



Cite this: *Phys. Chem. Chem. Phys.*,
2015, 17, 11958

Received 1st April 2015,
Accepted 15th April 2015

DOI: 10.1039/c5cp01910a

www.rsc.org/pccp

Observable-targeting global cluster structure optimization†

Johannes M. Dieterich‡ and Bernd Hartke*

Traditionally, global cluster structure optimization is done by minimizing energy. As an alternative, we propose minimizing the difference between actual experimental observables and their simulated counterparts. To validate and explain this approach, test cases for small clusters are shown. Additionally, an application to real-life data for a larger cluster illustrates the advantages of this method: it provides direct links between properties and structure, and avoids problems both with insufficient accuracy in theoretical energy-ordering and with non-equilibrium conditions in experiment.

The experimental and theoretical studies of clusters and their properties have enjoyed a fruitful relationship.¹ One of the fundamental recurring questions concerns the geometric structure of the studied cluster as it influences most physical and chemical properties. Finding the globally minimal geometry on the electronic or free energy hypersurface is in itself an extremely challenging undertaking.² Using these theoretical results for analysis of experimental data adds significant additional complexity. First, the experimental observable is typically not structural data but physical properties, such as spectral data or ionization potentials. Second, while there is the thermodynamical argument that the cluster structure of globally minimal free energy is of importance under equilibrated experimental conditions, experimental setups may produce kinetically trapped cluster structures. This in turn makes theoretical results relying on equilibrium energy ordering problematic. In this contribution, we propose a different approach that solves the problem of kinetically trapped clusters and links theoretical results directly with the experimental observable.

We will begin by discussing how an energy-targeting global cluster structure optimization can be transformed into an observable-targeting one, within the framework of a genetic

algorithm implementation. Subsequently, we will give a detailed overview of the steps required for directly targeting the infrared (IR) spectrum as one possible observable. We will continue by demonstrating that our novel observable-targeting algorithm is capable to both reproduce solutions of benchmark problems and to directly analyze experimental data in cases where the energy-targeting approach is problematic. Finally, a brief summary and a discussion of advantages and disadvantages of this optimization strategy will conclude this contribution.

Within this contribution, we will use the theoretical framework of genetic algorithms as a global optimization strategy. However, there is no fundamental reason why other global optimization protocols could not be redesigned to implement the algorithmic changes laid out. An overview of global optimization strategies as commonly employed in cluster structure optimization can be found in ref. 3 and 4.

Genetic algorithms (GAs) belong to the broad class of non-deterministic global optimization algorithms.⁵ For historical reasons, they borrow most of their terminology from natural evolution. Every solution candidate to a given optimization problem is encoded into a genome. The candidate possesses a fitness, a metric expressing the quality of this solution, in relation to the objective function. If the best solution to the problem is known, the fitness has the character of a difference, otherwise it is unbounded, with lower fitness values by definition representing better solutions. A population of different solution candidates is subjected to a repeating process of genetic selection, crossover, mutation, fitness evaluation and substitution. These iterations will improve both the best fitness in, and the average fitness of, the population. For reviews of GA implementations and their application to global cluster structure optimization, we refer to ref. 3 and 4. Details of our OGOLEM GA package used here, featuring an abstract, object-oriented concept and a highly scalable pool concept,⁶ can be found in ref. 7–9.

In the well established case of targeting the electronic energy, the fitness is readily obtained as (a function of) the energy after a local optimization of the candidate structure. To target an observable instead requires replacement of the energy-based fitness by a metric for the difference to a reference observable.

Institute for Physical Chemistry, Christian-Albrechts-University, 24098 Kiel, Germany. E-mail: hartke@pctc.uni-kiel.de; Tel: +49-431-8802753

† Electronic supplementary information (ESI) available: Technical details of the calculations, comparison of MD-based and Hessian-based simulated IR spectra for (H₂O)₆, example optimization input, and cluster structure coordinates. See DOI: 10.1039/c5cp01910a

‡ Present address: Scientific Computing & Modelling, Amsterdam, The Netherlands.



