

# Using Quantum Mechanical Approaches to Study Biological Systems

Kenneth M. Merz, Jr.

Department of Chemistry and the Department of Biochemistry and Molecular Biology, Michigan State University, 578 S. Shaw Lane, East Lansing Michigan 48824-1322, United States

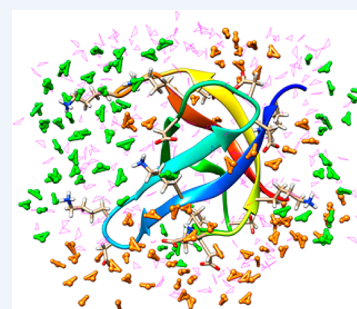
**CONSPECTUS:** Quantum mechanics (QM) has revolutionized our understanding of the structure and reactivity of small molecular systems. Given the tremendous impact of QM in this research area, it is attractive to believe that this could also be brought into the biological realm where systems of a few thousand atoms and beyond are routine. Applying QM methods to biological problems brings an improved representation to these systems by the direct inclusion of inherently QM effects such as polarization and charge transfer. Because of the improved representation, novel insights can be gleaned from the application of QM tools to biomacromolecules in aqueous solution.

To achieve this goal, the computational bottlenecks of QM methods had to be addressed. In semiempirical theory, matrix diagonalization is rate limiting, while in density functional theory or Hartree–Fock theory electron repulsion integral computation is rate-limiting. In this Account, we primarily focus on semiempirical models where the divide and conquer (D&C) approach linearizes the matrix diagonalization step with respect to the system size. Through the D&C approach, a number of applications to biological problems became tractable. Herein, we provide examples of QM studies on biological systems that focus on protein solvation as viewed by QM, QM enabled structure-based drug design, and NMR and X-ray biological structure refinement using QM derived restraints.

Through the examples chosen, we show the power of QM to provide novel insights into biological systems, while also impacting practical applications such as structure refinement. While these methods can be more expensive than classical approaches, they make up for this deficiency by the more realistic modeling of the electronic nature of biological systems and in their ability to be broadly applied. Of the tools and applications discussed in this Account, X-ray structure refinement using QM models is now generally available to the community in the refinement package Phenix.

While the power of this approach is manifest, challenges still remain. In particular, QM models are generally applied to static structures, so ways in which to include sampling is an ongoing challenge. Car–Parrinello or Born–Oppenheimer molecular dynamics approaches address the short time scale sampling issue, but how to effectively use QM to study phenomenon covering longer time scales will be the focus of future research. Finally, how to accurately and efficiently include electron correlation effects to facilitate the modeling of, for example, dispersive interactions, is also a major hurdle that a broad range of groups are addressing.

The use of QM models in biology is in its infancy, leading to the expectation that the most significant use of these tools to address biological problems will be seen in the coming years. It is hoped that while this Account summarizes where we have been, it will also help set the stage for future research directions at the interface of quantum mechanics and biology.



## ■ INTRODUCTION

Quantum mechanics (QM) has revolutionized our understanding of the structure and reactivity of small molecular systems. For example, QM based methods provide structural information in excellent agreement with experiment, match experimental barrier heights for chemical reactions, and provide chemically accurate interaction energies for hydrogen-bonded or dispersive systems.<sup>1,2</sup> Given the tremendous impact QM has had for systems of <100 atoms, it is attractive to believe that this impact could also be brought into the biological realm where systems of a few thousand atoms and beyond are standard. To achieve this goal, the bottlenecks of QM methods have to be addressed. Depending on the method employed, various steps in a QM calculation can be rate determining. In semiempirical methods, matrix diagonalization is rate limiting, while in density functional theory (DFT) or Hartree–Fock (HF) theory electron repulsion integral computation is rate-

limiting.<sup>3</sup> These theories neglect the correlation energy, which is important to account for to obtain highly accurate results including barrier heights and interaction energies (especially dispersion dominated ones). In this case, the correlation treatment of choice is rate-limiting be it Møller–Plesset (MP) theory or coupled-cluster (CC) methods.<sup>3</sup>

State-of-the-art linear-scaling algorithms, which linearize the computational cost with the system size, have attracted much attention.<sup>4–6</sup> Significant effort has been devoted to the development of linear-scaling methods to compute the total energy of large molecular systems at the HF or DFT level.<sup>7–14</sup> One of the challenges is to speed up the calculation of long-

**Special Issue:** Beyond QM/MM: Fragment Quantum Mechanical Methods

**Received:** March 4, 2014

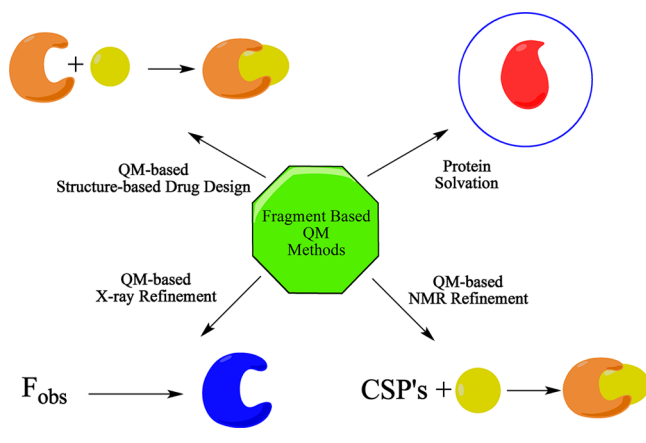
**Published:** June 6, 2014

range Coulomb interactions when assembling the Fock matrix elements. Fast multipole based approaches have successfully reduced the scaling in system size to linear<sup>11,12,15–17</sup> and made HF and DFT calculations affordable for larger systems when small to moderate sized basis sets are utilized. There are also fragment-based methods for QM calculation of protein systems including the divide and conquer (D&C) method of Yang,<sup>8</sup> Yang and Lee,<sup>9</sup> Dixon and Merz,<sup>18,19</sup> Gogonea et al.,<sup>20</sup> Shaw and St-Amant,<sup>21</sup> and Nakai and co-workers,<sup>22–25</sup> the adjustable density matrix assembler (ADMA) approach method of Exner and Mezey,<sup>13,26–28</sup> the fragment molecular orbital (FMO) method of Kitaura and co-workers,<sup>29–31</sup> the X-pol model of Gao and Xie,<sup>32</sup> and the molecular fractionation with conjugate caps (MFCC) approach developed by Zhang and co-workers.<sup>33,34</sup> Most applications of these methods to protein systems have been largely limited to semiempirical, HF, and DFT calculations. Among these approaches, FMO has been applied to higher level ab initio calculations such as second-order Møller–Plesset perturbation theory (MP2)<sup>35</sup> and coupled cluster theory (CC).<sup>36</sup> Nakai and co-workers have described D&C-MP2<sup>22,25,37</sup> and D&C-CCSD<sup>38</sup> approaches and applied them to linear or near-linear systems.

