J. S. Redinha
J. da Providência
A. J. C. Varandas
Editors

Quantal Aspects in Chemistry and Physics

A tribute to the memory of Professor Couceiro da Costa

IMPRENSA DA UNIVERSIDADE DE COIMBRA
COIMBRA UNIVERSITY PRESS

COIMBRA 2011

8. Computational Proteomics – From Methodological Developments to Biological Applications

Irina S. Moreira, Natercia F. Bras[†], Alexandra T. P. Carvalho[†], Nuno M. F. S. A. Cerqueira[†], Daniel F. Dourado[†], Marta A. S. Perez[†], Antonio J. M. Ribeiro[†], Sergio F. Sousa[†], Pedro A. Fernandes, and Maria J. Ramos[*]

REQUIMTE/Departamento de Química, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto - Portugal

Proteomics, a chimera of proteins and genomics, involves the study of the proteins expressed in a cell, organism, or tissue. Proteins are essential in all aspects of life, and so the computational study of proteomics is becoming a vital element in understanding the underlying concepts. In this review we are going to address some of the challenges and latest developments focusing in four different aspects that are thematic in our group: (a) Molecular Dynamics Simulations; (b) Drug Design; (c) Enzymatic Mechanisms; and (d) Benchmarking of DFT functionals.

8.1 Introduction

Proteomics, a chimera of **prote**ins and gen**omics**, was invented by Professor Mark Wilkins in the early 1990s and involves the study of the proteins expressed in a cell, organism, or tissue. This includes protein identification and quantification, protein-protein interactions, protein complexes prediction, protein modifications and protein localization in the cell. As proteins are essential for all life, proteomics is crucial in biomedical applications, and although more recent, the computational study of proteomics is becoming a key element in this biological field.

Computational proteomics involves the computational methods, algorithms, databases and methodologies used to model protein structure, dynamics and

[*]Email address: mjramos@fc.up.pt
[†]Equal participation.

function. In this review we are going to focus on four different aspects that are thematic in our group, and range from methodological developments to their biological application: (*a*) Molecular Dynamics Simulations; (*b*) Drug Design; (*c*) enzymatic Mechanisms; and (*d*) Benchmarking of DFT functionals.

8.2 Classical Molecular Dynamics Simulations

Classical molecular dynamics (MD) simulations have become during the past two decades, a particularly important discipline within the field of computational biochemistry, allowing the computationally efficient evaluation of a variety of properties in the study of biological molecules for which atomic and molecular motion is vital.

Generically speaking, MD simulations follow the time evolution of a system through the numerical integration of the equations of motion of the corresponding particles. In particular, MD simulations are based on the application of classical mechanics, with intra and inter-molecular interactions described by a sum of different contributions described by mathematical formulations of simple physics phenomena. The corresponding mathematical formulae and the accompanying parameters, which are typically fitted to reproduce experimental data or high-level *ab initio* calculations, are normally described under the generic designation of force field. A variety of different force fields is presently available, differing in features such as scope, accuracy and cost associated.

The range of application of MD simulations is remarkably wide and encompasses the study of phenomena such as protein and/or small molecule conformational changes, molecular association and recognition, folding, ion transport, etc. Over the last few years we have employed MD simulations in the study of several biological systems of interest, often in combination with other computational methods. In this section, we highlight 5 particularly different applications of this very powerful methodology, selected from our own work on biological systems. These include our MD studies on the elusive metalloenzyme farnesyltransferase (FTase), the medically critical HIV reverse transcriptase enzyme (RT), and the economically appealing carbohydrate-binding modules (CBMs)

from family 11.

8.2.1 Farnesyltransferase

Farnesyltransferase is a Zinc metalloenzyme that catalyses the addition of farnesyl groups from farnesyl diphosphate (FPP) to protein cysteine residues present in characteristic carboxyl terminal −CAAX motifs. In this motif C represents the cysteine residue that is farnesylated, A is an aliphatic amino acid, and X represents the terminal amino acid residue [1]. Proteins substrates bearing a CAAX motif include a number of biologically relevant protein targets, most notably the Ras family of proteins known to be implicated in something like 30% of all human cancers [2].

Performing MD simulations on FTase is particularly challenging, compared with the typical enzymes that are comprised simply by standard amino acid residues, because of the presence of a covalently bound Zinc atom with a metal coordination sphere that changes during catalysis. Metal atoms, and the corresponding bonds, angles, dihedrals, charges, and van der Waals parameters are normally absent in the typical biomolecular force fields such as AMBER, CHARMM or OPLS. Their inclusion involves not only the parameterization of the metal atom itself, but also of the directly interacting amino acid residues, and naturally a subsequent process of validation against experimental data. We have parameterized the three different Zn coordination spheres that are formed during the catalytic mechanism of this enzyme using quantum calculations and experimental data and have validated the new parameters against EXAFS and X-Ray crystallographic information [3].

Following this process, we have performed comprehensive MD simulations with the AMBER software package on the several intermediate states formed during the catalytic mechanism of this enzyme, in an attempt to understand the way this enzyme works at a molecular level, and taking into consideration features such as the effect of the solvent and the dynamic effects arising from the interaction of the enzyme, solvent and the substrate/product molecules [4,5]. Starting from extensive 10 ns MD simulations on the enzyme resting state, binary

complex (FTase-FPP), ternary complex (FTase-FPP-CAAX substrate) and product complex (FTase-Product), we have performed comparative analysis of the amino acid flexibility along the FTase sequence, radial distribution functions of water molecules around catalytically relevant atoms, statistic variations on key catalytic distances, detailed analysis on the conformation and orientation adopted by the substrate and product molecules in the presence of the enzyme, and hydrogen bonding analysis on the most important molecular recognition sites.

These results provided very useful information for the subsequent modeling of the catalytic mechanism of this enzyme, guiding and supporting the choice of the models used from QM or QM/MM calculations, and of the several approximations adopted.

8.2.2 Reverse Transcriptase

Reverse transcriptase (RT) is the human immunodeficiency virus (HIV) enzyme whose function is to copy the viral RNA into double-stranded DNA suitable to be integrated in the host cell genome. Several combinations of different RT inhibitors are currently used in antiretroviral therapy. Our study focused on nucleoside reverse transcriptase inhibitors (NRTIs). These are substrate analogues that compete for binding and incorporation into the nascent DNA chain. However, because they lack a 3'OH, after they are incorporated they do not allow the addition of the next incoming nucleoside blocking DNA synthesis. It is presently known that the long-term failure in the treatment of AIDS with the currently available NRTIs is related to the development of resistance by RT at the binding or incorporation level, or subsequent to the nucleotide incorporation (excision).

We have conducted a series of MD simulations of RT with different inhibitors in explicit solvent in order to correlate the structural characteristics of the inhibitors with the stage at which RT resistance emerges. To achieve a greater insight on how RT discrimination gets established we compared incorporation of a normal substrate (dNTP) with incorporation of two very similar inhibitors for which resistance emerges by different mechanisms: phosphorylated zalcitabine,