THE JOURNAL OF PHYSICAL CHEMISTRY A

pubs.acs.org/JPCA Article

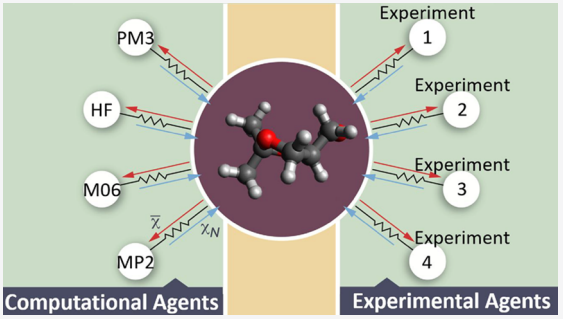
Multiagent Consensus Equilibrium in Molecular Structure Determination

James R.W. Ulcickas, Ziyi Cao, Jiayue Rong, Charles A. Bouman, Lyudmila V. Slipchenko, Gregery T. Buzzard, and Garth J. Simpson*

Cite This: J. Phys. Chem. A 2020, 124, 9105–9112 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Multiagent consensus equilibrium (MACE) is demonstrated for the integration of experimental observables as constraints in molecular structure determination and for the systematic merging of multiple computational architectures. MACE is founded on simultaneously determining the equilibrium point between multiple experimental and/or computational agents; the returned state description (e.g., atomic coordinates for molecular structure) represents the intersection of each manifold and is not equivalent to the average optimum state for each agent. The moment of inertia, determined directly from microwave spectroscopy measurements, serves to illustrate the mechanism through which MACE evaluations merge experimental and quantum chemical modeling. MACE results reported combine gradient descent optimization of each ab initio agent with an agent that predicts the chemical structure based on root-mean-square deviation of the predicted inertia tensor with experimentally measured moments of inertia. Successful model fusion for several small molecules was achieved as well as the larger molecule solketal. Fusing a model of moment of inertia, an underdetermined predictor of structure, with low cost computational methods yielded structure determination performance comparable to standard computational methods such as MP2/cc-pVTZ and greater agreement with experimental observables.



INTRODUCTION

Computational chemistry can provide causal mechanistic understanding of chemical phenomena. In particular, quantum chemical calculations that agree with independent experimental observables can provide key molecular-scale insights challenging to glean from experiments alone. This general architecture of modeling as an explanatory method has had great success in a variety of applications, from chemical structure analysis to ionization energy prediction and modeling of thermodynamic quantities. However, as the complexity of the experimental observable increases, it becomes a more difficult process. Spectroscopic matching, e.g., in vibrational analysis of flexible gas-phase molecules, has found success in identifying conformational structures.[1−3] Unfortunately, such studies are limited in their throughput due to the necessity for modeling many discrete configurations of the system in question and utilizing manual matching methods to identify conformations most closely resembling experimental spectra. Such methods are costly in both computational and human resources. Other experimental methods, such as probes of excited-state electronic structures for extended conjugated systems, semiconductors, and photosynthetic protein complexes, encounter difficulty in the theoretical description of the experimental outcome.[4] Computational scaling becomes a large obstacle in such cases; often, low-level methods are able to describe structures but struggle to predict more complex properties due to inaccurate description of electron correlation or incompleteness of a one-electron basis set. The classical paradigm of "compare and contrast" for ab initio theoretical studies with experiment provides no clear metric for success or failure with respect to the nature of differences (i.e., comparisons are generally qualitative in nature).

The limitations of "compare and contrast" are particularly evident for measurements of low-dimensional observables dependent on structures with large degrees of freedom. This case includes rotational spectroscopy in which no more than three principal moments are produced for a single isotopomer, which depend on typically a large number of atomic degrees of freedom. Agreement or disagreement between the observables and predicted results is not always trivial to interpret in such instances. In addition, computational modeling will necessarily not capture real-world subtleties not integrated into the calculations. Electronic and nuclear coupling and/or rotational

Received: August 9, 2020
Revised: September 22, 2020
Published: September 25, 2020



Downloaded via PURDUE UNIV on April 23, 2021 at 13:16:38 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ACS Publications © 2020 American Chemical Society

https://dx.doi.org/10.1021/acs.jpca.0c07282
J. Phys. Chem. A 2020, 124, 9105–9112

and vibrational coupling can significantly influence experimental results but can be challenging to accurately integrate into computational evaluations.

Model fusion provides one avenue to leverage existing theoretical frameworks as constraints upon models in computational chemistry, bridging the gap. Fusion in the context of modeling refers to an approach that leverages multiple descriptions of the same system in order to generate a single description that incorporates the nuances of each model into the result. For example, integration of the physics of nuclear coupling for interpreting multidimensional NMR data to produce a molecular structure would **not** be considered fusion as interpreted herein. However, integration of multidimensional NMR measurements as an additional constraint in multialgorithm ab initio/semi-empirical calculations of structure **would** be considered fusion. Applications of model fusion methods have had broad ranges of impact, from medical image reconstruction and denoising methods in X-ray computed tomography to sensor fusion for autonomous vehicle piloting and image reconstruction in celestial imaging.[5−7] Largely, these efforts have hinged upon creating an optimal reconstruction of the system measured. In more conventional chemical analyses, data and model fusion have been utilized for creation of electronic tongues, systems biology analysis on the basis of proteomics and metabolomics, and grape varietal classification, among many other applications.[8−10] The actual modality of fusion is as varied as the application, from naïve stacking of feature vectors to Bayesian inference frameworks; many methods are viable to combine disparate descriptions of the same chemistry.[9,10] Fragmentation-based methods provide one mechanism by which multiple levels of theory can be integrated to model a complex chemical system.[11] Slipchenko and co-workers have demonstrated the use of effective fragment potentials in which relatively high-level ab initio quantum chemical calculations are performed on sections (fragments) of a larger molecular assembly. Interactions between fragments are integrated using effective fragment potentials to describe the overall behavior of the larger system. Many of the fragment-based fusion strategies either combine multiple experimental characterizations or bring together several methods in computational chemistry but are often not directly useful for fusing experimental with theoretical models.

