UNIVERSIDADE FEDERAL DA BAHIA

INSTITUTO DE FÍSICA

Programa de Pós-graduação em Física

Tese de Doutorado

ION–RARE-GAS MICROSOLVATION STUDIES BY EMPLOYING AN EVOLUTIONARY ALGORITHM

Wanderson Silva de Jesus

2019

UNIVERSIDADE FEDERAL DA BAHIA

INSTITUTO DE FÍSICA

PROGRAMA DE PÓS-GRADUAÇÃO

ION–RARE-GAS MICROSOLVATION STUDIES BY EMPLOYING AN EVOLUTIONARY ALGORITHM

Wanderson Silva de Jesus

Orientador: Prof. Dr. Frederico Vascocellos Prudente

Co-orientador: Prof. Dr. Jorge Manuel Campos Marques

Tese apresentada ao Instituto de Física da Universidade Federal da Bahia como parte dos requisitos para a obtenção do título de Doutor em Física.

Salvador-BA, 2019

ACKNOWLEDGEMENTS

I can not measure the gratitude I have for the people who have somehow contributed to the accomplishment of this work. Facing such deep this reflection I would like to thank:

God, for life.

for all discernment, motivation and focus;

To the Advisors, especially to Prof. Frederico Prudente and Prof. Jorge Manuel Campos Marques, who guided me on the road of knowledge, always seeking the best ways, advice and teaching.

To the collaborators, especially Prof. Francisco B. Pereira and Nuno Lourenço. Your contribution was extremely important at each stage of this research. Thank you.

My Family, who did not let me give up; fundamental support to complete this work.

My mother, Rita de Cássia Santos Silva.

My father Waldir Conceição Araújo de Jesus.

My brother Uálex Silva de Jesus.

My grandmother Noemia Bispo dos Santos.

To my friends: Luis Gomes, Arthur Daltro, Leo Passos, Eval, Manuela and Tansir. Thank you, friends, for all the moments of sadness and joy, for sharing with me every minute of anguish and happiness. The word of each one of you was essential in my walk.

I am also grateful for my friends of Physics, in special for Robenilson, Andreia and Mariana. Thanks for all the support and encouragement.

I have great pleasure in thanking you for all the experience you had in Portugal. And on this journey I met people who were a gift from God: Luís Filipe Ribeiro Claro, Bruno, Bruna, Heitor and Raquel.

I am also grateful for all the financial support given by Brazilian CAPES for the doctoral sandwich fellowship to the Coimbra University and I also thank the financial and

scientific support given by CAPES and UFBA for this work.

ABSTRACT

The microsolvation has been subject of intense theoretical and experimental study with implications in diverse areas such as atmospheric chemistry, biological processes, permeation and transport membrane, among others. Therefore, this thesis aimed to develop a study of the microsolvatation of ions through noble gases. One of our specific goals is to survey the microsolvatation of ion $Li^+$by argon, kripton and mixture of these atoms. For this construction, analyses involving the interaction construction of two and three-bodies metal-alkaline ions with noble gases, from the fitting of potential functions ab initios on the eletronic energy using the methodology CCSD(T). The determination of stable clusters structures will be given by the application of the evolutionary algorithm (EA). The evolutionary algorithm will search of these low energy structures, both in global and local minima. In order to strengthen this study, pos-optimization calculations using the ab initios CCSD(T) and MP2 methods are also performed. Treatment of clusters with noble gases of higher numbers, a DFT methodology is also used in the stage of pos-optimization, with the goal to perform a Benchmark analysis of these systems. Lastly, from diverse local minima generated by evolutionary algorithm we applied a Machine Learning technique to enable the determination of choose rules of the better EA minima that present an efficient description of the energy landscape from the clusters length.

Keywords: Evolutionary Algorithm, Cluster Optimization, Microsolvation, DFT, Machine Learning.

RESUMO

A microsolvatação tem sido sujeito de intenso estudo teórico e experimental com implicações em diversas áreas como, por exemplo, química atmosférica, processos biológicos, membrana de permeação e transporte, dentre outros. Portanto, nesta Tese será desenvolvido um estudo da microsolvatação de íons por gases nobres. Especificamente, será estudada a microsolvatação do íon Li^+ por argônios, kriptônios e mistura desses átomos. Para tal construção serão desenvolvidas análises que envolvem a construção da interação de dois e três corpos de íons metais-alcalinos com gases nobres, a partir do ajuste de funções potenciais a partir de cálculos *ab initios* da energia eletrônica usando a metodologia CCSD(T). A determinação das estruturas estáveis dos clusters se dará pela aplicação de um algoritmo evolutivo (EA). O algoritmo evolutivo fará a busca dessas estruturas de baixas energias, tanto do mínimo global como dos mínimos locais. No intuito de fortalecer o estudo, cálculos de pós-otimização usando os métodos *ab initios* CCSD(T) e MP2 são também realizados. Para o tratamento de agregados com número de gases nobres maiores, uma metodologia DFT também é utilizada na etapa de pós-otimização, com o objetivo de realizar uma análise Benchmark desses sistemas. Por fim, a partir de diversos mínimos locais gerados pelo algoritmo evolutivo empregamos a técnica de Machine Learning para possibilitar a determinação de regras de escolha dos melhores mínimos EA que apresente uma descrição eficiente da paisagem energética dos diversos tamanhos de clusters.

Palavras-chave: Algoritmo Evolutivo, Otimização de Cluster, Microsolvatação, DFT, Machine Learning.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cluster analysis has provided an important opportunity for studies on the evolution of physical and chemical properties of solvated systems with solvent size increasing in volume [1]. This type is known as microsolvation; solvent molecules are succinctly added to the system. They present a high computational cost, but also provide a detailed knowledge of the phenomenon at the molecular level [2, 3, 4, 5]. The use of solvent microsolvation by atoms instead of molecules makes the complexity smaller (by adding atoms step by step) [6, 7, 8]. An example of this is the chromophore atoms in clusters of rare gases, which has had considerable attention, to be considered as a model system to know the solvation at the microscopic level [9].

Spectroscopy of elementary systems such as atoms or molecules trapped or interacting with clusters has been the subject of continuous investigations [1, 10, 11]. One of the areas of cluster study that presented a great relevance from the theoretical point of view, is the alkaline atoms in mixture with rare gases. One of the reasons is the fact that its electronic structure is easy to model [9].

The result of research into clusters of metal ions and noble gases can help understand many fundamental problems in the fields of atomic and molecular physics, organic and inorganic chemistry and biochemistry, such as nucleation phenomena and phase transitions. These interactions gained significant interest because of the aspects of noble gases related to

their closed valence shields that make them attractive as a model system [7].

It is important to highlight that there are also experimental studies that are used as reference for analysis of theoretical results. One of them is the mass spectroscopy; an experimental technique to detect the stability of clusters to from magic numbers [7]. An example is the ion cluster mass spectrum $(H_2)_nLi^+$ constructed from the Li^+ bombardment on the static target H_2.

One of the important fields of the clusters is the structures of low energies. A global optimization process, which can be directly through *ab initios* or DFT calculations, accomplishes the realization of these but the computational cost is high. An alternative of lower cost is the use of evolutionary algorithms (EA), which has been shown success in the search of low energy clusters [12].

From a theoretical point of view, mixed clusters have a greater structural complexity than homogeneous ones, so it is more challenging to optimize geometry and its importance in many contexts, resulting mainly in technological applications. Some case to be studied in this work is an aggregate of the heterogeneous type. These complexities have to do with the existence of isomers in the same geometry and structure, but they show different distributions in the two types of atoms, of which they are not present in homogeneous clusters.

The mixture of rare gas clusters has received theoretical and experimental attention, since they can be seen as prototype systems that comprise complex phenomena that occur at the mesoscale level and therefore with relevance in nanotechnology. Some experimental studies have been developed for the problem of the formation of heterogeneous clusters involving argon and krypton [13, 2, 14, 15].

Therefore, in this thesis we intend to develop a detailed study of the energetic and geometric properties of the structures of clusters of low energies of alkaline-metal ions with atoms of noble gases. The constructions of the clusters will be carried out by a global optimization generated by an evolutionary algorithm constructed by a portuguese researchers group [16]. This process depends on *ab initio* calculations to obtain the adjusted potential energy surfaces, which in this work take into account the cases with interaction between atoms to two and three bodies.

The achievement of global and local minima by EA, will allow understanding the landscape of low energies that encompass the structures. A complementary part in this work, a study of cluster optimizations of the Cu-Rh transition metals, and also of a new strategy of a "self-adaptive" algorithm that looks for the clusters of lower energy and at the same time develop structures that autonomously adapt to the addressed problem.

This work also aims at an analysis in two aspects; the first one refers to a more comprehensive study of the cluster size that becomes limited due to computational time, so the strategy to perform the re-optimization study by the DFT method and the second one refers to an innovative study because, it was not found in the literature studies related to energy calculation and mixture structure of Argon and Krypton clusters [17]. Hence, one of the motivations of this work, to seek to study something still unexplored, in this aspect.

The thesis is structured as follows. In Chapter 2, we review the theory of the methods: Pertubation of Mollet-Plesset, Coupled Cluster and Density Functional Theory beyocnd of the Evolutionary Algorithm. In Chapter 3 we discuss the study of alkali-ion microsolvatation with argon and for discovering global miminum structures of transition metal Rh-Cu binary cluster in a interation two-bodys. This chapter was inspired by the book chapter " Revealing energy landscapes of atomic clusters by applying adaptive bio-inspired algorithms" published in the Apple Academic Press with the respective authors: Jorge Manuel C. Marques, Wanderson S. de Jesus, Frederico V. Prudente, Francisco B. Pereira and Nuno Lourenço [18]. In Chapter 4 we establish the comparison between the structures global and local minimus of the Li^+Ar_n and Li^+Kr_n, by using potential energy new faces that including two and three-bodys. This chapter was inspired by the paper "Exploring the first-shell and second-shell structures arising in the microsolvation of Li^+ by rare gases" published in the International Journal of Quantum Chemistry with the respective authors: Wanderson S. Jesus, Jorge Manuel C. Marques, Frederico V. Prudente and Francisco B. Pereira [13]. In Chapter 5, we have perfomance a research innovative; the solvatation of an alkali-ion by a binary mixture solvent of rare gas atoms. Specifically, we lian the interest to know where we search discovery, what noble gas atom occupy the first solvatation shell. This chapter was inspired by the paper "Microsolvation of Li^+ in a mixture of argon and krypton: unveiling the most

stable structures of the clusters" published in the Journal Physical Chemistry A with the respective authors: Wanderson S. Jesus, Frederico V. Prudente and Jorge Manuel C. Marques [19]. In Chapter 6 we are performing a study with the goal of catch up the knowledge to big clusters. For this we use o theory DFT and the tecnhiques Machine Learning for the clusters $Li^{+}Kr_{n}$. This one was inspired by the paper that are in submission process "Modeling microsolvation clusters with electronic-structure calculations guided by analytical potentials and machine learning techniques" with the respective authors: Wanderson S. Jesus, Frederico V. Prudente, Jorge Manuel C. Marques, Francisco B. Pereira and Nuno Lourenço [20]. In Chaper 7 we present the conclusions of this thesis, as well as the perspectives of future work.

Chapter 2

Methodology

The purpose of this work is to demonstrate the development of techniques capable of constructing optimized clusters efficiently, with the functionality of obtaining information about geometric and energetic properties. To do so, we will outline the stages to be developed in this thesis.

In (i) we aim to obtain a function that describes the potential energy surface to be used in the evolutionary algorithm. According to figure 2.1, the cluster structures are initially unknown as shown at the top of the diagram, they are of the type: binary with an alkaline ion and varying the number of rare, tertiary gases with an alkaline ion and two rare gases varying the number of atoms and binary transition metals with the same amount of atoms in each cluster. Of these structures is taken into account the two and three-body interaction between the atoms. In view of this, ab initio calculations are performed using the CCSD(T) method by the package of the Gamess. The single-point results are used in the potential function for an adjustment of the same, through Powell or GAfit [21] . Finally, from one of these techniques we construct the potential energy surface to be used in the evolutionary algorithm.

