER AIDED DRUG DESIGN STRATEGIES

Drug discovery is an extended and time consuming process, which can take several years to translate a compound into a drug molecule. Therefore development of a drug discovery process with the ability for rapid identification of potential binders to the target of therapeutic interest is of great importance in the biotech and pharmaceutical companies. In this regard computational methods enable rapid screening of huge libraries of pharmacologically interesting compounds for identifying potential binders through modelling and simulation. Strategies for CADD vary depending on the availability of structural and other information regarding the target (enzyme/receptor) and the drug (ligand). Two major modelling strategies “indirect” and “direct” are currently used in the drug discovery process (Fig. **1**). In the indirect approach, also known as “Ligand based” the design is based on a comparative analysis of the structural features of compounds with known activity. The direct approach, also known as “Structure based”, utilizes the three-dimensional structural features of the target molecule of interest. We now examine these two in some detail.

CHAPTER 2

Computational Chemistry Assisted Design and Screening of Ligand-Solvent Systems for Metal Ion Separation

Sk. Musharaf Ali*, Anil Boda, Ashish Kumar Singha Deb, Pooja Sahu and **Kalsanka Trivikram Shenoy**

Chemical Engineering Division, Bhabha Atomic Research Centre, Mumbai, 40085-India

Abstract: Computational chemistry that comprises of Quantum mechanics (QM), Monte-Carlo (MC), molecular dynamics (MD) and *ab initio* MD (AIMD) simulation has emerged as a prospective tool for the calculation of various molecular properties by capturing the complex molecular interactions and is being used extensively in the fields of science and engineering. In that context, design and screening of suitable ligands for efficient separation of metal ions is imperative and demanding in view of the safe removal of metal ions from nuclear and chemical waste stream employing the much practiced separation processes namely: solvent extraction and ion exchange. In solvent extraction, an organic solvent containing one or more ligands and in ion exchange, a solid matrix called stationary phase, the surface of which is decorated by functional group/ligands are widely used to remove the desired metal ions from the aqueous medium.

This chapter will focus on designing new ligands or improving the existing ones for understanding the accurate nature of host-guest interactions together with the bonding nature with respect to the donor atom, complexation thermodynamics, conformational features of the ligands and solvent effect. Computational chemistry can play a crucial role for the quantitative predictions of the structural parameters, coordination number, complexation stability and selectivity of metal ion –ligand complexation with respect to the solvent extraction and ion exchange processes. Use of ionic liquid as a novel and alternative solvent to common organic solvent has also been discussed along with its

* **Corresponding author Sk. Musharaf Ali:** Chemical Engineering Division, Bhabha Atomic Research Centre, Mumbai, 40085-India; Tel: +91-022-25591991; Fax: +91-022-25505151; E-mail: musharaf@barc.gov.in

Zaheer Ul-Haq and Jeffry D. Madura (Eds.)

dual extraction mechanism. Further, how Quantum electronic structure calculation can be fruitfully used for selecting the suitable ligand anchored on a solid matrix for a complex separation process has been demonstrated.

Keywords: AIMD, BE, Born-Haber Thermodynamic cycle, Coordination number, COSMO, DFT, Entropy, Free energy, Geometry, Ion exchange, MP2, Partition coefficients, Radial distribution function, Separation factor, Solvent extraction.

1. INTRODUCTION

The design of suitable molecular ligands for efficient separation of metal ions is imperative and demanding in view of the secure discarding of nuclear and chemical waste in particular and value recovery of metals in general by means of the available extraction technologies [1, 2]. One of the most regularly used technologies for metal ion separations is the extraction of the metal ions from the aqueous solutions with an organic solvent containing the ligand. From the viewpoint of designing new ligands or improving the existing ones, understanding the accurate nature of ion-ligand interactions together with the binding energies, conformational features of the ligands and solvent effect would be very helpful. In this regard, quantum chemical computational software with the aid of high performance parallel computers plays a crucial role for the quantitative predictions of the stability, structural parameters, selectivity and is being applied fruitfully in the various separation processes.

