

Interpolation for Molecular Dynamics Simulations: From Ions in Gas Phase to Proteins in Solution

Young Min Rhee^{*[a,b]} and Jae Woo Park^[a,b]

The interpolation technique has shown many promises for simulating chemical dynamics with quantum chemical accuracy at molecular mechanics speed. This is achieved by constructing analytic potential energy surfaces with quantum chemical information at multiple conformational points, without assuming any functional form for the potentials. Here, we briefly review the course the method was developed over the past few decades, with a special focus on the activities in

Korea. We also describe its strengths and weaknesses toward describing condensed phase chemical dynamics with the present implementations. Perspectives for future developments toward increasing applicability are discussed as concluding remarks. © 2015 Wiley Periodicals, Inc.

DOI: 10.1002/qua.25064

Introduction

Chemists aim to study diverse changes in time in many situations. Such changes, which are often referred to as chemical dynamics, can be followed either with experiments or with computations. For reliable computational descriptions, chemists need to have accurate depiction of the associated potential energy surfaces. Traditionally, analytic models of potential surfaces were applied for this purpose. With increasing complexity of the target system, modeling becomes trickier and trickier and obtaining even a qualitatively accurate potential surface may become difficult. The most reliable and unbiased way of generating the potential energy information of an arbitrary molecular system is to adopt quantum chemical approaches. It is also possible to employ quantum chemical approaches during dynamics simulations in an “on-the-fly” manner. Indeed, such direct dynamics has been successfully adopted in a wide range of applications from gas phase dynamics of small molecules to complex changes in proteins. However, using a direct dynamics method for simulating complex systems for a long-enough duration becomes practically intractable especially when statistical analyses are sought for. Accordingly, modeling surfaces is essentially unavoidable in many chemical systems at least with the present day’s conventional computational power. This fact generates a necessity of having a methodology that can calculate the surface information in an efficient manner with the accuracy that can closely match the one obtained through quantum chemical means.

A peculiar approach with such capability is with the concept of Shepard interpolation pioneered by Collins and co-workers from early 1990’s.^[1–3] The approach was soon adopted by many researchers toward studying both classical molecular dynamics and quantum wavepacket dynamics.^[4–8] In this perspective, we will review on the directions the interpolation approaches have been developed in Korea with the purpose of eventually discussing what can also be examined in future studies. Due to the length limitation, we will only briefly cover the course and the background that

the interpolation approach was developed in the country, along with its contributions to understanding varying aspects of chemistry. We will also discuss recent developmental efforts and the related limitations. This will naturally lead to concluding considerations on potential future developments that are desired to extend the concept of interpolation. Due to the nature of the perspective, we cannot avoid omitting references to numerous relevant literatures, especially many key papers along the developmental history of interpolation itself. Curious readers on this topic are suggested to read further on reviewing parts in recent articles^[3,9] to access more comprehensive literature list.

The Concept

How the interpolation works toward generating global potential energy surfaces is actually quite simple and can be understood easily with Figure 1. With quantum chemical means, in most cases if not all, one can straightforwardly obtain pieces of local harmonic surfaces. By mathematically stitching such local surfaces together, one can also build a global potential model. Indeed, its reliability can be very high if there are enough local harmonic representations and the adopted level of quantum chemical means is high enough. Overall, the potential energy V can be written as

$$V(\mathbf{X}) = \sum_a^N w_a(\mathbf{X}) V_a(\mathbf{X}) \quad (1)$$

where $w_a(\mathbf{X})$ is the interpolation weight and $V_a(\mathbf{X})$ is the local harmonic expansion at \mathbf{X} computed with the energy (E_a), the

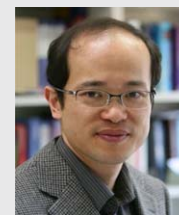
[a] Y. M. Rhee, J. W. Park
 Center for Self-Assembly and Complexity, Institute for Basic Science (IBS), Pohang, 37673, Korea.

[b] Y. M. Rhee, J. W. Park
 Department of Chemistry, Pohang University of Science and Technology (POSTECH), Pohang, 37673, Korea
 E-mail: ymrhee@postech.ac.kr

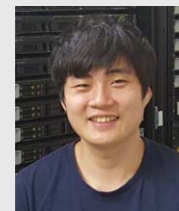
Contract grant sponsor: Institute for Basic Science (IBS), Korea.

© 2015 Wiley Periodicals, Inc.