Step (ii) aims to optimize the clusters structures as represented at the top of figure 2.1 and from the function that describes the potential energy surface as represented at the bottom of the same diagram. For this, in figure 2.2 the green box represents the techniques

used by the evolutionary algorithm to obtain optimized clusters. The algorithm generates the optimized structures of lower energy called global minimum and still higher energy, the local minimum. In addition, with this we extract information of structural and energetic properties.

With the structures we intend to obtain a more detailed study, so in (iii) a POS-OPTIMIZATION of the various minima (global and local) is carried out. According to figure 2.3 the post-optimization box represents a variety of information that it aims to obtain among them structural and energy properties to be compared with GA and an efficient choice of local and global minima (BEST MINIMA). For this purpose an MP2 optimization is performed and in sequence energy single points calculations by the CCSD(T) method to clusters with N = 5 atoms and MP2 with N = 8 atoms. Up to this step, we obtain information about the energy and structural properties of the clusters that are used as comparisons with EA.

It is noticeable from the diagram that there is a computational limitation to perform a CCSD(T)/MP2 study for larger clusters (N >8). That is why a path is necessary in order to overcome this limitation. Therefore, it is necessary to return the structure of and perform ab initios calculations by the DFT method since it demands lower computational cost. However, in this process it is necessary to obtain an efficient functional/base, so it is used as a parameter to choose the information of the method CCSD(T) "[5]" and MP2 "[8]" (See the Fig. 2.3). Thus, with the choice of the best functional/base, the DFT calculation is given by the global and local minima, which can now be for clusters larger than N = 8. Due to large number of minima, and aiming to reduce the computational cost of the post-optimization by the DFT method. It is necessary to use the Machine Learning method that will reduce the number of minima to be analyzed in detail and thus obtain the best local and global minima that can represent each size of clusters.

