Discovering Unique, Low-Energy Pure Water Isomers: Memetic Exploration, Optimization, and Landscape Analysis

Harold Soh, Yew-Soon Ong, Quoc Chinh Nguyen, Quang Huy Nguyen,
Mohamed Salahuddin Habibullah, Terence Hung, and Jer-Lai Kuo

***Abstract*—The discovery of low-energy stable and meta-stable molecular structures remains an important and unsolved problem in search and optimization. In this paper, we contribute two stochastic algorithms, the archiving molecular memetic algorithm (AMMA) and the archiving basin hopping algorithm (ABHA) for sampling low-energy isomers on the landscapes of pure water clusters $(H_2O)_n$. We applied our methods to two sophisticated empirical water cluster models, TTM2.1-F and OSS2, and generated archives of low-energy water isomers $(H_2O)_n$ $n = 3-15$. Our algorithms not only reproduced previously-found best minima, but also discovered new global minima candidates for sizes 9–15 on OSS2. Further numerical results show that AMMA and ABHA outperformed a baseline stochastic multistart local search algorithm in terms of convergence and isomer archival. Noting a performance differential between TTM2.1-F and OSS2, we analyzed both model landscapes to reveal that the global and local correlation properties of the empirical models differ significantly. In particular, the OSS2 landscape was less correlated and hence, more difficult to explore and optimize. Guided by our landscape analyses, we proposed and demonstrated the effectiveness of a hybrid local search algorithm, which significantly improved the sampling performance of AMMA on the larger OSS2 landscapes. Although applied to pure water clusters in this paper, AMMA and ABHA can be easily modified for subsequent studies in computational chemistry and biology. Moreover, the landscape analyses conducted in this paper can be replicated for other molecular systems to uncover landscape properties and provide insights to both physical chemists and evolutionary algorithmists.**

***Index Terms*—Basin hopping, isomer sampling, landscape analysis, memetic algorithm, molecular optimization.**

I. Introduction

Water CLUSTERS are important for understanding the enigmatic properties of water. In physical chemistry, water clusters are extensively studied to characterize the fundamental molecular interactions and collective effects of the condensed phase (liquid and ice) [1]–[3]. In biology, water clusters are used to elucidate water's role in biochemical processes, including protein folding and ligand docking, and to study hydrophobic and hydrophilic interactions.

At the heart of computational studies involving water clusters and their interactions are the water models used to calculate properties such as potential energy and electrostatic forces. Among the most accurate water models currently available are first principle quantum mechanical computations and semi-empirical methods, for example second-order Møller–Plesset (MP2) and density functional theory [4]. However, these methods are computationally expensive, limiting their use to simulations involving only a small numbers of atoms. To overcome this limitation, specialized cost-effective empirical models have been developed for use in large-scale simulation and optimization studies. Advanced empirical water models, which are fitted to experimental data or *ab initio* results, can reproduce water's fundamental properties with impressive accuracy. However, despite rapid progress, no empirical model to date is able to quantitatively account for all of water's characteristics nor reproduce all the ground-state structures of *ab initio* calculations.

Validation and comparison of empirical water models are generally performed by comparing the structural characteristics of the global minima. Consequently, there has been considerable work in global optimization algorithms. Among the more effective methods developed are the simulated annealing (SA) method used by Lee *et al.* [5] to optimize water clusters up to $n = 20$ using the Cieplak, Kollman, and Lybrand model, the basin hopping algorithm used by Wales *et al.* to optimize

Manuscript received September 8, 2008; revised August 4, 2009. First version published January 26, 2010; current version published May 28, 2010. This work was supported in part under the Agency for Science, Technology, and Research (A*STAR) Science and Engineering Research Council Grant 052 015 0024 administered through the National Grid Office, Nanyang Technological University, and the Ministry of Education of Singapore, under University Research Council Grants RG34/05 and RG57/05.

H. Soh is with the Imperial College London, South Kensington Campus, London SW7 2AZ, U.K. He was formerly with the Institute of High Performance Computing, A*STAR, Singapore 138632, where this work was done (e-mail: haroldsoh@imperial.ac.uk).

Y.-S. Ong and Q. H. Nguyen are with the Center for Computational Intelligence, School of Computer Engineering, Nanyang Technological University, Singapore 639798, Singapore (e-mail: asysong@ntu.edu.sg; Huynq@pmail.ntu.edu.sg).

Q. C. Nguyen is with the School of Mathematical and Physical Sciences, Nanyang Technological University, Singapore 639798, Singapore (e-mail: chinh@pmail.ntu.edu.sg).

M. S. Habibullah is with the Institute of High Performance Computing, A*STAR, Singapore 138632, Singapore (e-mail: mohdsh@ihpc.a-star.edu.sg).

T. Hung is with the Institute of High Performance Computing, A*STAR, Singapore 138632, Singapore (e-mail: terence@ihpc.a-star.edu.sg).

J.-L. Kuo is with the Institute of Atomic and Molecular Science, Academia Sinica, Taipei 106, Taiwan (e-mail: jlkuo@pub.iams.sinica.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TEVC.2009.2033584

the rigid TIP4P [6] and TIP5P [2] potential models for $n \leq 21$ and the genetic algorithm used by Bandow and Harke to optimize water clusters on TIP4P and TTM2-F potentials for $n \leq 34$ [7].

Unfortunately, recent work has concentrated solely on global optimization at the expense of locating other low-lying isomers (local minima). Isomers not only provide key insights into resultant properties but also a statistical comparison between isomers represents a more robust methodology for comparing models and determining fit to the quantum mechanical calculations.

In this paper, we focus on discovering *unique, low-energy isomers, including the global minimum*. The contribution of this paper is multifold. First, we propose two stochastic algorithms based on the highly successful methodologies of evolutionary computation and SA: the archiving molecular memetic algorithm (AMMA) and archiving basin hopping algorithm (ABHA). Both algorithms were developed from the ground up to address three key challenges associated with low-energy isomer sampling on the potential energy landscapes.

1) The potential energy landscapes of molecular clusters, including water clusters, are high-dimensional [8], [9]. In this paper, we evaluated water clusters consisting up to 15 molecules which are represented by 135 epistatic real-valued variables.
2) Prior work on Lennard–Jones clusters and water clusters showed that the number of isomers grows exponentially with cluster size [9]–[12].
3) Distinguishing unique minima is nontrivial because of rotational and translational symmetries. Failure to detect duplicate structures will result in large isomer databases with redundant copies.

Although the memetic algorithm (MA) and basin hopping are different approaches, our algorithms share three core features: the ultrafast shape recognition (USR) algorithm [13] for fast structural comparisons, specially designed operators for traversing energy landscapes and an efficient molecular structure archive for isomer storage.

We demonstrated the efficacy of AMMA and ABHA on two sophisticated empirical water cluster models, TTM2.1-F and OSS2, and performed extensive numerical tests (totaling more than 31 days of CPU time) for pure water cluster $(H_2O)_n$ n = 3–15. Experimental results show that AMMA and ABHA performed similarly but were superior to a baseline stochastic multistart local search (SMSL) algorithm in terms of convergence and number of low-energy isomers archived.

Our second contribution arose from our desire to gain insights into the properties of water model landscapes and to elucidate how these properties influenced the performance of proposed algorithms. This issue is fundamentally important to the field of evolutionary computation and water model research but has remained largely unresolved due to the high-complexity involved with such analyses.

In this paper, experimental results indicated that the performance of the algorithms differed significantly on TTM2.1-F and OSS2, despite both models being developed for the similar purpose of calculating the binding energy of water clusters. To illuminate the reasons behind this observation, we performed a large-scale study of the TTM2.1-F and OSS2 landscapes using the tens of thousands of isomers gathered during our experiments. With our developed landscape formulation and correlation tests, we demonstrated that the global and local "roughness" of OSS2 is significantly higher than TTM2.1-F, resulting in poorer convergence and smaller isomer archives. Not only did our analysis reveal how algorithm performance was related to landscape properties but also highlighted the landscape differences between TTM2.1-F and OSS2 despite similar global minima for small clusters.

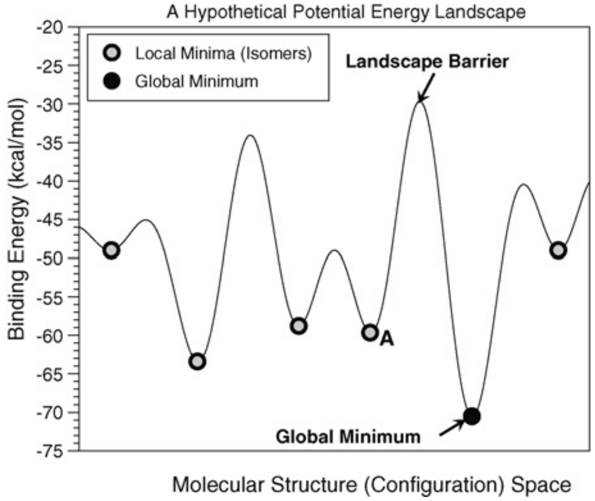


Fig. 1. Hypothetical potential energy landscape. Even though isomer A is close to the global minimum, it has to overcome a landscape barrier to reach it.

Guided by the results of our landscape analyses, our third contribution is a hybrid local search (HLS), which introduces a stochastic element to the deterministic Broyden–Fletcher–Goldfarb–Shannon (BFGS) local search method. When applied to the larger water clusters n = 13–15 on OSS2, we observed a significant improvement in AMMA's sampling capability.

The remaining portions of this paper are organized as follows. Section II describes the problem of isomer sampling from a landscape perspective, giving specifics on the configuration space, fitness/energy functions, and the structural distance measure. Section III details AMMA, ABHA, and SMSL, detailing the operators, replacement, and archival methods used. We present our empirical results in Section IV, comparing the convergence and isomer sampling abilities of AMMA, ABHA, and SMSL. Section V presents our analysis into the global and local properties of the TTM2.1-F and OSS2 landscapes. This is followed by Section VI, which describes the HLS method. Finally, Section VII summarizes our main findings and explores avenues for future work.

II. Problem Definition: A Landscape Perspective

The fitness or energy landscape has proven to be a useful conceptual framework in various fields, from biological evolution and protein folding to combinatorial and molecular optimization [8], [9], and [14]. Intuitively, an optimization process can be visualized as a search across this landscape to find a minimum or maximum point (Fig. 1). During the search, the process may encounter peaks and troughs which may impede progress. Without loss of generality, we only consider a minimization process throughout this paper.

We can formally define a landscape as an ordered set of three components $L = (X, f, d)$ where X is the set of possible solutions or configurations, f is the fitness or energy function, and d is a distance measure between two points in X. In the following subsections, we describe in detail each of these components as they relate to water cluster optimization.

A. *Configuration/Representation Space*

The configuration space X is the set of physically consistent water clusters, denoted as $(H_2O)_n$ where n is the number of water molecules in the cluster. Each member $x \in X$ is a vector of $9n$ real numbers representing each atom's coordinates in 3-D space measured in Ångstroms (Å), that is, $x \in \mathbb{R}^{9n}$. We note that there are other possible methods for representing water molecules, such as with Eulerian angles [7], but the Cartesian coordinate representation allows for fully flexible clusters.

B. *Fitness or Potential Energy Function*

The fitness or potential energy function, $f = f(x) : X \rightarrow \mathbb{R}$, gives the height of the landscape. In this paper, we work with empirical pure water models that calculate the binding energies of water clusters. In chemistry, water clusters have been extensively studied and used as prototypes to study solvation of ions in great detail [15]–[17]. Empirical water models are computationally less demanding compared to *ab initio* methods and hence, are used extensively in physical chemistry and computational biology for simulation and optimization.

Empirical models have undergone extensive development during the past decade and at the time of writing, there exist more than 50 empirical models for water [18]. Popular empirical models include SPC, the TIP family (TIP3P, TIP4P, TIP4P-Ew, TIP5P-Ew, etc.), GCPM and QCT [14], [19]–[22]. In this paper, we sought to expose the global and local minima of pure water clusters of the recently developed TTM2.1-F [23] and OSS2 [24] flexible models.

1) *TTM2.1-F:* Since its introduction in 2004, the flexible, polarizable, Thole-type interaction potential for pure water (TTM2-F) has been the subject of several optimization studies and was demonstrated to possess global minima for a wide range of cluster sizes that agree with *ab initio* MP2 calculations. TTM2-F extends the rigid version (TTM2-R) with an intra-molecular charge redistribution scheme which involves coupling the Partridge–Schwenke monomer potential energy and the dipole moment surfaces to the intermolecular component of the total interaction [25]. The TTM2-F model was recently updated in 2007 to correctly account for the individual water dipole movement [23] and this revised model, TTM2.1-F, maintains the accuracy of the original TTM2-F but prevents the inaccuracies that arise at short intermolecular separations. In this paper, we intended to verify if TTM2.1-F possessed the same global minima as TTM2-F for n = 3–15.