Young Min Rhee is an Associate Professor of Chemistry in POSTECH and a Group Leader of Institute for Basic Science in Korea. He got his PhD degree from Stanford University in 2005 and worked as a postdoctoral researcher in University of California, Berkeley before joining POSTECH faculty in 2008. His main research interest lies on studying dynamics of complex systems especially with electronic excitations.



Jae Woo Park obtained his BS degree in Chemistry at Pohang University of Science and Technology (POSTECH) in 2010, and his PhD degree in the same university in 2015 under the supervision of Prof. Rhee. He is currently a Postdoctoral Researcher at the Center for Self-assembly and Complexity of the Institute for Basic Science (IBS) in Korea.



gradient (\mathbf{g}_a), and the Hessian (\mathbf{H}_a) at a displacement Δ_a from \mathbf{X} :

$$V_a(\mathbf{X}) = E_a + \mathbf{g}_a \cdot \Delta_a + \frac{1}{2} \Delta_a \cdot \mathbf{H}_a \cdot \Delta_a \quad (2)$$

The weight is basically determined from the distance between \mathbf{X} and \mathbf{X}_a . Once the procedures of collecting the geometries of the N data points (indexed by the dummy variable a) are defined, the surface $V(\mathbf{X})$ eventually becomes mathematically well-defined. The collection can be performed in a manner to systematically improve the surface quality, potentially reaching an exact accord with the reference quantum chemical method in an ideal situation.

In the actual implementation, a variety of coordinate systems can be applied for the Taylor expansion and the weight.^[1,6,10–12] In fact, there is no need to employ the same set of coordinates for them. Then, Eq. (1) actually reads as

$$V(\mathbf{X}) = \sum_a^N w_a(\mathbf{R}_w(\mathbf{X})) V_a(\mathbf{R}_V(\mathbf{X})) \quad (3)$$

For example, using Cartesian coordinates for weight (\mathbf{R}_w) and the internal coordinates for the expansion (\mathbf{R}_V) is possible. Also, for the internal, a set of interatomic distances $\{r_{ij}\}$ or its functional transforms such as $\{1/r_{ij}\}$ or $\{e^{-\beta r_{ij}}\}$ have been utilized. For large molecules, interatomic distances become highly redundant and valence internal coordinates with bond lengths and bending/torsional angles can be adopted. Of course, in the actual implementation with the derivative $\nabla V(\mathbf{X})$, the Jacobian factors for the coordinate transform from the Cartesian should be properly handled. The displacement vector should also correspond to the coordinates adopted for the derivative properties in Eq. (2). For the weight, a simple function that is inversely proportional to an integral power of the Euclidean distance, for example,

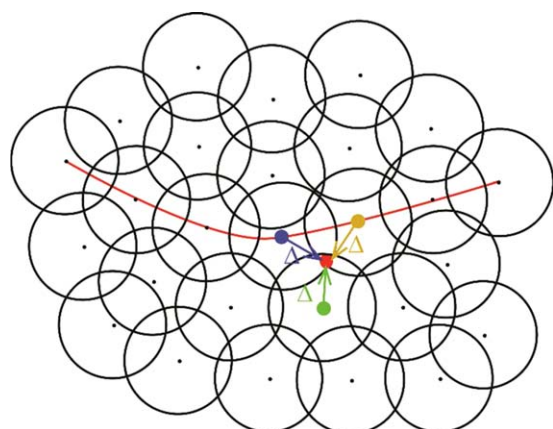
$$v_a(\mathbf{X}) = \frac{1}{|\mathbf{R}_w(\mathbf{X}) - \mathbf{R}_w(\mathbf{X}_a)|^{2p}} \quad (4)$$

can be utilized after normalization, $w_a = v_a / \sum_i v_i$.

Early Developments: Gas Phase Reactions

Interestingly enough, the interpolation approach in Korea was first developed from an experimental group about two decades ago for studying unimolecular decomposition reaction dynamics of energized polyatomic ionic systems in the gas phase.^[13] Even though constructing analytical model surfaces was more of a fashion at that time for studying polyatomic reactions, it was apparent that such surfaces were often not adequate for predicting detailed aspects such as exit channel energy distribution (e.g., kinetic energy releases from ion fragmentations^[13]). Even though the *ab initio* direct dynamics approach, which was just at its infancy around that time, was superior in reproducing such details,^[14] it was practically too costly especially for treating polyatomic systems.