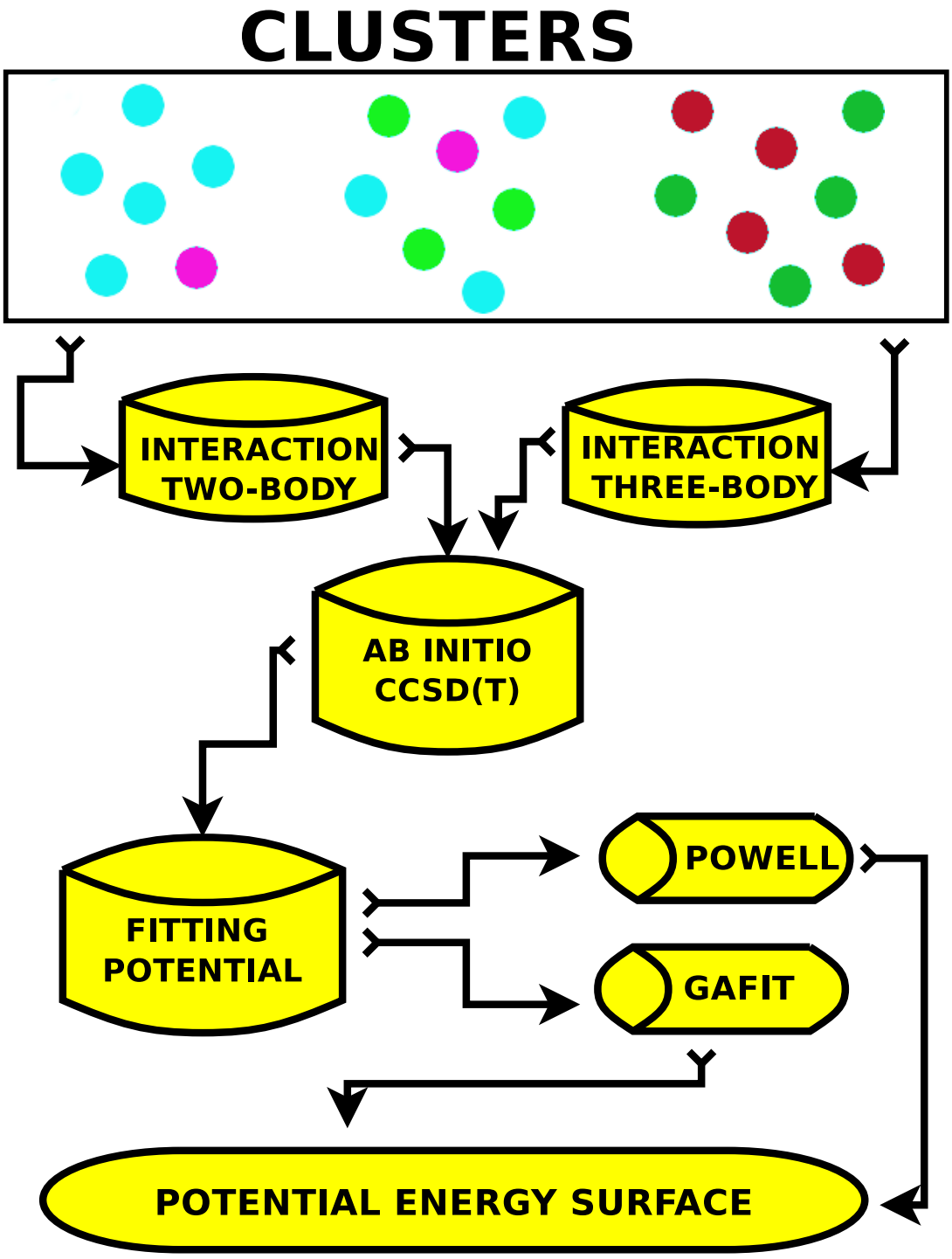


Figure 2.1: Diagram - Obtaining Surface Energy Potential

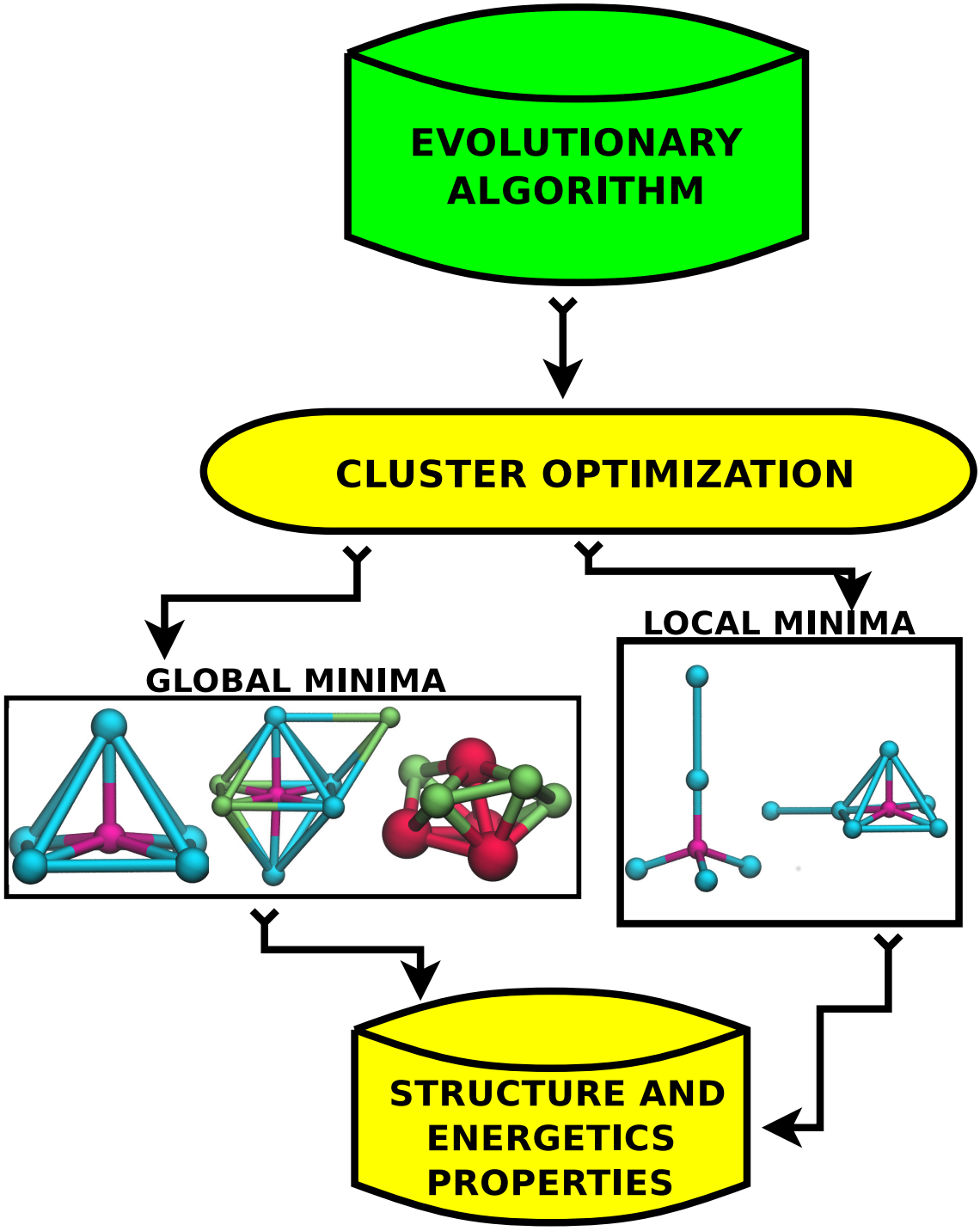


Figure 2.2: Diagram - Optimization cluster by Evolutionary Algorithm

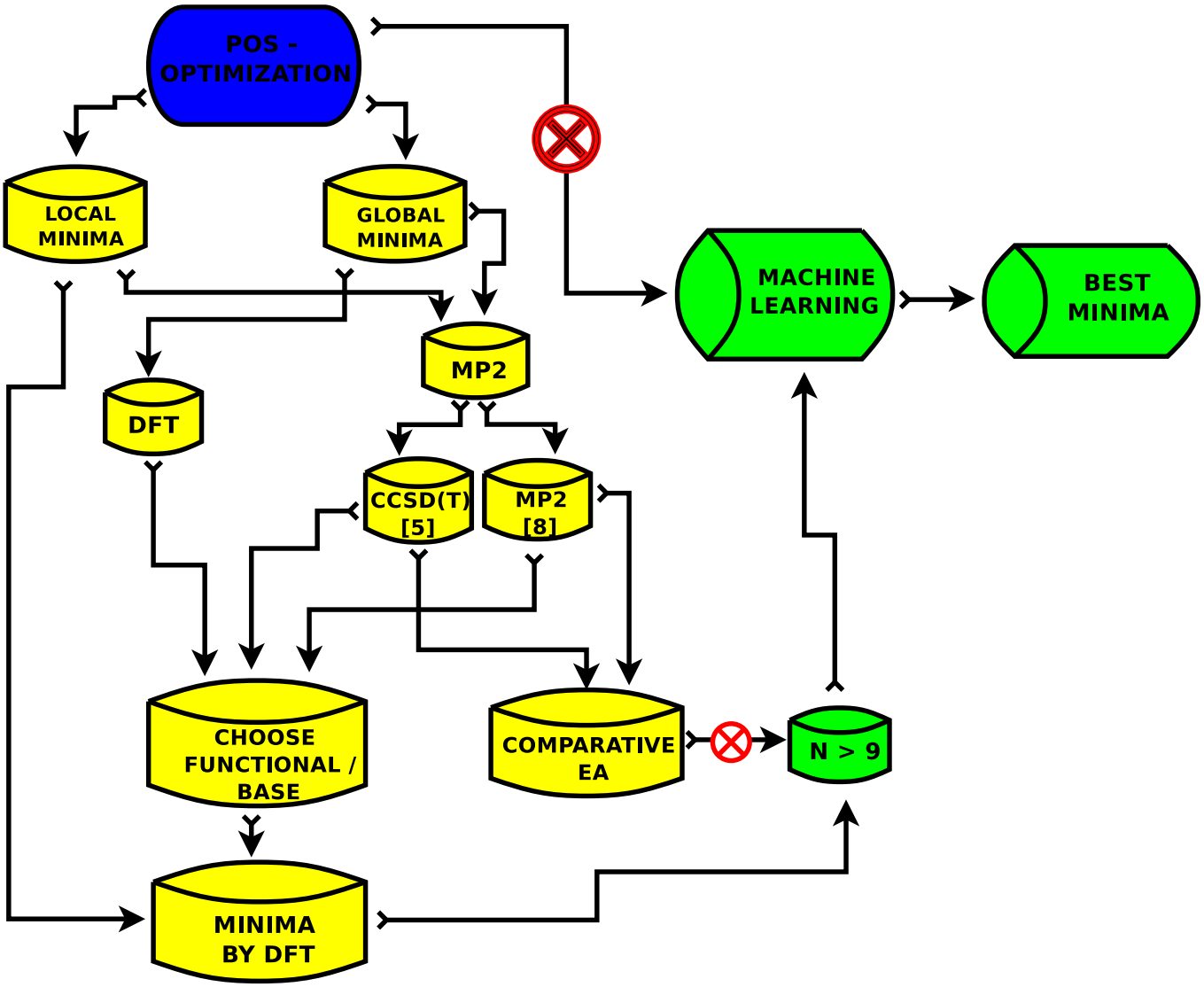


Figure 2.3: Diagram - Pos optmization by CCSD(T)/MP2 and by DFT

The diagrams give a brief summary of what is to be presented in detail in the course of the thesis. In Chapter 3 that corresponds to steps (i) and (ii), we will see that according to figure 2.1 the path to be demonstrated is the structure of clusters of alkaline-metal ions with Ar atoms; Li^+Ar, Na^+Ar e K^+Ar (at the top left of the figure), and Cu-Rh transition metals (at the top right of the figure) initially unknown, where only the two-bodies interaction path is considered of the potential energy surface. In addition, figure 2.2 is represented for this chapter the whole process of using the Evolutionary Algorithm with an additive of a test of an original self-adaptive algorithm, which seeks to develop structures that adapt autonomously to the addressed problem. The work will demonstrate that GA improves the set of genetic

operators and the distance measure that effectively maintains diversity by run optimization. In order to obtain the structural and energetic properties of the above-mentioned structures, a comparative analysis was performed between the three cluster types generated by GA with two bodies of an ion with Ar. Moreover, in the transition metals $(Rh\text{-}Cu)_n$, to verify the distribution tendency of the atoms in the optimized structures.

Chapter 4 consists of steps (i), (ii) and partially of (iii). It has as its proposal to present the researches of low energy structure for such, starting from alkaline metal structures Li^+Ar_n and Li^+Kr_n for n=2 - 14, as the top one on the left side of figure 2.1, the interaction paths of two and three atoms are considered to obtain the potential energy surface. In the next process indicated by figure 2.2, the EA finds structures of global and local minima. The generated clusters produce structural and energy information of the solvents Ar and Kr. The comparison in the formations of the clusters described in the course of the chapter indicate sizes of clusters that present difference between their structures. In addition, the analysis of the global and local minima by the EA, will allow understanding the landscape of low energy that encompasses the structures. In step (iii), the post-optimization process (figure 2.3) will be performed by MP2 optimization and energy single point calculations CCSD (T) and MP2, which results in structural and energy and that in this chapter ends in the comparison with the results of the EA in which the clusters optimized by the EA and by a re-optimization MP2, will guarantee a greater difference between the minima of energy re-optimized by MP2.

For chapter 5 steps (i) and (ii) are effective, with the clusters initially unknown as exemplified by the central geometry at the top of figure 2.1 in which a binary solvent $(Ar(_n\text{-}Kr_m)$, interacts with the Li^+ ion. One of the great curiosities in this research is to identify which noble gas tends to form the first layer of solvation of the cluster, this formation will provide the understanding of structural and energetic properties, on the way solvents surround the solute. EA generates geometries with all combinations of atoms of Ar and Kr between N=1-10 and for N= 12-20 only structures with atoms of Ar = Kr. The study of these clusters also seeks to present the energetic behavior of the structures when changing a Kr and Ar atom between the first and second layers of solvation.

Chapther 6 will represent the culmination of the three stages, shown by figures 2.1,

2.2 and 2.3. Since having a variety of results already presented in the previous chapters, this will come with the intention of carrying out studies of larger clusters of $Li^{+}Kr_{n}$, however as observed in the diagram of figure 2.3 in which there is a computational cost limitation of N=5 atoms for the CCSD (T) method and N=8 para MP2, for MP2, perform comparative studies with the EA. We follow the path of the DFT method that will enable the study of clusters larger than N=8 and with the aid of the Machine Learning technique will enable an efficient choice of minima.

The sections that follow in this chapter are responsible for describing the theory of each method that justifies its use in the steps of the steps described by the diagrams shown above.

2.1 Mollet-Plesset Pertubation Theory

The study of pertubation theory is developed in formal terms of the Hamiltonian and the full wave function and not only applied to one- or two-body systems, but is also valid for many-body systems, where it presents a wide variety of approximations, notations, and variations that have been used. Therefore, we must derive the equations from the basic pertubation theory and its representations of many bodies of various complementary forms. Pertubative methods are schemes used to obtain electron correlation energy and are one of the powerful tools for dealing with problems of many bodies [22, 23].

Among the pertubative methods, the most popular is the Rayleigh-Schrödinger pertubation theory - RSPT, which presents as a central idea to separate the Hamiltonian into two parts. The first one is the main one because it has known auto functions called the unpertubed Hamiltonian and the second part is the perturbed, characterized by a small pertubation in a way to alter the first one. Since the pertubation is small, it is expected that there is a tendency of convergence of the pertubative series so that the contributions of lower order are included. Therefore, to begin the development of the Rayleigh-Schrödinger theory, we present a Schrödinger equation of unknown solution.

$$H\Psi_n = E_n\Psi_n \tag{2.1}$$

e

$$H^{(0)}\Psi_n^{(0)} = E_n^{(0)}\Psi_0^{(0)} \tag{2.2}$$

where $H^{(0)}$ can be solved exactly and differs little from H, so

$$H = H^{(0)} + \lambda V, \tag{2.3}$$

where V is the second part mentioned at the beginning of the theory, the pertubation, and λ is a facilitator parameter for the ordering of the corrections in the energy and in the wave function. And the eigenvalues and auto functions of H are obtained by $H^{(0)}$ and by the matrix elements of V on the basis of the autofunctions of $H^{(0)}$.

By (2.3), both the energy $E_n = E_0(\lambda)$ and the perturbed wave autofunction $\Psi_n = \Psi_0(\lambda, q)$, depend on λ taking q as the coordinates.

Therefore, by expanding $E_0(\lambda)$ around λ=0,

$$\begin{aligned} E_0(\lambda) &= E_0(0) + \left(\frac{dE}{d\lambda}\right)_{\lambda=0}\lambda + \frac{1}{2!}\left(\frac{d^2E}{d\lambda^2}\right)_{\lambda=0}\lambda^2 + ... \\ &= E_0^{(0)} + \lambda E_0^{(1)} + \lambda^2 E_0^{(2)} + ... \end{aligned} \tag{2.4}$$

And in the same way by expanding the wave function we have,

$$\Psi_0 = \Psi_0^{(0)} + \lambda\Psi_0^{(1)} + \lambda^2\Psi_0^{(2)} + ... \tag{2.5}$$

Let us consider that the pertubed system is non-degenerate, that H and $H^{(0)}$ are Hermitian, and the pertubed auto functions are orthonormal,

$$< \Psi_i^{(0)}|\Psi_j^{(0)} >= \delta_{ij}. \tag{2.6}$$

Therefore, replacing equations (2.3), (2.4) and (2.5) in (2.1), results,

$$(H^{(0)} + \lambda V)\Psi_n = E_n\Psi_n$$
$$(H^{(0)} + \lambda V)(\Psi_0^{(0)} + \lambda\Psi_0^{(1)} + \lambda^2\Psi_0^{(2)} + ...) = (E_0^{(0)} + \lambda E_0^{(1)} + \lambda^2 E_0^{(2)} + ...), \tag{2.7}$$

separating the equations in terms that accompany the order of λ,

$$\lambda^0 \rightarrow H^{(0)}\Psi_0^{(0)} = E_0^{(0)}\Psi_0^{(0)}, \tag{2.8}$$

$$\lambda^1 \rightarrow H^{(0)}\Psi_0^{(1)} + V\Psi_0^{(0)} = E_0^{(1)}\Psi_0^{(0)} + E_0^{(0)}\Psi_0^{(1)}$$
$$(H^{(0)} - E_0^{(0)})\Psi_0^{(1)} = (E_0^{(1)} - V)\Psi_0^{(0)}, \tag{2.9}$$

$$\lambda^2 \rightarrow H^{(0)}\Psi_0^{(2)} + V\Psi_0^{(1)} = E_0^{(0)}\Psi_0^{(2)} + E_0^{(1)}\Psi_0^{(1)} + E_0^{(2)}\Psi_0^{(0)}$$
$$(H^{(0)} - E_0^{(0)})\Psi_0^{(2)} = (E_0^{(1)} - V)\Psi_0^{(1)} + E_0^{(2)}\Psi_0^{(0)}. \tag{2.10}$$

Therefore, in a generalized way,

$$\lambda^n \rightarrow H^{(0)}\Psi_0^{(n)} + V\Psi_0^{(n-1)} = E_0^{(0)}\Psi_0^{(n)} + E_0^{(1)}\Psi_0^{(n-1)} + E_0^{(2)}\Psi_0^{(n-2)} + ... + E_0^{(n)}\Psi_0^{(0)}$$
$$(H^{(0)} - E_0^{(0)})\Psi_0^{(n)} = (E_0^{(1)} - V)\Psi_0^{(n-1)} + E_0^{(2)}\Psi_0^{(n-2)} + ... + E_0^{(n)}\Psi_0^{(0)}. \tag{2.11}$$

Multiplying equations (2.8), (2.9), (2.10) and (2.11) by $< \Psi_0^{(0)}|$, equation (2.8) be-

comes null and consequently,

$$\begin{aligned} E_0^{(1)} &=< \Psi_0^{(0)}|V|\Psi_0^{(0)} > \\ E_0^{(2)} &=< \Psi_0^{(0)}|V - E_0^{(1)}|\Psi_0^{(1)} > \\ E_0^{(3)} &=< \Psi_0^{(0)}|V - E_0^{(1)}|\Psi_0^{(2)} > -E_0^{(2)} < \Psi_0^{(0)}|\Psi_0^{(1)} > \\ &\vdots \\ E_0^{(n)} &=< \Psi_0^{(0)}|V - E_0^{(1)}|\Psi_0^{(n-1)} > -... - E_0^{(n-1)} < \Psi_0^{(0)}|\Psi_0^{(1)} > \end{aligned} \tag{2.12}$$

Equations (2.12) can be represented in simplified form using the intermediate normalization, where equations

$$< \Psi_0^{(0)}|\Psi_0^{(0)} >= 1 \tag{2.13}$$

and substituting the expansion of equation (2.5) in (2.13),

$$< \Psi_0^{(0)}|\Psi_0^{(0)} > +\lambda < \Psi_0^{(0)}|\Psi_0^{(1)} > +\lambda^2 < \Psi_0^{(0)}|\Psi_0^{(2)} > +... = 1 \quad \text{onde,} \quad < \Psi_0^{(0)}|\Psi_0^{(n)} >= \delta_{n0}. \tag{2.14}$$