QM excels at the static modeling of a molecular system. However, in chemistry and biology, fluctuations are important to describe a broad range of effects. Thus, the ensemble picture of a molecular system is more appropriate, and while molecular dynamics (MD) methods are beginning to address dynamical issues using classical potentials, how to address this “sampling” issue when using more expensive QM based approaches will need to be addressed.

Herein we briefly review work carried out in my laboratory over the past decade in exploiting fully QM methods in the study of biological systems (see Scheme 1). In particular, we

**Scheme 1. Application of QM Methods to Biological Problems**

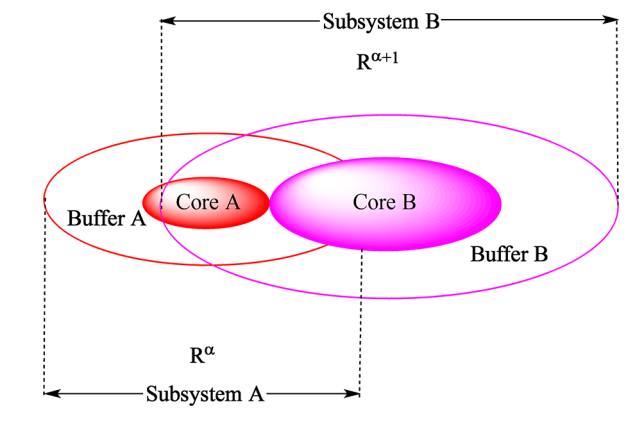


will discuss the use of D&C techniques to address the matrix diagonalization step in semiempirical methods and in HF and DT approaches. After summarizing the technical aspects of how to compute the QM energy of many thousands of atoms, we will review several fully QM application studies carried out for the first time in my laboratory on biological systems. Finally, I will discuss the future outlook of using QM to solve biological problems and the ongoing challenges of using a fully QM model on large systems.

## ■ DIVIDE-AND-CONQUER APPROACH FOR HARTREE–FOCK BASED CALCULATIONS

In biological systems, the D&C approach<sup>8</sup> takes advantage of chemical locality. Details of the implementations of this approach at the semiempirical,<sup>9,18,19</sup> HF and post-HF<sup>21–25,39</sup> levels of theory are given in the literature, so below we only give a brief overview. In this approximation, it is assumed that the atoms that are far away from the region of interest only weakly influence local regions of a protein. Hence, the entire system is divided into fragments called core regions (Core<sup>α</sup>) with an associated buffer region (Buffer<sup>α</sup>) assigned to each core region to account for the local environmental effects. The combination of every core region and its buffer region constitutes a

**Scheme 2. D&C Subsetting (Fragmentation) Scheme**



subsystem ( $R^\alpha$ ) (see Scheme 2). Local MOs of each subsystem are solved by the Roothaan–Hall equation

$$F^\alpha C^\alpha = S^\alpha C^\alpha E^\alpha \quad (1)$$

where  $F^\alpha$  and  $S^\alpha$  are local Fock overlap matrices, respectively.

$$F_{\mu\nu}^\alpha = \begin{cases} F_{\mu\nu} & \text{if } \chi_\mu \in R^\alpha \text{ and } \chi_\nu \in R^\alpha \\ 0 & \text{elsewhere} \end{cases} \quad (2)$$

After the local MO coefficient matrices  $C^\alpha$  are obtained, the total density matrix system is given by

$$P_{\mu\nu} = \sum_{\alpha=1}^{N_{\text{sub}}} P_{\mu\nu}^\alpha = \sum_{\alpha=1}^{N_{\text{sub}}} D_{\mu\nu}^\alpha p_{\mu\nu}^\alpha \quad (3)$$

where  $D_{\mu\nu}^\alpha$  is the partition matrix,

$$D_{\mu\nu}^\alpha = \begin{cases} 1 & \phi_\mu \in \text{Core}^\alpha \text{ and } \phi_\nu \in \text{Core}^\alpha \\ 1/2 & \phi_\mu \in \text{Core}^\alpha \text{ and } \phi_\nu \in \text{Buffer}^\alpha \text{ or } \phi_\mu \in \text{Buffer}^\alpha \text{ and } \phi_\nu \in \text{Core}^\alpha \\ 0 & \phi_\mu \notin \text{Core}^\alpha \text{ and } \phi_\nu \notin \text{Core}^\alpha \end{cases} \quad (4)$$

and  $p_{\mu\nu}^\alpha$  is the local density matrix defined by

$$p_{\mu\nu}^\alpha = \sum_i^{\text{LMOs}} n_i^\alpha C_{\mu i}^\alpha C_{\nu i}^{\alpha*} \quad (5)$$

where  $n_i^\alpha$  is a smooth approximation to the Heaviside step function:

$$n_i^\alpha = \frac{2}{1 + \exp[(\epsilon_i^\alpha - \epsilon_F)/kT]} \quad (6)$$

$\epsilon_F$  is determined through the normalization of the total number of electrons in the entire system.

$$N_{\text{elec}} = \sum_{\alpha} \sum_{\mu} (P^{\alpha} S^{\alpha})_{\mu\mu} \quad (7)$$

After the density matrix is converged, the total HF energy is given as

$$E_{\text{HF}}^{\text{DC}} = \frac{1}{2} \sum_{\alpha} \sum_{\mu\nu} P_{\mu\nu}^{\alpha} (H_{\mu\nu}^{\alpha} + F_{\mu\nu}^{\alpha}) \quad (8)$$

where  $H_{\mu\nu}^{\alpha}$  is the local one-electron core Hamiltonian matrix defined similarly to the local Fock matrix (see eq 2).