Hence, the algorithm must now evaluate the observable for the candidate solution first. This obviously increases the cost of each step in the optimization by the cost of the observable evaluation. Therefore, it may be worthwhile to filter by the electronic energy of the cluster if the subsequent property evaluation is computationally expensive and high-energy minima can be ruled out as valid solutions. Other filters are possible, depending on the physical problem at hand. Additionally, changing the objective function may alter the characteristics and the difficulty of the search space. We will investigate this fundamental question in future applications.

As an example for an observable-targeted cluster structure optimization, we will discuss how to target an experimental IR spectrum. We wish to put emphasis on this property-computing algorithm being just one possibility for an observable-targeting optimization. Following the local optimization of the cluster into a near energy minimum, a micro-canonical molecular dynamics propagation is carried out. Randomized, Gaussian-distributed initial velocities are assigned to all atoms to mimic the desired temperature (corresponding to the effective temperature in the experimental setup). As explicitly investigated by Thomas *et al.*,¹⁰ only short propagation times of a few picoseconds are needed to obtain reasonable spectra. Compared to the standard normal-mode analysis, the advantage of this MD approach is that fully anharmonic, coupled spectra¹⁰ are obtained. During the propagation, the total dipole moment of the cluster is calculated and recorded at each MD step. The resulting trajectory of dipole moments is autocorrelated. To avoid artificial, additional peaks (“ringing”, Gibbs phenomenon) at the ensuing Fourier transformation, induced by the discontinuity of the autocorrelation function at the end of the propagation interval, the autocorrelation function is folded with a Gaussian filter function. This is a standard procedure in signal processing, and also well established for the present purpose of obtaining spectra from autocorrelation functions.¹¹ The spectrum is obtained through a subsequent Fourier-transformation of the folded autocorrelation function. Using the dipole autocorrelation function will yield only the IR spectrum of the cluster; the power spectrum could be obtained from the velocity autocorrelation function.¹⁰

At this stage, the fitness metric must compare the reference spectrum and the simulated spectrum, and summarize the differences between these two into a scalar, the fitness. The difference integral of the two discretized spectra in a defined wavenumber interval is a natural choice for this. Before computing the fitness from this difference integral, possible overall intensity differences between reference and simulated spectra are compensated *via* a shift to zero at user-defined points and a one-dimensional optimization of an overall intensity scaling factor for the simulated spectrum, with the coefficient being bound to smaller than 10.0.

It is important to note that no gradient with respect to the property is required at any stage of this simulation protocol. Only energy and nuclear gradients are needed to complete all steps. Also, no experimental knowledge needs to be used other than the target observable.

To demonstrate the above ideas and their implementation, we have selected pure neutral water clusters as an application

example, modeled with the TTM3-F potential.¹² With varying potential models, these clusters enjoy the status of a benchmark system in the theoretical cluster structure optimization community.^{6,13–18} Size-selective experiments on pure neutral water clusters $(\text{H}_2\text{O})_n$ were limited to $n \leq 12$ ^{19–22} for a long time, but recent developments^{23–25} may have pushed this upper limit to $n = 70$ or even higher. The directly observed property in the latter experiments is the OH-stretch IR spectrum in the so-called “fingerprint” region ($2800\text{--}3800\text{ cm}^{-1}$). Interpretation of these experiments currently is somewhat speculative,²⁵ and links between theory and experiment need to be strengthened.^{18,23,24} Also, the currently practical upper limit of energy-based global structure optimization is unclear.^{6,18} Thus, new theoretical developments in this area are needed.

A basic requirement for the IR-spectra-targeted global cluster structure optimization advertised here is that it can differentiate between different structural isomers. This is indeed possible, as shown in Fig. 1 for $(\text{H}_2\text{O})_6$. This is the smallest water cluster exhibiting several qualitatively different, low-energy, three-dimensional structural isomers,²⁶ among others the “cage” and the “prism”. The IR-spectra of the cage and of the prism isomers are rather different in the fingerprint region. Therefore, we should expect that our optimization strategy should find the cage when a cage-spectrum is provided as reference and the prism when a prism-spectrum is the reference, respectively. This is exactly what happens. Since this is a proof-of-principle demonstration, the reference spectra were not generated experimentally but rather with the same MD-based algorithm described above. Nevertheless, the reference spectra and the spectra of the optimized clusters differ slightly. This is due to the dependence of IR intensities on the MD-trajectory length, which is present for short MD-trajectories but disappears for longer trajectories.¹⁰ Of course, longer MD-trajectories take more computing time, for their generation but also for the ensuing steps (autocorrelation, folding, Fourier transform). In a global optimization setting, one wants to keep this computational effort limited.