The equations (2.12) are then identified by,

$$\begin{aligned} E_0^{(1)} &=< \Psi_0^{(0)}|V|\Psi_0^{(0)} > \\ E_0^{(2)} &=< \Psi_0^{(0)}|V|\Psi_0^{(1)} > \\ E_0^{(3)} &=< \Psi_0^{(0)}|V|\Psi_0^{(2)} > \\ &\vdots \\ E_0^{(n)} &=< \Psi_0^{(0)}|V|\Psi_0^{(n-1)} > . \end{aligned} \tag{2.15}$$

In many cases, it is still difficult to solve them. One option is to perform an expansion of the function $\Psi_0^{(n)}$ based on the unpertubed problem $\{\ \Psi_i^{(0)}\ \}$ which constitute a complete set. Thus, the order 1 wave function in a completed way is,

$$\Psi_0^{(1)} = \sum_{i \neq 0} c_i^{(1)} \Psi_i^{(0)} \tag{2.16}$$

substituting (2.16) in (2.9),

$$\sum_{i\neq 0} c_i^{(1)}(H^{(0)} - E_0^{(0)})\Psi_i^{(0)} = (E_0^{(0)} - V)\Psi_0^{(0)} \tag{2.17}$$

and applying $< \Psi_j^{(0)}|$,

$$< \Psi_j^{(0)}| \sum_{i\neq 0} c_i^{(1)}(H^{(0)} - E_0^{(0)})|\Psi_i^{(0)} >=< \Psi_j^{(0)}|(E_0^{(0)} - V)|\Psi_0^{(0)} >$$

$$c_j^{(1)} = \frac{< \Psi_j^{(0)}|V|\Psi_0^{(0)} >}{E_0^{(0)} - E_j^{(0)}}. \tag{2.18}$$

The complete wave function of the first order is described by,

$$\Psi_0^{(1)} = \sum_{i\neq 0} \frac{< \Psi_i^{(0)}|V|\Psi_0^{(0)} >}{E_0^{(0)} - E_i^{(0)}}\Psi_i^{(0)} \tag{2.19}$$

and consequently the energy of correction in a second order

$$E_0^{(2)} = \sum_{i\neq 0} \frac{| < \Psi_0^{(0)}|V|\Psi_i^{(0)} > |^2}{E_0^{(0)} - E_i^{(0)}}. \tag{2.20}$$

Thus, we have the correction in energy and wave function that can be obtained from information about a unperturbed problem. In addition, by the representation of the second order energy correction, we can observe that the energy for this case will be negative, taking the ground state as reference.

So far, the development of the discussion presents results of up to second order energy corrections, considering that this is our focus of study in this work.

Considering the information presented so far, we can present the theory of pertubation of many bodies of Moller-Plesset, better known as MBPT.

The MBPT, as already foreseen, aims to obtain the electron correlation energy,

$$E_{corr} = E_{exata} - E_{HF} \tag{2.21}$$

defined by the difference of the exact energy and energy obtained by the Hartree Fock (HF) method.

The established proposal by Moller Plesset was to choose a partition from the electronic Hamiltonian, the unpertubed Hamiltonian, as a summation of the Fock operator,

$$H^{(0)} = \sum_{i=1}^{n} F(i) \tag{2.22}$$

and the sum in i is on the electrons of the system. In the case of the HF wave function, Φ_0, which is an auto function of $H^{(0)}$, we have

$$H^{(0)}\Phi_0 = \sum_{c} \epsilon_c \Phi_0 \tag{2.23}$$

where c represents the sum of the spin-orbitals present in the determinant HF.

To obtain the complete spectrum of $H^{(0)}$, we will use the concept of excited determinant or substituted determinant, where one or more occupied spin-orbitals represented by a,b,c are exchanged by one or more virtual spin-orbitals r,s,t. As an example, applying the Hamiltonian operator not disturbed in a simply substituted determinant, and in a simply substituted determinant,

$$H^{(0)}\Phi_a^r = (\sum_{c} \epsilon_c - \epsilon_a + \epsilon_r)\Phi_a^r \tag{2.24}$$

and in a doubly substituted determinant

$$H^{(0)}\Phi_{ab}^{rs} = (\sum_{c} \epsilon_c - \epsilon_a - \epsilon_b + \epsilon_r + \epsilon_s)\Phi_{ab}^{rs}. \tag{2.25}$$

Therefore, pertubative corrections in energy and wave function can be obtained from these determinants. In addition, the Hamiltonian pertubed part is described as being

$$V = H - H^{(0)} = \sum_{i} h(i) + \sum_{i<j} \frac{1}{r_{ij}} - \sum_{i} F(i) \tag{2.26}$$

being the first term of the right the representation of the operator of 1 electron, that de-

scribes the kinetic energy of the electrons and the interaction of the electron with the nucleus of the atom, the second term represents the electron-electron interaction and the latter is characterized by the operator of Fock , defined by

$$F(i) = \sum_i h(i) + \sum_c [J_c(i) - K_c(i)] \tag{2.27}$$

where J_c and K_c are respectively identified as Coulomb and exchange operators [22].

So

$$V = \sum_{i<j} \frac{1}{r_{ij}} - \sum_i v^{HF}(i) \tag{2.28}$$

$v^{HF}(i) = \sum_c [J_c(i) - K_c(i)]$ where the interaction equation is represented as 2 times in the average, and the first term in the left equation, the exact electron-electron to be defined in a simplified way by

$$V = O_2 - O_1 \tag{2.29}$$

explaining O_1 and O_2 as an interaction of 1 and 2 electrons.

Hence, in addition to having a role of correcting the double counting of the electron-electron interaction in $H^{(0)}$, V also gives a detailed description of the individual interactions between the electrons.

To calculate the pertubative corrections in energy it is necessary to calculate the matrix elements of the types $< \Phi|O_1|\widetilde{\Phi} >$ and $< \Phi|O_2|\widetilde{\Phi} >$. By using Slater's rule, (Szabo e Ostlund, 1989) in the case that $\Phi = \widetilde{\Phi} = \Phi_0$

$$< \Phi_0|O_1|\widetilde{\Phi}_0 >= \sum_{a,c} < ac||ac > \tag{2.30}$$

being $< ac||ac >$ an alternative notation to represent the permutations in the notation of Dirac [22], where a is the occupied spin-orbital, and

$$< \Phi_0|O_2|\widetilde{\Phi}_0 >= \frac{1}{2} \sum_{a,c} < ac||ac > . \tag{2.31}$$

Therefore,

$$< \Phi_0|V|\widetilde{\Phi}_0 >=< \Phi_0|O_2 - O_1|\widetilde{\Phi}_0 >= -\frac{1}{2}\sum_{a,c} < ac||ac > . \tag{2.32}$$

In matrix elements of V, between an HF determinant and a simply substituted determinant we have

$$< \Phi_0|O_1|\Phi_a^r >=< \Phi_0|O_2|\Phi_a^r >= \sum_c < ac||rc > \tag{2.33}$$

we conclude that

$$< \Phi_0|V|\Phi_a^r >=< \Phi_0|O_1 - O_2|\Phi_a^r >= 0. \tag{2.34}$$

Using the same theory one can also obtain the calculation of the matrix elements of V between the HF determinant and the doubly substituted one, so

$$< \Phi_0|O_1|\Phi_{ab}^{rs} >= 0 \tag{2.35}$$

this matrix element having a null value because the operator, when applied to the doubly substituted array, is only 1 electron. And

$$< \Phi_0|O_2|\Phi_{ab}^{rs} >=< ab||rs >, \tag{2.36}$$

so

$$< \Phi_0|V|\Phi_{ab}^{rs} >=< ab||rs > . \tag{2.37}$$

In situations where the prevailing differ by 3 or more orbital the solutions are always zero. Matrix elements between doubly substituted determinants, with the simply, triplicate or quadruple substituted, are not necessarily null because they may differ by two spin-orbitals or more.

We can then establish the MBPT power corrections as follows:

- zero-order energy

$$E_0^{(0)} = \sum_c \epsilon_c \tag{2.38}$$

- correction up to the first order, given by equation (2.32)

$$E_0^{(0)} + E_0^{(1)} = \sum_c \epsilon_c - \frac{1}{2}\sum_{a,c} < ac||ac > \tag{2.39}$$

- second-order correction, given by equation (2.20)

$$E_0^{(2)} = \sum_{a<b}\sum_{r<s} \frac{| < \Psi_0^{(0)}|V|\Psi_{ab}^{rs} > |^2}{E_0^{(0)} - E_{ab}^{rs}} \tag{2.40}$$

since only the doubly substituted determinants interact with the Φ_0 determinant.

As we can observe and define by the theories, the equation up to the first order energy is exactly the Hartree Fock energy, so the first-order energy MBPT has no correction. Thus, justifying the fact that it only makes use of second order energy corrections in order to obtain the electron correlation energy.

The second order correction energy, equation (2.40) can also be represented by the substituting equation (2.37), (2.23) and (2.24):

$$E_0^{(2)} = \sum_{a<b}\sum_{r<s} \frac{| < ab||rs > |^2}{\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s}. \tag{2.41}$$

We therefore conclude that for the case of interest in this work, the MBPT(2) calculation, corresponding to the energy calculation of electron correlation of Moller-Plesset perturbation theory for the second order energy correction bodies, it can be defined generalized by

$$E_{MBPT(2)} = E_{HF} + E_0^{(2)} \tag{2.42}$$

which as we already know to be a negative energy, will be smaller than the HF energy.

The MBPT(2), better known as MP2 has shown today that it is the most popular pertubative approach. This is due to its outstanding performance with good results and

modest computational resources[22].

2.2 Coupled Cluster

It is a complementary method to the MBPT (Moller-Plesset Pertubation Theory), able to provide electron correlation energy in a systematic and quite efficient way. It has more subtle effects of electronic correlation and results with high level of competitiveness. The proposal of this method is to treat the many electrons by separating them into several clusters with few electrons. In the clusters, the interaction between their electrons is calculated; this process is done with each of the clusters and then the interaction between the clusters is calculated. This type of approach can be performed by systems of non-interacting atoms [22].

In order to represent a mathematical mechanism that demonstrates the validity of this theory, we will begin the discussion by approaching two helium atoms [23].