The power of molecular modelling is growing rapidly with the continuing development of computer power, robust algorithms, and the availability of software [3]. Molecular modelling can provide useful estimates of the properties and behaviour of materials-even before they have been synthesized and useful estimates of the parameters and behaviour needed to do traditional chemical engineering process development and design [4]. From experimental point of view, choosing a suitable solvent/extractant from a myriad of solvents and extractants is a very time consuming and tedious affair. It will be of great help if the screening is done beforehand by means of other less time consuming and easy techniques. Hence, the solubility and partition coefficients of various extractants

with different donor atom, substituent, conformation and type of diluents can be easily estimated.

In this chapter, we have critically evaluated the structures, energetic and thermodynamics of complex solution phase pertaining to solvent extraction and ion exchange processes using varieties of *ab initio* density functional theory and MD simulation with an aim to demonstrate the selection of suitable computational methods for practical metal ion separation process of interest. We mainly focus on the results of our recently studied molecular systems pertaining to the separation of metal ions which are important for nuclear establishment with the aid of different macrocyclic and chelating ligands and ligand functionalized CNTs. We mainly discuss the complexation of ions which are of importance in nuclear technology: *viz.* Lithium, Sodium, Cesium, Strontium, Thorium, Uranium, Plutonium, Zirconium, Hafnium, Europium and Americium with ligand/solvent like crown ether, calix-crown, diglycolamide, DGA functionalized CNT, dodecane, ionic liquid *etc.*

2. COMPUTATIONAL METHODOLOGY

There is a wide variety of computational methods starting from semi-empirical to *ab initio*, each having its own merits and demerits in terms of cost, time and accuracy for a particular application [5]. Semiempirical methods can be useful for initial screening of the molecular system of interest [6]. Among *ab initio* methods, though Hartree Fock (HF) is considered to be the cheapest it has some serious limitation due to inability to handle the electron correlation [7]. The Moller Plesset perturbation (MP_n) [8] and couple cluster singles and doubles (CCSD) [9] methods are quite accurate but heavily expensive and hence can be restricted to small molecular system. However, density functional theory (DFT) [10] based methods, which earlier was considered to be *ab initio* has partly lost its *ab initio* credential due to large number of parameterization of the exchange-correlation functional but still is the work horse for large molecular system. There is a wide variety of DFT functional one can select for a specific interest of application. Similarly, the size of the basis set can be chosen depending on the molecular properties to be evaluated. Clearly, there is a trade off between the functional and size of the basis set depending on the molecular size and chemical properties. In

CHAPTER 3

Molecular Mechanisms of Cellular Transport, Resistance and Cytotoxic Side Effects of Platinum and Adjuvant Anti-cancer Drugs – A Molecular Orbital Study

Clifford W. Fong*

Eigenenergy, Adelaide, South Australia, Australia

Abstract: The side effects, acquired resistance, reversal of resistance, and combination of Pt drugs with adjuvant drugs has been examined using an extensive review of the literature and molecular orbital computations. It is concluded that Pt chemo-therapeutical regimes are dominated by side reactions, particularly hydrolysis in blood serum and delivery efficiency. For example, it is shown that transplatin is therapeutically inactive because it hydrolyses faster in blood serum than cisplatin, so little transplatin reaches its target DNA. The reactivity of charged hydrolysis products determines the severity of side effect. The reactivity determining properties of the approved Pt drugs and their various hydrolysed species are calculated. The cellular uptake of Pt and adjuvant drugs is fastest for neutral species, with high lipophilicity, since their desolvation penalties for crossing the cell membrane are lowest. Pt resistant cell lines generally have lower levels of drug uptake, indicating this is a dominant first order cause of cellular resistance. Resistance can be caused by complexation of charged Pt species to phosphatidylserine (PS) headgroups, or by Pt complexation to the inner terminii of the trans-membrane pore of the hCtr1 transporter. Drugs that are known to reverse induced Pt resistance can decomplex the Pt-PS or Pt-hCtr1 complexes, restoring normal PS or hCtr1 functions. Molecular biophysical properties of adjuvant drugs used in combination with Pt drugs *(e.g.,* paclitaxel, doxorubicin, gemcitabine, Folfox) can assist the clinical evaluation of combinatorial Pt based chemotherapeutic regimes. These drugs have similar properties to the approved Pt drugs, and fit into the same "therapeutic window" in sterms of their likely cellular uptakes and reduction