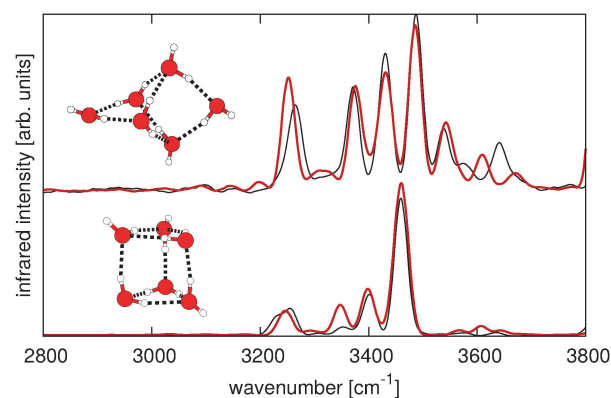


Fig. 1 Feasibility demonstration of IR-spectra-targeted cluster structure optimization for $(\text{H}_2\text{O})_6$: Comparison of IR spectra of the given reference cluster structure (black lines) and of the best cluster structure from global optimization (red lines), started from random seeds. (top) Cage (shifted upwards by an arbitrary amount, for clarity), (bottom) prism. The corresponding cluster structures are shown as insets.



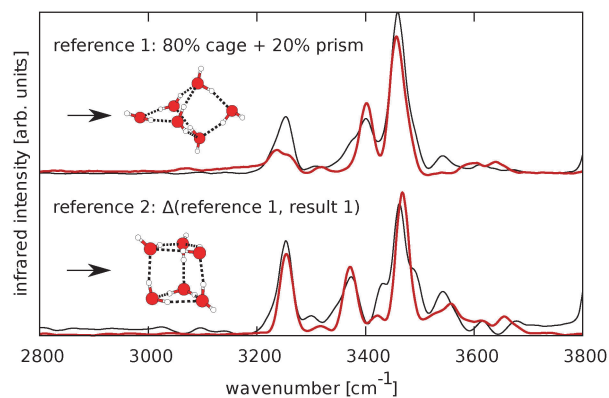


Fig. 2 Illustration of a successive N -stage algorithm for disentangling mixed IR-spectra arising from N different isomers, here for $N = 2$. (top) In a first stage, an 80 : 20 mixture of cage- and prism-spectra is the reference, and the cage structure is found. (bottom) In a second stage, the first-stage cage spectrum is subtracted from the reference spectrum, and then the remaining prism contribution is detected.

Hence, we have deliberately taken MD-trajectories that are significantly shorter than advocated in ref. 10, to demonstrate that this is not a severe problem.

Depending on the experimental procedure and the system size, experimental spectra may arise from more than one cluster isomer, with different abundances. Therefore, we have set up a similar situation for our $(\text{H}_2\text{O})_6$ test case, too: We have generated a mixed IR spectrum as reference, consisting of 80% cage and 20% prism. For this reference, our global structure optimization finds the cage structure as best fit. We have then subtracted the IR spectrum corresponding to this optimized cage, generating a new reference spectrum as the remainder. A second global structure optimization run then finds the prism (*cf.* Fig. 2). Obviously, this sequential procedure can be extended towards more than two mixture components.

Although this contribution is only meant as a proof-of-principle demonstration of the basic idea, we already provide first outlooks to real-life applications. For this, we utilize experimental IR spectra data from ref. 23 for $(\text{H}_2\text{O})_{25}$. To emulate the best possible traditional approach of simulating this spectrum, we have collected the 500 lowest-energy structures from all our energy-targeted global optimization runs with the TTM3-F potential for $(\text{H}_2\text{O})_{25}$. For all of these, we have simulated the IR spectra using the MD-based autocorrelation approach sketched above. Assuming that the experimental spectrum can be approximated by a linear combination of these 500 spectra but that the search space for the best linear combination of this type may well be multimodal and non-convex, we have employed the global optimization power of OGOLEM to find the best such linear combination. Fig. 3 depicts this best superposition, in direct comparison to the experimental spectrum. Despite using excessive settings in this optimization (*e.g.*, 42 million steps), agreement with the experimental spectrum is far from perfect. One would expect that a fairly simple, one-dimensional function should be representable far better with 500 basis functions, in particular if many basis functions contribute (37 with a coefficient above 0.1). Therefore, barring