Methods fusing experimental measurement with theoretical descriptions have been explored to the greatest extent within the field of chemical structure elucidation. As molecular systems of interest grow larger and more complex, predicting structure ab initio and finding agreement between theoretical results and empirical structures become more challenging. Large strides have been made in protein structure elucidation leveraging both experimental feedback with theoretical descriptions, primarily in the fields of solid-state NMR and cryo-electron microscopy. These efforts have both leveraged computational modeling to improve low-resolution experimental probes and used theoretical descriptions of dynamics to provide multiple poses for structure refinement.[12,13] Recent work using augmented Markov models (AMMs) has met with some success in direct integration of computational and experimental modeling.[14] Olsson et al. demonstrated a method for using AMMs to statistically model atomistic protein dynamics via empirical thermodynamic data.[14] These Markov state models (MSMs) fuse experimental averages with state-based transitions from atomistic MSM dynamical simulations, guiding the outcome of simulations based on known experimental observables. This work has already been extended to other applications in fusing multiple MSMs with reaction−diffusion simulations. However, MSMs are not trivially extended to applications utilizing quantum chemical modeling due to the mathematical distinction between quantum and classical Markov processes. Furthermore, the architecture of MSMs is uniquely suited to modeling thermodynamic equilibria, and modeling other types of phenomena often requires recasting the problem into a Markov-compatible architecture. There remains a need for flexible methods that can fuse a variety of models, including those that span quantum chemical descriptions with classical descriptions.

In this work, multiagent consensus equilibrium is described and demonstrated for direct model fusion of ab initio chemical calculations and experimentally determined empirical models of chemical structure. A proof-of-concept model fusion for a test-bed system of H_2 as a simple harmonic diatomic was performed. MACE was then utilized to fuse the results from multiple inexpensive computational chemistry methods including Møller−Plesset second order perturbation theory (MP2), density functional theory (DFT), Hartree−Fock (HF), and semiempirical PM3 methods for a suite of small organic molecules. Finally, fusion of empirical methods with computational modeling was demonstrated, combining an empirical model of the moment of inertia (MoI) with four quantum chemical models. Results are presented for several small organics, and the potential merits of MACE assessed by consideration of the root mean squared displacement (RMSD) between with the MACE-optimized structures and those produced from either experimental observables or high-level CCSD(T)/cc-pVTZ optimized structures serving as the ground truth. Assessment of the strengths and limitations of MACE are provided, along with the anticipated scope of use.

THEORETICAL FOUNDATION AND ALGORITHM

Before formally defining the MACE algorithm, several definitions will clarify the concept. The column vector $\mathbf{X} = [X_1^T \cdots X_N^T]^T$ denotes the concatenated set of N state vectors describing the state of the system (i.e., the complete set of nuclear positions), where N is the number of computational/experimental agents to be integrated (e.g., quantum chemical methods). In the present case of geometry optimization, each state vector defines the set of atomic coordinates for each atom in a molecule. For an input state vector of X_n, the next iteration of the n-th computational method will result in a new output state vector $F_n(X_n)$. Considering the collective set of all N computational and/or experimental agents, the overall operator $\mathbf{F}$ for the set of input structures $\mathbf{X}$ is given by stacking of the N agent operators in the MACE computation, as shown below.

$$\mathbf{FX} = \begin{pmatrix} F_1(X_1) \\ \vdots \\ F_N(X_N) \end{pmatrix} \tag{1}$$

The broadcast mean operator $\mathbf{G}(\mathbf{X})$ is defined below, which takes the arithmetic mean $\bar{X}$ of all agents' state (in this case, $\bar{X}$ corresponds to the mean molecular structure) and stacks it N times to produce an array of the same dimensionality as $\mathbf{X}$. Introducing $\mathbf{G}(\mathbf{X})$ ensures that the same arithmetic mean $\bar{X}$ can be broadcasted to every agent. Note that the arithmetic

https://dx.doi.org/10.1021/acs.jpca.0c07282
J. Phys. Chem. A 2020, 124, 9105−9112