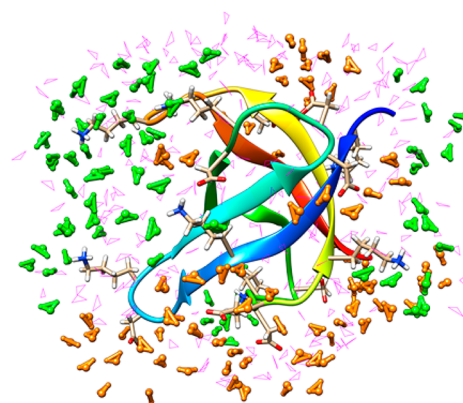
For HF calculations, the construction of the Coulomb matrix and exchange matrix along with the diagonalization of the Fock matrix are the three most time-consuming steps, while the diagonalization of the Hamiltonian matrix scales as  $O(N^3)$ . In the D&C scheme, the diagonalization calculation is performed for each subsystem, linearizing the diagonalization step as a function of the number of subsystems. However, it is important to realize that the D&C algorithm does not help to reduce the scale of computation of the Coulomb matrix and exchange matrix. Other approaches<sup>18</sup> provide ways to linearize these steps. However, for semiempirical methods where matrix diagonalization is rate-limiting, the D&C approach greatly enhances the capabilities of these methods.<sup>9,18,19</sup>

## ■ APPLICATIONS OF FRAGMENT QM TO BIOLOGICAL PROBLEMS

We have used our fragment QM methods (largely at the semiempirical level) to study biological problems with a focus on protein structure and solvation,<sup>40–44</sup> structure-based drug design<sup>45,46</sup> and biological structure refinement by NMR and X-ray spectroscopy.<sup>47–64</sup> Below we will briefly highlight work done in our laboratory using these methods to address specific biological problems.

## ■ A NEW VIEW OF THE PROTEIN–WATER INTERFACE

The interface between a biological molecule and water affects its stability and function.<sup>65</sup> The classical picture of this interface does not allow for the commingling of the electron density (the so-called charge transfer (CT) effect) between the protein and the surrounding aqueous environment. We examined the QM nature of a small protein (*E. coli* cold-shock protein A; CspA) in aqueous solution through the use of semiempirical D&C calculations over 100 snapshots. The surprising result was the observation that two units of charge were transferred from the protein surface to that of the water molecules. Thus, water molecules, at the protein interface, are perturbed not only by the rugged protein interface but also electronically. Via integration over the entire protein surface, the net results of numerous small net transfer of charge yield the observed value of 2 units of charge. Not surprisingly, the main participants in this charge transfer effect are charged amino acids at the protein–water interface (see Figure 1). The magnitude of the CT transfer effect on the energetics of hydrogen bond interactions has been debated,<sup>44,66–68</sup> but we reported<sup>44</sup> a value of  $\sim 30\%$  for a series of dimers using a Kitaura–Morokuma analysis<sup>69</sup> (this latter paper also reports a similar CT effect for the water dimer), a value later supported by more sophisticated calculations.<sup>67</sup> We carried out extensive validation studies of the nature of CT<sup>44</sup> because our pioneering studies employed



**Figure 1.** *E. coli* cold-shock protein A (CspA) surrounded by a water layer. The green water molecules are near Lys residues and experience a slight loss of electron density ( $\sim 0.06e$  in aggregate), while the orange water molecules are near Glu/Asp residues gain electron density ( $\sim 0.2e$  in aggregate). The magenta water molecules are not within 5 Å of the charged residues.

semiempirical models. Through these studies, we concluded that in terms of the magnitude of the CT effect the semiempirical models employed were reliable. Subsequent work using ab initio FMO<sup>70</sup> and ab initio MD calculations<sup>71</sup> reconfirmed our initial predictions.

In related work, charge transfer in receptor ligand interactions in the context of SBDD has been examined by my laboratory.<sup>45</sup> In a study of 165 noncovalent protein–ligand complexes, we found that 11% of the complexes had more than 0.1e of charge transferred from the protein to ligand. Not surprisingly, in 49 metalloenzyme complexes, on average, there is 0.6e transferred between the protein and the ligand. The direction of the CT effect depended on the nature of the protein–ligand complex. For example, in matrix metalloproteases (MMP), charge was transferred from the protein to the ligand, while for human carbonic anhydrase (HCA) and carboxypeptidases (CPA) charge was transferred in the opposite direction.

The view that the interface between a protein, water, or a ligand is electronically altered via polarization and CT effects is now generally accepted. Indeed, the observation of CT at the protein–water interface highlights the need to include this effect in modern force field methods, and work is being pursued along these lines.<sup>72,73</sup> However, without the aid of QM based methods, this effect would not have been discovered and its importance might have gone unappreciated for some time.

## ■ QM IN STRUCTURE-BASED DRUG DESIGN (SBDD)

The docking and subsequent scoring of small molecules bound to receptors has been a mainstay of modern SBDD tools.<sup>74,75</sup> However, the reliability of these approaches has been debated.<sup>75–77</sup> Most tools in this class utilize approximate potentials to study the interaction between the protein and its ligand. We reasoned that the use of more advanced QM based methods might improve the outcome.<sup>45,46</sup>

We reported the first application of linear-scaling methods to SBDD where we calculated the binding affinity of ligands bound to the HCA and CPA with reasonable accuracy.<sup>46</sup> The free energy of binding in solution was calculated using the following set of equations:



$$\Delta G_{\text{bind}} = \Delta G_{\text{b}}^{\text{g}} + \Delta G_{\text{solv}}^{\text{protein-ligand}} - \Delta G_{\text{solv}}^{\text{protein}} - \Delta G_{\text{solv}}^{\text{ligand}}$$

$$\Delta G_{\text{b}}^{\text{g}} = \Delta H_{\text{b}}^{\text{g}} - T\Delta S_{\text{b}}^{\text{g}}$$

$$\Delta H_{\text{b}}^{\text{g}} = \Delta H_{\text{f}} + \left(\frac{1}{R^6}\right)LJ$$

$$\Delta S_{\text{g}} = \Delta SA_{\text{C,N,O,S}} + \text{num}(\text{rot\_bonds})$$
(9)

where the free energy of binding in solution was calculated as the sum of the gas-phase interaction energy ( $\Delta G_{\text{b}}^{\text{g}}$ ) and a solvation correction ( $\Delta G_{\text{solv}}$  terms). The gas-phase interaction energy consisted of enthalpic (semiempirical computed  $\Delta H_{\text{f}}$ 's) and entropic components. The electrostatic part of the enthalpic component was calculated using the DivCon program, employing semiempirical Hamiltonians.<sup>46</sup> A dispersion correction was also utilized (the  $(1/R^6)LJ$  term). The solvation correction was calculated as a difference of the solvation free energies of the protein–ligand complex (PL), the protein (P), and the ligand (L). The solvation free energy was calculated using a Poisson–Boltzmann (PB) based self-consistent reaction field (PB/SCRF) method.<sup>78</sup> A major advantage of PB is the internal dielectric of the protein need not be preset as in classical treatments.<sup>78</sup> Finally, an empirical entropy term depending on the change in solvent accessible area ( $\Delta SA$ ) and a ligand rotatable bond count.