* **Corresponding author Clifford W. Fong:** Eigenenergy, Adelaide, South Australia, Australia; E-mail: cwfong@internode.on.net

Zaheer Ul-Haq and Jeffry D. Madura (Eds.)

potentials. The role of free radical species involved in Pt drug induced apoptosis *via* electron transfer from the guanine base of DNA, and the reactions of oxidizing hydroxyl radicals and reducing hydrated electrons reactions with cisplatin, transplatin and carboplatin under physiological pH and Cl ion condition have been examined.

Free radical sensitisers such as TMPD and TETA when combined with Pt drugs may allow targeting of the more hypoxic environments in solid cancerous tumours (and less damage to normal cells). The calculated biophysical parameters ionization energy and electron affinity, which are related to the redox environments found in cells, may be useful predictors of likely cytotoxic efficacy of combination therapies involving sensitisers and Pt drugs.

Keywords: Adjuvant drugs, Anti-cancer, Cellular uptake, Combinatorial chemotherapies, Platinum drugs, Resistance, Reversal of resistance, Side effects.

OBJECTIVES

Cytotoxic side effects of antineoplastic chemotherapeutic Pt drugs: to examine biophysical molecular properties that could be used to predict the propensity of Pt drugs to cause unwanted side effects.

Resistance to Pt drugs: to examine mechanisms involved in acquired resistance, particularly the cellular accumulation mechanisms: (a) passive diffusion, active transport including copper transporters, organic cation transporters; (b) active efflux and intracellular reductive processes.

Reversal of platinum resistance: to develop predictive criteria for drugs that can reverse acquired resistance.

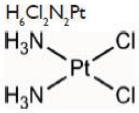
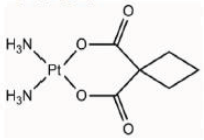
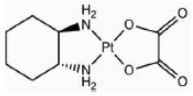
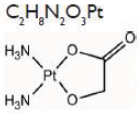
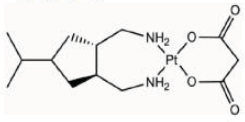
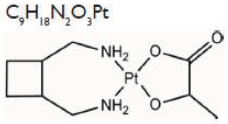
Combination of Pt chemotherapy with other adjuvant anti-tumour drugs and sensitisers: to develop predictive criteria that may serve to guide overall therapeutic efficacy.

1. INTRODUCTION

Chemotherapy, surgery and radiation are the main treatments for cancer. Platinum chemotherapeutics are the most widely prescribed drugs in modern oncology (administered to about 50 % of all cancer patients), either alone or in combination

with other anti-cancer drugs and/or radiation therapy. Platinum drugs are used for the treatment of a broad spectrum of specific cancers, including testicular, ovarian, bladder, head and neck, esophageal, small and non-small cell lung, breast, cervical, stomach and prostate cancers, as well as Hodgkin's and non-Hodgkin's lymphomas, neuroblastoma, sarcomas, multiple myeloma, melanoma, and mesothelioma. One of the greatest successes in chemotherapeutics was the advent of cisplatin treatment of metastatic testicular cancer where the survival rate in 1970 was about 5% of young men, whereas now greater than 80% of such cases are cured [1].

Table 1. Structure and clinical properties of approved Pt anti-cancer drugs.