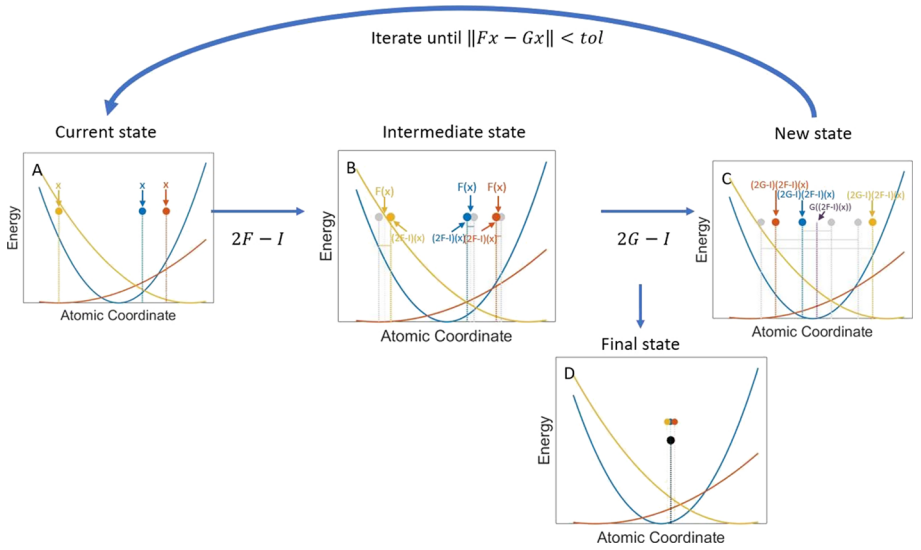


Figure 1. Illustrative example of the MACE algorithm. (A) Current state; each model begins in a different state, with each color representing a different potential produced by a computational agent. (B) Intermediate state; gray balls represent states before 2F-I operation, and colored balls represent states after 2F-I operation. Application of the 2F-I operation for each agent (potential) advances the atomic position along the potential to F(x) and then reflects the current state x about F(x) to (2F-I)(x). (C) New state; gray balls represent states before 2G-I operation, and colored balls represent states after 2G-I operation. The mean of the (2F-I)(x) positions is calculated as G(2F-I)(x), and the updated positions are evaluated by reflection about that location for each agent as (2G-I)(2F-I)(x). (D) Final state; when the difference of F(x) (shown using colored ball) and G(x) shown using black balls is smaller than tolerance, the consensus equilibrium conditions are met and the calculation is completed.

mean $\bar{X}$ can be weighted using the set of weights $\bar{w} = [w_1 \cdots w_N]^T$ to place a greater emphasis on the contributions of some agents over others.

$$\mathbf{GX} = \begin{pmatrix} \bar{X} \\ \vdots \\ \bar{X} \end{pmatrix}; \bar{X} = \sum_{n=1}^{N} w_n X_n; \sum_{n=1}^{N} w_n = 1 \tag{2}$$

Before describing the iterative process taken in MACE, it is useful to understand the endpoint goal defining the consensus equilibrium condition. Formally, consensus equilibrium arises when the operation **FX** returns a vector of states equal to the result of the broadcast mean, as shown in eq 3 below. Note that here the dimensions of **FX** and **GX** are consistent since we define both **FX** and **GX** to have the same dimensions as **X**.

$$\mathbf{FX} = \mathbf{GX} \tag{3}$$

In the present context of geometry optimization, each computational agent directs a change in the atomic position. Consensus equilibrium is reached when the application of the change for each agent for the preceding structure of each agent X_i all produces $\bar{X}$ as an output. In effect, the MACE approach identifies the structures for which one additional iteration of each agent produces a common output. The degree of change in the atomic position induced by a given agent can be interpreted qualitatively as a "spring constant" in the equilibrium evaluation, with larger changes corresponding to tighter springs (Figure 1B, the larger gradient correspond to larger change in the position in MACE). The consensus equilibrium corresponds to the location in atomic position-space in which all the tensions are balanced.

This balance of "spring constants" can similarly be interpreted as the minimum position on the sum of combined potential energy surfaces produced by each computational method. This result will generally be different from that recovered by determining the minimum of each potential energy surface independently and then averaging the results. Summing the potentials will retain differences in curvatures that pull or push the MACE structure away from this simple average result by different degrees depending on the steepness of the corresponding algorithm potential.

When each agent F_n corresponds to a proximal map, the solution to the consensus eq 3 is equivalent to the solution of the optimization problem formulated by the sum of all agent functions.[15] The proximal map of the function $f_n(v)$ is defined below, where σ is a general weighting parameter.

$$F_n(X) = \arg\min_v \left\{ \frac{\|v - X\|^2}{2\sigma^2} + f_n(v) \right\} \tag{4}$$

Proximal maps can be estimated from the numerically calculated gradient, which provide convenient transformation for optimizations in which derivatives are difficult to calculate analytically.[16] Although originally formulated for use with proximal maps, subsequent work extended MACE to more general maps that start from some input position and take a step toward the minimum of an arbitrary convex function.[17] In the present study, F_n is the gradient descent update operator, given below. However, operators that produce a new state vector of the same dimensionality are compatible with the MACE architecture (e.g., denoising algorithms in 4D X-ray CT).[18]

$$F_n(X_n) = X_n - \gamma \nabla f_n(X_n) \tag{5}$$

In the above expression, $\nabla f_n(X_n)$ is the gradient of an objective function to be optimized, and γ is a fixed parameter corresponding to the step size. Furthermore, the equilibrium state can be written as a fixed point on the map **T** as shown in eq 6 below. At the equilibrium state, the equation **FX** = **GX** holds, and simple algebraic manipulation allows the reformulation to the form **TX** = **X**. A more complete description of this procedure can be found in prior work.[15]

https://dx.doi.org/10.1021/acs.jpca.0c07282
J. Phys. Chem. A 2020, 124, 9105−9112

