Molecular computations for reactions and phase transitions: applications to protein stabilization, hydrates, and catalysis

Brian J. Anderson, Brian M. Baynes, Cynthia S. Lo, Jhih-Wei Chu, Bin Pan, Daniel I. C. Wang, Bernhardt L. Trout
Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, 66-458, Cambridge, Massachusetts 02139

Abstract – We compute the thermodynamics and kinetics of systems that undergo ordering and disordering, namely protein degradation via aggregation and oxidation, clathrate-hydrate equilibria, and reactions in the presence of zeolite catalysts. We examined the stabilization of therapeutic proteins against aggregation through the use of a novel stabilization mechanism, the gap effect. We have developed approaches to predict the change in the chemical potential of proteins upon changing the solvent environment by adding excipients. We have also determined the mechanism for oxidation of methionine sites in proteins and discovered a structural property of a protein system that correlates well to rates of oxidation that we have measured. We have used quantum mechanical calculations to compute potential energy surfaces for the guest–host interactions within a clathrate-hydrate and model reference parameters with unprecedented accuracy. We also developed analytical methods that allow for the prediction of the thermodynamic properties of mixed clathrate-hydrates that are much simpler and powerful than the currently used numerical methods. Finally, we studied the coupling reaction of two methanol molecules to form ethanol and water in the zeolite chabazite. From this study we have computed a new mechanism for the methanol coupling reaction.

Index Terms – aggregation, clathrate, hydrate, oxidation, therapeutic proteins, zeolite catalysis

I. INTRODUCTION

WE develop and apply molecular computational methods for solving problems in the chemical and pharmaceutical industries. The paradigm of chemical engineering has shifted in the past decade from making bulk chemicals inexpensively to making high-value added products and processes. Our research addresses the latter by revealing a molecular-level understanding of chemical processes and giving us insight into how chemical systems can be manipulated on the molecular level to achieve unprecedented specificity. Molecular computational methods allow us to identify and quantify the thermodynamics and kinetics of elementary reaction steps in complex chemical systems, something that is often beyond current experimental capabilities. Not only do our techniques allow us to compute these quantities, but also, they lead us to insight into how chemical systems can be changed to increase to decrease thermodynamic and kinetic parameters. In terms of areas of application, we have generally focused on projects in the areas of the environment and human therapeutics.

II. METHODS

In order to aid in solving chemical problems, we use a variety of molecular computational methods, including density functional theory (CPMD and DACAPO), molecular dynamics (CHARMM), Car-Parrinello methods (CPMD), Monte Carlo methods, and transition path sampling in addition to ab initio quantum mechanical calculations. Often, we modify these methods to make them more efficient or as needed for the problem of interest. We also have ongoing work in the development of ways of computing free energies and rates of reactions in complex systems.
III. PROTEIN STABILIZATION

Therapeutic proteins, such as insulin, interferon, and EPO (erythropoietin), represent an important and rapidly growing class of pharmaceuticals, presently accounting for $35B/yr in revenue worldwide. Proteins are useful as therapeutics because they have a wide range of physiological functions and are extremely potent. Natural proteins in the body, as well as man-made proteins, can often carry out their functions at extremely low concentrations, such as $10^{-9}$M, $10^{-12}$M, or even lower. Unfortunately, proteins are also only marginally stable, and are degraded and inactivated rapidly.

In industry, the inherent instability of proteins presents a serious problem, and a disadvantage relative to small molecule therapeutics. To optimally serve patients, it is desirable to store proteins at high purity and for long times, often for up to two years after manufacture. Thus, proteins must not only be removed from their natural cellular environment, but they must also be stable against degradation for unnaturally long periods of time. This is the challenge faced by researchers and practitioners in the area of protein stabilization.

Specific degradation routes that must be addressed include aggregation, deamidation, oxidation, and hydrolysis.

A. Aggregation

Empirically, it has been observed that by adding low molecular weight components, such as salts, sugars, or polyols, to protein solutions, the propensity of the protein to aggregate (as well as degrade by other routes) can often be significantly affected. Unfortunately, because proteins are tremendously diverse in chemistry and structure, additives that work well for a particular protein generally do not work universally. In addition, current understanding of the mechanisms by which additives confer stability on proteins is limited. Thus, there is often no theoretical guidance to aid selection of optimal additives.