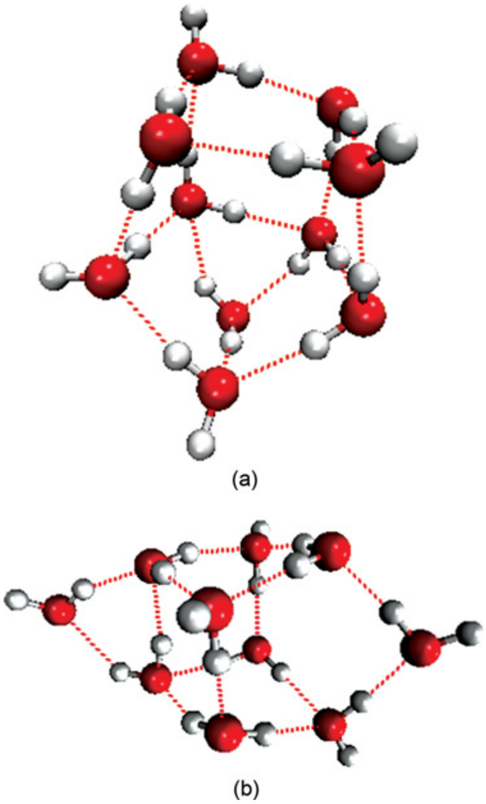


Fig. 2. Two structurally distinct $(H_2O)_{10}$ TTM2.1-F isomers which are almost iso-energetic with a binding energy difference of ≈0.0001 kcal/mol. (a) $E = -91.104$ kcal/mol. (b) $E = -91.1041$ kcal/mol.

2) *OSS2:* The second model considered in our paper is the OSS2 model by Ojamae, Shavitt, and Singer [24]. Unlike TTM2.1-F, OSS2 was developed to describe water as a participant in ionic chemistry, such as in biological processes. Although primarily developed for describing protonated water $H^+(H_2O)_n$, OSS2 can potentially model all three species of water, i.e., protonated, deprotonated, and pure water clusters. Compared to sister models (OSS1 and OSS3), OSS2 gave the "best overall performance with regard to structure and energetics of larger neutral and protonated water clusters" [24]. Specifically for small water clusters, OSS2 produced results for clusters $(H_2O)_n$ ($n < 6$) that are in agreement with MP2 calculations. To the best of our knowledge, global minima for sizes larger than $n = 8$ had not been investigated.

C. *Structural Distance Measure*

The structural distance measure, $d = d(x_i, x_j) : X \times X \Rightarrow \mathbb{R}$, is an important component of potential energy landscapes that was often overlooked in prior work. Without a structural metric, it is not possible to clearly define distinct structures. Previous optimization studies [2], [6], [7], and [26] filtered structures with similar energies, assuming implicitly that similarity in binding energies implies similarity in structure. Unfortunately, this assumption is not true in general (see Fig. 2) and simply disregarding structures with similar energies will dismiss potentially interesting isomers and pathways to other low-lying regions of the landscape.

Molecular structure comparison metrics can be broadly categorized into superposition and nonsuperposition methods [13]. Superposition methods optimize the overlay of compared molecules through a variety of metrics such as volume overlap [27], grid point counts [28] or Gaussian approximations [29]. These methods are generally accurate but are computationally expensive. Since we sought to compare thousands of structures, superposition methods were infeasible.

Instead, we used a computationally efficient nonsuperposition method with demonstrated accuracy: the USR [13]. Unlike superposition methods, USR measures a molecular structure's shape using a signature vector of 12 atomic distance statistics, U. This signature captures the mean, standard deviation, and asymmetry of the distances from each atom in the structure to four anchor points. The anchor points are a (the structure's centroid), b (the atom closest to a), c (the atom furthest from a), and d (the atom furthest from c).

This signature has nice properties in that it is invariant to translational and rotational symmetries. As such, we can easily define the similarity, $s(x_i, x_j)$ between two molecular structures x_i and x_j as inverses of distances between the signatures, for example, the inverse-scaled Manhattan distance

$$s(x_i, x_j) = \frac{1}{1 + \frac{1}{12}\sum_{k=1}^{12} |U_k^{x_i} - U_k^{x_j}|} \tag{1}$$

where $U_k^{x_i}$ denotes the kth component of x_i's USR signature. From (1), we can naturally define the distance or dissimilarity between structures as

$$d(x_i, x_j) = 1 - s(x_i, x_j). \tag{2}$$

In this form, $d(x_i, x_j)$ conveniently maps to $[0, 1)$ with 0 indicating maximum similarity. Note that $d(x_i, x_j)$ is also symmetric since $s(x_i, x_j) = s(x_j, x_i)$. Our tests with USR on pure water clusters indicated that it was effective at identifying duplicates and distinguishing dissimilar clusters. For the structures shown in Fig. 2, $d = 0.3192$ or 31.92%. From a computational perspective, USR is highly efficient and at least three orders of magnitude faster than the previously most efficient method, ROCS [30]. The signature computations are $O(n)$ for each pure water cluster of size n.

D. Isomer Sampling and Optimization

Now that we have defined the concept of a landscape, we define the problem of isomer sampling as a search for minima on a landscape. In particular, we wish to find $X_m \subseteq X$

$$X_m = \{x_i \in X | (|\nabla f(x_i)| = 0) \wedge \big(H_i \text{ is positive definite}^1\big)\} \tag{3}$$

where $|\nabla f(x_i)|$ is the magnitude of the gradient at x_i and H_i is the Hessian matrix evaluated at x_i.

For the real-world potential energy functions used in this paper, it was not feasible to numerically optimize a solution until $|\nabla f(x_i)| = 0$. As such, we approximated this requirement with $|\nabla f(x_i)| < \epsilon$ where $\epsilon = 5 \times 10^{-6}\,\text{kcal}\,\text{mol}^{-1}\,\text{Å}^{-1}$. Finite memory also mandated the limitation of isomers to a sample that was a subset of X_m for large search spaces. To formalize the concept of a "good" sample, we define the *ideal* sample $X_m^* \subseteq X_m$ which has the following properties:

1) contains all structures with energies below a user-defined threshold, i.e., $X_m^* = \{x_k \in X_m | f(x_k) \leq E_c\}$;
2) contains the global minima, i.e., $x^* \in X_m^*$ where for all $x_k \in X_m$, $f(x^*) \leq f(x_k)$;
3) contains no duplicates, i.e., there do not exist any $x_i, x_j \in X_m^*$ s.t. $(|f(x_i) - f(x_j)| < e_{dup}) \wedge (d(x_i, x_j) < d_{dup})$ where e_{dup} and d_{dup} are the maximum tolerable similarities in the energy and configuration spaces.

```
Archiving Molecular Memetic Algorithm($n$, $M$, $I$, $I_{LS}$, $w$, $F_T$, $e_c$, $\epsilon$, $e_{dup}$, $d_{dup}$, $p_i$, $p_c$, $p_p$, $p_r$, $n_g$)
Require: $c < 1.0$
$P \Leftarrow$ InitializePopulation($M$, $n$, $w$, $F_T$)
$A \Leftarrow$ InitializeArchive($P$)
for $i = 1$ to $I$ do
    $x' \Leftarrow$ GenerateChild($P$, $p_i$, $p_c$, $p_p$, $p_r$, $n_g$)
    $x' \Leftarrow$ LocalSearch($x'$, $\epsilon$, $I_{LS}$)
    $e \Leftarrow f(x')$
    $P \Leftarrow$ Replacement($x'$, $e$, $P$, $e_{dup}$, $d_{dup}$)
    $A \Leftarrow$ Archive($x$, $A$, $e_c$, $\epsilon$, $e_{dup}$, $d_{dup}$)
end for
```

Fig. 3. Archiving molecular memetic algorithm.

It follows naturally that the binding energies of structures in the ideal sample are bounded by E_c and $f(x^*)$. A good sample should approximate the ideal sample and although X_m^* is not known, any sample set X_k (without duplicates and with energies bounded by E_c) is a subset of X_m^*. Since the cardinality of X_k must be smaller or equal to that of X_m^*, we can test for closeness to X_m^* by measuring the size of X_k; the larger the size, the closer it is to the ideal set. Furthermore, X_m^* contains the global minima and as such, we can use the lowest energy isomer(s) in each set as another indication of similarity to X_m^*; the lower the energies of the best structures, the closer the sample to the ideal sample.

III. Stochastic Search and Optimization Methods

In this section, we discuss in detail the two stochastic methods developed in this paper: AMMA and ABHA. We first give an outline of both algorithms and then proceed to discuss each component in detail. Although the memetic algorithm and basin hopping approaches are different, both our algorithms share operators and the archival method. Both algorithms are also *asynchronous* and are easily parallelized for high-performance compute clusters using a master-slave framework, such as in [7] and [32]. As a reference, the parameters used by the algorithms with associated notation are shown in Table I.

A. Archiving Molecular Memetic Algorithm

Inspired by the myriad of complex organisms shaped by biological evolution, the evolutionary algorithm (EA) was

[1] H_i is a positive definite excluding the six modes associated with rotation and translation of the entire molecular cluster. The six modes were identified and removed with vibrational analysis [31].

TABLE I
AMMA AND ABHA PARAMETERS

General Parameters	
n	Water cluster size.
I	Maximum number of iterations.
I_{LS}	Maximum number of local search iterations.
Initialization Parameters	
w	Distance step-size.
F_T	Maximum failed attempts per distance step.
Child Generation Parameters	
p_i	Initialization probability.
p_c	Crossover probability.
p_p	Perturbation probability.
p_r	Relocation probability.
n_g	Maximum number of molecules to affect.
Archival Parameters	
e_c	Energy requirement.
ϵ	Gradient requirement.
Duplicate-check Parameters	
e_{dup}	Binding energy difference.
d_{dup}	USR dissimilarity.
AMMA-Specific Parameters	
M	Population size.
ABHA-Specific Parameters	
Q	Probabilistic acceptance function.
T	Temperature.

proposed as general method for search and optimization. Since its inception, an abundance of specific algorithms based on the evolutionary approach have been developed and demonstrated to solve difficult test and real-world problems. In contrast to conventional optimization methods, EAs use a population of solutions to iteratively sample the search space using competitive selection, crossover, and mutation operators. Although EAs are capable of exploring and exploiting promising regions of the search space, they can take a relatively long time to locate a minimum. Furthermore, EAs may not optimize a solution to the required precision, as compared to other search methods such as gradient descent.

The recently developed MA combines the evolutionary algorithm with individual learning procedures capable of performing local refinements to better explore and exploit the search landscape. Over the years, the MA has received increasing interest from researchers with many recent works revealing the ability of MAs to converge to high-quality solutions more efficiently than their conventional evolutionary counterparts [33]–[37]. In the context of complex optimization, many different instantiations of MAs have been reported across a wide variety of application domains [38]–[43], including water cluster optimization [7], [12], and [26].

In this paper, we developed an archiving memetic algorithm for collecting isomers or local minima while converging to the global minimum. It is worth noting that AMMA operates in the configuration space of *fully flexible* molecular structures described in Section II-A.The pseudo-code for AMMA is shown in Fig. 3.

```
**Archiving Basin Hopping Algorithm**($n$, $I$, $I_{LS}$, $Q$, $T$, $w$, $F_T$, $e_c$, $\epsilon$, $e_{dup}$, $d_{dup}$, $p_i$, $p_c$, $p_p$, $p_r$, $n_g$)
Require: $c + i < 1.0$
$x \Leftarrow$ InitializeWaterCluster($n$, $w$, $F_T$)
$x \Leftarrow$ LocalSearch($x$, $\epsilon$, $I_{LS}$)
$e \Leftarrow f(x)$
$A \Leftarrow$ InitializeArchive($\{x\}$)
for $i = 1$ to $I$ do
    $x' \Leftarrow$ GenerateChild($\{x\}$, $p_i$, $p_c$, $p_p$, $p_r$, $n_g$)
    $x' \Leftarrow$ LocalSearch($x'$, $\epsilon$, $I_{LS}$)
    $e' \Leftarrow f(x')$
    if $e' < e$ then {New Solution is better}
        $x \Leftarrow x'$
    else
        if Random(0,1) $\leq$ $Q(x', x, T)$ then {Accept bad trade}
            $x \Leftarrow x'$
        end if
    end if
    $A \Leftarrow$ Archive($x$, $A$, $e_c$, $\epsilon$, $e_{dup}$, $d_{dup}$)
end for
```

Fig. 4. Archiving basin hopping algorithm.

B. Archiving Basin Hopping Algorithm

Analogous to the MA that combines the evolutionary algorithm with local search, the basin hopping algorithm combines SA and local optimization [44]. SA was inspired from annealing in metallurgy and mimics the process undergone by atoms when a metal is heated and slowly cooled. The heat causes the atoms to free themselves and the slow cooling increases the probability of finding better configurations. Likewise, each step of the SA algorithm always makes a move to a better solution but also allows for bad trades, which are decided using an acceptance probability function [45].

The combination of SA with local search can be seen as a variant of the standard iterated local search (ILS) algorithm [46] with the extended capability of making bad trades to escape local minima. Because our intent is to sample low-lying minima and not merely locate the global minimum, ABHA does not use an annealing schedule, akin to the Metropolis algorithm [47]. That said, an annealing schedule could be added with minimal changes to the algorithm. Fig. 4 illustrates the pseudo-code for our proposed ABHA.