The expression $(2\mathbf{G} - \mathbf{I})$ is involutory and thus is its own inverse (see the Supplementary Information for proof), enabling the following fixed-point formulation.

$$\begin{aligned}
&\mathbf{FX} = \mathbf{GX}\\
&2\mathbf{FX} = 2\mathbf{GX}\\
&2\mathbf{FX} - \mathbf{X} = 2\mathbf{GX} - \mathbf{X}\\
&(2\mathbf{F} - \mathbf{I})\mathbf{X} = (2\mathbf{G} - \mathbf{I})\mathbf{X}\\
&(2\mathbf{G} - \mathbf{I})(2\mathbf{F} - \mathbf{I})\mathbf{X} = \mathbf{X};\\
&\mathbf{T} = (2\mathbf{G} - \mathbf{I})(2\mathbf{F} - \mathbf{I})\\
&\mathbf{TX} = \mathbf{X}
\end{aligned} \tag{6}$$

Practically, it is convenient to split the $\mathbf{T}$ operator into several lines of code. The general iterative strategy illustrated in Figure 1 is mathematically implemented using the pseudo-code represented in eq 7.

$$\begin{aligned}
&\text{while}(\mathbf{FX} - \mathbf{GX}) < \text{tolerance, do:}\\
&q \leftarrow \mathbf{FX}\\
&v \leftarrow 2q - \mathbf{X}\\
&\mathbf{X} \leftarrow 2\mathbf{G}v - v
\end{aligned} \tag{7}$$

Figure 1 illustrates the progression of the MACE algorithm, beginning from a single state for all agents. Each state is updated as a function of both the individual agent (line 1 of the loop in eq 7) and the consensus of all agents (line 3 of the loop in eq 7). At equilibrium, one further MACE step places all agents at the consensus mean state position. This one-step-removed procedure is mathematically analogous to balancing the forces of each agent in directing the final position. In practice, numerical differentiation (i.e., Mann iteration) can be used for determining the gradient descent direction in such an evaluation.[19]

MACE AND UNDERDETERMINED MODELS

The use of underdetermined agents in which the dimensionality of the agent is less than the dimensionality of the object is a key innovation developed herein to support merging of quantum chemical and experimental models. Previous implementations of MACE integrated multiple fully determined agents (e.g., image denoising algorithms) in which each agent operated in an identical parameter-space.[18] Underdetermined agents inform results in a lower-dimensional space rather than the full desired parameter-space (in this case, informing on the moments of inertia, MoI, rather than molecular structure-space). In the broader context of model fusion, we demonstrate the use of MACE with underdetermined models in combination with at least one determined model (e.g., quantum chemical calculation).

To utilize MoI as a constraint, the cost function $U = \text{trace}\left[\left(I - \sum_{k=1}^{K} m_k r_k^2\right)^2\right]$ was implemented within the inertia agent, expressed with the molecule in a coordinate system defined by the principal moments of I with the center of mass at the origin. In this function, m_k is the mass of the k^{th} constituent atom, R_k describes the atomic coordinates for the k^{th} constituent atom, and I is the empirically derived inertia tensor. The inertia agent takes a single gradient descent step upon the cost function of the forward model predicting the MoI from a given set of atomic coordinates. Analogous to the approach described in Figure 1, gradients and gradient descent steps for ab initio quantum chemical methods were also computed, and MACE steps were taken for each agent. The MACE step updated the state describing the molecular structure for each agent individually, until the consensus equilibrium condition was achieved. For simple molecular systems, such as H_2 or H_2O, the molecular degrees of freedom provide a sufficiently determined mapping to uniquely recover structure from MoI measurements, and the use of computational agents is not essential for accurate reconstruction of the molecule. For low-symmetry nonlinear molecules with more than three atoms (not considering isotopomers), this cost function alone is typically insufficient to determine the molecular structure, as the fitting from MoI to atomic coordinates is generally underdetermined. In these cases, the computational agents play the larger role of constraining the total set of states to physically reasonable structures.

COMPUTATIONAL METHODS

All ab initio calculations were conducted with the GAMESS electronic structure package, version 2016 R1.[20] All initial molecule structure was drawn by Avogadro except for solketal. The solketal structure was first drawn by Qmol and then optimized by UFF. Upon drawing the structure, geometry optimizations were performed via three distinct methods: (i) single agent, gradient descent optimizations done solely in GAMESS; (ii) 4 agent, utilizing the four computational agents Hartree−Fock (HF)/cc-pVDZ, MP2/cc-pVTZ, M06/cc-pVTZ, and PM3; (iii) 5 agent, utilizing the same four computational agents and an additional experimental agent. The experimental agent was the gradient-descent optimizer of an empirical model for the MoI. Additional geometry optimizations were performed with coupled clusters with singles, doubles, and perturbative triples (CCSD(T))/cc-pVTZ for all systems for comparative purposes. To calculate the root mean square deviation (RMSD) between molecular structures, the Kabsch algorithm was utilized as implemented through the PyPI package.[14,21] To implement the MACE framework, a suite of functions in Python were written, enabling parallel execution of individual gradient computations and calculation of the MACE step during the equilibrium computation.