The description of the state of this system can be defined by two orthonormal functions of two electrons $\phi_0(i)$ and $\chi(i)$ respectively representing the zero-order function and the excitation function of 2 electrons on $\phi_0(i)$, where the atoms are non-interacting and their functions are non-orthogonal, that is,

$$\widehat{t}_i\phi_0(i) = \tau\chi(i) \tag{2.43}$$

and

$$\widehat{t}_i\chi(i) = 0 \tag{2.44}$$

where $\widehat{t}_i$ is the excitation operator of two electrons in a single atom and has no effect on the base functions of other atoms,

$$\widehat{t}_i\phi_0(j) = \phi_0(j) \qquad e \qquad \widehat{t}_i\chi(j) = \chi(j). \tag{2.45}$$

Generally, in the wave function of N atoms, we define the operator,

$$\widehat{T}_2 = \sum_i \widehat{t}_i \tag{2.46}$$

which excites the two electrons at the same time. Thus by applying the operator in a fundamental wave function,

$$\widehat{T}_2\Phi_0 = \sum_i \widehat{t}_i\Phi_0 = \tau \sum_i \Phi_i, \tag{2.47}$$

$$\widehat{T}_2\Phi_i = \sum_j \widehat{t}_j\Phi_i = \tau \sum_{j(j\neq i)} \Phi_{ij} = \tau \sum_j \Phi_{ij}; \qquad \Phi_{ii} = 0 \tag{2.48}$$

being Φ_i the double excitation function, and the others are a consequence of the same logic. Soon,

$$\widehat{T}_2^2\Phi_0 = \tau \sum_i \widehat{T}_2\Phi_i = \tau^2 \sum_{ij} \Phi_{ij}, \tag{2.49}$$

$$\widehat{T}_2^3\Phi_0 = \tau^2 \sum_{ij} \widehat{T}_2\Phi_{ij} = \tau^3 \sum_{ijk} \Phi_{ijk}, \tag{2.50}$$

so the exact solution of N non-interacting He atoms can be described by,

$$\begin{aligned}\Psi &= \Phi_0 + \sum_i \tau\Phi_i + \frac{1}{2!}\sum_{ij} \tau^2\Phi_{ij} + \frac{1}{3!}\sum_{ijk} \tau^3\Phi_{ijk} \\ &= \Phi_0 + \widehat{T}_2\Phi_0 + \frac{1}{2!}\widehat{T}_2^2\Phi_0 + \frac{1}{3!}\widehat{T}_2^3\Phi_0 + ... \\ &= e^{\widehat{T}_2}\Phi_0.\end{aligned} \tag{2.51}$$

The expression of Ansatz $\Psi = e^{\widehat{T}_2}\Phi_0$ is characteristic of the coupled cluster approximation and τ is the unknown coefficient of $\widehat{T}_2$. To obtain an equation that determines τ and the corresponding energy, we apply the Schrödinger equation, in the form

$$(\widehat{H} - E)e^{\widehat{T}_2}\Phi_0 = 0. \tag{2.52}$$

Applying $< \Phi_0|$ e $< \Phi_i|$,

$$\begin{aligned} < \Phi_0|(\widehat{H} - E)e^{\widehat{T}_2}|\Phi_0 >= 0, \\ < \Phi_i|(\widehat{H} - E)e^{\widehat{T}_2}|\Phi_0 >= 0 \end{aligned} \tag{2.53}$$

using (2.47), (2.48), (2.49) and (2.50) and mathematical mechanisms cited by reference [23] we have,

$$\begin{aligned} < \Phi_0|Ee^{\widehat{T}_2}|\Phi_0 >= E < \Phi_0|\Phi_0 > +E < \Phi_0|\widehat{T}_2|\Phi_0 > +... = E, \\ < \Phi_i|Ee^{\widehat{T}_2}|\Phi_0 >= E < \Phi_i|\Phi_0 > +E < \Phi_i|\widehat{T}_2|\Phi_0 > +\frac{1}{2}E\sum_{jk} < \Phi_i|\tau^2|\Phi_{jk} > +... \\ = 0 + E\tau < \Phi_i|\Phi_j > +0 + ... = E\tau. \end{aligned} \tag{2.54}$$

So,

$$\begin{aligned} < \Phi_0|\widehat{H}e^{\widehat{T}_2}|\Phi_0 >=< \Phi_0|\widehat{H}|\Phi_0 > + < \Phi_0|\widehat{H}\widehat{T}_2|\Phi_0 > +\frac{1}{2} < \Phi_0|\widehat{H}\widehat{T}_2^2|\Phi_0 > +... \\ = E_0 + \tau\sum_i < \Phi_0|\widehat{H}|\Phi_i > +0 \\ = N\epsilon_0 + N\beta\tau, \qquad \beta =< \Phi_0|\widehat{H}|\Phi_i >; \quad and \quad E_0 = N\epsilon_0. \end{aligned} \tag{2.55}$$

E_0 and ϵ_0 represent the fundamental energy respectively of N atoms and a single atom. And

$$\begin{aligned} < \Phi_i|\widehat{H}e^{\widehat{T}_2}|\Phi_0 >=< \Phi_i|\widehat{H}|\Phi_0 > + < \Phi_i|\widehat{H}\widehat{T}_2|\Phi_0 > +\frac{1}{2} < \Phi_i|\widehat{H}\widehat{T}_2^2|\Phi_0 > +... \\ = \beta * +\tau\sum_j < \Phi_i|\widehat{H}|\Phi_j > +\frac{1}{2}\tau^2\sum_{jk} < \Phi_i|\widehat{H}|\Phi jk > +... \\ = \beta * +\tau\sum_j < \Phi_i|\widehat{H}|\Phi_i > +\frac{1}{2}\tau^2\sum_j < \Phi_i|\widehat{H}|(\Phi ij + \Phi ji) > +... \\ = \beta * +\tau[(N-1)\epsilon_0 + \alpha] + \tau^2(N-1)\beta. \end{aligned} \tag{2.56}$$

Substituing in equation (2.53), we obtain the pairs of simultaneous equation for τ

and E,

$$< \Phi_0|\widehat{H}e^{\widehat{T}_2}|\Phi_0 >= E < \Phi_0|e^{\widehat{T}_2}|\Phi_0 >$$
$$N(\epsilon_0 + \beta\tau) = E, \tag{2.57}$$

$$< \Phi_i|\widehat{H}e^{\widehat{T}_2}|\Phi_0 >= E < \Phi_i|e^{\widehat{T}_2}|\Phi_0 >$$
$$\beta * +\tau[(N-1)\epsilon_0 + \alpha] + \tau^2(N-1)\beta = E\tau. \tag{2.58}$$

Applying operations between the equations (2.57) and (2.58), we have

$$\beta * +\tau(\alpha - \epsilon_0) - \beta\tau^2 = 0 \tag{2.59}$$

Then we conclude that τ is independent of N atoms, and, by eq. (2.57), E depends on N as required by the extensity theory. Thus, the coupled cluster solution is in fact exact for N atoms problems, because it compares the CI, configuration, and therefore determines the exact response in the form $e^{\widehat{T}_2}\Phi_0$ for the non-interacting helium atoms.

For a generalized system the exact wave function can be described by,

$$\Psi = e^{\widehat{T}}\Phi_0 \tag{2.60}$$

where $\widehat{T}$ is the excitation operator. The Ansatz exponential is the operator that converts the zero-wave function into an exact function and manifests the extensity property in the wave operator, for this it is considered two non-interacting subsystems A and B, and expressed in terms of localized orbitals of 2 subsystems. If the zero order wave function for the system is separable,

$$\Phi_0(A...B) = \Phi_0(A) + \Phi_0(B) \tag{2.61}$$

and if for the system, $\widehat{T}$ is additive,

$$\widehat{T}(A...B) = \widehat{T}(A) + \widehat{T}(B). \tag{2.62}$$

The total wave function is:

$$\begin{aligned}\Psi(A...B) = e^{\widehat{T}(A...B)}\Phi_0(A...B) &= e^{\widehat{T}(A)+\widehat{T}(B)} \\ &= e^{\widehat{T}(A)}\Phi_0(A)e^{\widehat{T}(B)}\Phi_0(B) \\ &= \Psi(A)\Psi(B).\end{aligned} \tag{2.63}$$

This separation of the wave function guarantees the additivity of the energy, therefore

$$\begin{aligned}\widehat{H}(A...B)\Psi(A...B) &= [\widehat{H}(A) + \widehat{H}(B)]\Psi(A)\Psi(B) \\ &= [\widehat{H}(A)\Psi(A)\Psi(B) + \Psi(A)\widehat{H}(B)\Psi(B)] \\ &= [E(A) + E(B)]\Psi(A...B).\end{aligned} \tag{2.64}$$

This is a weaker condition than the previous one and is preferred in the definition of extensity.

With the Coupled Cluster wave function defined, the excitation operator can be described by,

$$\widehat{T} = \widehat{T}_1 + \widehat{T}_2 + \widehat{T}_3 + ... \tag{2.65}$$

where $\widehat{T}_1$ corresponds to the excitation of a body, $\widehat{T}_2$ of two bodies, etc., so that the cluster operators are defined by

$$\begin{aligned}\widehat{T}_1 &= \sum_{ia} t_i^a \hat{a}^t\hat{i} \\ \widehat{T}_2 = \frac{1}{(2!)^2}\sum_{ijab} t_{ij}^{ab}\hat{a}^t\hat{b}^t\hat{j}\hat{i} &= \frac{1}{4}\sum_{ijab} t_{ij}^{ab}\{\hat{a}^t\hat{i}\hat{b}^t\hat{j}\} \\ \widehat{T}_3 = \frac{1}{(3!)^2}\sum_{ijkabc} t_{ijk}^{abc}\hat{a}^t\hat{i}\hat{b}^t\hat{j}\hat{c}^t\hat{k} &= \frac{1}{36}\sum_{ijkabc} t_{ijk}^{abc}\{\hat{a}^t\hat{i}\hat{b}^t\hat{j}\hat{c}^t\hat{k}\}\end{aligned} \tag{2.66}$$

etc, where $t_{ij...}^{ab...}$ the coefficients to be determined, referenced as the amplitudes of the corresponding operators and the strings of creation and annihilation operators are automatically in normal-ordered form. In general the operator is:

$$\widehat{T}_m = \sum_{(m!)^2} t_{ij...}^{ab...} \{\hat{a}^t \hat{i} \hat{b}^t \hat{j} ...\}. \tag{2.67}$$

If the exponential wave operator is expanded in Taylor series,

$$e^{\hat{T}} = 1 + \hat{T} + \frac{1}{2}\hat{T}^2 + \frac{1}{3!}\hat{T}^3 + ... \tag{2.68}$$

using the expansion in equation (2.65), we obtain

$$\Psi = \Phi_0 + \hat{T}_1\Phi_0 + \hat{T}_2\Phi_0 + ... + \frac{1}{2}\hat{T}_1^2\Phi_0 + \hat{T}_1\hat{T}_2\Phi_0 + \frac{1}{2}\hat{T}_2^2 + ... \tag{2.69}$$

and it is notable that the different operators of $\widehat{T}_m$ commute. The contributions of the wave function in the form $\widehat{T}_m\Phi_0$ are called connected clusters while those involving products of cluster operators such as $\frac{1}{2}\widehat{T}_2^2\Phi_0$ or $\widehat{T}_1\widehat{T}_2\Phi_0$ are called disconnected clusters.

The term $\widehat{T}_2\Phi_0$ is the most important cluster-connected contribution, due to the nature of two Hamiltonian electrons. While in terms of contribution of disconnected clusters the most important is $\frac{1}{2}\widehat{T}_2^2\Phi_0$, where $\widehat{T}_2$ can become more important with increasing system. The term connected to $\widehat{T}_3\Phi_0$ is important in systems with high electron density. And the terms $\widehat{T}_4\Phi_0$, $\widehat{T}_5\Phi_0$, $\widehat{T}_6\Phi_0$ etc, are generally of small importance. May be relevant in special cases.

Among the Coupled Cluster approximations, the simplest is the Double Coupled Cluster (CCD), where $\widehat{T}$ is truncated for

$$\widehat{T}_{CCD} = \widehat{T}_2 \tag{2.70}$$

Thus the CCD wave function includes only the connected and disconnected contributions of

clusters involving $\widehat{T}_2$, so

$$\Psi_{CCD} = e^{\widehat{T}_2}\Phi_0 = \Phi_0 + \widehat{T}_2\Phi_0 + \frac{1}{2}\widehat{T}_2^2\Phi_0 + \frac{1}{3!}\widehat{T}_2^3\Phi_0 + ... \tag{2.71}$$

The most common extension of this model is simple and double coupled-cluster (CCSD),

$$\widehat{T}_{CCSD} = \widehat{T}_1 + \widehat{T}_2 \tag{2.72}$$

that it was even necessary to use it in some calculations.

And the extension CCSDT - coupled cluster of single, double and triple excitations, recognized as an excellent approximation of the exact wave function,

$$\widehat{T}_{CCSDT} = \widehat{T}_1 + \widehat{T}_2 + \widehat{T}_3 \tag{2.73}$$

Although it is a more sophisticated approach, the CCSDT has some computational issues that makes it a rather labor-intensive computational cost method.

Therefore, it is necessary to use approximate and accessible models, which in the case of this work was used the CCSD(T), known to be a method of partial inclusion of triple substitutions, where there is no great increase in computational cost. The method is propesed by Gauss and Cremer (1988), which is given by the inclusion of triple partial substitutions in CCSD [22, 24, 25].

