SAR and QSAR in Environmental Research

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gsar20

Molecular dynamics simulations of the enzyme Catechol-O-Methyltransferase: methodological issues

A Bunker a , PT Männistö b , J.-F. St. Pierre b c , T Róg d e , P Pomorski f , L Stimson f & M Karttunen e f
a Drug Discovery and Development Technology Center, Faculty of Pharmacy, University of Helsinki, Finland
b Division of Pharmacology and Toxicology, Faculty of Pharmacy, University of Helsinki Finland, Québec, Canada
c Département de Physique and Regroupement Québécois sur les Matériaux de Pointe, Université de Montréal Montreal, Québec, Canada
d Laboratory of Physics, Helsinki University of Technology Finland, Kraków, Poland
e Department of Biophysics, Faculty of Biotechnology, Jagiellonian University, Kraków, Poland
f Department of Applied Mathematics, The University of Western Ontario, London, Ontario, Canada

Published online: 04 Dec 2010.


To link to this article: http://dx.doi.org/10.1080/10629360701843318

PLEASE SCROLL DOWN FOR ARTICLE
Molecular dynamics simulations of the enzyme Catechol-O-Methyltransferase: methodological issues

A. Bunker a,*, P.T. Männistö b, J.-F. St. Pierre b,c, T. Róg d,e, P. Pomorski f, L. Stimson f, and M. Karttunen d,f

a Drug Discovery and Development Technology Center, Faculty of Pharmacy, University of Helsinki, Finland; b Division of Pharmacology and Toxicology, Faculty of Pharmacy, University of Helsinki Finland; c Département de Physique and Regroupement Québécois sur les Matériaux de Pointe, Université de Montréal Montreal, Québec, Canada; d Laboratory of Physics, Helsinki University of Technology Finland; e Department of Biophysics, Faculty of Biotechnology, Jagiellonian University, Kraków, Poland; f Department of Applied Mathematics, The University of Western Ontario, London, Ontario, Canada

(Received 5 September 2007 In final form 26 November 2007)

Results from extensive 70 ns all-atom molecular dynamics simulations of catechol-O-methyltransferase (COMT) enzyme are reported. The simulations were performed with explicit TIP3P water and Mg²⁺ ions. Four different crystal structures of COMT, with and without different ligands, were used. These simulations are among the most extensive of their kind and as such served as a stability test for such simulations. On the methodological side we found that the initial energy minimization procedure may be a crucial step: particular hydrogen bonds may break, and this can initiate an irreversible loss of protein structure that becomes observable in longer time scales of the order of tens of nanoseconds. This has important implications for both molecular dynamics and quantum mechanics–molecular mechanics simulations.

Keywords: catechol-o-methyltransferase; comt; molecular dynamics; hydrogen bonds; stability; protein dynamics

1. Introduction

Catechol-O-methyltransferase (COMT) is an enzyme responsible for the inactivation, through O-methylation, of catecholamine neurotransmitters, including dopamine, epinephrine and norepinephrine, as well as catechol steroids [1]. COMT also inactivates a wide variety of xenobiotic (foreign) molecules containing catechol groups, i.e., a benzene ring with two OH groups, see Figure 1, including possible drug molecules. The most important of these drug molecules, and the main driver for the development of inhibitors for COMT activity, is L-dopa, the leading drug for the treatment of Parkinson’s disease [2]. L-dopa must pass through the blood-brain barrier intact before becoming metabolized
into dopamine, thus raising dopamine levels in the brain. Current knowledge of the biochemistry and molecular biology of COMT, and the pharmacology and clinical efficacy of the inhibitors is covered by the review of Männistö and Kaakkola [3].

COMT catalyzes the transfer of a methyl group (CH₃) from a molecule of S-adenosyl methionine (ADOMET), see Figure 2, to one of the two aromatic OH groups of its catechol (see Figure 1). That occurs in the presence of an Mg²⁺ ion, whose role is to bind the substrate. COMT exists in two forms, the membrane bound MB-COMT, and the soluble S-COMT, both performing the same aforementioned function. Our study focuses on S-COMT, the form that has been successfully crystallized [4–7]. S-COMT is a monomeric protein composed of 221 residues. The methylation occurs through several steps. The structure of COMT contains a binding site that attracts the ADOMET molecule. Once ADOMET is in place, a binding site is created for the Mg²⁺ ion. The entry of the Mg²⁺ ion, in turn, creates a binding pocket for the substrate.