In further studies, we carried out a larger scale validation of this QM based scoring function for predicting binding affinity. We calculated interaction energies for a diverse range of protein–ligand complexes comprising of 165 noncovalent complexes and 49 metalloenzyme complexes.<sup>45</sup> For the 165 noncovalent complexes, the interaction energies, without any fitting, agreed with experimental binding affinity within 2.5 kcal/mol. When different parts of the scoring function were fit to experimental free energy of binding using regression methods, the agreement was within 2.0 kcal/mol. For metalloenzymes, the agreement with experiments without fitting was within 1.7 kcal/mol, and with fitting was within 1.4 kcal/mol. Overall, the correlation with experiment was satisfactory with  $R^2$  values of 0.69<sup>46</sup> and 0.55,<sup>45</sup> respectively. However, while the method performed reasonably it was not a “quantum” improvement over traditional methods

Much further work has been done to explore the use of QM and QM/MM approaches to predict protein–ligand binding affinities<sup>79,80</sup> with the general conclusion that, while QM does help, it is not a panacea, at least in the current incarnation. Beyond simply having a better level of theory, many issues remain to be resolved to produce robust computed binding affinities including the inclusion of explicit water molecules<sup>81</sup> and incorporation of receptor flexibility. These effects plague both simpler and more sophisticated model Hamiltonians and is the subject of much on ongoing research.<sup>82</sup>

## ■ QM BASED STRUCTURE REFINEMENT

### QM NMR Refinement of Protein/Ligand Complexes

NMR spectroscopy has proven itself to be a powerful and versatile tool for the study of protein–ligand interactions. The three-dimensional structures of protein–ligand complexes are determined by combining interproton distance restraints derived from nuclear Overhauser effect (NOE) with other restraints from J coupling constants, hydrogen bonds, and/or residual dipolar couplings. Since Fesik and co-workers

introduced SAR (structure–activity relationship) by NMR,<sup>83</sup> many NMR-based screening methods have been developed to identify potential drug molecules in pharmaceutical research.<sup>84–88</sup> All these techniques take advantage of the fact that, upon ligand binding, significant perturbations can be observed in NMR parameters of either the receptor or the ligand. These perturbations can be utilized qualitatively to detect complex formation or quantitatively to measure the binding affinity.

Among these NMR parameters, chemical shifts are exquisitely sensitive on the chemical environments of compounds. Therefore, theoretical calculation of chemical shift perturbations (CSP) upon ligand binding provides insights into protein–ligand interactions at the molecular level.

There are two categories of computational approaches to calculate NMR chemical shifts: classical/empirical models and QM. The classical/empirical models<sup>89–95</sup> are parametrized to experimental data or DFT results. These approaches are fast so that they can be easily applied to proteins, but are less general being largely focused on protein chemical shifts. In this case, the generality of a QM based approach offers a significant advantage.

We developed a fast and accurate approach<sup>96</sup> to calculate NMR chemical shifts using the D&C method at the semiempirical level of theory. Along with this approach, we have reported the automatic fragmentation QM/MM (AF-QM/MM) method that computes chemical shifts for large systems using ab initio methods.<sup>54</sup> This approach was first applied to the FKBP-GPI complex (see Figures 2 and 3).<sup>48</sup> By

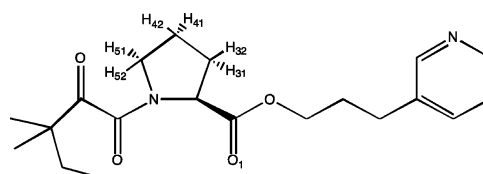


Figure 2. Structure of GPI.

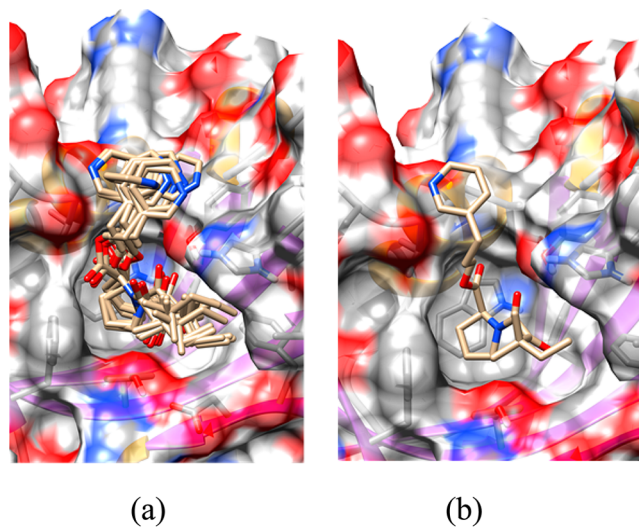


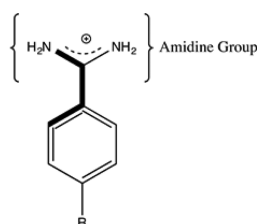
Figure 3. By comparing CSP RMSDs the preferred pose for the FKBP-GPI complex can be determined. Left hand panel (a) shows 10 possible poses, while after analysis by CSP RMSDs between experimental and computed chemical shifts, one pose gives the best agreement between theory and experiment.

comparing calculated proton chemical shifts of the ligand to experimental data, it was possible to determine the best binding site pose. To further validate this approach, we generated several hundred poses of GPI using different docking programs and then scored them by calculating CSPs and then comparing them to experiment.<sup>97</sup> We have found that the deviation of the computed CSPs from experiment better differentiate decoy poses from native poses than scoring functions used in docking studies. This demonstrates that CSP based approaches can provide an accurate way in which to predict protein/ligand complex structure using in silico NMR approaches.