This lack of understanding necessitates that protein stabilization be carried out on a case-by-case basis using heuristic experimental screens. This limits the additive search space and the possible formulation patent protection to those additive combinations which are explicitly tested. In some cases, additives that confer a useful level of stability cannot be identified.

As protein therapeutics branch out into new routes of administration, such as inhalers, implants, and stents, significant new stability challenges are presented. These new routes of administration involve protein-damaging factors such as atomization, elevated temperature, and high protein concentration, all of which can contribute in an unfavorable way to aggregation and other routes of degradation. Thus, there is an ever-increasing need to understand how to control these degradation processes to ensure that the full potential of proteins as therapeutics can be realized.

Our research addresses these needs by developing mechanistic understanding of additive function, thereby paving the way for rational design and selection of additives to stabilize proteins against aggregation.

Early work focused on calculating thermodynamic properties in mixed solvent systems (systems in which water, protein, and at least one more component were present in solution) [1]. In this study, we developed a statistical-mechanical method to model the thermodynamic effects of additives in molecular-level detail. This method was validated by quantitative comparison with experimental data on proteins in glycerol and urea solutions. It was the first method demonstrated to predict additive effects for real protein systems without the use of adjustable parameters.

We then applied our molecular simulation technique to study the mechanism by which arginine, a common refolding buffer additive, deters protein aggregation. We propose that arginine is a member of a class of anti-aggregation additives, which we term “neutral crowders,” characterized by their (1) negligible effect on the free energy of isolated protein molecules and (2) large size relative to water. With a simplified statistical-mechanical model, we have shown that such additives selectively increase the free energy of protein-protein encounter complexes by being preferentially-excluded from the gap between the protein molecules in such complexes. This free energy effect, which we call “the gap effect,” slows protein association reactions [2]. The gap effect is a novel mechanism by which solution additives can affect a broad range of association processes such as aggregation and crystallization.

In accordance with the gap effect model predictions, we showed experimentally that arginine slows the association of model globular proteins (antibodies and antigens) and of folding intermediates and aggregates of carbonic anhydrase II. We predict that neutral crowders larger than arginine will be superior anti-aggregation additives [3].

B. Oxidation

It is also well-known that proteins can be degraded through chemical pathways under various stresses encountered in aqueous solution. One of the major chemical degradation pathways is methionine oxidation due to many possible reasons, e.g. the presence of reaction oxygen species (ROS) such as hydrogen peroxide, hydroxyl radicals, superoxide radicals etc [4]. We are interested in preventing or hindering the oxidation of methionine residues whose chemical modification significantly changes the bioactivity and structure of protein which is shown below. We focus on hydrogen peroxide induced oxidation of methionine in which an oxygen atom is added covalently to the reactant methionine sulfur site to form methionine sulfoxide.

The oxidation mechanism has been focus of many
studies since 1968, such as the $S_n2$ type displacement reaction mechanism [5], or recently proposed acid-mediated mechanism or with both acid-catalyzed and protonated water mediated mechanism [6]. But these mechanisms cannot explain all of the available experimental data such as the reaction activation energy and the pH dependence of the oxidation rate. Thus, we tried to resolve this discrepancy by using molecular computational methods. First, we used ab initio quantum chemistry calculation to determine the reaction pathway in gas phase with 2 or 3 waters present around the sulfur site of a single methionine. We found a more reasonable reaction mechanism is that 2 or 3 water molecules stabilize the transition complex via specific interactions including formation of hydrogen bonds with $\text{H}_2\text{O}_2$ but not proton transfer as previously assumed in literature. Also we found the appropriate reaction coordinate leading to oxidation of methionine is the separation of O-O bond together with the formation of S-O bond. This new mechanism meets all available experimental data [7].

In order to understand the connection between the conformation of a protein molecule and the oxidation of its methionine residues, the structural properties of G-CSF (Granulocyte Colony Stimulating Factor) were investigated via molecular dynamics (MD) simulations and were correlated with the rate of oxidation of methionine residues at different pH [8]. The simulation results indicate that the solvent accessible area (SAA), traditionally used to measure solvent accessibility of a protein site, of the sulfur atom of methionine residues does not correlate well with the rate of oxidation. Instead, we identified a structural property, averaged two-shell water coordination number (2SWCN) that can be used to correlate well with the measured oxidation rates.