	Molecular formula and structure	Dose-limiting toxicity	Clinical status and indications
Cisplatin	$H_6Cl_2N_2Pt$ 	Nephrotoxicity	Approved worldwide (sarcomas, small cell lung cancer, ovarian cancer, lymphomas, and germ cell tumors)
Carboplatin	$C_6H_{12}N_2O_4Pt$ 	Myelosuppression	Approved worldwide (ovarian carcinoma, lung, head and neck cancers)
Oxaliplatin	$C_8H_{14}N_2O_4Pt$ 	Neurotoxicity	Approved worldwide (colorectal cancer, advanced gastric and ovarian cancers)
Nedaplatin	$C_2H_8N_2O_3Pt$ 	Myelosuppression	Approved in Japan (head and neck, lung small cell, bladder, ovary, esophagus and cervix cancer)
Heptaplatin	$C_{11}H_{20}N_2O_6Pt$ 	Nephrotoxicity, intra-abdominal bleeding	Approved in Korea (gastric cancer)
Lobaplatin	$C_9H_{18}N_2O_3Pt$ 	Thrombocytopenia	Approved in the People's Republic of China (chronic myelogenous leukemia, inoperable, metastatic breast, small cell lung cancer)

CHAPTER 4

Elucidating Allosteric Communications in Proteins *via* Computational Methods

Burak Alakent[1,*] and **Z. Nevin Gerek Ince**[2]

[1] *Department of Chemical Engineering, Bogazici University, Istanbul, Turkey*

[2] *Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, USA*

Abstract: Cellular functions are primarily facilitated by biomolecular interactions with proteins, and ligand binding synchronizes the function of a protein to the requirements of its surroundings. Consequences of ligand binding to a protein may range from subtle perturbations in the side chain conformations in the vicinity of the binding region to large-scale global conformational changes. Coupling of a change in conformation with that in activity of a protein is traditionally referred to as allostery. In the recent years, however, the conventional allostery concept has been challenged to include perturbations in dynamics of a large number of proteins even in the absence of detectable changes in their backbone structure. Although it can evidently be suggested that binding produces a signal which can propagate to distant sites of a protein to achieve the observed conformational and/or dynamical perturbations, revealing a detailed mechanism of signal propagation is still an elusive task. In order to elucidate this mechanism, the following two questions demand to be answered: i) How do different regions of the protein respond? ii) How does the protein "sense" and transmit the local perturbation? The former question, being relatively easier to handle, has been tackled with Normal Mode Analysis (NMA), Elastic Network Models (ENMs), and statistical analyses of Monte Carlo (MC) and Molecular Dynamics (MD) simulation trajectories for the last ~30 years in the literature. The latter question, on the other hand, is currently a hot research topic in research community. Allosteric signals are generally suggested to propagate through "energy transport channels" (residue networks, or signaling pathways) formed by bonded and nonbonded contacts of residues, and experimental methods, such as double-mutant analysis and NMR relaxation methods, are used to identify residues participating to these intraprotein signaling pathways. For the last 10-15 years, there has been a tremendous interest in

[*] **Corresponding author Burak Alakent:** Department of Chemical Engineering, Bogazici University, Istanbul, Turkey; Tel: +90 212 359 6433; E-mail: burak.alakent@boun.edu.tr

Zaheer Ul-Haq and Jeffry D. Madura (Eds.)

utilizing computational techniques to elucidate allostericity in proteins. While elastic network models and molecular simulations have continued to be resourceful methods, the most important novel contributions, presumably, have come from the graph theory, perturbation methods, and the statistical coupling method. In this chapter of Frontiers in Computational Chemistry, various computational techniques used to elucidate allosteric mechanisms in proteins are to be discussed with various examples.

Keywords: Conformational change, Communication pathway, Crystal structure, Database, Elastic network model, Frequency, Graph theory, Induced fit, Information theory, Ligand binding, Molecular dynamics, Monte carlo simulation, Perturbation, Population shift, principal component analysis, protein dynamics, residue network, Signal propagation, Statistical coupling analysis, Web-server.