In this sense, at that time, it was natural to attempt to build a method with which the dynamics simulations would be performed based on *ab initio* or other quantum chemical data without discarding previously calculated potential surface information. The interpolation approach exactly fitted this purpose.^[1,6] After successfully implementing this approach, it was shown with an analytic surface model^[15] that the dynamics simulations with an interpolated surface could closely mimic the simulation results from the original analytic surface. From the initiation, the method was developed to handle polyatomic ions such as CH_2OH^+ .^[6,16,17] To some surprise, only a set of hundreds of data points for N in Eq. (1) was enough to have good exit channel energy distribution pattern with classical trajectory calculations.^[6,18] This is first because the original approach^[6] was developed by focusing on reproducing the surface reliably near the minimum energy path (intrinsic reaction coordinate)^[19,20] and also because ~ 300 data points with



$$\begin{aligned}
 V(\bullet) &= w(\bullet)V(\bullet) + w(\bullet)V(\bullet) + w(\bullet)V(\bullet) + \dots \\
 V(\bullet) &= V(\bullet) + \Delta_e(\bullet)g(\bullet) + 0.5\Delta_e(\bullet)H(\bullet)\Delta_e(\bullet) \\
 w(\bullet) &= 1/(\Delta_w)^{2p}
 \end{aligned}$$

Figure 1. Schematic illustration of the concept of potential energy interpolation. The locally harmonic surfaces (solid circles) are generated as Taylor expansions ($V(\cdot)$) from data points (yellow/green/blue dots) and then “stitched” together to generate a globally reliable energy representation at an arbitrary configuration (red dot). The actual stitching is achieved through Shepard interpolation with distance-based weights, w . The displacement vectors for the expansion (Δ_e) and the weighting (Δ_w) can be based on different coordinate systems.

up to Hessian information actually consist of $\sim 40,000$ independent numerical data. (Of course, the data size grows quadratically with the system size.) Additionally, an attempt to build a very reliable and smooth surface that could be used for quantum nuclear wavepacket dynamics was also made with a successful demonstration with the dynamics of the $\text{Cl} + \text{H}_2$ system.^[21]

Recent Progresses: Solution Phase Simulations

More recently, the focus has been shifted to extending the approach to studying condensed phase chemical dynamics.^[22] Indeed, there should not be any serious hurdle against using the interpolation in the condensed phase situation. As interpolating all molecules including solvent would really require infinitely large data set in the practical sense and would be excessively time consuming, a system needs to be divided into an interpolation region and its surrounding. Interfacing these two regions^[23] becomes conceptually similar to what have been developed in more conventional quantum mechanics/molecular mechanics (QM/MM)^[24] approaches. In an analogy to this QM/MM approach, we will call the interpolated region of the system as “interpolation mechanics (IM) region” and the combined multiscale approach as IM/MM.^[23] Just like in QM/MM, we can calculate (1) the electrostatic interaction between the IM and MM regions by computing the electric field from the IM region and (2) the dispersion interaction between the two regions by assigning Lennard–Jones parameters to the IM atoms.^[25] If we ignore the polarizability of the IM part, we can calculate the electric field based on atomic partial charges that are obtained by fitting to electrostatic potential map of the IM

part. This simple idea, supplemented by additional tweaks for extending the simulation timescales such as adding bounding potentials for stabilizing trajectory integrations,^[22] was shown to be quite satisfactory in reproducing spectroscopic features of a solvated chromophore^[22] with quantitatively reliable excited state vibrational frequencies and good agreements in their dephasing timescales.

The method was then extended to a protein-embedded chromophore with an example with the blue fluorescent protein (BFP).^[23] In this case, the chromophore is covalently bound to the protein backbone. Thus, a method of extracting the chromophore as a whole molecule with completed valence was needed. There, again, an algorithm was constructed in an analogy to the QM/MM approach by introducing link atoms^[24] that could cap the IM part for valence completion. In BFP, the chromophore twisting in its electronically excited state is important in deciding its fluorescent lifetime (Figure 2). Because this twisting from the Franck–Condon point after vertical excitation happens in a downhill manner with a rather well-defined pathway to the stable geometry on the excited state surface, it was again shown that ~ 300 data points were enough to reproduce features that were experimentally observed, especially the fluorescent lifetimes of BFP variants.^[23] Most importantly, numerous trajectory simulations could be performed to reach an aggregate simulation time of $10 \mu\text{s}$. Reaching this timescale with more standard QM/MM direct dynamics simulations would be impractical with present days’ computing power. With the many trajectories, of course, statistically meaningful analyses could be performed.^[26] For example, the electrostatic interactions between the twisting chromophore and neighboring protein residues were analyzed