### QM X-ray Refinement

We have also explored the use of QM based methods in X-ray refinement. We utilized D&C QM calculations to refine an entire protein,<sup>98</sup> and we have carried out several studies looking at the QM/MM refinement of small molecules.<sup>59–62</sup> One of the strengths of QM is the ability of these methods to provide high quality structural models of both the protein and the small molecule. Modern X-ray refinement of protein structures typically uses a pseudoenergy formalism where a chemical model ( $E_{\text{Chem}}$ ) is combined with the X-ray signal ( $E_{\text{X-ray}}$ ) for which the gradients are obtained and the energy minimized.<sup>99</sup> In this formulation, the quality of the results depends on both the chemical model term and the quality of the experimental data. For proteins, highly parametrized expressions are available for the  $E_{\text{Chem}}$  term, but the quality of this term for small molecules is variable. However, QM models can treat both the protein and ligand equally well with no parametrization.

As an example of how a QM (in this case a QM/MM refinement) can improve structure quality, I briefly describe our work on benzamidines ligands bound to a number of receptors.<sup>55</sup> In the case of benzamidines, the amidine group (see Figure 4) prefers to be out of the plane of the benzene ring

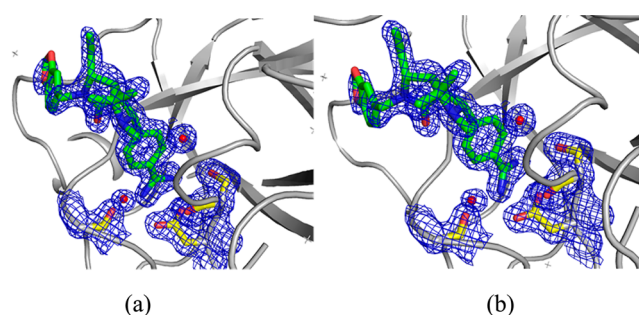


**Figure 4.** Benzamidinium with the amidine group indicated.

by  $\sim \pm 40^\circ$ ; however, many of the reported protein X-ray structures model this group as planar. Using QM/MM X-ray refinement approaches, we demonstrated that this moiety was indeed nonplanar and that the models employed in the earlier refinements did not accurately represent this functional group. Figure 5 shows an example of this for PDBID 1Y3X, where we show the result before and after QM/MM X-ray refinement. Moreover, the standard  $R$ ,  $R_{\text{free}}$  and real-space  $R$  standard structure quality metrics all improved in the QM/MM X-ray refined structures.<sup>55</sup>

### The Future of QM in Biology

QM has already had a major impact on the study of biological systems primarily through its use to build sophisticated classical force field representations of biological macromolecules. Via QM/MM methods,<sup>100</sup> enzymatic catalysis can be routinely studied and this development was awarded the Nobel Prize in Chemistry in 2013. However, it is important to keep in mind that the computational study of biological systems at the



**Figure 5.** Left-hand panel (a) shows the benzamidinium derivative refined using standard approaches, while the right-hand panel (b) shows the results from QM/MM X-ray refinement. Note the out of plane twist of the amidine group.

molecular-level faces two daunting challenges: We must both accurately<sup>101–106</sup> calculate the energies and forces involved, but we must also sample all relevant states of a system. QM largely addresses the former, but how to extensively sample biological systems at the QM level of theory remains a challenging issue. Beyond the creation of faster and more accurate QM models, strategies to address the sampling issue will also have to be devised. This will likely be addressed via a combination of classical and QM models<sup>82</sup> and remains an active area of research.

### AUTHOR INFORMATION

#### Notes

The authors declare no competing financial interest.

#### Biography

**Kenneth M. Merz, Jr.** is the Joseph Zichis Chair in Chemistry and Director of the Institute for Cyber Enabled Research (iCER) at Michigan State University. His research interest lies in the development of theoretical and computational tools and their application to biological problems including structure and ligand based drug design, mechanistic enzymology, and methodological verification and validation. He has received a number of honors including election as an ACS Fellow, the 2010 ACS Award for Computers in Chemical and Pharmaceutical Research, election as a fellow of the American Association for the Advancement of Science, and a John Simon Guggenheim Fellowship.

### ACKNOWLEDGMENTS

This research was generously supported by the NSF and the NIH. I would like to thank the many group members who contributed to the work described herein, whose names and specific contributions can be found in the references cited.

### REFERENCES

- (1) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley & Sons: New York, 1986.
- (2) Hohenstein, E. G.; Sherrill, C. D. Wavefunction methods for noncovalent interactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 304–326.
- (3) Cramer, C. J.: *Essentials of Computational Chemistry: Theories and Models*; 2 ed.; John Wiley & Sons: Chichester, England, 2006.
- (4) Kussmann, J.; Beer, M.; Ochsenfeld, C. Linear-scaling self-consistent field methods for large molecules. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 614–636.