COMT has been successfully crystallized with four different inhibitors in its binding site. The first successful structure was obtained by Vidgren et al. [4] with Mg²⁺ ions, ADOMET, and the inhibitor 3-5 dinitrocatechol (DNC) (PDB database structure identification 1VID). Then, Lerner et al. [5] crystallized COMT with Mg²⁺ ions and the bisubstrate inhibitor blocking both the substrate and co-enzyme docking sites (PDB database structure identification 1JR4). The third structure is due to Bonifacio et al. [6] who obtained a crystal structure for COMT with Mg²⁺ ions, ADOMET, and the inhibitor BIA 3-335 (BIA) (PDB database structure identification 1H1D). Finally, and most recently, Palma et al. [7] successfully crystallized COMT with Mg²⁺ ions, ADOMET, and BIA 8-176 (BIE) (PDB structure identification 2CL5). BIA 3-335 and BIA 8-176 are both comprised of a DNC core structure with two different added tails. In all of these cases rat S-COMT has been used. S-COMT possesses 81% sequence identity and 86% sequence similarity to the human form of the enzyme [8]. A crystal structure of human COMT with only ADOMET has recently been obtained, and is available on the PDB database [9] (PDB structure identification 2AVD). This structure was not, however, used in our study.

A reasonably large amount of computational work has been performed on COMT. This has mostly, however, been simulations using the QM-MM method (Quantum Mechanics–Molecular Mechanics) to focus on the details of the catalytic reaction itself [10–15] or QSAR-type analysis [16]. In addition to the interest in the activity and

Figure 1. The catechol group before (a) and after methylation (b).
inhibition of COMT, a second reason for the extensive study of this particular catalytic reaction is that it can be seen to form an almost ideal model system for study of catalysis using QM-MM. The COMT protein provides both the environment for the interaction to occur and positions the ligand, Mg$^{2+}$, and ADOMET in what is known as a “near attack conformation”, i.e., all the atoms are positioned for the chemical reaction (methylation) to occur. The COMT protein is however not itself directly involved in this reaction. Quantum mechanics must thus only be taken into account for the smaller molecules (ADOMET, Mg$^{2+}$, and ligands). In other words, one can save considerable computer time by limiting the atoms that need to be taken into account quantum mechanically. By doing that, observing the entire reaction path becomes tractable, which would not be possible if one used a purely quantum mechanical algorithm.

What has not been studied is longer time dynamics using the molecular mechanics paradigm. Quantum mechanical simulation can only access very short time periods; see, e.g., Vattulainen and Karttunen [17] for an overview of the different molecular modeling methods, and as such one can only directly observe the details of the catalyzed reaction. Any question relating to protein flexibility and structure cannot be addressed with a simulation on the time scale accessible to a quantum mechanical simulation. In particular, molecular mechanics can be used to study the effect on the structure of COMT of the presence of the ADOMET co-enzyme and Mg$^{2+}$ ion. Whether or not the COMT enzyme structure is dependent on the presence of the ADOMET is unknown, as crystallization is the only means to accurately determine protein structure and no crystal structure for COMT without ADOMET exists. The structure of an enzyme being dependent on a co-enzyme is by no means without precedent. Two clear examples are the protein hexokinase which has two domains that are held together by a sugar molecule, and pyrophosphatases, proteins which exhibit a variety of closed conformations, held together by metal and sulphate ions [18, 19].

From the four crystal structures of COMT (as described above), we have extracted five configurations of the COMT protein (2CL5 is a crystal structure with a unit cell containing two proteins so we were able to obtain two crystal structures from this). We have
removed all ligands and performed 70 ns of atomistic molecular dynamics simulations under buffer saline conditions using potassium chloride as the salt to best approximate the intra-cellular environment for each of them. For comparison, we have also simulated the 1VID structure with all ligands, also for 70 ns.

In this paper we focus on methodological issues. We found that the initial energy minimization procedure may be a crucial step, as particular hydrogen bonds may break, leading to an irreversible loss of protein structure which only becomes observable on time scales of the order of tens of nanoseconds. This emphasizes the importance of performing longer simulations to determine the stability of the structures, important on the shorter time scales, over which simulations are generally carried out, as well.

The rest of this paper is organized as follows: In the next section we describe the details of our simulation method, Section 3 summarizes the results, and finally in Section 4 we present a discussion and conclusions.

2. Methods

We performed atomic scale molecular dynamics (MD) simulations using the five crystal structures for COMT [4–7] obtained from the PDB database [20]. One crystal structure each was obtained from structures 1VID, 1H1D, and 1JR4. Two structures were obtained from 2CL5 since this crystal structure had a two protein unit cell. For all proteins there exist short tails at either ends of the protein sequence that are not fixed in the crystal structure. Since in all cases they are all very short (<8 residues), and most probably do not possess stable secondary structure to begin with, inserting them in a random coil configuration and allowing them to equilibrate will not produce any significant structural errors.