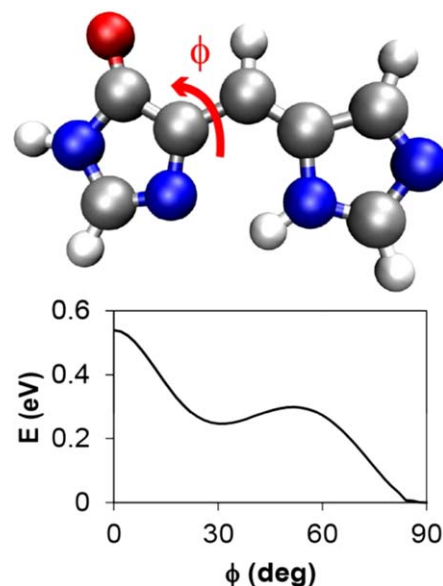


Figure 2. The dominant twisting pathway of the blue fluorescent protein chromophore in its emissive state. The potential along this twisting pathway is highly anharmonic and interpolation performed quite well in describing this surface characteristic. The modulations by the protein on the twisting timescales were also well described by interpolation, when combined with the molecular mechanics approach (IM/MM). See Ref. [24] for details.

to show that the proteins with longer lasting fluorescence possess charged residues that tend to repel the chromophore's twisting moiety.^[23] The ability of performing statistical analyses will be crucial in understanding solution phase dynamics as the related diversities and noises will likely prevail in such systems. In this sense, extending the simulation efficiency through the interpolation will be an important key for studying similar systems in the future.

Outlook to the Future

Even with the various recent developments, the interpolation is still rather far from being a black box tool, especially in studying complex systems in the condensed phase. One notorious issue will be the negative curvature problem, which can be easily imagined from the structure of Eq. (2). Namely, when the Hessian at a data point has a negative eigen-component, the Taylor expansion at \mathbf{X} may become strongly negative at a large distance Δ_a . If the interpolating weight decays faster than a quadratic function, and if there are other points at closer distances that can correctly describe the potential at \mathbf{X} , this does not pose any problem. The issue may become serious when the trajectory visits a distant conformation away from all available data points. In such a case, the potential may become incorrectly too low and the atoms in the simulated system may gain excessively high kinetic energy. This will of course induce instabilities to the simulations. Indeed, such distant conformations are more frequently visited in condensed phase situations with larger energy fluctuations and longer simulation times. Different flavors of remedies have been proposed such as adding simple stabilizing bounding potential^[22] or adopting more sophisticated weighting scheme with a Bayesian analysis.^[27] Even still, these schemes tend to be somewhat system specific and developing a more generalizable approach will be beneficial for future applications. In many cases, the rarely visited conformations do not contribute importantly to the overall dynamics in the average sense. Besides, if they are important to the dynamics, then their neighboring areas should be handled properly by adding more data points to those areas. Thus, a practical approach may be designed to avoid the instability without affecting the simulation results in the statistical sense.

The next important issue is the intrinsic difference between the IM/MM and QM/MM approaches. As alluded in the Introduction, what interpolation aims for is mimicking the reliability of the reference quantum chemical calculations. In the traditional QM/MM calculations, the QM part is naturally polarizable and can properly (or sometimes even overly) respond to the changes of the MM part. Polarizable MM models can also be employed and they can actually eliminate the over polarization issue of the QM part in principle. In the current implementation, the interpolation part is not polarizable at all and this aspect will intrinsically limit the accordance of IM/MM surface to the reference QM/MM surface. In the simplest sense, the polarization in the IM part should be handled from two different angles. First, the molecule that is described by interpolation will have different electron density at different

geometries. This is not exactly polarization per se, as it is not affected by external electric field. As such, it can be described not by the polarizability but by the geometrically varying atomic charges of the interpolated system. In fact, geometric dependence of any property may well be described by interpolation itself, and improvements in this area may become available in the near future. Next, the true polarization from external field should be treatable by borrowing ideas from the fast developing polarizable models.^[28,29] A polarizable IM part can be mixed with either a nonpolarizable or a polarizable MM part, just to mimic more classic nonpolarizable QM/MM or more recent polarizable QM/MM. Employing external electrostatic potential as an interpolating variable, as utilized for evaluating the off-diagonal Hamiltonian matrix terms in EE-MCMM method by Higashi and Truhlar^[30] may also be an amenable solution for describing polarizations.