- (5) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation methods: a route to accurate calculations on large systems. *Chem. Rev.* **2012**, *112*, 632–672.
- (6) Richard, R. M.; Herbert, J. M. A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory. *J. Chem. Phys.* **2012**, *137*, 064113.
- (7) Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **1999**, *71*, 1085–1123.
- (8) Yang, W. T. Direct Calculation of Electron-Density in Density-Functional Theory. *Phys. Rev. Lett.* **1991**, *66*, 1438–1441.
- (9) Yang, W. T.; Lee, T. S. A Density-Matrix Divide-and-Conquer Approach for Electronic-Structure Calculations of Large Molecules. *J. Chem. Phys.* **1995**, *103*, 5674–5678.
- (10) Kohn, W. Density functional and density matrix method scaling linearly with the number of atoms. *Phys. Rev. Lett.* **1996**, *76*, 3168–3171.
- (11) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. Achieving linear scaling for the electronic quantum coulomb problem. *Science* **1996**, *271*, 51–53.
- (12) Scuseria, G. E. Linear scaling density functional calculations with Gaussian orbitals. *J. Phys. Chem. A* **1999**, *103*, 4782–4790.
- (13) Exner, T. E.; Mezey, P. G. Ab initio-quality electrostatic potentials for proteins: An application of the ADMA approach. *J. Phys. Chem. A* **2002**, *106*, 11791–11800.
- (14) Friesner, R. A.; Murphy, R. B.; Beachy, M. D.; Ringnalda, M. N.; Pollard, W. T.; Dunietz, B. D.; Cao, Y. X. Correlated ab initio electronic structure calculations for large molecules. *J. Phys. Chem. A* **1999**, *103*, 1913–1928.
- (15) Challacombe, M.; Schwegler, E. Linear scaling computation of the Fock matrix. *J. Chem. Phys.* **1997**, *106*, 5526–5536.
- (16) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. The Continuous Fast Multipole Method. *Chem. Phys. Lett.* **1994**, *230*, 8–16.
- (17) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. Linear scaling density functional calculations via the continuous fast multipole method. *Chem. Phys. Lett.* **1996**, *253*, 268–278.
- (18) Dixon, S. L.; Merz, K. M., Jr. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.* **1996**, *104*, 6643–6649.
- (19) Dixon, S. L.; Merz, K. M., Jr. Fast, Accurate Semiempirical Molecular Orbital Calculations for Macromolecules. *J. Chem. Phys.* **1997**, *107*, 879–893.
- (20) Gogonea, V.; Westerhoff, L. M.; Merz, K. M. Quantum mechanical/quantum mechanical methods. I. A divide and conquer strategy for solving the Schrödinger equation for large molecular systems using a composite density functional-semiempirical Hamiltonian. *J. Chem. Phys.* **2000**, *113*, 5604–5613.
- (21) Shaw, D. M.; St-Amant, A. Linear scaling for density functional calculations on large molecules with the deft software package. *J. Theor. Comput. Chem.* **2004**, *3*, 419–442.
- (22) Kobayashi, M.; Nakai, H. Dual-Level Hierarchical Scheme for Linear-Scaling Divide-and-Conquer Correlation Theory. *Int. J. Quantum Chem.* **2009**, *109*, 2227–2237.
- (23) Akama, T.; Fujii, A.; Kobayashi, M.; Nakai, H. Is the divide-and-conquer Hartree-Fock method valid for calculations of delocalized systems? *Mol. Phys.* **2007**, *105*, 2799–2804.
- (24) Akama, T.; Kobayashi, M.; Nakai, H. Implementation of divide-and-conquer method including Hartree-Fock exchange interaction. *J. Comput. Chem.* **2007**, *28*, 2003–2012.
- (25) Kobayashi, M.; Akama, T.; Nakai, H. Second-order Møller-Plesset perturbation energy obtained from divide-and-conquer Hartree-Fock density matrix. *J. Chem. Phys.* **2006**, *125*, 204106.
- (26) Exner, T. E.; Mezey, P. G. Ab initio quality properties for macromolecules using the ADMA approach. *J. Comput. Chem.* **2003**, *24*, 1980–1986.
- (27) Exner, T. E.; Mezey, P. G. The field-adapted ADMA approach: Introducing point charges. *J. Phys. Chem. A* **2004**, *108*, 4301–4309.
- (28) Exner, T. E.; Mezey, P. G. Evaluation of the field-adapted ADMA approach: absolute and relative energies of crambin and derivatives. *Phys. Chem. Chem. Phys.* **2005**, *7*, 4061–4069.
- (29) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. Fragment molecular orbital method: use of approximate electrostatic potential. *Chem. Phys. Lett.* **2002**, *351*, 475–480.
- (30) Fedorov, D. G.; Kitaura, K. The three-body fragment molecular orbital method for accurate calculations of large systems. *Chem. Phys. Lett.* **2006**, *433*, 182–187.
- (31) Fedorov, D. G.; Kitaura, K. Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J. Phys. Chem. A* **2007**, *111*, 6904–6914.
- (32) Xie, W.; Gao, J. The Design of a Next Generation Force Field: The X-POL Potential. *J. Chem. Theory Comput.* **2007**, *3*, 1890–1900.
- (33) Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599–3605.
- (34) He, X.; Zhang, J. Z. H. A new method for direct calculation of total energy of protein. *J. Chem. Phys.* **2005**, *122*, 031103.
- (35) Fedorov, D. G.; Ishimura, K.; Ishida, T.; Kitaura, K.; Pulay, P.; Nagase, S. Accuracy of the three-body fragment molecular orbital method applied to Møller-Plesset perturbation theory. *J. Comput. Chem.* **2007**, *28*, 1476–1484.
- (36) Fedorov, D. G.; Kitaura, K. Coupled-cluster theory based upon the fragment molecular-orbital method. *J. Chem. Phys.* **2005**, *123*, 134103.
- (37) Kobayashi, M.; Imamura, Y.; Nakai, H. Alternative linear-scaling methodology for the second-order Møller-Plesset perturbation calculation based on the divide-and-conquer method. *J. Chem. Phys.* **2007**, *127*, 074103.
- (38) Kobayashi, M.; Nakai, H. Extension of linear-scaling divide-and-conquer-based correlation method to coupled cluster theory with singles and doubles excitations. *J. Chem. Phys.* **2008**, *129*, 044103.
- (39) He, X.; Merz, K. M., Jr. Divide-and-Conquer Hartree-Fock Calculations on Proteins. *J. Chem. Theory Comput.* **2010**, *6*, 405–411.
- (40) van der Vaart, A.; Merz, K. M., Jr. The Role of Polarization and Charge Transfer in the Solvation of Biomolecules. *J. Am. Chem. Soc.* **1999**, *121*, 9182–9190.
- (41) van der Vaart, A.; Merz, K. M., Jr. Divide and Conquer Interaction Energy Decomposition. *J. Phys. Chem. A* **1999**, *103*, 3321–3329.
- (42) Suárez, D.; Gogonea, V.; van der Vaart, A.; Merz, K. M., Jr. New Developments in Applying Quantum Mechanics to Proteins. *Curr. Opin. Struct. Biol.* **2001**, *11*, 217–223.
- (43) van der Vaart, A.; Bursulaya, B. D.; Brooks, C. L., III; Merz, K. M., Jr. Are Many-Body Effects Important in Protein Folding? *J. Phys. Chem. B* **2000**, *104*, 9554–9563.
- (44) Vaart, A. v. d.; Merz, J. K. M. Charge transfer in biologically important molecules: comparison of high-level ab initio and semiempirical methods. *Int. J. Quantum Chem.* **2000**, *77*, 27–43.
- (45) Raha, K.; Merz, K. M., Jr. Large-Scale Validation of a Quantum Mechanics Based Scoring Function: Predicting the Binding Affinity and the Binding Mode of a Diverse Set of Protein-Ligand Complexes. *J. Med. Chem.* **2005**, *48*, 4558–4575.
- (46) Raha, K.; Merz, K. M., Jr. A Quantum Mechanics Based Scoring Function: Study of Zinc-ion Mediated Ligand Binding. *J. Am. Chem. Soc.* **2004**, *126*, 1020–1021.
- (47) Wang, B.; Brothers, E. N.; van der Vaart, A.; Merz, K. M., Jr. Fast semiempirical calculations for nuclear magnetic resonance chemical shifts: a divide-and-conquer approach. *J. Chem. Phys.* **2004**, *120*, 11392–11400.
- (48) Wang, B.; Raha, K.; Merz, K. M., Jr. Pose Scoring by NMR. *J. Am. Chem. Soc.* **2004**, *126*, 11430–11431.
- (49) Wang, B.; Merz, K. M., Jr. Validation of the binding site structure of the cellular retinol-binding protein (CRBP) by ligand NMR chemical shift perturbations. *J. Am. Chem. Soc.* **2005**, *127*, 5310–5311.