C. Stochastic Multistart Local Search

As a baseline algorithm to benchmark AMMA and ABHA, we used the SMSL algorithm. Unlike AMMA and ABHA, SMSL does not attempt to bias the generation of new individuals. It generates a maximum of I locally optimized water clusters using the initialization operator followed by a local search. As such, SMSL is explorative in nature and does not attempt to favor a particular region of the landscape over another.

```
**InitializeWaterCluster**($n$, $w$, $F_T$)
if Random(0,1) $<$ 0.5 then {Start at the centroid}
    $w_c \Leftarrow 0$
else
    $w_c \Leftarrow w$
end if
$a \Leftarrow 0$ {Track number of attempts}
while Size($x$) $<>$ $n$ do
    $h \Leftarrow$ CreateWaterMolecule() {Initialize at the origin}
    $h \Leftarrow$ RandomTranslateByDistance($h$, $w_c$)
    if isValid(AddMolecule($h$, $x$)) then {Cluster is Valid}
        $x \Leftarrow$ AddMolecule($h$, $x$) {Add the molecule to $x$}
    else
        $a \Leftarrow a + 1$
    end if
    if $a > F_T$ then {Too many attempts}
        $w_c \Leftarrow w_c + w$ {Increase current distance}
        $a \Leftarrow 0$ {Reset attempts count}
    end if
end while
return $x$
```

Fig. 5. Water cluster initialization operator.

D. Child Generation with Landscape Traversal Operators

To traverse the landscape, AMMA and ABHA use a combination of five operators: an initialization operator to generate entirely new solutions, a local search operator for "drilling down" to minima, a perturbation operator for exploring nearby points, a molecular relocation operator for jumping large distances and finally, a crossover operator for combining good solutions. Each of these operators plays a crucial role in the search and optimization process.

1) *Local Search Operator:* The local search operator is essential because we require our algorithms to locate isomers with binding energy gradients of 5×10^{-6} or lower. As such, it is necessary to find minima with sufficient precision. Initially, we used the BFGS algorithm, one of the most widely used quasi-Newton methods for solving nonlinear optimization problems [48].

2) *Initialization Operator:* The initialization operator creates a new pure water cluster of a given size by iteratively adding molecules at increasing distances w from a central starting point at the origin (see Fig. 5 for pseudo-code). After initialization, the cluster is locally optimized to bring a solution to its local minimum. Our preliminary tests on $(H_2O)_6$ demonstrated our initialization method was effective at generating a wide range of clusters from across the energy spectrum with $w = 2.5$ Å and $F_T = 5$ (Fig. 6).

3) *Perturbation and Relocation Operators:* The perturbation operator [Fig. 7(a)] is a standard operator used in prior research [7], and [26], and explores the neighborhood around the parent cluster. The operator arbitrarily perturbs (translates and rotates) randomly selected molecules in a cluster. Molecular translation is achieved by adding a vector of three random real numbers (uniformly generated between 0 and 2.0 Å) to the coordinates of each atom in the selected molecule. Molecular rotation is performed by rotating a molecule by an arbitrary degree (uniformly generated between 0 and 2π radians) around the axis formed by the oxygen atom and a randomly generated point (obtained by adding a uniformly generated real number between 0 and 1 to each coordinate of the oxygen atom).

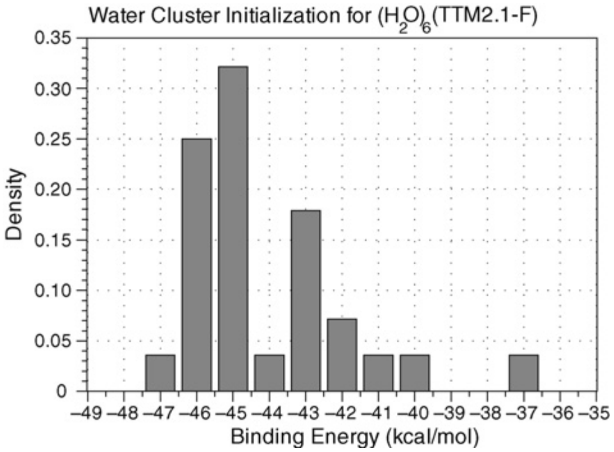


Fig. 6. Energy distribution of 600 water clusters of size 6 generated using the initialization operator and local optimization with BFGS. For this small cluster size, the initialization method generated samples from across the energy spectrum and was sufficient to locate the global minimum.

Unlike the perturbation operator, which explores nearby solutions, the relocation operator was formulated to be more drastic. As its name implies, the operator *relocates* randomly selected molecules to random locations on surface of the water cluster [Fig. 7(b)]. The relocation operator allows the search process to leap great distances to other regions of the landscape, which is useful for escaping deep minima.

Fig. 8 illustrates the effect of the perturbation and relocation operators on ten independently initialized $(H_2O)_{10}$ clusters. As expected, the mean USR dissimilarity from 30 generated solutions to the original water clusters for the pertubation operator is low (<20%) even when four molecules are perturbed (Fig. 8). In contrast, the USR dissimilarity between relocated solutions is significantly higher (>57%) even when only a single molecule is affected.

4) *Crossover Operator:* In contrast to the perturbation and relocation operators which are applied on a single cluster, the crossover operator merges two parent clusters, $x_k, x_l \in X$ to create a single child cluster. The merge process simulates a growth from the extreme ends of two clusters. The operator first randomly rotates both x_k and x_l around an arbitrary axis passing through the clusters' centroids, $(x_{k,c})$ for x_k. It then finds the furthest molecule from the centroid in x_k $(x_{k,f})$ and the furthest molecule from $x_{k,c}$ in x_l $(x_{l,f})$. Then, it locates the closest molecule to either $x_{l,f}$ or $x_{k,f}$ and adds it to the child cluster. The added molecule is marked so that it cannot be added again. This growth process continues until the child cluster meets the required size.

The crossover operator can also be used to generate a new cluster from only one parent by setting $x_k = x_l$. Since a random rotation is applied to the clusters before the growth process, the generated child is unlikely to identically match the parent cluster. We found 300 $(H_2O)_{10}$ clusters generated using the

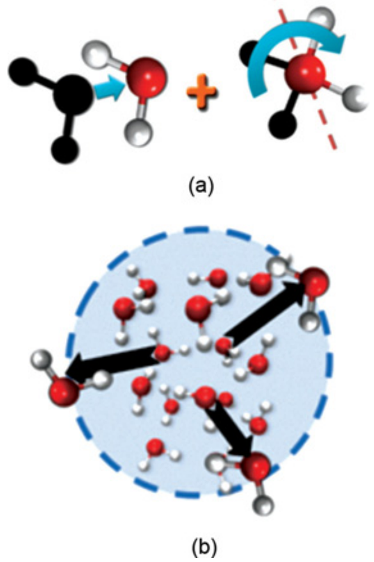


Fig. 7. Perturbation and relocation operators for traversing the landscape. (a) Perturbation operator. (b) Relocation operator.

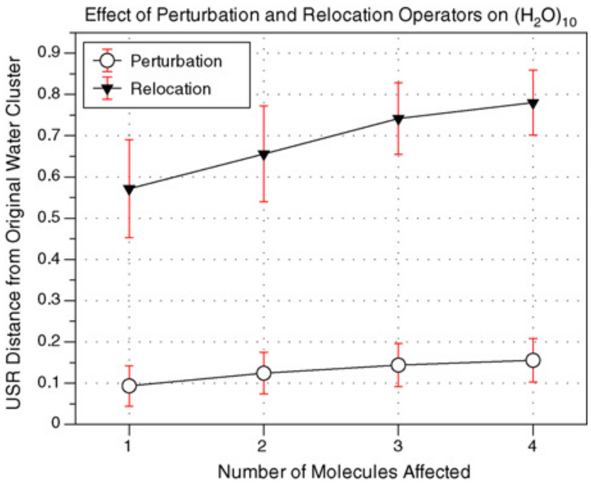


Fig. 8. USR dissimilarity of 30 new structures generated from ten independently initialized water clusters of size 10, using the perturbation and relocation operators. Both operators are able to create more distinct structures when a greater number of molecules are affected, but the relocation operator is clearly the more disruptive of the two.

crossover operator were similar to both parents ($d \approx 20\%$). Fig. 9 also shows that the USR dissimilarity remains fairly constant at 20% even as the cluster size is varied from seven to ten molecules.

5) *Random Initialization:* During our initial tests, we discovered that it was possible for algorithms to get "stuck" in a sub-optimal region that was not well-modeled by the empirical water models if the initial clusters were formed in that region. This issue was more detrimental to ABHA, especially on larger water cluster sizes. Since AMMA possessed a population of initial starting points, it was less dependent on any one starting solution. We solved this problem by randomly initializing solutions with a probability of 0.05 per iteration.

6) *Finalized Child Generation Algorithm:* The finalized child generation algorithm is shown in Fig. 10. In addition to a current sample of structures $\hat{X}$, the algorithm accepts five parameters, $(p_i, p_c, p_p, p_r, n_g)$, which control how often each operator is applied and the number of molecules affected by the perturbation/relocation operators. For parent selection, the method uses simple rank selection [49] where the current population members are ranked in the order of increasing fitness (individuals with the identical fitness values are given identical ranks). The parents are then picked with probability in proportion to their ranks. Note that because ABHA uses only a single search point, the selection function would always return the current point.

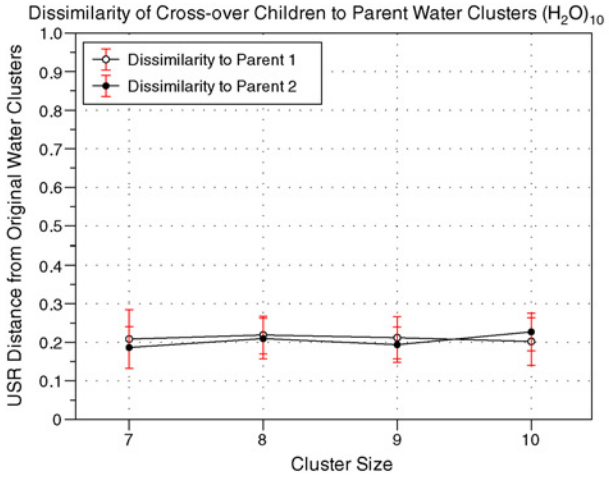


Fig. 9. USR dissimilarity of 300 new structures generated from ten independently initialized water clusters sizes $n = 7, 8, 9, 10$ using the crossover operator. The resultant structures are shown to be similar to both parents.

```
**GenerateChild**$(\hat{X}, p_i, p_c, p_p, p_r, n_g)$
**Require:** $p_i + p_c + p_p + p_r = 1.0$
$r \Leftarrow$ Random(0,1)
**if** $r \leq p_i$ **then**
    $x' \Leftarrow$ InitializeWaterCluster$(n, w, F_T)$
**else if** $r \leq p_i + p_c$ **then** {Perform Crossover}
    $(x_k, x_l) \Leftarrow$ SelectParents$(\hat{X})$
    $x' \Leftarrow$ Crossover$(x, x)$
**else if** $r \leq p_i + p_c + p_p$ **then** {Perform Perturbation}
    $x \Leftarrow$ SelectParent$(\hat{X})$
    $x' \Leftarrow$ Perturbation$(x, n_g)$
**else** {Perform Relocation}
    $x \Leftarrow$ SelectParent$(\hat{X})$
    $x' \Leftarrow$ Relocation$(x, n_g)$
**end if**
**return** $x'$
```

Fig. 10. Child generation with the perturbation, relocation, and crossover operators.

E. Replacement Strategy

A fundamental way in which both AMMA and ABHA differs is in the replacement strategy employed. Recall that as discussed in section III-B, ABHA always accepts good trades and bad trades are decided using an acceptance probability function. For this paper, we used the standard Boltzmann function

$$Q(x', x, T) = e^{\frac{-|f(x')-f(x)|}{T}}. \qquad (4)$$

Cluster x is only accepted if $Q(x', x, T) > R(0, 1)$ where $x, x' \in X$ and $R(0, 1)$ is a random number in the interval $[0, 1]$. T is a tuning parameter, which varies the probability that higher energy clusters are accepted and we used $T = 0.4$ as it worked well in our initial tests with small clusters.

Unlike ABHA which uses a single search point, AMMA manages a population of solutions. Diversity is a measure of the distinctiveness of the solutions/clusters in a population and is an important property in evolutionary optimization. Too low a diversity may lead to premature convergence and impedes the search for new isomers. On the other hand, too high a diversity may slow convergence. Diversity preservation is a well-researched topic in evolutionary computation as evident by the variety of strategies such as niching methods (e.g., fitness sharing and crowding), mating restriction and entropy-based methods [50]–[55].

Many of these methods rely on a quantitative distance measure either in the configuration or fitness (energy) spaces. Recall that in prior research [2], [6], [7], [26], the difference in the fitness space, specifically the binding energy, is often used as the sole distance measure. Because we have defined a suitable structural distance measure, $d(x_i, x_j)$, AMMA, and ABHA can better distinguish structures by using distances in *both* the configuration and energy spaces.

AMMA preserves diversity by preventing the duplication of structures in the population, which has been implicated as a cause for premature convergence [56]. Before a cluster is inserted into the population, it is checked against every population member. If the USR dissimilarity to any existing population member is below 4% and the binding energy difference between the two clusters is less than 0.01 kcal/mol, the cluster is classified as a duplicate and is prevented from entering the population. Otherwise, the new cluster replaces the highest energy water cluster in the population. The threshold values of 4% and 0.01 were chosen based on investigations performed in our prior work [12] but can be easily modified for other studies.