1. INTRODUCTION

Although the relation between protein structure and dynamics has long been established, one of the leading roles in replacing the static view of the protein with a "dynamic machine" paradigm has been played by computer modelling and simulations performed since 1980s. In two independent and simultaneous pioneering computational studies [1, 2], Normal Mode Analysis (NMA) was employed on the energy minimized crystal structure of bovine pancreatic trypsin inhibitor, and the resulting eigenvectors (modes) were found to be consistent with the directions of collective fluctuations. It is interesting to note that even in these earliest studies, the paradox between the harmonicity assumption of NMA, *i.e.* protein oscillates harmonically around a single minimum, and the existence of anharmonic motions particularly at low frequency modes was well recognized. Since then, these two "contradictory" views have been simultaneously and widely accepted in the literature, and a compromise has been aimed to be attained *via* various models.

Recognizing the multiminimum architecture of the protein energy landscape has been an important step in deciphering protein dynamics. In a leading study, experimental and modelling work on the binding of CO to myoglobin showed that protein dynamics evolve within conformational substates and cannot be explained by simple exponentials and Arrhenius kinetics [3]. The "rugged" energy landscape of myoglobin dictating its dynamics was described by a hierarchical

organization of "tier"s: Tier 0 conformations were suggested to correspond to three distinct orientations of the bound CO with respect to the heme group, while conformations at higher tiers, which stemmed from one of the conformations at the lower tier, were much greater in number and demanded a statistical treatment [4]. Molecular Dynamics (MD) simulations helped to clarify the issue to a certain extent. Structures separated by a time interval of >0.15-0.20 ps during the MD simulation of myoglobin at 300 K were observed to converge into different energy minima, and the multi mininum protein energy surface sampled during the simulation was suggested to be characterized by the conformational changes of the loop regions and rigid body motions of α-helical segments [5]. MD simulations also showed that the coupled local and global motions have essentially a nonlinear character, consisting of nonperiodic transitions between multiple minima [6]. Schematic representations of NMA and multiminimum models of the protein free energy landscape are shown in Fig. (**1**).

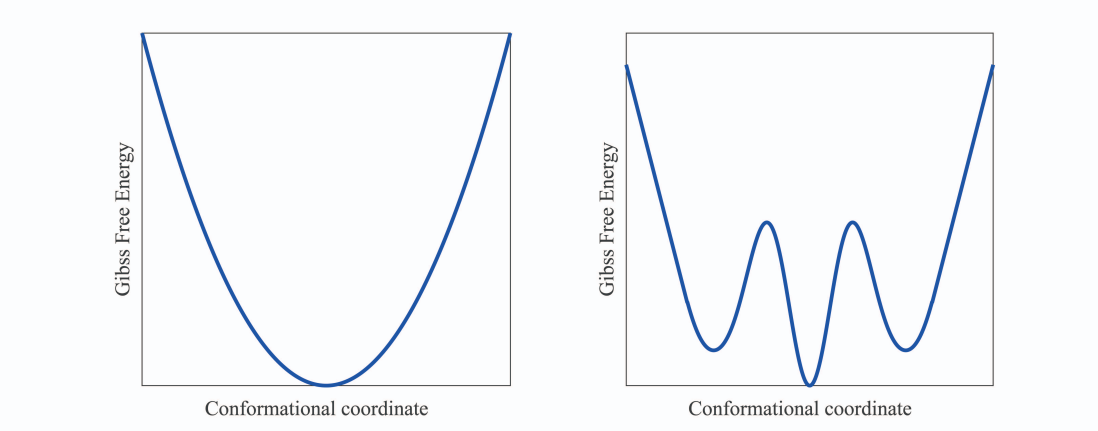


Fig. (1). Energy landscape of a protein viewed by two different models: **(A)** NMA: Harmonic fluctuations around a single minimum **(B)** Transitions between multiple minima: Here, Tier 0 corresponds to the whole "valley", while Tier 1 minima correspond to each local minimum.