We also tried to include other protein residues around meth and studied the effect of protein environment on met oxidation. Combined QM/MM methods and transition path sampling simulation were applied to do molecular dynamics and the free energy barrier calculation. It was found that the computed free energies of the oxidation of methionine residues in G-CSF indicate that the protein environment has significant effects on the reaction barrier of oxidation. It is also found the number of hydrogen bonds between the distal oxygen of hydrogen peroxide and the water molecules near the methionine increases along the reaction coordinate as oxidation progresses, indicating that the charge separation developed during the $S_n2$ oxidation by hydrogen peroxide is stabilized by specific interaction with water molecules such as hydrogen bonding.

In the protein oxidation study, a comprehensive picture of methioine oxidation by hydrogen peroxide is developed [9]. It allows an accurate prediction of protein oxidation, and provides a rationale for developing strategies to control oxidation, such as modulating protein conformation via adding excipients. This knowledge could aid in developing in a more rational manner solvent formulations that protect therapeutic proteins against oxidation.

IV. CLATHRATE-HYDRATES

Natural gas clathrate-hydrates (called gas hydrates) are systems of polyhedral cells formed by hydrogen-bonded water molecules forming nonstoichiometric inclusion compounds consisting of a three-dimensional host lattice of water molecules, in which guest molecules, such as methane and/or carbon dioxide, are encaged. The development of production and transmission operations for conventional natural gas hydrates depends on the ability to model and make quantitative predictions of methane hydrate behavior.

Since the late 50’s, the van der Waals and Platteeuw statistical mechanical model with the Lennard-Jones and Devonshire (LJD) spherical cell potential approximation has been the basis of most modeling efforts [10]. However, our studies [11] have demonstrated the inadequacy of the LJD approximation. Moreover, such inadequacy has long been surmised based on the fact that potential parameters computed from gas-hydrate phase data using the LJD approximation do not match those computed from other experimental data.

Additionally, large discrepancies exist in the values of the reference parameters, $\Delta \mu_i$ and $\Delta H_i^\theta$, used in the van der Waals and Platteeuw model [12]. Thus, present methods can be used to fit experimental data, but cannot be generalized to make accurate predictions. Proper determination of the form of the intermolecular interaction potential is necessary both to compute equilibria thermodynamic properties and to perform classical simulations of kinetic phenomena such as formation and dissociation. Experimental techniques, however, cannot be used easily to determine the physical interaction potential. Our objective is therefore to use methods of computational chemistry to develop an accurate and robust multi-dimensional potential between guest and host molecules and then apply it to phase equilibria and kinetics studies.

Ab initio intermolecular potentials between guest and host molecules of gas hydrates provide a means for making consistent and accurate predictions of phase equilibrium thermodynamics and transport properties. High level ab initio quantum mechanical calculations were used to determine the intermolecular potential energy surfaces for the argon–water and methane-water systems. The effects of multi-body interactions within the clathrate hydrate lattice have been evaluated and taken into account and these potentials are used to accurately predict monovariant and invariant phase equilibria for simple and mixed hydrate systems [13].

Structure I to structure II transitions are also predicted for the methane-argon mixed hydrate system. The phase equilibrium predictions for the mixed methane–argon hydrate system as well as structural transitions that are predicted to occur as the gas composition varies are shown.
Methane cage occupancy data is predicted with unprecedented accuracy and reference state parameters for structure I and structure II clathrates are computed for use in the van der Waals and Platteeuw statistical thermodynamic model, see Table 1. Methane cage occupancy predictions can be found in Table 2.

**TABLE 1: CALCULATED THERMODYNAMIC REFERENCE PROPERTIES FOR STRUCTURE I AND STRUCTURE II HYDRATES, WITH 95% CONFIDENCE INTERVALS**

<table>
<thead>
<tr>
<th></th>
<th>structure I</th>
<th>structure II</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \mu^\circ$ (J/mol)</td>
<td>1203±3</td>
<td>1077±5</td>
</tr>
<tr>
<td>$\Delta H^\circ$ (J/mol)</td>
<td>1170±19</td>
<td>1294±11</td>
</tr>
</tbody>
</table>