- (50) Yu, N.; Hayik, S. A.; Wang, B.; Liao, N.; Reynolds, C. H.; Merz, K. M. Assigning the protonation states of the key aspartates in beta-Secretase using QM/MM X-ray structure refinement. *J. Chem. Theory Comput.* **2006**, *2*, 1057–1069.
- (51) Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; Wollacott, A. M.; Westerhoff, L. M.; Merz, K. M., Jr. The role of quantum mechanics in structure-based drug design. *Drug Discovery Today* **2007**, *12*, 725–731.
- (52) Wang, B.; Westerhoff, L. M.; Merz, K. M., Jr. A critical assessment of the performance of protein-ligand scoring functions based on NMR chemical shift perturbations. *J. Med. Chem.* **2007**, *50*, 5128–5134.
- (53) Williams, D. E.; Peters, M. B.; Wang, B.; Merz, K. M., Jr. MNDO parameters for the prediction of <sup>19</sup>F NMR chemical shifts in biologically relevant compounds. *J. Phys. Chem. A* **2008**, *112*, 8829–8838.
- (54) He, X.; Wang, B.; Merz, K. M. Protein NMR Chemical Shift Calculations Based on the Automated Fragmentation QM/MM Approach. *J. Phys. Chem. B* **2009**, *113*, 10380–10388.
- (55) Li, X.; He, X.; Wang, B.; Merz, K., Jr. Conformational variability of benzamidinium-based inhibitors. *J. Am. Chem. Soc.* **2009**, *131*, 7742–7754.
- (56) Wang, B.; Dossey, A. T.; Walse, S. S.; Edison, A. S.; Merz, K. M., Jr. Relative configuration of natural products using NMR chemical shifts. *J. Nat. Prod.* **2009**, *72*, 709–713.
- (57) Williams, D. E.; Peters, M. B.; Wang, B.; Roitberg, A. E.; Merz, K. M. AM1 Parameters for the Prediction of H-1 and C-13 NMR Chemical Shifts in Proteins. *J. Phys. Chem. A* **2009**, *113*, 11550–11559.
- (58) Cui, G.; Li, X.; Merz, K. M., Jr. Understanding the substrate selectivity and the product regioselectivity of Orf2-catalyzed aromatic prenylations. *Biochemistry* **2007**, *46*, 1303–1311.
- (59) Fu, Z.; Li, X.; Merz, K. M., Jr. Accurate assessment of the strain energy in a protein-bound drug using QM/MM X-ray refinement and converged quantum chemistry. *J. Comput. Chem.* **2011**, *32*, 2587–2597.
- (60) Fu, Z.; Li, X.; Merz, K. M., Jr. Conformational Analysis of Free and Bound Retinoic Acid. *J. Chem. Theory Comput.* **2012**, *8*, 1436–1448.
- (61) Fu, Z.; Li, X.; Miao, Y.; Merz, K. M., Jr. Conformational analysis and parallel QM/MM X-ray refinement of protein bound anti-Alzheimer drug donepezil. *J. Chem. Theory Comput.* **2013**, *9*, 1686–1693.
- (62) Li, X.; Fu, Z.; Merz, K. M., Jr. QM/MM refinement and analysis of protein bound retinoic acid. *J. Comput. Chem.* **2012**, *33*, 301–310.
- (63) Li, X.; Hayik, S. A.; Merz, K. M., Jr. QM/MM X-ray refinement of zinc metalloenzymes. *J. Inorg. Biochem.* **2010**, *104*, 512–522.
- (64) Yu, N.; Li, X.; Cui, G.; Hayik, S. A.; Merz, K. M., 2nd Critical assessment of quantum mechanics based energy restraints in protein crystal structure refinement. *Protein Sci.* **2006**, *15*, 2773–2784.
- (65) Pal, S. K.; Peon, J.; Zewail, A. H. Biological water at the protein surface: Dynamical solvation probed directly with femtosecond resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1763–1768.
- (66) Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **1988**, *88*, 899–926.
- (67) Khaliullin, R. Z.; Bell, A. T.; Head-Gordon, M. Analysis of charge transfer effects in molecular complexes based on absolutely localized molecular orbitals. *J. Chem. Phys.* **2008**, *128*, 184112.
- (68) Mo, Y.; Gao, J. Polarization and Charge-Transfer Effects in Lewis Acid–Base Complexes. *J. Phys. Chem. A* **2001**, *105*, 6530–6536.
- (69) Kitaura, K.; Morokuma, K. A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *Int. J. Quantum Chem.* **1976**, *10*, 325–340.
- (70) Komeiji, Y.; Ishida, T.; Fedorov, D. G.; Kitaura, K. Change in a protein's electronic structure induced by an explicit solvent: An ab initio fragment molecular orbital study of ubiquitin. *J. Comput. Chem.* **2007**, *28*, 1750–1762.
- (71) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Charge Transfer and Polarization in Solvated Proteins from Ab Initio Molecular Dynamics. *J. Phys. Chem. Lett.* **2011**, *2*, 1789–1793.
- (72) Soniat, M.; Rick, S. W. The effects of charge transfer on the aqueous solvation of ions. *J. Chem. Phys.* **2012**, *137*, 044511.
- (73) Wick, C. D.; Lee, A. J.; Rick, S. W. How intermolecular charge transfer influences the air-water interface. *J. Chem. Phys.* **2012**, *137*, 154701.
- (74) Kuntz, I. D. Structure-Based Strategies for Drug Design and Discovery. *Science* **1992**, *257*, 1078–1082.
- (75) Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *J. Med. Chem.* **2006**, *49*, 5851–5855.
- (76) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
- (77) Merz, K. M., Jr. Limits of Free Energy Computation for Protein-Ligand Interactions. *J. Chem. Theory Comput.* **2010**, *6*, 1018–1027.
- (78) Gogonea, V.; Merz, K. M., Jr. Fully Quantum Mechanical Description of Proteins in Solution. Combining Linear Scaling Quantum Mechanical Methodologies with the Poisson-Boltzmann Equation. *J. Phys. Chem. A* **1999**, *103*, 5171–5178.
- (79) Lodola, A.; De Vivo, M. The increasing role of QM/MM in drug discovery. *Adv. Protein Chem. Struct. Biol.* **2012**, *87*, 337–362.
- (80) Peters, M. B.; Raha, K.; Merz, K. M., Jr. Quantum mechanics in structure-based drug design. *Curr. Opin. Drug Discovery Dev.* **2006**, *9*, 370–379.
- (81) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 808–813.
- (82) Ucisik, M. N.; Zheng, Z.; Faver, J. C.; Merz, K. M. Bringing Clarity to the Prediction of Protein–Ligand Binding Free Energies via “Blurring.” *J. Chem. Theory Comput.* **2014**, *10*, 1314–1325.
- (83) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534.
- (84) Homans, S. W. NMR spectroscopy tools for structure-aided drug design. *Angew. Chem., Int. Ed. Engl.* **2004**, *43*, 290–300.
- (85) Lepre, C. A.; Moore, J. M.; Peng, J. W. Theory and applications of NMR-based screening in pharmaceutical research. *Chem. Rev.* **2004**, *104*, 3641–3676.
- (86) Meyer, B.; Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 864–890.
- (87) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **2005**, *48*, 2518–2525.
- (88) Hajduk, P. J.; Huth, J. R.; Tse, C. Predicting protein druggability. *Drug Discovery Today* **2005**, *10*, 1675–1682.
- (89) Sitkoff, D.; Case, D. A. Density functional calculations of proton chemical shifts in model peptides. *J. Am. Chem. Soc.* **1997**, *119*, 12262–12273.
- (90) Wishart, D. S.; Watson, M. S.; Boyko, R. F.; Sykes, B. D. Automated <sup>1</sup>H and <sup>13</sup>C chemical shift prediction using the BioMagResBank. *J. Biomol. NMR* **1997**, *10*, 329–336.
- (91) Iwadata, M.; Asakura, T.; Williamson, M. P. C-alpha and C-beta carbon-13 chemical shifts in proteins from an empirical database. *J. Biomol. NMR* **1999**, *13*, 199–211.
- (92) Xu, X. P.; Case, D. A. Automated prediction of <sup>15</sup>N, <sup>13</sup>Calpha, <sup>13</sup>Cbeta and <sup>13</sup>C' chemical shifts in proteins using a density functional database. *J. Biomol. NMR* **2001**, *21*, 321–333.
- (93) McCoy, M. A.; Wyss, D. F. Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. *J. Am. Chem. Soc.* **2002**, *124*, 11758–11763.
- (94) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. Fast and Accurate Predictions of Protein NMR Chemical Shifts from Interatomic Distances. *J. Am. Chem. Soc.* **2009**, *131*, 13894–13895.