2.3 Density Functional Theory - DFT

The DFT method, also known as the density functional theory, has become a very efficient computational simulation strategy in the study of the electronic structure of atoms and molecules. Secondly, because it makes possible the calculation of systems and sizes of more than twenty atoms, with an acceptable physical and chemical precision, second, because it presents a computational cost that sometimes represents a fraction, compared to the MBPT and coupled cluster method. Things that have driven the development of this method are

more accurate results of exchange-correlation functionals and numerical integration efficient algorithms. Still emphasizing the importance of the implementation of DFT, it is worth noting that commercial use contributes significantly to the popularization and expansion of its use in the field of atomic and molecular physics such as GAMESS and NWCHEM [22].

Its formal development arose in 1964, in the publication of the theorems of Hohenberg and Konh [26]. Its use is applicable in many electron systems, which with some degree of approximation can become a solvable problem. The purpose of the method is to have as a fundamental object. The total electronic density $\rho(r)$, considering that the Schrödinger equation of $3N$ variables (without considering the spin) can be described as an electronic density equation of only three variables.

Konh-Shan's Theorem proposes two ideas for the construction of theory. First, the external potential $v(\vec{r})$ sense by the electrons is a unique functional of the electronic density $\rho(\vec{r})$, and the electronic density of a system can determine the external potential and the number of electrons N. To prove this, we start from an electronic Hamiltonian,

$$\widehat{H} = \widehat{T} + \widehat{U} + \widehat{V}, \tag{2.74}$$

being $\widehat{T}$ the kinetic energy, $\widehat{U}$ the potential electron-electron energy and $\widehat{V}$ the external potential energy. And there is still a independence of the terms in study, $v(\vec{r})$, $\rho(r)$, N and $\widehat{H}$.

Suppose then that there is an external potential $v'(\vec{r})$, which results in the Hamiltonian $\widehat{H}'$ and the wave function $\widehat{\phi}'$, and leads to the same density $\rho(\vec{r})$. By the variational theorem, the energy E starting from the wave function ϕ of the system under study can be described by,

$$\begin{aligned} E &= \langle\phi|\widehat{T} + \widehat{U} + \widehat{V}|\phi\rangle < \langle\phi'|\widehat{T} + \widehat{U} + \widehat{V}|\phi'\rangle \qquad e \\ E' &= \langle\phi'|\widehat{T} + \widehat{U} + \widehat{V}'|\phi'\rangle < \langle\phi|\widehat{T} + \widehat{U} + \widehat{V}'|\phi\rangle \end{aligned} \tag{2.75}$$

or,

$$\langle\phi|\widehat{H}|\phi\rangle < \langle\phi'|\widehat{H}|\phi'\rangle = \langle\phi'|\widehat{H}'|\phi'\rangle < \langle\phi'|\widehat{V} - \widehat{V}'|\phi'\rangle. \tag{2.76}$$

Thus, the external potential matrix is

$$\langle\phi|\widehat{V}|\phi\rangle = \int \rho(\vec{r})v(\vec{r})d^3r. \tag{2.77}$$

In which $\rho(\vec{r}) = \langle\phi|\sum_{i=1}^{N}\delta(\vec{r}-\vec{r}_i)|\phi\rangle$ e $\widehat{V} = \sum_{i=1}^{N} v(\vec{r}_i)$. So, from the equation (2.75) we obtain,

$$\begin{aligned} E &< E' + \int [v(\vec{r}) - v'(\vec{r})]\rho(\vec{r})d^3r \\ E' &< E + \int [v'(\vec{r}) - v(\vec{r})]\rho(\vec{r})d^3r \end{aligned} \tag{2.78}$$

then,

$$E + E' < E' + E. \tag{2.79}$$

As we define that the electron density is the same for $v(\vec{r}) = v'(\vec{r})$ we conclude something absurd $\phi \neq \phi'$. Thus for this to occur, the unit of $\rho(r)$ requires to consider $\phi = \phi'$. Hence the first theorem states that the density of $\rho(r)$ of the ground state must contain the same information of the wave function of the state in question. We can then represent a unique function of density by the representation of a physical observable,

$$O = \langle\phi|\widehat{O}|\phi\rangle = O[\rho(r)]. \tag{2.80}$$

A second theorem established by Kohn-Sham states that: the fundamental state energy $E_0[\rho]$ is minimal for the exact density $\rho(r)$ that is, from the electronic density of the ground state we obtain the minimum value of $E[\rho]$ which is a functional of $\rho(\vec{r})$,

$$E[\rho] = \langle\phi|\widehat{T} + \widehat{U} + \widehat{V}|\phi\rangle. \tag{2.81}$$

To justify this, it is assumed that ρ_0 is a density of $\widehat{H} = \widehat{T} + \widehat{U} + \widehat{V}$ and which may

have a density $\rho(\vec{r})$, only determined by the state ϕ. Soon

$$\begin{aligned} \rho \neq \rho_0; \phi = \phi_0, \quad E > E_0 \\ and if \rho = \rho_0; \phi = \phi_0, \quad E = E_0. \end{aligned} \tag{2.82}$$

We rewrite equation (2.81) as,

$$E[\rho] = F[\rho] + \langle\phi|\widehat{V}|\phi\rangle; \qquad F[\rho] = \langle\phi|\widehat{T} + \widehat{U}|\phi\rangle \tag{2.83}$$

In analogy we have

$$E[\rho_0] = F[\rho_0] + \langle\phi_0|\widehat{V}|\phi_0\rangle \tag{2.84}$$

in which ϕ_0 is the fundamental state wave function and $F[\rho]$ is defined as the universal functional. Knowing that both electronic densities are determined by external potential, we apply the variational theorem,

$$E[\phi_0] < E[\phi] \tag{2.85}$$

$$F[\rho_0] + \langle\phi_0|\widehat{V}|\phi_0\rangle < F[\rho] + \langle\phi|\widehat{V}|\phi\rangle. \tag{2.86}$$

Assuming then the affirmation of the second theorem,

$$E[\rho_0] < E[\rho]. \tag{2.87}$$

In this way we demonstrate in both theorems how to determine the fundamental state, from a given external potential, using the electronic density instead of the wave function. For most problems of atomic and molecular physics, including for this work the external potential is already established, aiming to perform calculations of the electronic properties, making use of artifices proposed by the two theorems.

In the last 10 years, the problems of DFT have become more evident [27], because they are not able to describe the dispersion caused by non-covalent interactions. Therefore, the corrections of Grimme D3, in functional DFT has emerged as the most popular tool to

describe non-covalent interactions (DFT-D) because of the low cost of corrections calculations and because they are in accordance with ab initio and experimental results [27, 28].

DFT-D takes into account a pair of potential atoms that are added to standard Kohn-Sham results. There are some versions of the Grimme fixes, which will be used in this work is the DFT-D3. There is a more up-to-date correction to DFT-D3 (BJ), but as we will use NWCHEM, this does not allow us to perform DFT calculations with the proposed functional and bases for this type of correction, but as will be presented by theory and by the results in this work, an efficient and satisfactory method for calculations is shown. Therefore according to theory, the correction of Grimme DFT-D3, has the total energy (E_{DFT-D3}) described by,

$$E_{DFT-D3} = E_{KS-DFT} - E_{disp}. \tag{2.88}$$

Where E_{KS-DFT} EKS-DFT is the self-consistent Kohn-Sham energy described by equation (2.86) and,

$$E_{disp} = E^{(2)} + E^{(3)} \tag{2.89}$$

in which, $E^{(2)}$ is a two-body and $E^{(3)}$ three-body term. Specifically,

$$E^{(2)} = \frac{-1}{2} \sum_{A \neq B} \sum_{n=6,8,10} S_n \frac{C_n^{AB}}{R_{AB}^n} f_{damp}(R_{AB}), \tag{2.90}$$

where the sum is over all the pairs of atoms of the C_n^{AB}, system, which is the mean of the n-order scattering coefficient (n=6,8,10...) for atomic pairs AB and R_{AB} and the internuclear distance . The S_n factor is of global scale (functional dependent) that can be used to adjust the repulsive behavior correction of the density functional exchange (DF) correlation choice. But, the key ingredient of DFT-D3 is the damping function (f_{damp}), which determines the behavior of small variation of the dispersion correction and is necessary to avoid singularities in small R_{AB} and effects of double counting of correlations of electrons of intermediate distances,

$$f_{damp}(R_{AB}) = \frac{1}{1 + 6(\frac{R_{AB}}{S_{R,n} R_{0AB}})^{-\alpha_n}} \tag{2.91}$$

where $S_{R,n}$ is a function dependent on the radius of cut R_{0AB} and α is the parameter that

determines the slope of the function for small R_{AB}.

And on the other hand the terms of three bodies is,

$$E^{(3)} = \sum ABC f_{d,3}(\overline{R}_{ABC}) E^{ABC} \tag{2.92}$$

where the sum is over all triple ABC atoms of the system, and E^{ABC} is

$$E^{ABC} = \frac{C_9^{ABC}(3cos\theta_a cos\theta_b cos\theta_c + 1)}{(R_{AB}R_{AC}R_{BC})^3}, \tag{2.93}$$

called a mutual Axilrod-Teller or triple dipole, whose cos θ_a, cos θ_b and cos θ_c are the internal angles of the triangle formed by R_{AB}, R_{AC} and R_{BC} and C_9^{ABC} is the triple dipole constant.

According to the reference [29] the DFT-D method is considered as robust and has been tested meticulously and applied sucetively now in millions of different systems, it has been what combines the best properties [30, 31, 32, 33].

Comparing revisions [31, 34] and other DFT-D3 implementations [35, 36] versions have the following advantages and properties according to [29]:

- We calculate the least empirical and the most important parameter are first by Kohn-Sham standards.
- Approach is asymptotically correct with all DFT's for finite systems or infinite non-metallic systems. It determines the almost exact energy dispersion for a weakly interacting neutral atom gas and small interpolations for molecular regions.
- It performs a consistent description of all chemically relevant elements of the periodic system (nuclear charge = 1-94).
- The dispersion of coefficients of specific atomic pairs and cutting radius are explicitly calculated.
- The dispersion depends on the number of coordinates; these do not rely on the information of the connectivity of atoms.

- It performs similar or better accuracy for lightweight molecules and better description of heavier metal systems.

2.3.1 The funtionals and basis to DFT

The development of new and better correlation functional has become a promising pursuit in the Density Functional theory - DFT [37] . With this in view, we went in search of functional literature that could be tested and could represent the studied system well. In particular, we chose to work with 5 functional ones that have presented good results.

The first is the PW91 functional that makes use of the generalized gradient approximations (GGA). In the GGA, it is considered that the densities vary little or moderately over the space so that the exchange and correlation functional around the electron has the first principles of expansion in the density gradient. These expansions are realistic close to the electrons, but not distant. GGA is based on two limiting strategies: adjustment of some experimental datasets and realization of a number of known physical constraints. In this case, the functional to be tested is xperdew91 (named by NWCHEM), and its choice is because it does not present semiempirical parameters and because it adds theoretical refinements in the limits of small and large gradients [38, 39].

The PBE is a functional GGA introduced by Perdew, Burke and Ernzerhof [40] in which all parameters are fundamental constants. It is constituted by the Perdew-Wang functional correlation and a new exchange contribution. This functional is expected to provide comparable quality results for different amplitudes of chemical systems. There are several publications that attest to the good results of this functional [41, 42, 43].

The functional B3LYP is a hybrid method of functional combining, among them the exact exchange energy[44], Becke's gradient correction [44] and Perdew-Wang's gradient correction [38]. Besides, it is a functional composed of three semi-empirical coefficients to be determined by an appropriate adjustment of experimental data [45]. In particular, it is one of the most commonly used functionalities in the literature.