F. Isomer Archival and Vibrational Analysis

Any generated structure was archived if and only if it met the user-defined gradient requirement (i.e., $|\nabla f(x_i)| < \epsilon$ as defined in Section II-D) and was not already present in the archive. To ensure that comparisons and duplicate checking could be performed efficiently, AMMA and ABHA store clusters in a multimap. Multimaps are associative data structures that store elements indexed by keys (which need not be unique). This permits for fast access and retrieval based on key values, provided the elements are fairly well-distributed across the keys. Given M elements for a particular key, worst case access time is $O(M + 1)$ and worst case insertion time is $O(\log |A|)$. In this paper, we indexed structures by binding energies reduced to two decimal points.

After the optimization process is completed, the archive is further reduced with vibrational analysis. Vibrational analysis ensures that a given molecular cluster has converged to a minimum on the energy landscape by computing second derivatives and removing symmetries. Briefly, the method consists of six steps: computing the Hessian, H, of the water cluster coordinates matrix, mass weighting H, determining the principal components of H (principal axes of inertia), generating a transformation with separated rotation and translation modes, transforming H into the new internal coordinates, H', and finally, computing H''s eigenvalues. We refer readers desiring more detail on vibrational analysis to [31].

```
**Archive**($x$, $A$, $e_c$, $\epsilon$, $e_{dup}$, $d_{dup}$)
**Require:** $e_{dup} = 10^{-\alpha}$ for some integer $\alpha$
if ($|\nabla f(x)| < \epsilon$) $\wedge$ ($f(x) \leq E$) then {Structure is a potential isomer}
    $k \Leftarrow$ Integer($f(x) \times 1/e_{dup}$) {Compute key}
    $X_D \Leftarrow$ GetStructuresInRange($k-1$,$k+1$)
    Duplicate $\Leftarrow$ false
    for $x_i$ in $X_D$ do
        if ($|f(x) - f(x_i)| < e_{dup}$) $\wedge$ ($d(x, x_i) < d_{dup}$) then {$x$ is a duplicate of an existing archived structure}
            Duplicate $\Leftarrow$ true
            break
        end if
    end for
    if Duplicate = false then
        $A' \Leftarrow$AddStructure($x$, $k$, $A$) {Add $x$ to archive $A$ with key $k$}
    end if
end if
return $A'$
```

Fig. 11. Isomer archival algorithm.

G. Dissociative Clusters

During this paper, we found that locally searching the empirical functions would occasionally result in "dissociative" clusters possessing lower than reasonable binding energies. These dissociative clusters were characterized by two or more disconnected pieces. We hypothesize that these broken clusters resulted from the gradient-based local search algorithms exploring regions that were "off the map" and not well-modeled by the empirical fits. As a solution, we modified the evaluation function to ensure that the cluster was a single connected graph (with the maximum distance between any two connected atoms set at 6 Å). Any cluster which did not pass this check was disregarded.

H. Worst-Case Computational Complexity

The worst-case computational complexity of AMMA and ABHA is dependent on the computational costs of the different functions including child generation and isomer archival. For a given cluster of size n, each child generated requires at most $O(n)$ time while isomer archival requires at most $O(\log I)$ time where I is the maximum number of global iterations and hence, the maximum size of the archive.

TABLE II
EXPERIMENTAL PARAMETERS FOR AMMA AND ABHA

Parameter	Value
I	$200n$ (n = water cluster size.)
I_{LS}	2500
w	2.5 Å
F_T	5
p_i	0.05
p_c	0.15
p_p	0.64
p_r	0.16
n_g	$\lceil 0.2n \rceil$
e_c	0 kcals/mol
ϵ	5×10^{-6} kcal mol^{-1} Å^{-1}
e_{dup}	0.01 kcals/mol
d_{dup}	0.04 (96% Similarity)
M	10
Q	$e^{\frac{-\lvert f(x')-f(x)\rvert}{T}}$ (Boltzmann)
T	0.4

However, these costs are often eclipsed by the computational complexity of the potential energy and gradient functions and the number of calls to these functions made by the local search. Since AMMA and ABHA generate a single structure per iteration, the maximum number of function and gradient evaluations in a single local search is $O(I_{LS})$ where I_{LS} is the maximum number of local iterations.

In general, if we let $c_f(n)$ and $c_g(n)$ be the computational cost of arbitrary potential energy and gradient functions, respectively, the computational cost for a single run of either algorithm is $O(I \cdot [I_{LS}(c_f(n)+c_g(n))+n+\log I])$. In the typical case where $c_g(n) > c_f(n)$ and $c_g(n) = O(n^2)$ (or larger), and I is a constant picked depending on the maximum number of isomers desired, we can drop the lower order terms to yield $O(I_{LS}c_g(n))$.

IV. EXPERIMENTS

A. Experimental Setup

To test the effectiveness of the AMMA and ABHA algorithms, we conducted computational experiments on pure water clusters $(H_2O)_n$, n = 3–15 with the parameters in Table II on a 512-processor ×86 cluster. The average CPU time required for each run is shown in Fig. 12. We observed that TTM2.1-F required approximately twice the computation time of OSS2. Because of the computational expense associated with these experiments, our tests were limited to ten independent runs per cluster size per algorithm. For small clusters (<6), we conducted an additional ten runs per test with no significant change in the results presented. Furthermore, statistical results presented are significant at the 0.01 level.

In the following subsections, we compare the convergence and isomer sampling abilities of ABHA, AMMA, and SMSL on both the TTM2.1-F and OSS2 landscapes.

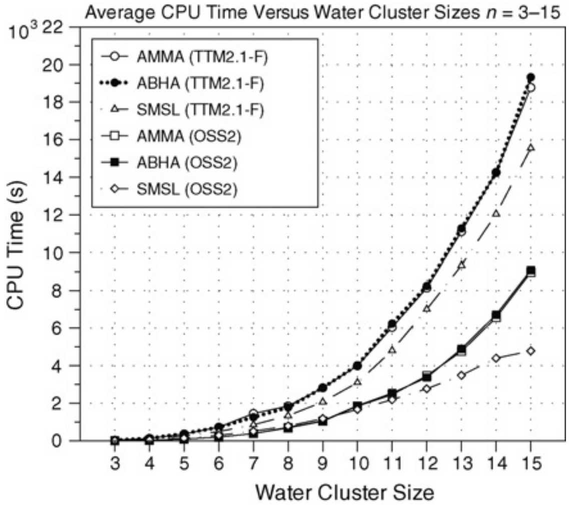


Fig. 12. Mean CPU time required for AMMA, ABHA, and SMSL for water cluster sizes n = 3–15 (averaged over ten independent runs).

B. Global Convergence

Table IV-B shows the best overall minima located during our paper. The lowest binding energy isomers located by AMMA and ABHA on TTM2.1-F corresponded to the global minima found on the older TTM2-F landscape [7]. For small clusters $n \leq 8$ on the OSS2 landscape, the best minima match the structures located in our previous work [12]. To the best of our knowledge, no global minima structures for OSS2 have been produced for $(H_2O)_n$ $n > 8$ and we submit the structures found in this paper as global minima candidates.

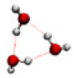
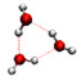
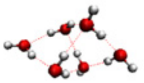
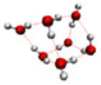
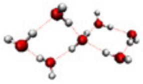
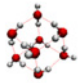
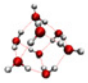
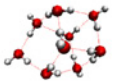
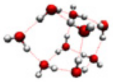
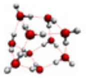
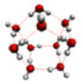
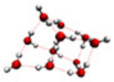
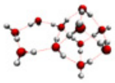
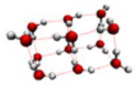
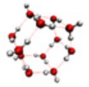
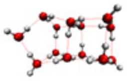
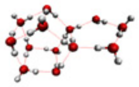
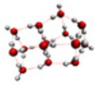
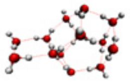
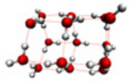
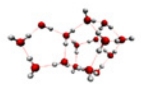
Comparing the best minima both visually and using the dissimilarity measure, we observed that TTM2.1-F and OSS2 have similar minima only for smaller clusters n = 3–5, 8–10. To ensure that the best minima were indeed different, we locally searched the TTM2.1-F best minima using the OSS2 potential energy model (and vice versa) and verified that the resulting local minima had higher energies than the structures shown in Table IV-B.

Furthermore, the frequency that the algorithms attained the best minima differed significantly for TTM2.1-F and OSS2 [Fig. 13(a)]. On TTM2.1-F, the three algorithms converged for all ten runs in the allotted number of iterations for small water clusters $(H_2O)_n$ $n \leq 9$. For $n \geq 10$, AMMA converged with the highest frequency, followed by ABHA, except the largest size of n = 15 where ABHA converged once out of the ten runs. Given the larger problem size and the expected exponential increase in local minima, it was not surprising that AMMA and ABHA did not converge for every run within the number of iterations used in our experiments. In fact, even when the best minima were not found, AMMA and ABHA algorithms found solutions near (≤1.0091 kcal/mol on average) to the global minimum with a small standard deviation (≤0.7 kcal/mol) as shown in Fig. 13(b). On the other hand, SMSL fared poorer with greater average distances from the global minimum (≤2.427 kcal/mol) and larger standard deviations of up to 1.2 kcals/mol.

On the OSS2 landscape however, AMMA and ABHA did not achieve a convergence rate of 100% even for small water

TABLE III

GLOBAL MINIMA CANDIDATES FOR $(H_2O)_{n,\, n=3-15\, (TTM2.1-FandOSS2)}$

	Binding Energy (kcal/mol)		Dissimilarity		Binding Energy (kcal/mol)		Dissimilarity
Cluster Size	TTM2.1-F	OSS2	d	Cluster Size	TTM2.1-F	OSS2	d
3	 −15.9423	 −18.2660	**2.54%**	4	 −27.6254	 −30.0699	**2.78%**
5	 −36.8065	 −39.2643	**4.02%**	6	 −46.5334	 −48.8647	25.96%
7	 −57.8342	 −57.7663	43.11%	8	 −73.3289	 −69.1407	**12.82%**
9	 −83.4193	 −77.9108	**6.32%**	10	 −94.6743	 −88.2630	**7.88%**
11	 −104.818	 −97.8907	29.02%	12	 −120.151	 −109.937	28.36%
13	 −130.540	 −118.777	22.63%	14	 −142.929	 −129.698	24.74%
15	 −154.090	 −139.663	32.51%				

Molecular structures were visualized using VMD [57]. Small dissimilarity scores ($d \leq 15\%$) are in **bold**.

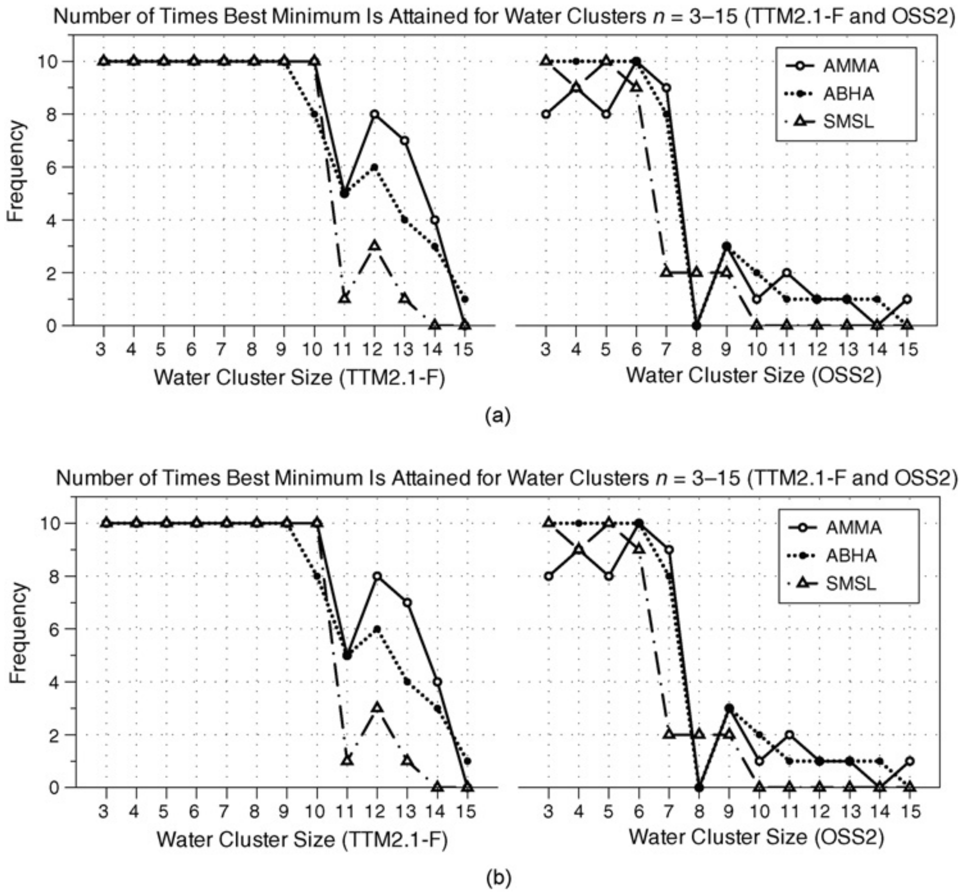


Fig. 13. Convergence results for AMMA, ABHA, and SMSL on the TTM2.1-F and OSS2 empirical water models for $(H_2O)_n n$ = 3–15. (a) Convergence frequency of AMMA, ABHA, and SMSL on TTM2.1-F and OSS2. (b) Mean convergence of AMMA, ABHA, and SMSL on TTM2.1-F and OSS2.

clusters. The convergence rate of all three algorithms fell drastically from 80–100% to 10–30% for water clusters larger than $n = 7$. Although AMMA and ABHA appear to still outperform SMSL for water clusters $n \geq 9$, the difference is less apparent than on TTM2.1-F. Furthermore, the mean energy differences of the solutions located to the best minima were double that for TTM2.1-F with larger, more erratic, standard deviations [Fig. 13(b)].