As for force field, the OPLS-AA (Optimized Parameters for Liquid Simulations; AA stands for all-atom) potential set was used [21]. All ligands were removed including the Mg$^{2+}$ ion as well as the ADOMET and inhibitor molecules. The remaining COMT molecule was solvated in a box of water of size 8 x 8 x 8 nanometers, and potassium and chlorine ions were added to neutralize the system and create buffer saline conditions (140 mM salt concentration). Potassium was used instead of sodium to more realistically simulate the conditions inside the cell where potassium is the dominant anion. In addition to the five COMT structures simulated without ligands, the structure 1VID with DNC, ADOMET and Mg$^{2+}$ was also simulated. All of the structures obtained from the PDB database contain crystallized water molecules. For the structure simulated with the ligands in place these have an integral structural role, so they were kept in place. When the ligands are removed so is the structural role of the water molecules, and all hydrophobic cavities within the COMT become easily accessible to the solvent. Thus for the all cases of COMT simulated without ligands, all crystal water molecules were removed before the protein was resolvated.

The simulation protocol follows that of our previous protein simulations [22,23]. All molecules (protein and ligand) were parameterized with OPLS-AA forcefield [21]. For ADOMET, the missing partial charges were obtained by fitting the electrostatic potential of the molecule using the Restrained ElectroStatic Potential (RESP) method [24]. The electronic structure and the electrostatic potential were calculated with the 6-31G$^*$ basis set, compatible with the OPLS force field. This was carried out using the GAMESS software suite for electronic structure calculations [25]. For water, the TIP3P model was used [26].
The MD simulations were performed using the GROMACS 3.3.1 simulation package [27]. The simulations were performed for 70 ns at constant pressure and temperature for all structures. The Nosé-Hoover thermostat [28] was used in combination with the Parrinello-Rahman barostat [29]. Following standard practice the protein and solvent were connected to separate Nosé-Hoover thermostats but the same parameters were used for both: the coupling time was set to $\tau_t = 0.1\, \text{ps}$ and the reference temperature was set to 310 K. For pressure, the coupling time was set to $\tau_p = 1.0\, \text{ps}$. The simulations were performed at atmospheric pressure. Electrostatic interactions were calculated without any cutoff using the PME method [30]. The PME order used was 5 and the Fourier spacing was 0.1 for the simulation of the 1VID. For each simulation, a configuration was output every 10 ps, and thus for each of the 70 ns runs there are 7000 output configurations.

Once the full output for each of the MD runs was obtained, a set of analysis programs were applied to each of them. These included calculating the radius of gyration ($R_g$), i.e., the square of the average distance from the center of mass and a lowest order measure of molecular structure,

$$R_g = \frac{1}{N} \sum_{k=1}^{N} (r_k - r_{\text{mean}})^2.$$  \hspace{1cm} (1)

Another quantity of interest is the root mean square deviation of particles from the start position, RMSD at time $t$, given by,

$$\text{RMSD}(t) = \sqrt{\frac{\sum_{k=1}^{N} (r_k(t) - r_k(t=0))^2}{N}},$$  \hspace{1cm} (2)

where $r_k(t)$ is the position of the $k$th particle at time $t$. Our analysis also included the fluctuation of root mean square deviation, solvent accessible surface area [31] (SASA), i.e., the surface area of the protein accessible to the solvent. It is defined as the surface area that is 1.4 Å away from the nearest atom in the molecule, also known as the Connolly surface [32, 33]. To complete the analysis, we also made use of the DSSP (Definition of Secondary Structure of Protein) program [34]. The DSSP program, developed by Kabsch and Sander, calculates the secondary structure, geometrical features, and solvent exposure of a protein given its atomic coordinates in PDB format. The secondary structure is defined by the network of hydrogen bonds detected. The dynamic runs were then visualized and a structurally stable area of the protein was determined. Using software written in house, we performed a visualization with the protein centered in the simulation box and all angular momentum removed from the stable structural element of the protein. Only once the data was filtered this way was it possible to use visualization to examine the development of the protein structure throughout the simulation. The final visualization of the protein dynamics was performed using the VMD (Visual Molecular Dynamics) package [35].

3. Results

Figure 3 shows snapshots of the 2CL5-A structure at the three main steps of our modeling procedure. First, the initial crystal structure, i.e., before performing energy minimization, with all ligands (BIE, ADOMET and Mg$^{2+}$) is shown in Figure 3a. Then, the structure
of the protein with the ligands removed after the initial energy minimization is shown in Figure 3b, and the structure of the protein after 70 ns of MD simulation is shown in Figure 3c.

As can be seen from Figure 3, the overall structure of the protein is preserved during the entire simulation run but the energy minimization alters the structure somewhat. In the case of our four remaining models, the minimized structure was distinctly different from the crystal structure and as a consequence the resulting final configurations of these systems also significantly differ from the crystal structure. By careful examination of all the trajectories, we have identified these observations as due to hydrogen bonding and the minimization procedure, which was the almost universally used conjugate gradient method. We discuss the reasons for this behavior in more detail in Discussion and Conclusions (Section 4). As the other structures, apart from 2CL5-A, did not turn out to be stable, we have only presented the results for the system 2CL5-A.