RESULTS AND DISCUSSION

Evaluation of MACE with Model Small Molecules. Direct optimization was implemented by evaluating the gradient of the electronic energy with respect to atomic coordinates for each agent independently. The set of gradients was then utilized to calculate the next set of state vectors X_k (i.e., a single MACE step).

Prior to implementation with experimental inputs, initial assessments were performed to evaluate the reliability of the novel MACE algorithm integrating one or more underdetermined agents. In these evaluations, known "ground truth" structure-space references were generated by high-level CCSD(T)/cc-pVTZ calculations, from which the moments of inertia were calculated as the corresponding underdetermined agents. The MACE-optimized geometries for the five test molecules H_2O, HCN, SO_2, pyridine, and urea were then compared with the ground truth structures from the CCSD(T)/cc-pVTZ calculations. The inertia agent minimized the squared deviation between the predicted inertia tensor and the inertia tensor obtained from the CCSD(T) result. Figure 2

https://dx.doi.org/10.1021/acs.jpca.0c07282
J. Phys. Chem. A 2020, 124, 9105−9112

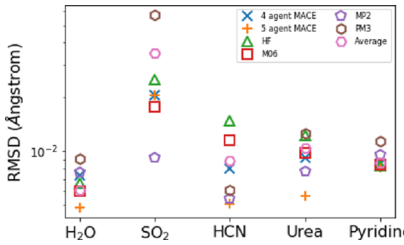


Figure 2. RMSD (Å) of individual agents with the CCSD(T)/cc-pVTZ structure plotted on a log scale. 4 agent MACE utilizes only computational agents, while 5 agent MACE also includes the inertia agent.

contrasts MACE results for 4 agent MACE (HF/cc-pVDZ, MP2/cc-pVTZ, M06/cc-pVTZ, and PM3) and 5 agent MACE (HF/cc-pVDZ, MP2/cc-pVTZ, M06/cc-pVTZ, PM3, and the inertia agent) with direct optimization of a single agent (e.g., gradient descent upon the Hartree−Fock potential energy surface).

From Figure 2, several trends can be observed. First, as anticipated, the MACE result is not equivalent to averaging the optimum structure for each agent individually. In the above plot, the weighting of each agent is equal − as such, inclusion of the inertia agent provides a relatively small change to the recovered structure. The spread of RMSD values obtained from direct optimization of the ab initio and semi-empirical structures is variable depending on chemical identity. In Figure 2, H_2O, HCN, urea, and pyridine indicate a lower RMSD for the MACE-generated structure (relative to the ground truth CCSD(T)/cc-pVTZ calculations) than any single agent alone. With urea containing oxygen and nitrogen as a noteworthy example, the large spread is at least in part due to the PM3 agent, which exhibits an RMSD on average at least two times greater than that of the MP2 and MO6 agents. Interestingly, the 5 agent MACE constrained by the moments of inertia for SO_2 does not generate a structure closer to the CCSD(T) structure than the MP2 or MO6 agent individually and yet generates a smaller RMSD than the average of the agents. We speculate that the outlying behavior of SO_2 relative to the other small molecules considered arose from the presence of π-bonding and greater electron correlation effects not captured by the low-cost computational methods.

Implementation of MACE with Experimental Constraints. To illustrate the potential utility of MACE in enhancing the structural sensitivity of experimental measurements, the molecule solketal was modeled using a variety of MACE configurations, as well as ab initio methods. Solketal contains nine heavy atoms and consequently is nearing the size of molecules in which high-level computations, e.g., CCSD(T)/cc-pVTZ, become costly to run on benchtop machines. The rotational constants of solketal have been previously published.[22] Microwave spectroscopy provides precise spectral measurements (typical uncertainties reported for rotational constants in the order of 10^3 MHz are in the order of 10^{-3} MHz) but is insufficient as a sole probe to predict molecular structures for molecules with more than three atoms in the absence of isotopomers or nontrivial molecular symmetry. The present work is the first to our knowledge to implement MACE with one of the agents underdetermined (i.e., produces an output with lower dimensionality than the desired outcome).

Given the novelty of integrating MACE with underdetermined agents, the applicability of MACE calculations with simulated moments of inertia in larger molecules first was evaluated relative to an assumed known ground truth [in this case, the moments of inertia of solketal derived from M06-2X/6311++G(d,p)]. MACE performance for structural elucidation was evaluated via RMSD of atomic coordinates relative to the computational structure reported by Lobsiger et al.[22] MACE jobs conducted for solketal included the following three combinations of agents: (i) PM3 + M06 + MP2, (ii) MP2 + M06 + PM3 + Inertia, and (iii) PM3 + Inertia. Figure 3

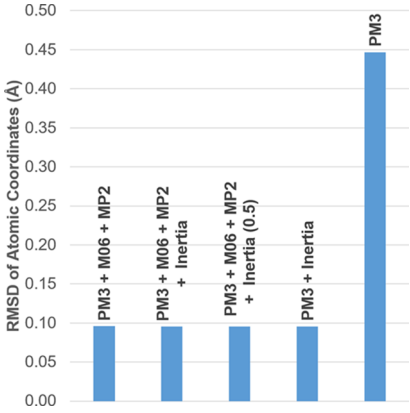


Figure 3. RMSD of molecular structures relative to the literature reported M06-2X/6311++G(d,p) for several MACE configurations and the PM3 method.