C. Isomer Archive Sizes

Fig. 14 shows the mean and standard deviation of the isomer archive sizes for AMMA, ABHA, and SMSL. We applied the nonparametric Mann–Whitney *U*-test and found no statistical difference ($P < 0.01$) between the archive sizes generated by AMMA and ABHA on the TTM2.1-F landscape. Surprisingly, the SMSL algorithm appeared very effective at sampling isomers on TTM2.1-F, generating archives statistically larger ($P < 0.01$) than AMMA and ABHA. However, upon closer inspection, we observed that the SMSL archives were biased toward higher energy structures. On the other hand, AMMA and ABHA sampled more low-energy structures, clearly shown by the plotted energy distributions in Fig. 15.

Similar to the TTM2.1-F landscape, we observed no statistical difference between AMMA and ABHA on the OSS2 landscape. We also noted that SMSL's performance was not replicated on the OSS2 landscape. On the contrary, the AMMA and ABHA algorithms far surpassed the SMSL algorithm for water cluster sizes larger than $(H_2O)_8$, sampling up to 840% more isomers. Unlike AMMA and ABHA, which continued to gather more isomers on the higher dimensional landscapes of larger water clusters, we observed a *falling* trend in SMSL's ability to sample new structures. Furthermore, all three algorithms gathered more isomers on TTM2.1-F than on OSS2 for water clusters larger than $(H_2O)_{10}$.

V. Landscape Analysis and Discussion

Our empirical results show that AMMA and ABHA were comparable in terms of isomer sampling and global convergence. However, we observed that both algorithms found OSS2 more difficult to explore and optimize. Since both TTM2.1-F and OSS2 were developed to model water clusters and possessed the same degree of freedom for each water cluster size, the significant performance disparity between the two landscapes was unexpected. To uncover the reasons for this, we probed the underlying global and local properties of both landscapes.

We first combined all the isomers archived during this paper into two archives; one for TTM2.1-F and one for OSS2. All duplicates were filtered during the process and the retained

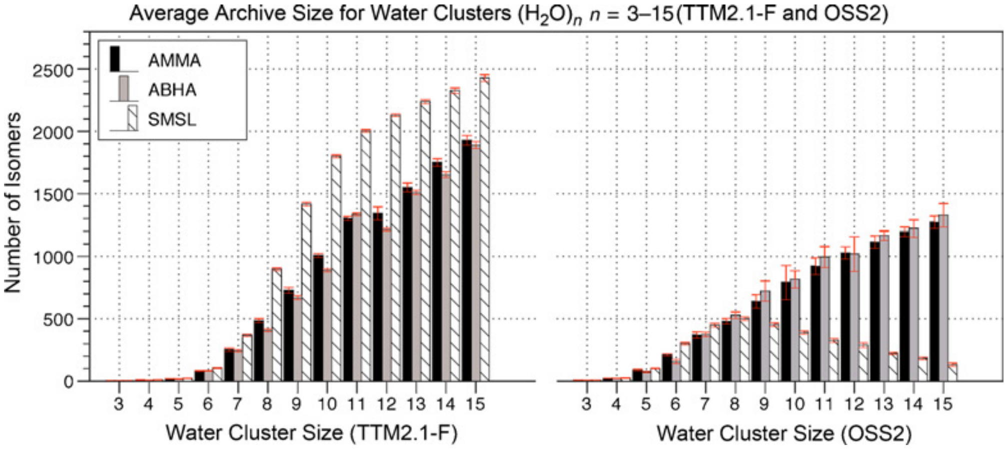


Fig. 14. Isomer archive sizes for AMMA, ABHA, and SMSL on the TTM2.1-F and OSS2 empirical water models for $(H_2O)_n n$ = 3–15.

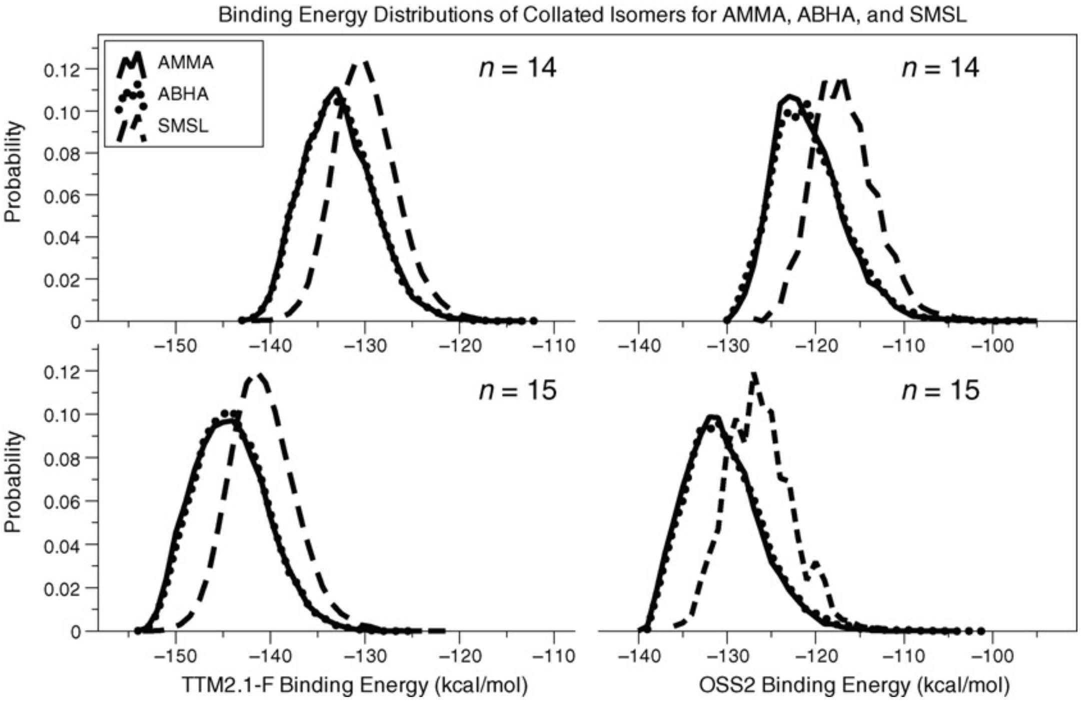


Fig. 15. Binding energy distributions for isomers located by AMMA, ABHA, and SMSL on TTM2.1-F and OSS2 (sizes 14 and 15).

solutions were re-verified with vibrational analysis. Fig. 16 shows the total number of isomers (log-scale) used for the following landscape analysis. The largest archive size was for $(H_2O)_{15}$ with 65 597 isomers.

A. Global Landscape Correlation Measures

Landscape correlation is an indication of problem difficulty [58]. Intuitively, a high-correlation (>0.6) indicates the minima are well-ordered and an optimization method can easily roll downward toward the global minimum. An uncorrelated landscape (≈0) may mislead an optimization algorithm to sub-optimal regions and is considered "rough." A landscape with negative correlation is said to be "deceptive" as the global minimum is located among high-energy solutions. We computed two metrics, the fitness-distance correlation metric (FDC) [58] and the FDC-tau, to measure the global correlation of the TTM2.1-F and OSS2 landscapes.

1) *Fitness-Distance Correlation (FDC):* The FDC is the Pearson product moment correlation between the energy differences and the structural differences of the samples to the lowest energy isomer

$$FDC = \frac{cov(\delta E, \delta D)}{\sigma(\delta E)\sigma(\delta D)} \tag{5}$$

where $cov()$ is the covariance function, δE and δD are the energy difference and USR dissimilarity between each solution and the lowest energy solution, respectively. Likewise, $\sigma(\delta E)$ and $\sigma(\delta D)$ represent the standard deviations of the energy differences and the structural dissimilarity.

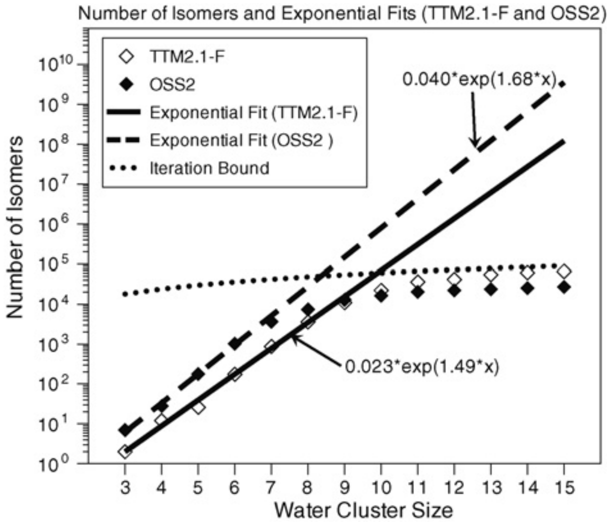


Fig. 16. Total number of archived isomers for water cluster sizes n = 3–15 on TTM2.1-F and OSS2.

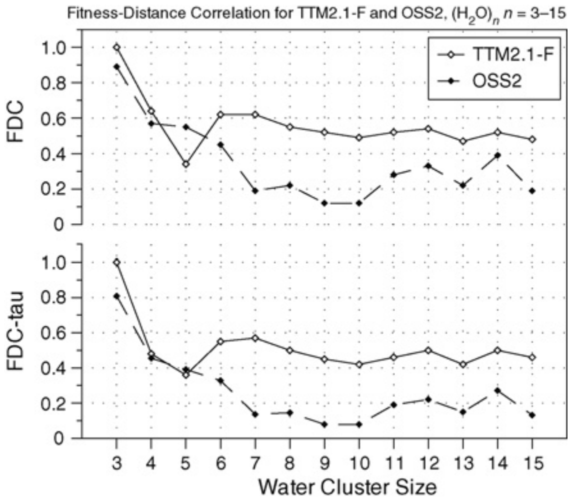


Fig. 17. Fitness-distance correlation for water cluster sizes n = 3–15 on TTM2.1-F and OSS2.

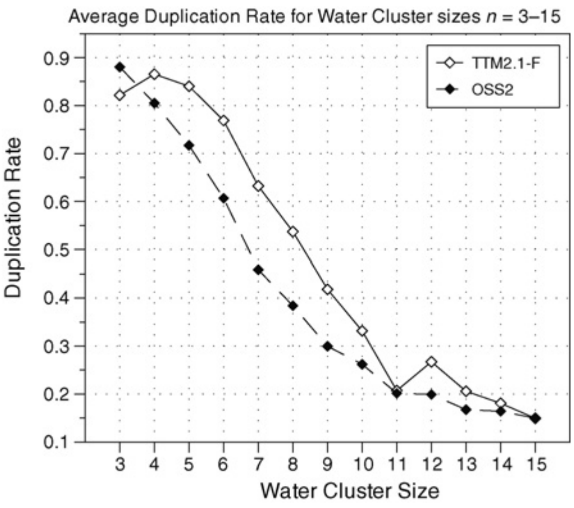


Fig. 18. Duplication rate for water cluster sizes n = 3–15 on TTM2.1-F and OSS2.

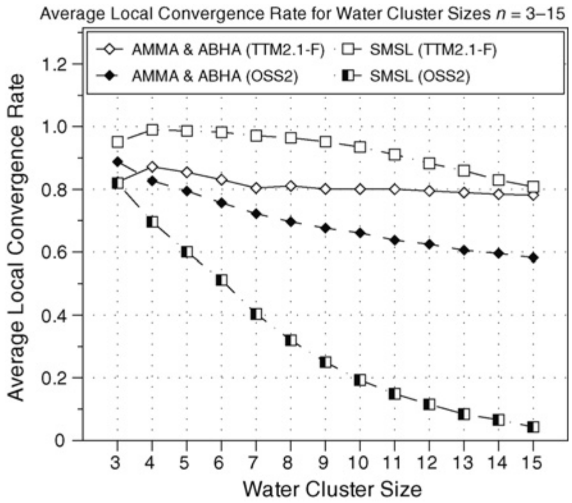


Fig. 19. Local convergence rate of AMMA, ABHA, and SMSL for water cluster sizes n = 3–15 on TTM2.1-F and OSS2 using the BFGS algorithm.

2) *Fitness-Distance Correlation-tau (FDC-tau):* Because the FDC assumes normally distributed data, we propose the use of an additional metric, the FDC-tau, which uses the nonparametric Kendall's tau measure of correlation

$$FDC{-}tau = \frac{n_c - n_d}{n(n-1)/2} \qquad (6)$$

where ranks are used instead of raw binding energy values; n represents the number of samples, n_c represents the number of concordant pairs, and n_d represents the number of discordant pairs. When the assumptions of normality and linear relationship are broken, the FDC-tau is a more robust metric compared to the FDC.

B. Global Landscape Correlation of TTM2.1-F and OSS2

The FDC and FDC-tau plots for TTM2.1-F and OSS2 are shown in Fig. 17. Although the global correlations of both landscapes decrease with increasing problem size, OSS2's FDC and FDC-tau scores rapidly fall to less than 0.3 (low-correlation region) for water cluster sizes $n \geq 7$. In contrast, TTM2.1-F's correlation scores remain greater than 0.4, in the moderate correlation region. We also observed that despite the similar best minima, TTM2.1-F and OSS2 have different FDC/FDC-tau scores for water clusters sized n = 8–10, suggesting differing landscapes.