To assess the stability of the simulations, and whether the system had reached equilibrium, we monitored a set of parameters describing the protein structure over the course of each simulation run. The results for the system 2CL5-A are shown in Figure 4. First, Figure 4a shows the time development of the root mean square displacement parameter (RMSD). As can be seen, RMSD stabilizes to the value of 0.42 ± 0.01 nm after a comparatively long time of 15 ns. Importantly, the drift of RMSD between 15 ns to 70 ns is small suggesting that this part of the trajectory is stable. Figure 4b shows the time dependence of the protein radius of gyration ($R_g$) which has stabilized after 30 ns. The solvent accessible surface area (SASA) is shown in Figure 4c. It is stable over the whole simulation run.

We also monitored the secondary structure of the protein. It turned out that the secondary structure (i.e., the number of amino acids in helical and beta structures) remains almost constant over the whole simulation run. These results indicate that after the comparatively long simulation time of about 30 ns, the simulated system had reached equilibrium. Since the different parameters took varying times to equilibrate, this demonstrates that as many global parameters as possible should be monitored for any protein dynamics simulation in order to properly judge if the system has reached equilibrium. Similar analyses performed for the three remaining systems (data not shown) show proteins losing their secondary structures and unfolding after about 40–50 ns.
in particular in the regions of residues 170–221 and 30–60. Those simulations can thus be regarded as unstable.

For a better insight into changes of protein structure that occur during the simulation run, the secondary structure of the 2CL5-A model is presented as a function of time in Figure 5 following the Kabsch and Sander procedure [34]. As can be seen from Figure 5, most of the secondary structure elements of the protein are preserved over the simulation run. We can observe temporal changes of the structure that are typically seen to occur in protein dynamics. To describe protein flexibility, we calculated the root mean square fluctuations of the C-alpha atoms averaged over the whole equilibrated simulation run (Figure 6). As can be seen from Figure 6, both ends of the protein are characterized by high mobility. This is not unexpected since both ends were not resolved in the crystal structure, 2–3 residues at C-end and 4–5 at N-end were different in different structures and they were added prior to the initial minimization step. Highly mobile protein fragments without fixed structure remain often unresolved in the crystal structure. Higher mobility is also observed for the loops regions. Particularly prominent on the graph are the loops centered at residues 200, 40, and 175.

4. Discussion and conclusions
In this paper we have presented the first results of our extensive all-atom MD simulations of COMT. The simulation times reached 70 ns which is far beyond the current standards
for simulations of this kind. The long duration of these simulations raised several interesting questions about stability and equilibration of protein systems as will be discussed below.

As described in Section 3, we made several interesting observations concerning methodological issues. First of all, equilibration time of 30 ns is comparatively long, far beyond what is normally done in protein dynamics simulations; typical MD simulations are of 10–20 ns with equilibration times of a couple of nanoseconds. On the other hand, in simulations of other biological systems, such as those involving membranes, it has long been established that equilibration may indeed take a long time – that is particularly important in systems containing charges [36, 37]. The observations here are
also relevant for QM-MM simulations; longer simulations such as these are needed to assess the structural stability which is often, as is the case here, related to hydrogen bonding. The length of the equilibration is related to the studied system’s properties, such as formation of hydrogen bonds, as well as the presence and strength of charges, and cannot be predicted prior to the simulation.

The second observation we made concerns the relation between the initial structure and simulation stability. In procedure we used in these studies in first step we removed ligands and then performed energy minimization of the structure. Already this step was influencing structure significantly. During energy minimization certain particular hydrogen bonds can be seen to break and this can initiate an irreversible (on the time-scale accessible to molecular dynamics) loss of protein structure. What makes this observation particularly noteworthy is the fact that the energy minimization technique we used is the standard accepted method, namely the conjugate gradient method. The difference between our simulations and the others is the long simulation time; we simulated long enough for the structure to either stabilize or to break down (40–50 ns), a period well beyond the vast majority of simulations of single proteins in solution that have been published. This has important methodological implications and will be the focus of the continued research of this project. Further work is also in progress to study these systems with ligands using both the MD and QM-MM methods.

Acknowledgments
This work has been supported by Emil Aaltonen foundation (M.K.) and the Natural Sciences and Engineering Research Council of Canada (M.K.). Tomasz Róg has been supported by a Marie Curie Intra-European Fellowship 024612-Glychol. Jean-Francois St. Pierre is supported through the GALENOS program. We thank the Finnish IT Center for Science (CSC) and the Southern Ontario SharCNet grid computing facility (www.sharcnet.ca) for computer resources.

References