A compromise between the description of NMA, MD simulations and functional protein dynamics has been reached by a combination of experimental and computational studies. Vibrational motions in fs-ps scale (fast dynamics) determined by NMA, and the inter-minima transitions known to have functional importance (slower dynamics) determined by experimental studies were shown to be in agreement for ubiquitin [7] and adenylate kinase (ADK) [8], while MD

CHAPTER 5

Information-Theoretic Representation of the Chemical Space of Many Electron Systems

R.O. Esquivel[a,b,*], S. López-Rosa[b,c], M. Molina-Espíritu[a], C. Soriano-Correa[d], J.C. Angulo[b,e] and **J.S. Dehesa[b,e]**

[a] *Departamento de Química, Universidad Autónoma Metropolitana, 09340-México, CDMX, México*

[b] *Instituto Carlos I de Física Teórica y Computacional, Universidad de Granada, 18071-Granada, Spain*

[c] *Departamento de Física Aplicada II, Universidad de Sevilla, 41012-Sevilla, Spain*

[d] *Química Computacional, FES-Zaragoza, Universidad Nacional Autónoma de México, 09230-Iztapalapa, México, CDMX, México*

[e] *Departamento de Física Atómica, Molecular y Nuclear, Universidad de Granada, 18071-Granada, Spain*

Abstract: In this chapter we review the utility of an information-theoretic three-dimensional (IT-3D) space to unveil the unique physical, chemical and biological aspects of a great diversity of many electron systems, ranging from neutral and ionized atomic systems and simple molecules to much more complex ones such as amino-acids and pharmacological molecular ensembles. This space is generated from the Shannon entropy, the Fisher information and the disequilibrium measures along with their corresponding Fisher-Shannon and López-Ruíz-Mancini-Calvet (LMC) complexity measures. To achieve it we start from the theoretical ground that atoms and molecules can be described by means of the basic information-theoretical notions of delocalization, order, uniformity and complexity; thus, revealing the possible existence of an universal three-dimensional information-theoretic space for all systems in Nature. On the other hand, we discuss the abilities of the Shannon entropy, Fisher information and disequilibrium to capture the spatial spreading features of delocalizability, order and uniformity of biological molecules. Indeed, these three entropic measures are

* **Corresponding author R.O. Esquivel:** Departamento de Química, Universidad Autónoma Metropolitana, 09340-México, D.F., México; Tel/Fax: +525558044675; E-mail: esquivel@ xanum.uam.mx

Zaheer Ul-Haq and Jeffry D. Madura (Eds.)

found to uniquely characterize all amino acids, and some selected pharmacological systems, through a predominant information-theoretic quality scheme (PIQS) which gathers all chemical families by means of three major spreading features: delocalization, narrowness and uniformity. This scheme is shown to recognize 4 chemical groups characterized by this entropic scheme: delocalized (aliphatic and aromatic), narrowed (electro-attractive) and uniform (tiny). Chemical groups are differentiated according to their energy classifications. Also, it is shown that information planes produce interesting patterns associated to the PIQS scheme.

1. INTRODUCTION

Most physical theories pursue to describe the most basic aspects of the macroscopic world through simple models, predicting some parameters that are assumed or taken from experiments. In consequence, the prediction of these parameters cannot be predicted by simple theoretical models. Obviously, to gain insight of all physical features requires to analyse the features of the systems in smaller scales where the simplest processes correspond to the lowest level of knowledge. It is advantageous to go to a deeper level since it reduces the number of unspecified parameters, and hence the corresponding theory is considered to be complete and fairly adequate. A typical example of this kind of theories is Molecular Biology which is ultimately based on quantum chemistry and molecular dynamics. Notwithstanding that more comprehension of the lower level is achieved, it is practically impossible to attain a full description of the molecular processes taking place in living systems, hence the intricacy of the large set of parameters makes the endeavour a very difficult one. Considering an alternative approach to extracting the essential features of biological processes by use of Information Theory (IT) concepts has proven to be a succesful one. Moreover, the rapidly evolving field of Quantum-information biology [1 - 3], which employs information-theoretic concepts, is gaining wide attention to comprehend some of the most basic and yet unsolved questions of molecular biology.