contrasts RMSDs between these three MACE configurations and the RMSD of PM3 optimum geometry relative to the literature structure from M06-2X/6311++G(d,p), which served as the simulated ground truth structure for determining the principal moments of inertia. Each MACE structure including PM3 as an agent exhibited RMSDs near 0.096 Å, while the equilibrium geometry of PM3 alone exhibits a value of 0.447 Å, a nearly fivefold increase. If the literature reported structure is taken as the ground truth, no significant improvement in structural accuracy is obtained from using more computationally costly methods than PM3. This result is remarkable in that it suggests that with the inclusion of a single low-cost experimental constraint, structure prediction from PM3 can be significantly improved, reducing RMSD with M06-2X/6311++G(d,p) by a factor of nearly five. Although the simulated experimental agent gives uncertainty comparable to errors of some computational methods, incorporation of the inertia constraint substantially improved the PM3 result. It is also worth reiterating the power of MACE to enable recovery of the structure in situations where normally fitting to the structure or properties would be an underdetermined problem.

The preceding simulations with the computed moments of inertia provide a foundation for extension to integration of genuine experimental constraints. In practice, the ground truth structure from which the MoI are derived will generally not be known a priori. In such cases, the RMSD can be evaluated between the experimental MoI and the MoI of the MACE result, as shown in Figure 4 for several different MACE configurations. Within the plot, both the experimentally measured rotational constants, which serve as the reference, and the M06-2X/6−311++G(d,p) data were adapted from Lobsiger et al.,[22] while the remaining bars correspond to the

https://dx.doi.org/10.1021/acs.jpca.0c07282
J. Phys. Chem. A 2020, 124, 9105−9112

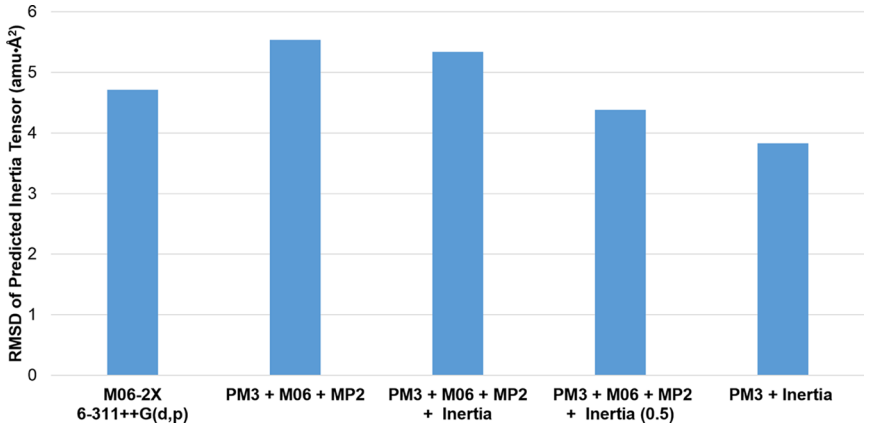


Figure 4. RMSD of predicted inertia tensors for structures previously reported (M06-2X) and those obtained from MACE with varying agents with respect to the experimental MoI. From left to right, the MACE results correspond to runs with computational agents only, computational and inertia agents, computational and inertia agents with the inertia agent weighted equal to 0.5, and PM3 plus the inertia agent.

same MACE configurations described in Figure 3 but evaluated using the experimentally measured moments of inertia.[22]

The performance of MACE integrating experimental moments of inertia can be evaluated relative to stand-alone computational methods, such as M06-2X. Integrating the experimental agent into PM3 + M06 + MP2 MACE reduced the RMSD between the predicted inertia tensor and measured value by 0.20 amuÅ^2, approximately 4%. The MACE architecture allows tuning of the system by utilizing differing weights for each agent. Changing the relative weight of each agent during the broadcast mean step of the MACE algorithm will favor certain agents over others in their contribution to the MACE step. Reweighting can be utilized to place greater significance on certain computational agents over others (e.g., to increase the contribution of the experimental agent). Further increasing the weighting of the inertia agent to be equal to the sum of computational agents resulted in further improvement in recovering the inertia tensor by 0.96 amuÅ^2, or about 18%. Similarly, reducing the total number of agents to 2, including only the experimental agent and the PM3 agent, yielded additional improvement of 0.55 amuÅ^2, or another 13%. Though not plotted, the RMSD of PM3 alone is 28.7 amuÅ^2. Thus, inclusion of the experimental agent reduces the PM3 RMSD by a factor of 7.5.

The improved performance of the Inertia + PM3 result is intriguing and can be ascribed to dual causes. First, reducing the number of agents to two gives an overall greater weight to the experimental agent, reducing discrepancy between the computationally predicted structure and the structure derived from experimental measurements; however, this is insufficient to describe the whole effect, as the MP2 + M06 + PM3 + Inertia (0.5) demonstrates reduced performance relative to the PM3 + Inertia result. Second, the greater contribution of the experimental agent to the equilibrium structure in the two-agent case suggests that the PM3 potential energy surface is softer than that of the other computational agents (i.e., weaker "spring constant"). This effect is analogous to the case illustrated in Figure 1 for simple harmonic potentials. The shallowness of the red potential relative to the blue and yellow potentials results in an equilibrium structure that is closer to the minima of the blue and yellow potentials. In this case, the PM3 is most similar to the red potential and contributes less to the equilibrium structure, while the inertia agent provides a steeper potential akin to the yellow and blue potentials.