**TABLE 2: OCCUPANCY RATIO, $\theta_L / \theta_S$, OF METHANE STRUCTURE I HYDRATES**

<table>
<thead>
<tr>
<th>temperature (K)</th>
<th>experimental value [14],[15]</th>
<th>CSMHYD[12]</th>
<th>this study</th>
</tr>
</thead>
<tbody>
<tr>
<td>260</td>
<td>0.916 ± 0.01</td>
<td>0.910</td>
<td>0.920</td>
</tr>
<tr>
<td>273.65</td>
<td>0.947 ± 0.02</td>
<td>0.900</td>
<td>0.906</td>
</tr>
<tr>
<td>274.65</td>
<td>0.925 ± 0.02</td>
<td>0.904</td>
<td>0.910</td>
</tr>
<tr>
<td>275.65</td>
<td>0.892 ± 0.03</td>
<td>0.908</td>
<td>0.914</td>
</tr>
<tr>
<td>276.65</td>
<td>0.890 ± 0.01</td>
<td>0.912</td>
<td>0.917</td>
</tr>
</tbody>
</table>

The potentials developed in this work, along with experimental phase data for single-cage occupying hydrates, are used to compute cell potentials between the guest molecules and the host cages [16]. These cell potentials have been developed via an exact solution to the Lennard-Jones and Devonshire approximation [17] and provide physical insight into the overall guest-host lattice interaction. Figure 2 shows the robustness of the form of the cell potentials of 13 different guest molecules. The van’t Hoff behavior of the Langmuir constants illustrated in Figure 2 provides an analytical form for the temperature dependence of the Langmuir constants, thus allowing us to solve the inverse problem for the form of the cell potential.

Predictions of phase equilibria for mixed hydrate systems are made using these computed cell potentials and experimentally measured structural changes in the cyclopropane, ethane-methane, and ethane-propane systems are predicted accurately [16]. Quintuple (L_{w}-sI-sII-L_{hc}-V) points have been predicted for the ethane-propane-water (277.3 K, 12.28 bar, and $x_{\text{eth,waterfree}} = 0.676$) and ethane-isobutane-water (274.7 K, 7.18 bar, and $x_{\text{eth,waterfree}} = 0.81$) systems.

### V. ZEOLITE CATALYSIS

#### A. Introduction

Zeolites are crystalline, porous aluminosilicates that are commonly used in industry for diverse applications such as solid acid catalysis, ion exchange, and size-selective separations. In the 1970’s, their use in the petrochemical industry was spurred by the energy crisis and subsequent development of the methanol-to-gasoline (MTG) and methanol-to-olefins (MTO) processes by Mobil Oil and UOP/Hydro, respectively.

Unfortunately, there are several reasons why zeolites have not been used to their full potential, compared to liquid-phase catalysts, in industry. One reason is that it is difficult for reactants and products to diffuse to and from the zeolite acid site without the aid of a solvent. More importantly, the individual steps comprising catalytic reactions are not well understood and are difficult to study experimentally.

Another question that needs to be addressed is the role of the zeolite in catalytic reactions. Since zeolites are Brønsted acids, perhaps they may simply serve as a convenient proton source, which would otherwise be unattainable in the gas phase, for acid-catalyzed reactions. Another possibility is that zeolites work by molecular shape selectivity, where the gas phase reactants are constrained within the zeolite cages and channels so that they are in close enough proximity to react. These are short-range repulsions. Recently, it has been thought that the primary role of zeolites is for confinement, which
consists of long-range and attractive interactions. These interactions increase the physisorption energies of the adsorbates up to values that are comparable to the activation energies for reactions. The zeolite framework may also act as a “solid solvent” in the case where the reactants and products are of comparable size to the diameter of the zeolite cages and channels, and work to stabilize the reactive transition state. Most likely, zeolites work by a combination of these factors.

To address these questions, we studied the coupling reaction of two methanol molecules to form ethanol and water in the zeolite chabazite. It is thought that the formation of the first C-C bond is the rate-limiting step for the MTG/MTO processes. Specifically, we focused on computing the reaction mechanism, identifying the correct reaction coordinates, and computing the free energy barriers to reaction.

**B. Methodology**

Chabazite was chosen because it has a small unit cell (36 atoms), which means that the zeolite can be modeled as a periodic system instead of just a cluster of atoms near the Brønsted site. This is more representative of the physical system, since it includes the interactions of the adsorbed methanol molecules with the entire zeolite framework. Also, chabazite has been shown to be catalytically active for the MTG reaction.

Density functional theory, with the PW91 GGA functional, was used to compute the energetics of the molecular system with as much accuracy as possible. Norm-conserving pseudopotentials were used to reduce the computational cost relative to all-electron calculations. Car-Parrinello molecular dynamics was used to simulate the time-evolution of the molecular system, without having to specify the empirical interatomic potentials beforehand. This allowed us to accurately model bond-breaking and bond-forming processes [18].