The Slater-VWN5 is a functional represented by the Slater's exchange, plus the

VWN5 correlation. The Slater's exchange that considers an atom around a sphere and within each of the spheres, replacing the exact potential by a spherical average and in the region between the spheres the potential is replaced by a constant; the same if it is a molecule. [46, 47].

M08-HX is the conformation that the optimization procedures work well for data outside the training set. This functional is an improvement in cases of atomic energies, large molecules and non-covalent interaction energies among others [37, 48, 49]. The M08-HX [50] is characterized by an improvement of the M06-2X [51], and is part of functional known as the functional ones of Minnesota, developed by the group of Prof. Donald Trulhar, from the University of Minnesota.

In an ab initio study of the molecule electron structure of the exact solution from the non-realist Schrödinger equation requires a complete treatment of N-electrons and a complete expansion of a set bases. The use of a complete set bases is impractical. The most powerful strategy is the approximation of the exact solution Schrödinger, equation, from a selection to a possible expanded set of bases. This will depend on the size of the molecular system and the treatment of N-electrons. In the case of the DFT method, a specific study of the Li^+Kr_n cluster will be carried out.

For this type of system, the bases denominated by the NWCHEM as: cc-pVDZ, aug-cc-pVDZ, aug-pcseg-1 and aug-pcseg-2 will be treated.

The set bases cc-pVDZ and aug-cc-pVDZ (D indicates the second polarization level) for the Lithium consists of: (i) generalized contraction of the HF orbitals, (ii) additional primitive functions step in relation to the corresponding s e p em relação aos correspondentes conjuntos HF (sp), (iii) spherical harmonic contractions of the sets of the highest angular functions ($dfgh...$), (iv) increased functions required to describe the effects of core-valence correlation and / or anionic characteristics (supplier of negative ions - our case study the Li^+) [52].

While the base cc-pVDZ is a primitive set of HF (sp) determined for optimizations of polarized functions to describe correlation effects, the base aug-cc-pVDZ (pointed above in item (iv)) is intended to cover more spaces wide features in the use of base functions with

additional diffuse characteristics. They are necessary to accurately describe the affinity of electrons, properties of electric fields such as permanent electrical moments and polarizability, and long-range interaction [52, 53, 54].

For the Kr (Krypton) atom, the bases follow the equivalent principle, in order to allow a better flexibility and in the description of the atomic orbitals, the general contraction method of Raffenitti [55] was used. These approximations have been used to develop consistent base sets for atoms such as Kr [56]. It is important to note that for *ab initios* MP2 and CCSD(T) calculations we will use a set of larger bases aug-cc-PVTZ and aug-cc-pVQZ [56, 52]

The basis Apcseg-1 and Apcseg-2 is based on the basis pcseg-n - polarization-consistent basis set indices 1 and 2 indicate the level of polarization. For Lithium (Li), the selection of atomic set bases s requires some considerations, such as the energy balance between the different types of functions of a given atom. The element Li has only occupied function orbitals s and p are, therefore, the first set of polarization exponents. Li has core orbitals, which are also high enough in energy to participate in molecular bonds in some extents. Thus, the construction of the increased Apcseg-n base includes the dependence of electronic densities far from the nucleus, for example, electric dipoles moments and polarizability in the base set convergence that can be significantly improved by additional diffuse functions with small exponents [57]

The article [58] extends these principles to the set bases for third row of the periodic table and consequently for our case of interest, the Krypton.

2.4 Evolutionary Algorithm

We consider the global minimization problem difficult and it is impossible to designate an algorithm that can respond in a linearly increasing time scale with the size of the system, so different methods appear in the literature. A solution found for this, was given by the evolutionary algorithm (EA) or also known as genetic algorithm, which has as one of the functionalities to discover a global minimum of small scope. The direct use of ab initio

calculations or functional density theory has been accessible for homogeneous clusters of up to 20 atoms whereas in heterogeneous clusters only 10 atoms have been possible [17].

Therefore, the study of larger heterogeneous clusters implies the use of analytical potentials. We also know that a rigorous knowledge of the potential energy surface (PES) of the clusters is required for global minimum energy optimization. Since they model the interaction and therefore, for the use of the algorithms to which it will be presented, these will be applied from curves or surfaces of potential energy, which can describe the interaction behavior of atoms, to two or three bodies.

The goal of optimizing cluster geometry is to determine a structural organization for a set of atoms or molecules that minimizes the total potential energy. PES is the function that models the interaction between the cluster atoms and contains the relevant information for the system. They are functions that generate high rugged landscapes of energy with a large number of local minima and a deeper characterized in this work as the global minimum.

2.4.1 The algorithm

In 1990, the first applications of the evolutionary algorithm (EA) for optimization of clusters appeared [59, 60]. In addition, it has now been a state-of-the-art in this area. The performance of the EAs improves when they are coupled with the local procedure using first order derivative to guide in the nearest local optimization. And according to studies [61, 62] maintaining diversity is a key issue to improve the effectiveness of populations.

The evolutionary algorithm is capable of carrying a two-level optimization: it searches the 3D optimization configuration that minimizes the potential energy of the cluster and at the same time involves the composition of aggregates (search for a more formidable proportion of atoms in each specific type). Some results showed that diversity was a key issue for the success of the quantization algorithm [16]. For, this characterizes including two-stage local research and efficiency depends on the maintenance of diversity throughout the optimization.

Based on this information the evolutionary algorithm to be used in this work will be accompanied by two techniques: the local optimization L-BFGS (Near Newton method

of local minimization)[63]. The L-BFGS is responsible for local research applied to each generated by the algorithm, and the steady-state method that guarantees greater diversity by keeping the population of parents and offspring competing for survival.

The figure 2.4 summarizes the logic applied by the algorithm in this work, but only 5 individuals of the population are represented in an illustrative way. Therefore the following descriptions will follow in accordance with the one described in the image.

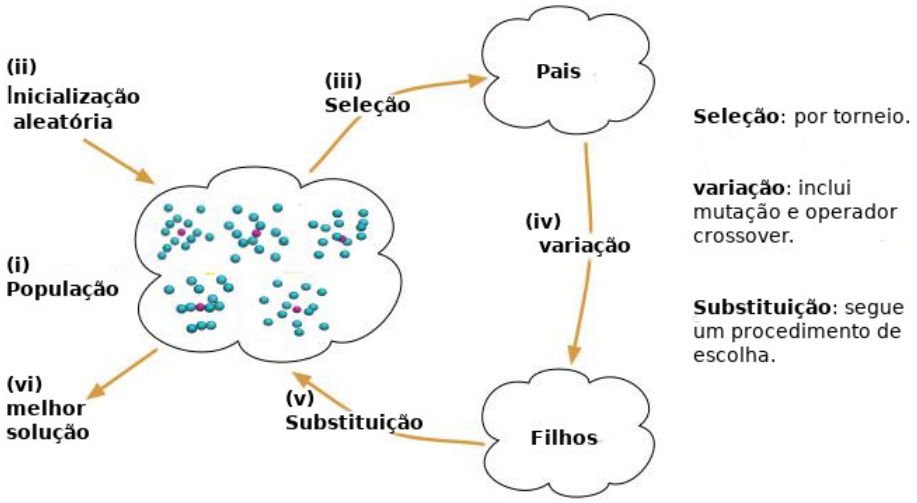


Figure 2.4: Algorithm structure.

As a first step in the algorithm structure (i) we choose a random initial population of 100 individuals of cluster structures, which can show possible solutions of the problem. Each individual composed of N atoms, $N = N_A + N_B$ in which $\mathrm{N}_A(\mathrm{N}_B)$ is the number of A(B), atoms, in figure 2.4 we can portray A as the blue atoms (noble gases) and as B a single atom in pink (Li^+).

The coordinates of the atoms are defined by an individual coding of $3XN$ real values, specifying the Cartesian coordinates of each particle, as shown in figure 2.5(the self-adaptive step to the left of figure 2.5 will only be used and analyzed in chapter 3). Dividing in two steps, the first would be to codify the Cartesian coordinates of the atoms belonging to the cluster to be optimized (*cluster coordinates*) and a second step (*self-estrategy*) is the choice of components and configurations that help in its own research strategy. The type *Cx_type*, corresponds to the type of crossover that the individual can adopt, *Mut_type* is the type of mutation that can be applied, for example, of sigma type adopted in this work, the individual

codes for his own standard deviation, or the choice of a random mutation. And distance M where the individual can estimate the similarity to other solutions using the set distance, structural distance, or distance from the center of mass [18].

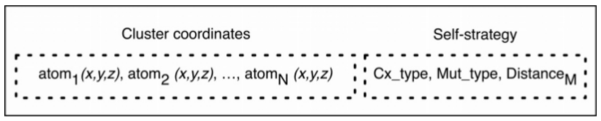


Figure 2.5: The algorithm coordinates and new strategy.

In step (ii) each individual of the initial population is then locally minimized with the L-BFGS algorithm. The minimum and maximum distances between the atoms are defined by parameters, taking into account that the potential becomes much more repulsive when two atoms approach more than one. This condition is established in the initial population and during the application of the genetic code. The adjustment of the global minimum results is obtained according to the energy values calculated with the appropriate potential function.

In step (iii), starting from the initial population, we try to find out the best configuration of the Cluster geometry, of the problems to be solved. There is a selection by tournament where the two individuals of the current population are chosen to generate the new individuals (descendants) (figure 2.6). We select each parent in two steps. First, several current population solutions (number is given by the size parameter of the tournament) are randomly chosen to participate in the tournament, so the best competitor of the tournament (ie the cluster with lower potential energy) is selected as a parent of the new solutions; the tournament is repeated until the population parents are completed. This selection tournament is considered one of the most effective selection mechanisms [64].

In the process of variation (iv), the genetic operators are applied to the selected parents, to create a new set of solutions, the children. The variation operators to be employed in the evolutionary genetic algorithm is the Generalized Cut method and Splice (GenC&S), an extended variation of the original Cut and Splice (C&S) operator. We design the C&S in general for homogeneous cluster on geometry optimization problems, in order to preserve some semantic properties of the parents involved in a crossover operation. It selects a random plane, passing through the center of mass of each parent cluster, and assembles the children through the combination of complementary parts. We take precautions to ensure that each

child contains a correct number of atoms.

GenC&S differentiates from C&S because it determines the sub clusters (subset of atoms in an individual) to be swapped and is specifically assigned to account of heterogeneous clusters. And it still tries to make sure that each parent's contribution creates a set of atoms that are nearby. The location of GenC&S is more in agreement with what is expected of the crossing [64]. Here is a brief description of how GenC&S operates.

Figures 2.6, 2.7, 2.8 depict each step to be described in this process. First, select a random number $X \epsilon$ [1, N-1]. Where N is the size of the cluster, of these, X atoms (in the example $x = 4$ atoms, in red), are chosen in the first parents and $N - X$ atoms (10-4 = 6 atoms in blue) are selected to be part of the second parent. The contribution of each parent is determined by a composition of neighboring atoms in 3D space.

According to figure 2.7 the parents $P1$ and $P2$ are crossed, each one representing a cluster with 10 atoms. The operation starts by selecting the random atoms of $P1$ (the cut-off point CP), determining the size of the sub-cluster $X \epsilon$ [2, N-1] which will be inherited from this parent. In figure 2.7 in a) X equal to 4 represents the choice of the sub cluster of $P1$ that will form the descendant. And in b) shows the chosen cutting atoms in $P1$. As observed by the image, there are $P1$ atoms that overlapped with atoms in $P2$, that is, they occupy approximately the same location, in order to avoid choosing atoms in $P2$ that already occupy this position then in $P2$ the 7 atoms listed. It is then necessary to choose 6 of these atoms, since the intention is to form a descendant whose cluster size has 10 atoms, since 4 have already been chosen in $P1$. The choice of these 6 is made according to the shortest distance from the cut atoms at $P1$, so the atoms numbered 1-6 are chosen. Finally, figure 2.8 shows the descendant formed by the intersection, composed of the combination of the two sub-clusters. The procedure is repeated in the operation of the EA, the same pair of parents, just changing the role of $P1$ and $P2$.