Recall that both AMMA's and ABHA's search processes are biased toward low-energy solutions, based on the intuition that the global minimum exists in low-energy regions. However, OSS2's low-global correlation scores indicate that its local minima are not well-ordered and the bias toward low-energy solutions was less likely to lead to the global minimum. In contrast, the bias toward low-energy solutions proved fruitful on TTM2.1-F, which is more correlated or "smoother" for the cluster sizes considered in this paper [with the single exception of $(H_2O)_5$]. These global correlation results provide a reason for the convergence disparity between TTM2.1-F and OSS2 but do not clarify the difference in archive sizes.

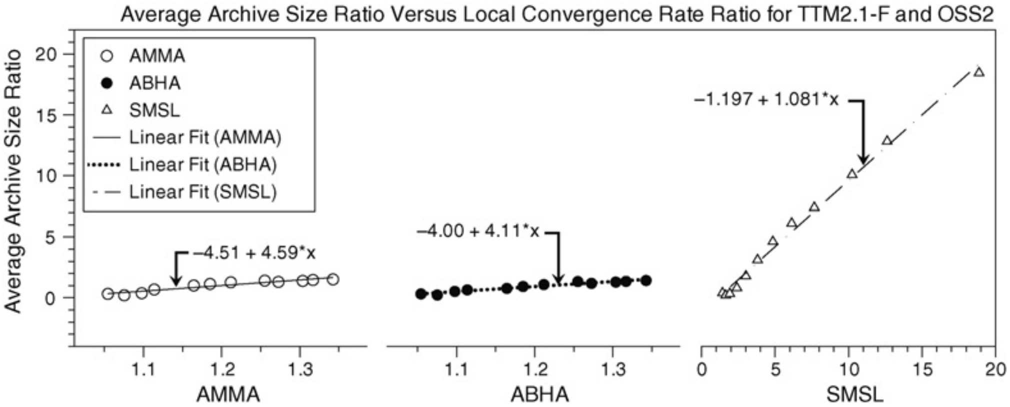


Fig. 20. Archive size ratio versus the local convergence rate of AMMA, ABHA, and SMSL for water cluster sizes n = 3–15 on TTM2.1-F and OSS2 using the BFGS algorithm.

C. Local Landscape Correlation of TTM2.1-F and OSS2

On average, AMMA and ABHA gathered up to 50% more isomers per run on the TTM2.1-F landscape compared to OSS2. While it was possible that the TTM2.1-F possessed more isomers than OSS2, we found this hypothesis unlikely. On the contrary, the total combined isomer archive for OSS2 was greater than TTM2.1-F for every cluster size up to $n = 9$ (Fig. 16). Indeed, the total number of unique isomers sampled appear to be bounded by the maximum iterations used in our experiments. Furthermore, exponential fits to the data up to the point of inflection ($n = 8$ for TTM2.1-F and $n = 6$ for OSS2) supported the notion that OSS2 possessed more isomers than TTM2.1-F. In addition, the average number of duplicate isomers generated during each run was consistently lower on OSS2 for every size except $(H_2O)_3$, suggesting the presence of more isomers compared to TTM2.1-F (Fig. 18).

To elucidate the reason behind the lower sampling rate on OSS2, we analyzed the local nature of the landscapes. As a proxy metric, we used the local convergence rate, which captured how often a local minimum was derived from a child solution generated during our experiment. When all operators were combined, such as in AMMA and ABHA, we observed the local convergence rate fell appreciably on OSS2 from 88% to 58% with increasing cluster size (Fig. 19). In contrast, the local convergence rate on TTM2.1-F remained relatively high at 78% even for the largest water cluster size of 15. Clearly, OSS2 was more difficult to locally optimize.

When we considered only the initialization operator (the sole operator used in SMSL), the difference in local convergence rates on both landscapes was more apparent. The local convergence rate on OSS2 fell dramatically from 81% to only 4% as water cluster size increased from 3 to 15, suggesting that the local search operator was not effective on the OSS2 landscape. By correlating the archive size ratio and local convergence ratio between TTM2.1-F and OSS2, we observed that all three algorithms were linearly impaired by lower convergence rates (Fig. 20). This impairment likely resulted in the observed difference in archive sizes between the two landscapes.

D. Discussion Summary

Despite differences in terms of formulation, both TTM2.1-F and OSS2 were designed for the purpose of computing the binding energies of water clusters and even share similar best minima for small cluster sizes. However, both the global and local landscape roughness conspired to make isomer sampling and global optimization more difficult on OSS2 compared to TTM2.1-F.

For the wider problem of isomer sampling on arbitrary potential energy landscapes, our landscape analysis has highlighted an interesting point; that landscapes of outwardly similar models may differ significantly. Therefore, one should not simply use identical methods (or parameters) to search and optimize models that may appear similar on the surface. We recommend that before initiating a search procedure, one should use the landscape analysis methods previously discussed, possibly on a smaller-scale with fewer isomers, to reveal global and local correlation properties.

If a landscape is revealed to possess low-global correlation, possible solutions to improve global convergence (for AMMA) include an increase in population size, multiple populations or a reduction of the selection pressure. For ABHA, a possible solution is to use a higher temperature, T, in the acceptance probability function, Q (4). These changes may encourage the exploration of other (perhaps higher energy) regions of the landscape, increasing the changes of locating the global minimum. To the algorithm designer, we postulate that parameter adaptation, such as in [59], [60], could play an important role in enabling algorithms to "fit" themselves to any arbitrary landscape, managing exploration, and exploitation as more landscape information becomes available.

Turning our attention to the local nature of the landscapes, our analysis suggested that OSS2 was difficult to locally optimize, limiting the isomer sampling abilities of our algorithms. In our implementation, BFGS returned when (1) the maximum number of iterations, $I_{LS} = 2500$, was reached, (2) a solution with a low-gradient, $|\nabla f(x_i)| < \epsilon$ where $\epsilon = 5 \times 10^{-6}$ kcal mol^{-1} Å^{-1}, was found or (3) when the line search along the (approximated) Newton direction did not yield a lower energy solution. Our tests revealed that (3) tended to occur

```
Hybrid Local Search Algorithm($x$, $I_{LS}$, $I_{\text{Pert}}$, $\epsilon$, $\sigma$)
  $x' \Leftarrow$ BFGS($x$, $\epsilon$, $I_{LS}$)
  if $|\nabla f(x')| \leq \epsilon$ then {Gradient Requirement met}
    return $x'$
  else {Gradient Requirement not met}
    for $i = 1$ to $I_{\text{Pert}}$ do
      $z \Leftarrow x' + \sigma$Random(-1,1) {Perturbation}
      $z' \Leftarrow$ BFGS($z$, $\epsilon$, $I_{LS}$)
      if $|\nabla f(z')| < \epsilon$ then {Found a local minimum}
        return $z'$
      end if
      if $f(z') < f(x')$ then {Found a solution with lower binding energy}
        $x' \Leftarrow z'$
      end if
    end for
  end if
  return $x'$
```

Fig. 21. Hybrid local search algorithm (HLS).

without returning a minimum, which we hypothesized to be a sign of discontinuities on OSS2's surface. This may be a problem with other empirical functions and to improve the general applicability of our algorithms, we sought to enhance AMMA with an improved local search method, described in the next section.

VI. A Hybrid Local Search

To handle possible discontinuities, we developed a *HLS* algorithm. At its core, HLS is an ILS variant [46] that introduces a stochastic element to the local search while maintaining the convergence precision of a gradient-based search (pseudo-code shown in Fig. 21).

The basic concept underlying HLS is straightforward and illustrated in Fig. 22: use BFGS until it arrives at a local minimum (G to D) or it encounters a difficulty, such as a discontinuity (A to B). Apply a simple perturbation to "jump" the discontinuity (B to C) and locate a new nearby starting point from which BFGS can reach the minimum (C to D). For simplicity, the perturbation step is uniformly generated between $(-\sigma, \sigma)$ where σ is a user-defined parameter. However, future work may look into varying σ automatically to adapt to the underlying landscape.

We integrated the HLS into AMMA (referred to as AMMA-HLS) and with our remaining computational budget, we were able to run AMMA-HLS on the larger pure water clusters $(H_2O)_n$ $n = 13, 14, 15$ using the OSS2 potential energy model. As before, our results are based on ten independent runs. To minimize the possibility of jumps to other basins, we set a small perturbation value of $\sigma = 0.05$ and $I_{\text{Pert}} = 10$.

When compared to AMMA using BFGS (AMMA-BFGS), AMMA-HLS produced statistically larger archives (Mann–Whitney U-test, $P < 0.01$), generating 38% to 47% more isomers on average (Fig. 24). As expected, this improvement was matched with an increase in computational cost (Fig. 23) due to the increase in successful local searches. In fact, with HLS, AMMA's isomer sampling performance on OSS2 was now on par with TTM2.1-F. While not definitive proof, our results support the notion of discontinuities on the OSS2 landscape.

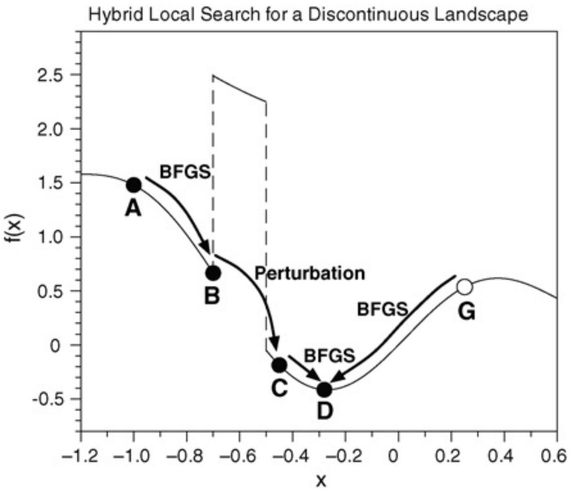


Fig. 22. Hybrid local search algorithm operating on a hypothetical potential energy landscape with two jump discontinuities where a small change in x leads to a large change in the function value $f(x)$.

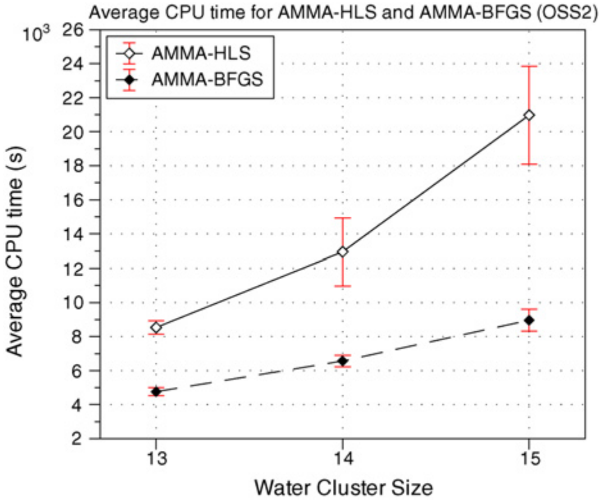


Fig. 23. CPU time required by AMMA-HLS and AMMA-BFGS on OSS2.

Although we developed HLS mainly to improve local convergence, we were curious about any possible global convergence effects. We analyzed the number of times AMMA-HLS converged to the structures in Table IV-B and of the ten runs, AMMA-HLS converged once for each cluster size, similar to the performance of AMMA-BFGS (as described in Section IV-B). This was not all-together surprising because although HLS improved local convergence, it was unlikely to improve AMMA's ability to locate the global minimum's basin, which is determined by global landscape properties and other algorithm parameters.

That said, the incorporation of HLS into AMMA achieved its purpose of improving isomer sampling. While we were not able to perform the same test with ABHA or SMSL due to computational budget constraints, we believe the two algorithms would be similarly improved.

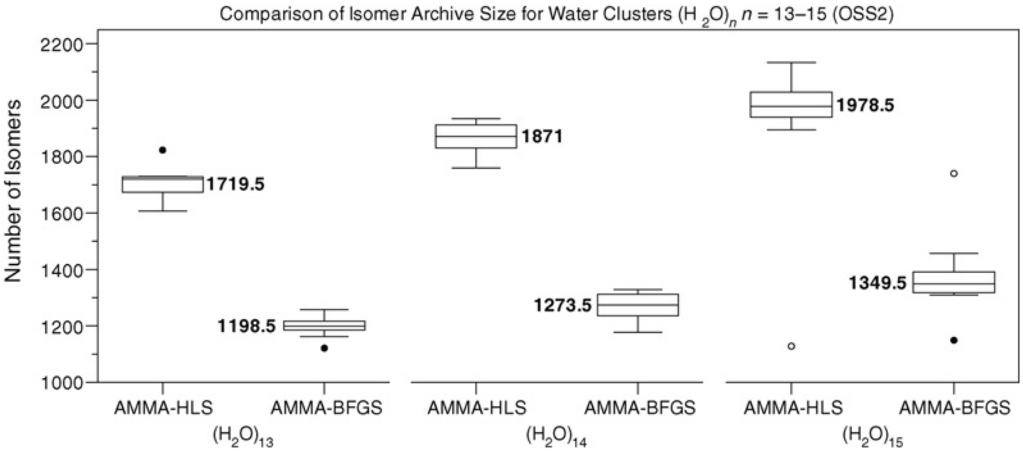


Fig. 24. Isomer archive sizes generated by AMMA with the HLS algorithm (AMMA-HLS) and with BFGS (AMMA-BFGS).