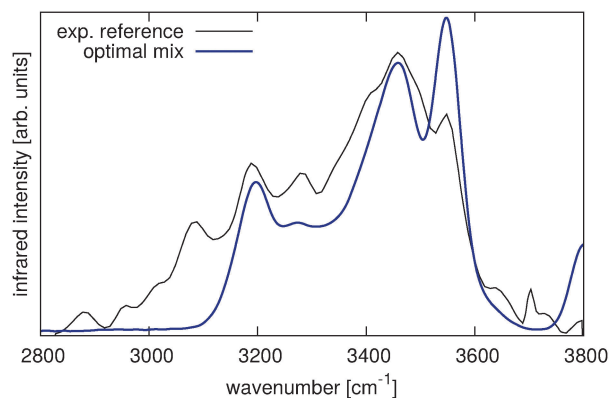


Fig. 3 Approximation of the experimental $(\text{H}_2\text{O})_{25}$ IR spectrum (black line) by a globally optimized superposition (blue line) of the simulated IR spectra of the 500 lowest-energy $(\text{H}_2\text{O})_{25}$ structures.

severe defects in the experiment and/or in the TTM3-F potential, the conclusion is that this 500-function basis is incomplete; decisive features are missing from it.

This is in stark contrast to what happens when we employ our new property-based global optimization, *i.e.*, when we globally optimize $(\text{H}_2\text{O})_{25}$ cluster structures for best fit with this experimental spectrum. Our best result at the present stage is shown in Fig. 4, produced with three orders of magnitude fewer global optimization steps (24 thousand). The agreement between experiment and theory in Fig. 4 is not decisively better than in Fig. 3; in both cases, spectral features are missing (*e.g.*, between 3000 and 3100 cm^{-1}). However, while 37 different structures contribute visibly to the simulated spectrum in Fig. 3, there only is one single structure responsible in Fig. 4, and no structure similar to this one occurs in the set of 500 low-energy structures used for Fig. 3. Hence, our new approach produces valuable structural information hypotheses from the experimental IR spectrum, while the traditional approach offers little help for this task.

In this contribution, we have shown that the traditional, two-step recipe of first optimizing cluster structures globally for lowest energy and then simulating experimental observables

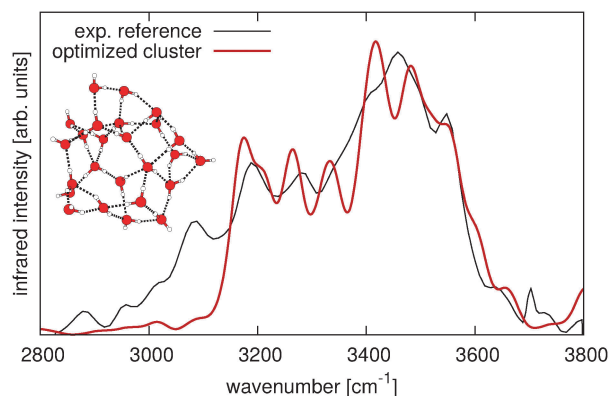


Fig. 4 With the same experimental $(\text{H}_2\text{O})_{25}$ IR spectrum (black line) as in Fig. 3 as reference, global cluster structure optimization yields a well-fitting simulated spectrum (red line) for the best structure (inset).



for these lowest-energy structures may be replaced by the more direct approach of optimizing cluster structures globally for smallest deviation between experimentally measured and theoretically simulated properties. We have shown this for the fairly complicated case of IR spectra as property, but clearly this recipe is far more general and can be used for any property or even collections of properties.

This new route has several advantages: The energy-based, two-step recipe requires a theoretical model that produces reliable energies and properties; the direct approach requires only a reliable property model. Furthermore, energy is a single number, frequently plagued by erratic error compensation effects. Using several properties instead (or an array or function property, as in the case of spectra) is inherently more reliable and more informative.