(95) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G.; Eletsky, A.; Wu, Y.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685–4690.

(96) Wang, B.; Brothers, E. N.; Van Der Vaart, A.; Merz, K. M. Fast semiempirical calculations for nuclear magnetic resonance chemical shifts: A divide-and-conquer approach. *J. Chem. Phys.* **2004**, *120*, 11392–11400.

(97) Wang, B.; Westerhoff, L. M.; Merz, K. M., Jr. A Critical Assessment of the Performance of Protein–Ligand Scoring Functions Based on NMR Chemical Shift Perturbations. *J. Med. Chem.* **2007**, *50*, 5128–5134.

(98) Yu, N.; Yennawar, H. P.; Merz, K. M., Jr. Refinement of protein crystal structures using energy restraints derived from linear-scaling quantum mechanics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2005**, *61*, 322–332.

(99) Jack, A.; Levitt, M. Refinement of large structures by simultaneous minimization of energy and R factor. *Acta Crystallogr. A* **1978**, *34*, 931–935.

(100) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbon Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(101) Faver, J. C.; Benson, M. L.; He, X.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Sherrill, C. D.; Merz, K. M. The Energy Computation Paradox and ab initio Protein Folding. *PLoS One* **2011**, *6*, e18868.

(102) Faver, J. C.; Benson, M. L.; He, X. A.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Kennedy, M. R.; Sherrill, C. D.; Merz, K. M. Formal Estimation of Errors in Computed Absolute Interaction Energies of Protein-Ligand Complexes. *J. Chem. Theory Comput.* **2011**, *7*, 790–797.

(103) Faver, J. C.; Merz, K. M., Jr. Fragment-based error estimation in biomolecular modeling. *Drug Discovery Today* **2014**, *19*, 45–50.

(104) Faver, J. C.; Ucisik, M. N.; Yang, W.; Merz, K. M., Jr. Computer-aided Drug Design: Using Numbers to your Advantage. *ACS Med. Chem. Lett.* **2013**, *4*, 812–814.

(105) Faver, J. C.; Yang, W.; Merz, K. M., Jr. The Effects of Computational Modeling Errors on the Estimation of Statistical Mechanical Variables. *J. Chem. Theory Comput.* **2012**, *8*, 3769–3776.

(106) Ucisik, M. N.; Dashti, D. S.; Faver, J. C.; Merz, K. M. Pairwise additivity of energy components in protein-ligand binding: The HIV II protease-Indinavir case. *J. Chem. Phys.* **2011**, *135*, 085101.