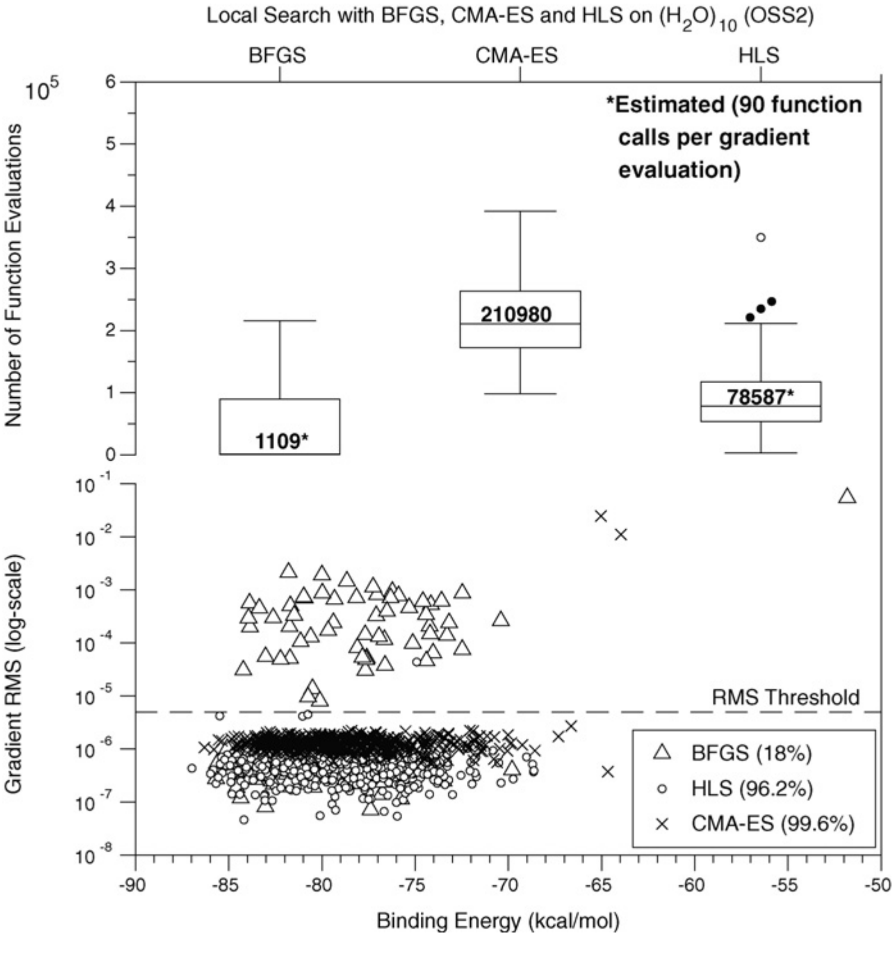


Fig. 25. Comparison of BFGS, CMA-ES, and HLS on 500 initialized $(H_2O)_{10}$ clusters.

A. Local Search with the Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES)

HLS was effective in our experiments due to the availability of relatively inexpensive gradient evaluations (approximately 2.5 times the computational cost of an energy evaluation in the case of OSS2). But other molecular models may lack an analytical gradient function and using numerical gradients may prove too expensive. As such, we explored the use of a leading nongradient-dependent stochastic local search algorithm, the CMA-ES [59], [61].

As a test, we applied CMA-ES, HLS, and BFGS to 500 initialized $(H_2O)_{10}$ clusters and compared the resulting structures in terms of binding energy and root-mean-square gradient (rms). We used the C version of the CMA-ES source code [62], and implemented the basic algorithm described in [61], using the standard normally distributed mutation and arithmetic recombination operators. To emphasize local-searching in

CMA-ES, we set the population size $k = 5$ and the initial coordinate-wise standard deviation $\phi = 0.01$. The algorithm was set to return when a solution with suitably low-rms value (5×10^{-6} kcal mol^{-1} Å^{-1}) was located or after 10^6 function evaluations.

We captured both the energy and rms value of the optimized solution as well as the number of evaluation calls needed to arrive at the solution. For BFGS and HLS, we estimated the number of evaluations by assuming that each gradient evaluation would require 90 potential energy function calls, as would be the case when estimating gradients with forward or backward finite differencing.

The experimental results in Fig. 25 clearly show that CMA-ES was the most robust local optimizer, yielding a minimum for 99.6% of the initial starting structures, closely followed by HLS (96.2%). In contrast, BFGS converged successfully for only 18% of the starting structures. Although CMA-ES was the best local optimization method in terms of convergence, it did require significantly more iterations—approximately 2.6 times more iterations compared to HLS on average.

We acknowledge our results are not conclusive but they suggest that CMA-ES is a robust nongradient dependent local search method for the problem of isomer discovery but further work may be necessary to improve its efficiency. Certain parameters sets or other CMA-ES variants [61] and [63] may yield superior results and outperform the standard method used in this paper.

VII. Conclusion and Future Work

In this paper, we presented and compared the AMMA and the ABHA for discovering isomers on the potential energy landscapes of fully flexible pure water clusters. AMMA and ABHA represent an enhancement of recent work which has focused solely on locating global minima. Empirical results on pure water clusters $(H_2O)_n$ $n = 3$–15 establish that both algorithms were comparable and effective in terms of convergence and isomer sampling.

AMMA and ABHA generated larger archives of low-energy isomers (up to 840% more isomers on OSS2) compared to SMSL and also verified that global minima for the TTM2.1-F empirical water model correspond to the older TTM2-F version for $(H_2O)_{3-15}$. In addition, the algorithms located new best minima for the OSS2 empirical model for water cluster sizes $n = 9$–15. Prior work has relied on global minima comparisons "by-eye" but we demonstrate how quantitative differences in structure can be assessed using an appropriate distance measure such as the USR.

Although global minima are important structures, they are nevertheless poor representatives for entire landscapes. As such, we conducted a landscape analysis using the large isomers archives generated during our experiments. To the best of our knowledge, this paper represents the first large-scale landscape study comparing two complex, sophisticated empirical water models, specifically, TTM2.1-F and OSS2. That said, our methods are sufficiently capable of being applied to alternative water models and extended to other molecular or atomic systems, from simple Lennard–Jones clusters to more complex nano-materials.

From the perspective of the evolutionary algorithmist, our landscape analysis revealed why our algorithms performed poorer on OSS2: OSS2 is rougher than TTM2.1-F, with low-global correlation (FDC and FDC-tau) scores of below 0.3 for $(H_2O)_n$ $n > 6$ which resulted in poorer convergence (in terms of frequency and mean energy difference). Moreover, local convergence rates were approximately 20% lower on OSS2, suggesting a less smooth local landscape. From the insights gained from our landscape analysis, we developed a HLS algorithm which substantially improved AMMA's isomer sampling capabilities, yielding statistically larger isomer archives on the OSS2 landscapes for $(H_2O)_n$ $n = 13 - 15$. We speculate that further information can be extracted from the landscape analysis, which can be performed "on the fly" in future algorithms to improve performance through parameter adaptation.

In addition, further study can be conducted on the mutation and crossover operators, possibly to better sample the search space. In particular, the random molecular rotation currently used is not uniformly distributed in 3-D space and could be improved using uniform random rotation matrices [64] to avoid biases. More research is also needed to address the problem of locally optimizing molecular clusters, particularly for models where analytical gradients may not be available. Our preliminary test with CMA-ES indicated that it is an effective at finding isomers but further research is necessary to improve its efficiency.

From the perspective of the physical chemist, landscape analysis is not only useful for understanding algorithm performance but also has the potential to significantly impact the scientific study of molecular systems. We believe the quantitative measurement and study of landscape properties is a move toward a more robust methodology for validating and improving models. Similar to TTM2.1-F and OSS2, the landscapes of other potential energy models may also vary substantially, despite similar best or global minima. The insights gained from similar landscape analysis could provide scientists with a better understanding of the underlying potential energy landscapes and aid future model creation, refinement, and calibration. One particular extension of our landscape analysis which we are investigating is an analysis of the Hessians of the discovered isomers to extract the equilibrium properties of water clusters [65] and [66].

Acknowledgment

The authors would like to thank A. Mak for his helpful comments on the manuscript and P. Hiew for his technical support on the high-performance systems used in this paper.

References

[1] S. S. Xantheas, "Interaction potentials for water from accurate cluster calculations," in *Intermolecular Forces and Clusters II* (Structure and Bonding Series 116), Berlin: Springer-Verlag, 2005, pp. 119–148.

[2] T. James, D. Wales, and J. Hernandez-Rojas, "Global minima for water clusters $(H_2O)_n$, $n \leq 21$ described by a five-site empirical potential," *Chem. Phys. Lett.*, vol. 415, nos. 4–6, pp. 302–307, Nov. 2005.

[3] M. Mella, J.-L. Kuo, D. C. Clary, and M. L. Klein, "Nuclear quantum effects on the structure and energetics of $(H_2O)_6H^+$," *Phys. Chem. Chem. Phys.*, vol. 7, no. 11, pp. 2324–2332, 2005.

[4] G. Robinson, "Theoretical description of water," in *Water in Biology, Chemistry and Physics: Experimental Overviews and Computational Methodologies*, vol. 9, Singapore: World Scientific, 1996, ch. 5, pp. 129–175.

[5] C. Lee, H. Chen, and G. Fitzgerald, "Chemical bonding in water clusters," *J. Chem. Phys.*, vol. 102, no. 3, pp. 1266–1269, 1995.

[6] D. Wales and M. Hodges, "Global minima of water clusters $(H_2O)_n$, $n \leq 21$, described by an empirical potential," *Chem. Phys. Lett.*, vol. 286, nos. 1–2, pp. 65–72, Apr. 1998.

[7] B. Bandow and B. Hartke, "Larger water clusters with edges and corners on their way to ice: Structural trends elucidated with an improved parallel evolutionary algorithm," *J. Phys. Chem. A*, vol. 110, no. 17, pp. 5809–5822, 2006.

[8] C. L. Brooks, N. J. Onuchic, and J. D. Wales, "Taking a walk on a landscape," *Science*, vol. 293, no. 5530, pp. 612–613, Jul. 2001.

[9] D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*, Cambridge, U.K.: Cambridge Univ. Press, 2004.

[10] C. J. Tsai and K. D. Jordan, "Theoretical study of small water clusters: Low-energy fused cubic structures for $(H_2O)_n$, n = 8, 12, 16, and 20," *J. Phys. Chem.*, vol. 97, no. 20, pp. 5208–5210, May 1993.

[11] F. H. Stillinger, "Exponential multiplicity of inherent structures," *Phys. Rev. E*, vol. 59, no. 1, pp. 48–51, Jan. 1999.

[12] Q. C. Nguyen, Y.-S. Ong, H. Soh, and J.-L. Kuo, "A multi-scale approach to explore the potential energy surface of water clusters $(H_2O)_n$ $n \leq 8$," *J. Phys. Chem. A*, vol. 112, no. 28, pp. 6257–6261, Jul. 2008.

[13] P. Ballester and W. Richards, "Ultrafast shape recognition to search compound databases for similar molecular shapes." *J. Comput. Chem.*, vol. 28, no. 10, pp. 1711–1723, Jul. 2007.

[14] S. Wright, "The roles of mutation, inbreeding, crossbreeding, and selection in evolution," in *Proc. 6th Int. Congr. Genet.*, 1932, pp. 355–366.

[15] M. Miyazaki, A. Fujii, T. Ebata, and N. Mikami, "Infrared spectroscopic evidence for protonated water clusters forming nanoscale cages," *Science*, vol. 304, no. 5674, pp. 1134–1137, May 2004.

[16] J.-W. Shin, N. I. Hammer, E. G. Diken, M. A. Johnson, R. S. Walters, T. D. Jaeger, M. A. Duncan, R. A. Christie, and K. D. Jordan, "Infrared signature of structures associated with the $H^+(H_2O)_n$ (n = 6 to 27) clusters," *Science*, vol. 304, no. 5674, pp. 1137–1140, May 2004.

[17] C.-K. Lin, C.-C. Wu, Y.-S. Wang, Y. T. Lee, H.-C. Chang, J.-L. Kuo, and M. L. Klein, "Vibrational predissociation spectra and hydrogen-bond topologies of $H^+(H_2O)_{9-11}$," *Phys. Chem. Chem. Phys.*, vol. 7, no. 5, pp. 938–944, 2005.

[18] A. Wallqvist and R. D. Mountain, "Molecular models of water: Derivation and description," in *Reviews in Computational Chemistry*. New York: Wiley, 2007, pp. 183–247.

[19] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, "Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew," *J. Chem. Phys.*, vol. 120, no. 20, pp. 9665–9678, May 2004.

[20] S. Liem, P. Popelier, and M. Leslie, "Simulation of liquid water using a high-rank quantum topological electrostatic potential," *Int. J. Quant. Chem.*, vol. 99, no. 5, pp. 685–694, Jan. 2004.

[21] M. Mahoney, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *J. Chem. Phys.*, vol. 112, no. 20, pp. 8910–8922, May 2000.

[22] S. Rick, "A reoptimization of the five-site water potential (TIP5P) for use with Ewald sums," *J. Chem. Phys.*, vol. 120, no. 13, pp. 6085–6093, Apr. 2004.