One additional issue that warrants some considerations is the algorithm for collecting data points. In the past, often, the data set efficiency was an important criterion during its construction. Namely, constructing a surface with a given reliability with less data points was considered to be beneficial as it will reduce the computational cost for both the data set construction and the subsequent dynamics simulations. In the case of condensed phase simulations with the IM/MM scheme, because there are a much larger number of particles in the MM part than in the IM part, computing interactions inside the MM part often dominates toward the overall cost. Thus, the interpolation efficiency during the dynamics simulations may not be of great concern unless the size of the data set becomes overwhelmingly large. Thus, reducing the collection time should be the central point of matter in designing the algorithm. Combined with the ever-developing computational power especially with highly parallelized architectures, new attempts will likely be made to improve the efficiency from this direction. As mixed-style approaches are also in fashion in the field of quantum chemistry (e.g., dual basis approach), the data collection may be performed in an analogous way to or even in conjunction with such quantum chemical approaches.

Keywords: molecular dynamics · interpolation · potential energy surface

How to cite this article: Y. M. Rhee, J. W. Park. *Int. J. Quantum Chem.* **2016**, *116*, 573–577. DOI: 10.1002/qua.25064

- [1] J. Ischtwan, M. A. Collins, *J Chem Phys.* **1994**, *100*, 8080.
- [2] M. J. T. Jordan, K. C. Thompson, M. A. Collins, *J Chem Phys.* **1995**, *102*, 5647.
- [3] M. A. Collins, *Theor Chem Acc.* **2002**, *108*, 313.
- [4] T. Ishida, G. C. Schatz, *J Chem Phys.* **1997**, *107*, 3558.
- [5] K. A. Nguyen, I. Rossi, D. G. Truhlar, *J Chem Phys.* **1995**, *103*, 5522.
- [6] Y. M. Rhee, T. G. Lee, S. C. Park, M. S. Kim, *J Chem Phys.* **1997**, *106*, 1003.
- [7] K. Thompson, T. J. Martinez, *J Chem Phys.* **1999**, *110*, 1376.
- [8] M. A. Collins, D. H. Zhang, *J Chem Phys.* **1999**, *111*, 9924.
- [9] J. W. Park, Y. M. Rhee, *J Chem Phys.* **2014**, *140*, 164112.

- [10] K. C. Thompson, M. J. T. Jordan, M. A. Collins, *J Chem Phys.* **1998**, *108*, 8302.
- [11] Y. M. Rhee, *J Chem Phys.* **2000**, *113*, 6021.
- [12] K. C. Thompson, M. J. T. Jordan, M. A. Collins, *J Chem Phys.* **1998**, *108*, 564.
- [13] M. S. Kim, J. H. Moon, *Int J Mass Spectrom* **2003**, *225*, 191.
- [14] E. Uggerud, T. Helgaker, *J Am Chem Soc.* **1992**, *114*, 4265.
- [15] T. G. Lee, M. S. Kim, S. C. Park, *J Chem Phys.* **1996**, *104*, 5472.
- [16] Y. M. Rhee, M. S. Kim, *J Chem Phys.* **1998**, *109*, 5363.
- [17] T. H. Choi, S. T. Park, M. S. Kim, *J Chem Phys.* **2001**, *114*, 6051.
- [18] B. J. Sung, M. S. Kim, *J Chem Phys.* **2000**, *113*, 3098.
- [19] K. Fukui, *J Phys Chem.* **1970**, *74*, 4161.
- [20] S. Maeda, Y. Harabuchi, Y. Ono, T. Taketsugu, K. Morokuma, *Int J Quantum Chem.* **2015**, *115*, 258.
- [21] S. Y. Lin, S. C. Park, M. S. Kim, *J Chem Phys.* **1999**, *111*, 3787.
- [22] J. W. Park, H. W. Kim, C. I. Song, Y. M. Rhee, *J Chem Phys.* **2011**, *135*, 014107.
- [23] J. W. Park, Y. M. Rhee, *J Phys Chem B* **2012**, *116*, 11137.
- [24] H. M. Senn, W. Thiel, *Angew Chem Int Ed* **2009**, *48*, 1198.
- [25] C. I. Song, Y. M. Rhee, *Int J Quantum Chem.* **2011**, *111*, 4091.
- [26] C. W. Kim, J. W. Park, Y. M. Rhee, *J Phys Chem Lett.* **2015**, *6*, 2875.
- [27] R. P. A. Bettens, M. A. Collins, *J Chem Phys.* **1999**, *111*, 816.
- [28] A. Warshel, M. Kato, A. V. Pislakov, *J Chem Theory Comput* **2007**, *3*, 2034.
- [29] C. M. Baker, *WIREs Comput Mol Sci.* **2015**, *5*, 241.
- [30] M. Higashi, D. G. Truhlar, *J Chem Theory Comput.* **2008**, *4*, 790.

Received: 30 September 2015

Revised: 23 November 2015

Accepted: 25 November 2015

Published online 22 December 2015