**C. Results**

Transition path sampling was used to generate an ensemble of reaction paths that connected the reactant and product states, while passing through the transition state at finite temperatures. All of the simulations were initially equilibrated at 400 °C, and then run for 1 ps.

The first step of the methanol coupling reaction starts with the chemisorption of the two methanol molecules to the chabazite acid site. The next step is the breakage of the C-O bond of the methoxonium cation, followed by a proton transfer from the second methanol to the methyl cation, resulting in the formation of the stable intermediates H$_2$O, CH$_4$, and CH$_3$OH$^+$ [19].

The second step of the methanol coupling reaction proceeds by the reorientation of the intermediates, a simultaneous proton transfer from methane to water and C-C bond formation, and the final proton transfer back to the chabazite acid site, resulting in the formation of the products CH$_3$CH$_2$OH and H$_2$O [19].

Based on this mechanism, we identified the reaction coordinates of the first step of the reaction to be the C$_1$-O distance and the [(C$_1$-H*) – (C$_2$-H*)] asymmetric stretch. The reaction coordinates of the second step of the reaction are the C$_1$-C$_2$ distance and the [(O-H*) – (C$_1$-H*)] asymmetric stretch. Using these coordinates, the free energy ($F$) barrier was computed using constrained molecular dynamics and thermodynamic integration:

$$\frac{\partial F}{\partial \xi_j} = \left\langle \left[ Z_\xi \right]^{1/2} \frac{\partial}{\partial \xi_j} \left( -\lambda \right) \right\rangle$$

where $\lambda$ is the Lagrange multiplier on the constraint, which is proportional to the negative of the force due to the constraint, $\xi$ is the value of the constrained reaction coordinate, and the elements of the matrix $Z$ are given by:

$$[Z_{ij}] = \sum_{i=1}^{N} \frac{1}{m_i} \frac{\partial^2 \xi_i}{\partial \chi_i \partial \chi_j}$$

The free energy surfaces for both steps of the methanol coupling reaction are shown in Figures 3 and 4.

![Figure 3: Free energy surface for the first step of the methanol coupling reaction](image)

The overall rate-limiting step, as determined by the height of the free energy barrier, is the formation of the C-C bond from the intermediate species. This height is 106.0 kJ/mol. Our previous work, sampling the free energy barrier across only the C-C distance reaction coordinate, gave us an upper bound to the free energy barrier of 223.5 kJ/mol. The computed free energy barrier is lower than those computed by other researchers for C-C bond forming processes involving the formation of surface methoxy groups.
VI. CONCLUSION

In this work we have made significant contributions in three different areas of interest: therapeutic protein stabilization, thermodynamics of natural gas clathrate-hydrates, and zeolite catalysis. In all three fields, using our various computational techniques, we have been able to elucidate phenomena that are difficult or impossible to explain experimentally. More specifically, in mixed solvent systems for proteins we developed a statistical-mechanical method to model the thermodynamic effects of additives in molecular-level detail. It was the first method demonstrated to have truly predictive (no adjustable parameters) capability for real protein systems. We also describe a novel mechanism that slows protein association reactions, called the “gap effect.”

We developed a comprehensive picture of methionine oxidation by hydrogen peroxide that allows for accurate prediction of protein oxidation and provides a rationale for developing strategies to control oxidation. The method of solvent accessible area (SAA) was shown not to correlate well with oxidation rates. A new property, averaged two-shell water coordination number (2SWCN) was identified and shown to correlate well with oxidation rates.

Reference parameters for the van der Waals Platteauw model of clathrate-hydrates were found for structure I and structure II. These reference parameters are independent of the potential form (unlike the commonly used parameters) and have been validated by calculating phase behavior and structural transitions for mixed hydrate systems. These calculations are validated with experimental data for both structures and for systems that undergo transitions from one structure to another. This is the first method of calculating hydrate thermodynamics to demonstrate predictive capability for phase equilibria, structural changes, and occupancy in pure and mixed hydrate systems.

We have computed a new mechanism for the methanol coupling reaction to form ethanol and water in the zeolite chabazite. The mechanism at 400 °C proceeds via stable intermediates of water, methane, and protonated formaldehyde.

REFERENCES