Last but not least, in many cluster experiments, it is hard to prove that the clusters carry no memory of their preparation but are thermally relaxed; hence it is convenient that the direct approach does not rely on this assumption. As possible disadvantage, one may argue that some of the physics is lost if this assumption is not made. However, if there is no thermal distribution in experiment, then this is the physics that needs to be captured. Our new approach can provide direct links between observed properties and underlying structures also in this case. Comparison to a thermal distribution of structures, and drawing physical conclusions from this comparison, can then be performed as a secondary step.

In any case, we are confident that property-targeting global cluster structure optimization is a useful addition to the theoretical toolbox, allowing for extended and more direct contact between theory and experiment in cluster research.

Acknowledgements

JMD is indebted to Scientific Computing & Modelling (SCM) for permission to carry out this work independently. Both authors would like to express their appreciation to Sascha Frick for tirelessly maintaining computational resources.

References

- 1 D. J. Wales, *Energy Landscapes*, Cambridge University Press, 2003.
- 2 L. T. Wille and J. Vennik, *J. Phys. A: Math. Gen.*, 1985, **18**, L419.
- 3 B. Hartke, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 879.
- 4 R. L. Johnston, *Dalton Trans.*, 2003, 4193.
- 5 T. Weise, *Global Optimization Algorithms – Theory and Application*, Self-published, 2009.
- 6 B. Bandow and B. Hartke, *J. Phys. Chem. A*, 2006, **110**, 5809.
- 7 J. M. Dieterich and B. Hartke, *Mol. Phys.*, 2010, **108**, 279.
- 8 J. M. Dieterich, U. Gerstel, J.-M. Schröder and B. Hartke, *J. Mol. Model.*, 2011, **17**, 3195.
- 9 J. M. Dieterich and B. Hartke, *J. Comput. Chem.*, 2011, **32**, 1377.
- 10 M. Thomas, M. Brehm, R. Fligg, P. Vöhringer and B. Kirchner, *Phys. Chem. Chem. Phys.*, 2013, **15**, 669.
- 11 J. Dai and J. Z. H. Zhang, *J. Chem. Phys.*, 1995, **103**, 1491.
- 12 G. S. Fanourgakis and S. S. Xantheas, *J. Phys. Chem.*, 2008, **128**, 074506.
- 13 D. J. Wales and M. P. Hodges, *Chem. Phys. Lett.*, 1998, **286**, 65.
- 14 B. Hartke, *Z. Phys. Chem.*, 2000, **214**, 1251.
- 15 B. Hartke, *Phys. Chem. Chem. Phys.*, 2003, **5**, 275.
- 16 A. Lagutschenkov, G. S. Fanourgakis, G. Niedner-Schatteburg and S. S. Xantheas, *J. Chem. Phys.*, 2005, **122**, 194310.
- 17 J. M. Dieterich and B. Hartke, *J. Comput. Chem.*, 2014, **35**, 1618.
- 18 S. Kazachenko and A. J. Thakkar, *J. Chem. Phys.*, 2013, **138**, 194302.
- 19 U. Buck, I. Ettischer, M. Melzer, V. Buch and J. Sadlej, *Phys. Rev. Lett.*, 1998, **80**, 2578.
- 20 J. Sadlej, V. Buch, J. K. Kaszimirski and U. Buck, *J. Phys. Chem. A*, 1999, **103**, 4933.
- 21 N. Pugliano and R. J. Saykally, *Science*, 1992, **257**, 1936.
- 22 K. Liu, M. G. Brown and R. J. Saykally, *J. Phys. Chem. A*, 1997, **101**, 8995.
- 23 U. Buck, C. C. Pradzynski, T. Zeuch, J. M. Dieterich and B. Hartke, *Phys. Chem. Chem. Phys.*, 2013, **16**, 6859.
- 24 C. C. Pradzynski, C. W. Dierking, F. Zurheide, R. M. Forck, U. Buck, T. Zeuch and S. S. Xantheas, *Phys. Chem. Chem. Phys.*, 2014, **16**, 26691.
- 25 F. Zurheide, C. W. Dierking, C. C. Pradzynski, R. M. Forck, F. Flüggen, U. Buck and T. Zeuch, *J. Phys. Chem. A*, 2015, **119**, 2709–2720.
- 26 B. Hartke, M. Schütz and H.-J. Werner, *Chem. Phys.*, 1998, **239**, 561.