Further insight into the intriguing result of dramatic improvement of PM3 by the experimental agent can be achieved by direct comparison of molecular structures. Figure 5 displays an image containing the superimposed PM3 structure, Inertia + PM3 MACE structure, and the reported structure from M06-2X.[22] The key structural difference between the PM3 and M06-2X result is most readily observable within the ring moiety. The plane defined by the three carbon atoms in the ring is considered; the literature M06-2X structure exhibits the oxygens above this plane, closer to the alcohol moiety. This upward "pucker" contrasts with the PM3 result, where the oxygens are puckered downward relative to the plane of the ring carbons and the alcohol moiety. Note that the M06-2X structure reported previously is in excellent agreement with the structure obtained by solving Kraitchman's equations.[22] While the PM3 optimization alone results in a distinct structural difference, inclusion of the inertia agent as a constraint on structure eliminates this failure, resulting in an upward pucker and oxygens oriented above the carbon plane. This outcome is illustrative of the potential power in fusing experimental measurements with theoretical values. In this case, not only is an underdetermined chemical problem in the form of full structure determination from the MoI achieved but also the experimental constraint served to improve upon deficiencies in the theoretical model.

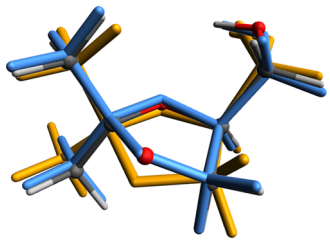


Figure 5. Overlay of the M06-2X structure (atom-type colored), PM3 structure (yellow), and 2 agent MACE structure (blue) of solketal. Inclusion of the inertia agent constrains recovered geometry to more physical results.

The example above comparing stand-alone PM3 calculations and PM3 + Inertial MACE illustrates an advantage likely to

https://dx.doi.org/10.1021/acs.jpca.0c07282
J. Phys. Chem. A 2020, 124, 9105−9112

translate to higher order calculations as well. With the exception of very simple molecular systems, all practical quantum chemical calculations implicitly introduce simplifying approximations that may not be reflected in the experimental outcomes. One well-known example is breakdown of the Born-Oppenheimer approximation for the separability of electronic and nuclear degrees of freedom. In rotational spectroscopy, rotational/vibrational coupling can be similarly challenging to accurately integrate into relatively low-level calculations. Integrating experimental constraints through MACE provides a means for visualizing and quantifying the impact of these perturbations to the recovered molecular structures and rotational constants.

It is worth briefly discussing the scaling attributes of the MACE framework. The addition of experimental constraints via MACE is a simple parallelizable process. Evaluation of the gradient of the experimental agent in the present study was computationally much cheaper than even the semiempirical PM3 methods used. However, this is a property of the model and will vary depending upon the target of study. During MACE, each computational agent is called in parallel; provided sufficient parallelization resources, computational time for each equilibrium step is limited by the agent with the greatest computational cost. As such, parallel processing with additional agents with lower computational cost will add no significant additional computational time.

CONCLUSIONS

MACE was demonstrated for formal fusion of ab initio methods with experimental measurement. Weighting showed a reduced RMSD for the equilibrium structure relative to individual agents. When compared to direct optimization of structures with the theoretically superior CCSD(T)/cc-pVTZ method, MACE achieved comparable accuracy as measured by the RMSD relative to the reported experimental structure for several small molecules while utilizing computational agents with lower scaling. Further tuning of weighting parameters and implementation of more sophisticated iterative reweighting approaches to reduce contribution of poorly performing agents are anticipated to improve the performance of MACE. In addition to quantifying MACE performance with respect to "ground truth" structures obtained from CCSD(T)/cc-pVTZ calculations, experimental reports of rotational spectroscopy of solketal were used to generate structures informed by experiment. The MACE result fusing previously reported inertia tensors with ab initio modeling generated a structure whose inertia tensor more closely resembled experiment than pure ab initio methods. Furthermore, the structure obtained from a single low-cost computational agent (PM3) was demonstrably improved via fusion with the experimental agent constraining structure to empirically determined moments of inertia. With the demonstration of successful fusion for the simple empirical parameters of the moments of inertia, future work can leverage more complex empirical feedback for structural determination, such as powder x-ray diffraction patterns or infrared spectra.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.0c07282.

Preliminary experimental/computational MACE benchmarking on structure determination of H2 and a mathematic derivation that the matrix $(2\mathbf{G} - \mathbf{I})$ is involutory (PDF)

AUTHOR INFORMATION

Corresponding Author

Garth J. Simpson − *Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States;* orcid.org/0000-0002-3932-848X; Email: gsimpson@purdue.edu

Authors

James R.W. Ulcickas − *Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States*

Ziyi Cao − *Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States*

Jiayue Rong − *Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States*

Charles A. Bouman − *Department of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907, United States*

Lyudmila V. Slipchenko − *Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States;* orcid.org/0000-0002-0445-2990

Gregery T. Buzzard − *Department of Mathematics, Purdue University, West Lafayette, Indiana 47907, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.0c07282

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from the NSF CIF program, award number CCF-1763896. This research was supported in part through computational resources provided by Information Technology at Purdue, West Lafayette, Indiana.