[23] G. Fanourgakis and S. Xantheas, "The flexible, polarizable, Thole-type interaction potential for water (TTM2-F) revisited," *J. Chem. Phys. A*, vol. 110, no. 11, pp. 4100–4106, Mar. 2006.

[24] L. Ojamae, "Potential models for simulations of the solvated proton in water," *J. Chem. Phys.*, vol. 109, no. 13, pp. 5547–5564, Oct. 1998.

[25] C. Burnham and S. Xantheas, "Development of transferable interaction models for water. IV. A flexible, all-atom polarizable potential (TTM2-F) based on geometry dependent charges derived from an ab initio monomer dipole moment surface," *J. Chem. Phys.*, vol. 116, no. 12, pp. 5115–5124, Mar. 2002.

[26] F. Guimarães, J. Belchior, R. Johnston, and C. Roberts, "Global optimization analysis of water clusters $(H_2O)_n$ ($11 \leq n \leq 13$) through a genetic evolutionary approach," *J. Chem. Phys.*, vol. 116, no. 19, pp. 8327–8333, May 2002.

[27] B. Masek, A. Merchant, and J. Matthew, "Molecular shape comparison of angiotensin II receptor antagonists," *J. Med. Chem.*, vol. 36, no. 9, pp. 1230–1238, Apr. 1993.

[28] A. Y. Meyer and W. G. Richards, "Similarity of molecular shape," *J. Computer-Aided Mol. Des.*, vol. 5, no. 5, pp. 427–439, Oct. 1991.

[29] J. Grant and B. Pickup, "A Gaussian description of molecular shape," *J. Phys. Chem.*, vol. 99, no. 11, pp. 3503–3510, Mar. 1995.

[30] A. Nicholls, N. E. MacCuish, and J. D. MacCuish, "Variable selection and model validation of 2-D and 3-D molecular descriptors," *J. Computer-Aided Mol. Des.*, vol. 18, nos. 7–9, pp. 451–474, Jul.–Sep. 2004.

[31] J. Ochterski, "Vibrational analysis in Gaussian," Gaussian, Inc., Wallingford, CT, Tech. Rep., 1999 [Online]. Available: http://www.gaussian.com/g_whitepap/vib.htm

[32] D. Lim, Y.-S. Ong, Y. Jin, B. Sendhoff, and B. S. Lee, "Efficient hierarchical parallel genetic algorithms using grid computing," *Future Generation Comput. Syst.*, vol. 23, no. 4, pp. 658–670, May 2007.

[33] N. Noman and H. Iba, "Accelerating differential evolution using an adaptive local search," *IEEE Trans. Evol. Comput.*, vol. 12, no. 1, pp. 107–125, Feb. 2008.

[34] Q. H. Nguyen, Y.-S. Ong, M. H. Lim, and N. Krasnogor, "Adaptive cellular memetic algorithms," *Evol. Comput.*, vol. 17, no. 2, pp. 231–256, May 2009.

[35] M.-H. Lim, S. Gustafson, N. Krasnogor, and Y.-S. Ong, "Editorial to the first issue," *Memetic Comput.*, vol. 1, no. 1, pp. 1–2, Mar. 2009.

[36] F. Neri and V. Tirronen, "Scale factor local search in differential evolution," *Memetic Comput.*, vol. 1, no. 2, pp. 153–171, Feb. 2009.

[37] X. Yu, K. Tang, T. Chen, and X. Yao, "Empirical analysis of evolutionary algorithms with immigrants schemes for dynamic optimization," *Memetic Comput.*, vol. 1, no. 1, pp. 3–24, Mar. 2009.

[38] Y.-S. Ong and A. Keane, "Meta-Lamarckian learning in memetic algorithms," *IEEE Trans. Evol. Comput.*, vol. 8, no. 2, pp. 99–110, Apr. 2004.

[39] S. Hasan, R. Sarker, D. Essam, and D. Cornforth, "Memetic algorithms for solving job-shop scheduling problems," *Memetic Comput.*, vol. 1, no. 1, pp. 69–83, Mar. 2009.

[40] Z. Zhu, Y. Ong, and M. Dash, "Wrapper–filter feature selection algorithm using a memetic framework," *IEEE Trans. Syst. Man Cybern. B: Cybern.*, vol. 37, no. 1, pp. 70–76, Feb. 2007.

[41] Z. Zhu, Y.-S. Ong, and J. M. Zurada, "Identification of full and partial class relevant genes," in *Proc. IEEE/Assoc. Comput. Mach. Trans. Comput. Biol. Bioinf.*, vol. 99, no. 1. 2009.

[42] C. Aranha and H. Iba, "The memetic tree-based genetic algorithm and its application to portfolio optimization," *Memetic Comput.*, vol. 1, no. 2, pp. 139–151, Apr. 2009.

[43] T. Fischer, K. Bauer, P. Merz, and K. Bauer, "Solving the routing and wavelength assignment problem with a multilevel distributed memetic algorithm," *Memetic Comput.*, vol. 1, no. 2, pp. 101–123, Feb. 2009.

[44] D. Wales and J. Doye, "Global optimization by basin-hopping and the lowest energy structures of Lennard–Jones clusters containing up to 110 Atoms," *J. Phys. Chem. A*, vol. 101, no. 28, pp. 5111–5116, Jul. 1997.

[45] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science, New Ser.*, vol. 220, no. 4598, pp. 671–680, May 1983.

[46] H. H. Hoos and T. Stützle, "SLS methods," in *Stochastic Local Search: Foundations and Applications*. San Mateo, CA: Morgan Kaufmann, 2004, ch. 2, pp. 61–111.

[47] N. Metropolis, A. W. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, Jun. 1953.

[48] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Minimization or maximization of functions," in *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2007, ch. 10, pp. 398–448.

[49] D. Whitley, "The GENITOR algorithm and selection pressure: Why rank-based allocation of reproductive trials is best," in *Proc. 3rd Int. Conf. Genet. Algorithms*, 1989, pp. 116–121.

[50] D. E. Goldberg, "Advanced operators and techniques in genetic search," in *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston, MA: Addison-Wesley Longman, 1989, ch. 5, pp. 147–217.

[51] T. Bäck, D. Fogel, and Z. Michalewicz, *Handbook of Evolutionary Computation*. New York: Taylor & Francis, 1997.

[52] L. Juan, C. Zixing, and L. Jianqin, "Premature convergence in genetic algorithm: Analysis and prevention based on chaos operator," in *Proc. 3rd World Congr. Intell. Control Automat.*, 2000, pp. 495–499.

[53] K. A. DeJong, “An analysis of the behavior of a class of genetic systems,” Ph.D. dissertation, Dept. Comput. Commun. Sci., Univ. Michigan, Ann Arbor, 1975.

[54] S. W. Mahfoud, “Niching methods for genetic algorithms,” Ph.D. dissertation, Dept. Comput. Sci., Univ. Illinois at Urbana-Champaign, Urbana, 1995.

[55] Y. Leung, Y. Gao, and Z. Xu, “Degree of population diversity: A perspective on premature convergence in genetic algorithms and its Markov chain analysis,” *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 1165–1176, Sep. 1997.

[56] S. Ronald, “Duplicate genotypes in a genetic algorithm,” in *Proc. IEEE World Congr. Comput. Intel.*, May 1998, pp. 793–798.

[57] W. Humphrey, A. Dalke, and K. Schulten, “VMD–Visual molecular dynamics,” *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, Feb.1996.

[58] T. Jones and S. Forrest, “Fitness distance correlation as a measure of problem difficulty for genetic algorithms,” in *Proc. 6th Int. Conf. Genet. Algorithms*, 1995, pp. 184–192.

[59] N. Hansen and A. Ostermeier, “Completely derandomized self-adaptation in evolution strategies,” *Evol. Comput.*, vol. 9, no. 2, pp. 159–195, Jun. 2001.

[60] H. Soh and M. Kirley, “moPGA: Toward a new generation of multiobjective genetic algorithms,” in *Proc. IEEE Congr. Evol. Comput.*, Vancouver, BC, 2006, pp. 1702–1709.

[61] N. Hansen, “The CMA evolution strategy: A comparing review,” in *Toward a New Evolutionary Computation* (Advances in Estimation of Distribution Algorithms Series). J. Lozano, P. Larranga, I. Inza, and E. Bengoetxea, Eds., New York: Springer-Verlag, 2006, pp. 75–102.

[62] N. Hansen. (2008). *CMA Evolution Strategy Source Code* [Online]. Available: http://www.bionik.tu-berlin.de/user/niko/

[63] A. Auger and N. Hansen, “A restart CMA evolution strategy with increasing population size,” in *Proc. IEEE Congr. Evol. Comput.*, vol. 2. Sep. 2–5, 2005, pp. 1769–1776.

[64] J. Arvo, “Fast random rotation matrices,” in *Graphic Gems III*, (Academic Press Graphic Gems Series). San Francisco, CA: Academic, 1992, pp. 117–120.

[65] T. V. Bogdan, D. J. Wales, and F. Calvo, “Equilibrium thermodynamics from basin-sampling,” *J. Phys. Chem.*, vol. 124, no. 4, p. 044102, Jan. 2006.

[66] F. Calvo, J. P. K. Doye, and D. J. Wales, “Equilibrium properties of clusters in the harmonic superposition approximation,” *Chem. Phys. Lett.*, vol. 366, nos. 1–2, pp. 176–183, Nov. 2002.

Harold Soh received the B.S. degree with majors in computer science and economics in 2004 from the University of California, Davis, where he was a Regents Scholar. He received the M.S. degree in software systems engineering from Melbourne University, Melbourne, Australia, in 2005. He is currently a Khazanah Global Scholar pursuing the Ph.D. degree from the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K.

He was with the Institute of High Performance Computing, the Agency for Science, Technology and Research, Singapore, where he worked on high-performance evolutionary algorithms and infectious disease spread on complex networks. His interests include machine learning, human-assistive robotics, and evolutionary computation.

Yew-Soon Ong received the B.S. and M.S. degrees in electrical and electronics engineering from Nanyang Technological University, Singapore, in 1998 and 1999, respectively. He received the Ph.D. degree in artificial intelligence in complex design from the Computational Engineering and Design Center, University of Southampton, Southampton, U.K., in 2002.

He is currently an Associate Professor and Director of the Center for Computational Intelligence at the School of Computer Engineering, Nanyang Technological University. His research interest in computational intelligence spans across memetic computing, evolutionary design, optinformatics, and grid computing.

Dr. Ong is the Technical Editor-in-Chief of the *Memetic Computing Journal*, the Chief Editor of a book series on studies in adaptation, learning, and optimization, the Associate Editor of the IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS—PART B, the *International Journal of System Science*, and the *Soft Computing Journal*. He is also Chair of the Task Force on Memetic Computing in the IEEE Computational Intelligence Society Emergent Technology Technical Committee, and he has served as a Guest Editor of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, the IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS—PART B, the *Journal of Genetic Programming Evolvable Machine*, as well as the *Soft Computing Journal*.

Quoc Chinh Nguyen received the B.S. degree in physics and applied physics from Vietnam National University, Hanoi, Vietnam, in 2006. Since 2006, he has been pursuing the Ph.D. degree at the School of Mathematical and Physical Sciences, Nanyang Technological University, Singapore.

His research interests include computational chemistry and molecular modeling via first-principles methods.

Quang Huy Nguyen received the B.Eng. (honors) degree from the School of Computer Engineering (SCE), Nanyang Technological University (NTU), Singapore, in 2005. He is currently pursuing the Ph.D. degree in memetic algorithms.

He is currently with the center for Computational Intelligence, SCE, NTU.

Mohamed Salahuddin Habibullah received the B.S. degree in mechanical engineering (first class honors) from the University of Leicester, Leicester, U.K., in 1999, and the Ph.D. degree in the development of computational methods for structures subjected to cyclic loading from the Mechanics of Materials Center, Department of Engineering, University of Leicester, in 2004.

Currently, he is a Senior Research Engineer with the Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research, Singapore. At IHPC, he works on many research and development projects in diverse computational engineering fields. His current research interests include the areas of optimization, safety and reliability, and system-level integration.

Terence Hung received the B.S. degree in computer engineering, M.S. and Ph.D. degrees in electrical engineering from the University of Illinois, Urbana-Champaign, in 1985, 1991, and 1993, respectively.

He currently holds the position of Program Manager at the Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore, and of Associate Professor at Nanyang Technological University, Singapore. His research interests include high-performance, grid, and cloud computing.

Dr. Hung is a Council Member of the Gerson Lehrman Group, New York, an Asia Pacific Director on the Hewlett-Packard Consortium for Advanced Scientific and Technical Board, and a Member of the Editorial Board for the Institute of Advanced Scientific Research. He also serves on the National Grid Advisory Council of Singapore. He has participated actively as Principal Investigator/Co-Principal Investigator for seven grant projects.

Jer-Lai Kuo was born in Quemoy, Republic of China. He received the B.S. and M.S. degrees in physics from National Taiwan University, Taipei, Taiwan, in 1995 and 1997, respectively. He received the Ph.D. degree from the Chemical Physics Program, Ohio State University, Columbus, where he completed a thesis on the development and application of graph invariant theory while studying H-boning systems, in 2002.

He is currently an Associate Research Fellow at the Institute of Atomic and Molecular Science, Academia Sinica, Taipei, Taiwan. His research interests include water and related problems.