There has been an increasing interest in characterizing and classifying different physical systems in terms of a few fundamental properties, not only in Physics but also in Chemistry and Biology. Perhaps, quantitative structure activity relation (QSAR) and quantitative structure properties relation (QSPR) constitute the most commonly approaches employed to relate molecular structures with physical

properties and biological activity. The beginning of these techniques could be remounted to mid 60's of the last century, when Hansch and Fujita proposed a connection between biological activity and chemical structure [4]. Furthermore, they argued that similar molecules share similar solubility, expecting that the relative polarity of molecules could be crucial in order to find a parameter relating structure and activity. Based on the ideas of Robert Muir and the Hammet equation [5], the structural changes might be correlated by means of parameters allusive with the partition coefficient to numerically analyse structure-activity problems [6] of biomolecules. Consequently, QSAR has evolved from simple regression methods to the analysis of very large sets of data comprising thousands of diverse molecular structures, and uses a wide variety of statistical and machine learning techniques. These advances have found broad application on QSAR methods in chemistry, material and nano-material, and life sciences to assess potential impacts on ecological systems [7]. One of the most promising application of this methodology resides in the chemical space [8]. The concept of chemical space emerges as a metaphor, and suggests the existence for a chemical universe which contains millions of organic compounds [9]. Although chemical space has not been well defined, it considers a multidimensional descriptor space in the sense of a region defined by a particular choice of descriptors to characterize as many chemical compounds as possible, and relate similar molecular structures with desired physicochemical properties and biological activity. In that respect, the relevance of any region of the chemical space must be judged by its ability to group compounds with similar bioactivity together [10].

The large number of physicochemical properties to be chosen as descriptors of the chemical space is an important disadvantage, due to the risk of employing irrelevant and redundant descriptors. Moreover, different systems could be wrongly misplaced at the same point of such a space if the descriptors selection is not well chosen [7]. A deeper understanding of this vast set of molecules will advance our knowledge of biological processes; therefore, the development of a systematic and rational classification of the chemical space is crucial for the progress of chemical applications. The analysis and exploration of this space represents a highly demanding computational task due to the immense number of possible stable molecules [11]. This challenge has led to several sophisticated

SUBJECT INDEX

D

E

F

G

H

I

L

M

N

O

P

Q

R

S

T

V

W

X

Zaheer Ul-Haq

Dr. Zaheer Ul-Haq is directing the Computational Chemistry group at the Dr. Panjwani Center for Molecular Medicine and Drug research, University of Karachi. He obtained his PhD under the supervision of Prof. Atta-ur-Rahman and completed his post-doctoral studies with Prof. Bernd M. Rode in Innsbruck, Austria. He is a recipient of Fulbright and Humboldt Fellowship from USA and Germany, respectively. He has published over 100 research articles in top international journals of computational chemistry. His area of interest includes in silico screening and Molecular Dynamics simulation of bio-molecules. He is currently serving as editorial board member to the Journal of Molecular Graphics and Modelling, and Current Computer-Aided Drug Design.

Jeffry D. Madura

Jeffry is the Lambert F. Minucci Endowed Chair in Engineering and Computational Sciences and Professor in the Department of Chemistry and Biochemistry at Duquesne University. He earned a B.A. from Thiel College, a Ph.D. in Physical Chemistry from Purdue University and was a postdoctoral fellow at the University of Houston. His research interests include the development and application of biomolecular simulation software, the study of neurotransmitter transporters, the electronic structure of solid-state materials, and the thermoresponsive behavior of smart polymers. He has published 100+ peer-reviewed papers in physical chemistry and received over $6M in external research funding. He was a recipient of a Dreyfus Teacher-Scholar Award, was the chair of the ACS COMP Division and is an ACS Fellow. Dr. Madura received the 2014 ACS Pittsburgh Local Section Award. He is a co-author to the textbook titled "General Chemistry: Principles and Modern Applications" as well as a co-author to a physical chemistry solutions manual. He received the Bayer School of Natural and Environmental Sciences and the Duquesne University Presidential Award for Excellence in Scholarship in 2007 and the Bayer School of Natural and Environmental Sciences Award for Excellence in Service in 2004. He is currently co-editor to the Journal of Molecular Graphics and Modelling.