ABBREVIATIONS

MoI,Moment of Inertia

REFERENCES

(1) Baquero, E. E.; James, W. H.; Choi, S. H.; Gellman, S. H.; Zwier, T. S. Single-Conformation Ultraviolet and Infrared Spectroscopy of Model Synthetic Foldamers: β-Peptides Ac-β3-hPhe-NHMe and Ac-β3-hTyr-NHMe. *J. Am. Chem. Soc.* **2008**, *130*, 4784−4794.

(2) McGlone, S. J.; Elmes, P. S.; Brown, R. D.; Godfrey, P. D. Molecular structure of a conformer of glycine by microwave spectroscopy. *J. Mol. Struct.* **1999**, *485-486*, 225−238.

(3) Pispisa, B.; Palleschi, A.; Mazzuca, C.; Stella, L.; Valeri, A.; Venanzi, M.; Formaggio, F.; Toniolo, C.; Broxterman, Q. B. The versatility of combining FRET measurements and molecular mechanics results for determining the structural features of ordered peptides in solution. *J. Fluoresc.* **2002**, *12*, 213−217.

(4) Curutchet, C.; Mennucci, B. Quantum Chemical Studies of Light Harvesting. *Chem. Rev.* **2017**, *117*, 294−343.

(5) Huang, C.; Ackerman, J. L.; Petibon, Y.; Brady, T. J.; El Fakhri, G.; Ouyang, J. MR-based motion correction for PET imaging using wired active MR microcoils in simultaneous PET-MR: Phantom study. *Med. Phys.* **2014**, *41*, No. 041910.
(6) Chang, B. R.; Tsai, H. F.; Young, C.-P. Intelligent data fusion system for predicting vehicle collision warning using vision/GPS sensing. *Expert Syst. Appl.* **2010**, *37*, 2439−2450.
(7) Bouman, K. L.; Johnson, M. D.; Zoran, D.; Fish, V. L.; Doeleman, S. S.; Freeman, W. T.. In *Computational imaging for vlbi image reconstruction*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016; 913−922.
(8) Gutiérrez, J. M.; Haddi, Z.; Amari, A.; Bouchikhi, B.; Mimendia, A.; Cetó, X.; del Valle, M. Hybrid electronic tongue based on multisensor data fusion for discrimination of beers. *Sens. Actuators B Chem.* **2013**, *177*, 989−996.
(9) Smilde, A. K.; van der Werf, M. J.; Bijlsma, S.; van der Werff-van der Vat, B. J. C.; Jellema, R. H. Fusion of mass spectrometry-based metabolomics data. *Anal. Chem.* **2005**, *77*, 6729−6736.
(10) Roussel, S.; Bellon-Maurel, V.; Roger, J.-M.; Grenier, P. Fusion of aroma, FT-IR and UV sensor data based on the Bayesian inference. Application to the discrimination of white grape varieties. *Chemometr. Intell. Lab. Syst.* **2003**, *65*, 209−219.
(11) 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.
(12) Miyashita, O.; Tama, F., Hybrid Methods for Macromolecular Modeling by Molecular Mechanics Simulations with Experimental Data. In *Integrative Structural Biology with Hybrid Methods*, Nakamura, H.; Kleywegt, G.; Burley, S. K.; Markley, J. L., Eds. Springer Singapore: Singapore, 2018; 199−217.
(13) Jonić, S. Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images. *Curr. Opin. Struct. Biol.* **2017**, *43*, 114−121.
(14) Olsson, S.; Wu, H.; Paul, F.; Clementi, C.; Noé, F. Combining experimental and simulation data of molecular processes via augmented Markov models. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 8265−8270.
(15) Buzzard, G. T.; Chan, S. H.; Sreehari, S.; Bouman, C. A. Plug-and-Play Unplugged: Optimization-Free Reconstruction Using Consensus Equilibrium. *SIAM J. Imaging Sci.* **2018**, *11*, 2001−2020.
(16) Rockafellar, R. T.. In Moreau's Proximal Mappings and Convexity in Hamilton-Jacobi Theory, *Nonsmooth Mechanics and Analysis, Boston, MA, 2006//*; Alart, P.; Maisonneuve, O.; Rockafellar, R. T., Eds. Springer US: Boston, MA, 2006; 3−12.
(17) Sridhar, V.; Wang, X.; Buzzard, G. T.; Bouman, C. A., Distributed Iterative CT Reconstruction using Multi-Agent Consensus Equilibrium. arXiv preprint arXiv:1911.09278 2019.
(18) Majee, S.; Balke, T.; Kemp, C.; Buzzard, G.; Bouman, C., *4D X-Ray CT Reconstruction using Multi-Slice Fusion*. 2019; 1−8.
(19) Giselsson, P.; Boyd, S. Metric selection in fast dual forward−backward splitting. *Automatica* **2015**, *62*, 1−10.
(20) Gordon, M. S.; Schmidt, M. W., Advances in electronic structure theory. In *Theory and applications of computational chemistry*, Elsevier: 2005; 1167−1189.
(21) Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sect. A* **1976**, *32*, 922−923.
(22) Lobsiger, S.; Perez, C.; Evangelisti, L.; Lehmann, K. K.; Pate, B. H. Molecular Structure and Chirality Detection by Fourier Transform Microwave Spectroscopy. *J. Phys. Chem. Lett.* **2015**, *6*, 196−200.

https://dx.doi.org/10.1021/acs.jpca.0c07282
J. Phys. Chem. A 2020, 124, 9105−9112