A specific case that will be pertinent at work because we are dealing with aggregates of heretogene atoms is HGenC&S - Heterogeneous Generalized Cut&Splice therefore, it guarantees that the descendants maintain the same proportion of atoms that are present in their parents [64].

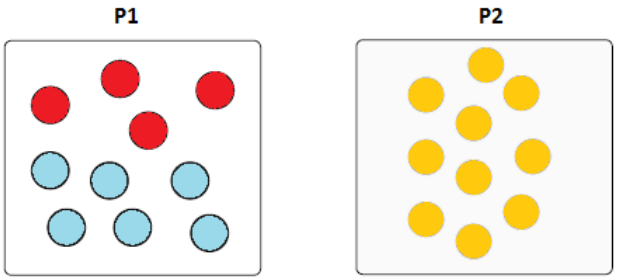


Figure 2.6: The Parents.

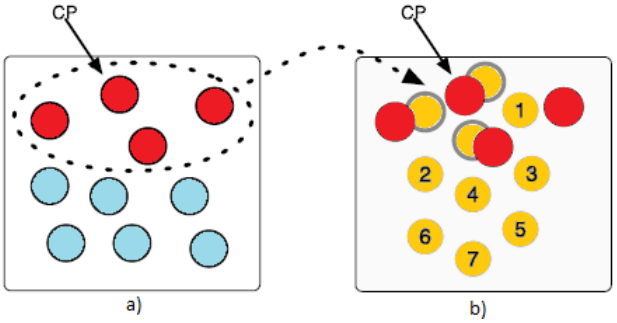


Figure 2.7: The crossover operation betwen the parents P1 e P2.

In some cases, given the minimum distance constraint, it may be impossible to select the sufficient number of particles in $P2$ to complete the descending. If this happens, the descendant is completed with atoms placed at random locations. In any case, the random conclusion usually occurs in less than 10% of all crossing operations and generally involves the addition of only 1 or 2 particles, being equivalent to the application of a mutation operator.

The sigma mutation is applied to solutions resulting from crossing. From a random atom, its location is slightly modified, as shown in figure 2.9. We obtain the new location, disturbing each coordinate with random value shown with small standard deviation and null Gaussian distribution.

The final step that defines the new current population and its main characteristic is to maintain a greater diversity of the population, is the steady-state model, in which the children compete for survival with their parents. In this model, only one or two individuals are generated at each step of time. After the creation of the descendants, it is necessary to decide whether they are allowed to enter into the present solution, that is, to be part of the population, and if so, the place of the descendant will occupy. The steady-state is a crucial step in EA's as it has a greater impact.

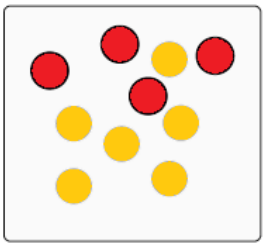


Figure 2.8: The offsrpings.

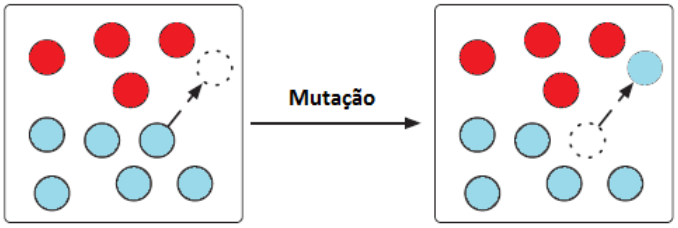


Figure 2.9: The mutation.

The terms of steady-state used in this work are the ones proposed by Lee.et. al. 2006 [65], *fitness-based*, where the quality of the individual is used to decide who will be replaced. This criterion may be associated with mechanisms that aim to maintain an appropriate level of diversity. An example of this is to remove the worst individual and insert the youngest child generated (with the condition that it is best).

In order to describe the proposed strategy, a programming logic is presented. Before the descendant D is allowed to join the current Pop population, the EA verifies whether a solution already exists, as shown below.

Algorithm 1 Replacement strategy adopted by the EA

```
Find X ∈ Pop such that d(X, D) is minimum
if [d(X, D) ≤ d_min and V_Morse(D) < V_Morse(X)] then
    D replaces X in Pop
else if [d(X, D) > d_min] then
    Select Y ∈ Pop, such that V_Morse(Y) is maximum
    if [V_Morse(D) < V_Morse(Y)] then
        D replaces Y in Pop
    end if
else
    D is discarded
end if
```

Figure 2.10: EA: Steady-state

The steady-state specifies that a descendant D is inserted into the population if one of the conditions is satisfied.

(i) if D is identical to solution $X \epsilon Pop$ and it has better adaptation than X, so D replaces X. (ii) if D is different from all solutions in Pop and D has the best fitness than

the worst X solution Pop, then D replaces X.

Among the measures, d_{min} is the parameter of the algorithm specifying the minimum distance of any two individuals belonging to the population. For d_{min} measurement, the center-of-mass distance method is used, since it presents better acceptance in the calculation of the distances [16], this is based on the distance of each atom in relation to the center of mass of the cluster.

Diversity is forced, because if there are no similar solutions, D replaces the worst individual in the population, otherwise, if the offspring are different from all the solutions of the current population, then it will replace the worst individual. If neither of these conditions is known, D is discarded. Starting from the initial random solution, EA intervenes between these three steps; selection, variation, and substitution, to vary the generations, and in the end returns the best solution found. So the best competitor in the cluster, that is, the parent with the lowest potential energy is selected as the parent of the new solutions.

Chapter 3

Revealing energy landscapes of atomic clusters by applying adaptive bio-inspired algorithms

We review our work on the development of evolutionary algorithms (EAs) for revealing low-energy structures of atomic and molecular clusters. The application of EAs on the study of the microsolvation of alkali-metal ions with argon and assessing the chemical ordering of binary clusters of transition-metal elements is discussed. Additionally, we discuss the application of novel self-adaptive bio-inspired algorithms to model cluster-systems. Several adaptive strategies dealing both with control parameters and algorithmic components will be presented and some preliminary results are described and analyzed.

3.1 Introduction

The structure of a chemical compound (e.g., a drug) in biological media is intrinsically related to its function. In particular, proteins are formed by chains of aminoacids that fold to acquire the specific native structure. Such an aggregate of coils (chains) may lead to several available minima, but there is just one structure that is effective to carry

out the biological function (role) and, indeed, the sequence of aminoacids forming a protein is able to fold into the corresponding native state [66]. Extra complexity in the folding process can be observed when competition between two native states is manifest [67, 68]. This connection between protein folding and its biological function has been well-established theoretically through the analysis of the energy landscapes [69], whose framework benefits of employing global geometry optimization methods. Indeed, several algorithms that search for low-energy configurations of atomic and molecular systems have been applied to the study of the structure of molecules with biological interest [70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80].

Among global optimization techniques [81, 82, 83, 62, 84, 85], Evolutionary Algorithms (EAs) [12] are state-of-the-art methods that have shown to be very successful for discovering low-energy structures of atomic and molecular cluster aggregates. In this work, we review the development of bio-inspired algorithms carried out by our group over the past ten years, including their application, for the first time, to discover putative global minima of some relevant aggregates (see Refs. [2] and [64] and references therein). We also propose a novel self-adaptive approach [86] that aims to enhance optimization effectiveness when exploring the energy landscape of atomic clusters with unknown properties. There are several reports in the literature that advocate the development of adaptive bio-inspired approaches, spanning from the early Evolutionary Strategies [87] efforts to the recent area of hyper-heuristics [88]. They all share the common goal of developing an optimization framework that autonomously adapts to the problem being addressed. This adaptation simplifies the task of the practitioner that does not have to rely on expert knowledge to tailor an algorithm to a specific situation and enhances the ability of the method to effectively sample the search space of the problem under study. Self-adaptation may occur at the parameters level and/or at the algorithmic components [88, 89]. Also, it may consider the selection of a subset of existing components or it may foster the discovery of a novel optimization strategy. Here we focus on a framework that selects EA algorithmic components (transformation operators and diversity measures) and favorable settings. Preliminary results presented in Section 5 suggest that the self-adaptive EA is able to autonomously identify the best components that enhance the likelihood of discovering high quality solutions. Our expectation is that, in the near future, such novel methods can ultimately be applied to the study of biochemical systems, like

protein folding or drug design.

The application of global optimization methods may be carried out together with the calculation of the interaction energy at either the ab initio or DFT levels of theory. This approach has been recently reviewed by Heiles and Johnston [90]. A less computational demanding alternative for calculating the interaction energy consists in the use of semi-empirical potential energy functions. We follow here the latter approach. Thus, some well-established potential models were employed: (i) sum of simple pair-potential functions for the clusters resulting from the microsolvation of alkali-metal ions with argon, and (ii) the usual Gupta potential for the transition-metal alloy, with the parameters taken from Clari and Rosato [91]. Specifically, we have studied $Li^+(Ar)_N$, $Na^+(Ar)_N$ and $K^+(Ar)_N$ clusters as well as bimetallic $(RhCu)_N$ aggregates. These are relevant chemical systems that can be also appropriate to show the flexibility and robustness of the EAs. Accordingly, the main features of our bio-inspired algorithms are described in Section 3.2, while examples of their application to the microsolvation of alkali-metal ions and $(RhCu)_N$ clusters are reported in Section 3.3 and Section 3.4, respectively. In Section 3.5, we present a self-adaptive framework that can autonomously identify the best components for an effective cluster geometry optimization EA. Finally, the main conclusions are given in Section 3.6.

3.2 Evolutionary Cluster Optimization

The first work describing the application of EAs to cluster geometry optimization was published in 1993. Hartke pioneered this area by applying what can be considered a naïf binary encoded algorithm to the optimization of Si_4 atomic clusters[59]. These first attempts were quite inefficient and were unable to compare with existing state-of-the-art techniques, which were mainly based in stochastic single point search methods (e.g., simulated annealing). In 1995, a breakthrough paper from Deaven and Ho completely changed the area of evolutionary cluster optimization and set the foundations for modern algorithms [92]. They describe the application of an EA to carbon clusters and propose several novel components that clearly enhanced the effectiveness of this class of methods: i) the coordinates of the

atoms composing the aggregate are codified as real values, thus departing from the binary encoding (this same modification was simultaneously proposed by Zeiri [93]); ii) genetic operators are applied directly in the 3D space, thus preserving some semantic properties of the solutions. The proposed crossover is considered the first cut and splice like operator; iii) conjugate gradient minimization is applied to all solutions generated by the EA, moving them to the nearest local optimum; iv) a simple mechanism prevents identical individuals from belonging to population. An energy-based similarity is considered. Over the last twenty years, the proposals of Deaven and Ho have been improved with the aim of further enhancing the effectiveness of EAs for cluster geometry optimization [94, 95, 16, 62]. Two key issues have been considered by researchers: the development of genetic operators that are aware of the properties of the solutions when transforming them [92, 96, 97, 98] and the definition of enhanced distance measures that better estimate the existence of duplicates/identical solutions in the population [99, 62, 100, 16, 101].

Our group has been developing cutting edge EAs for cluster optimization for over a decade [102, 103, 16, 104, 105]. The most distinguish feature of our approaches is the unbiased nature of the search performed by the algorithm. Although this might compromise absolute effectiveness, it allows for an increased robustness, thus fostering its application to novel chemical systems for which specific knowledge about the general organization of the global minimum is not available. Considering existing global optimization approaches, our algorithms comprise two key enhancements: improved set of genetic operators [16, 103] and a suite of distance measures that effectively maintain diversity throughout an optimization run [16]. The representation of solutions builds upon the proposals of Zeiri and Deaven and Ho. In accordance, to represent an atomic aggregate with N particles, the chromosome is composed by 3 X N real values codifying the Cartesian coordinates of each particle. In turn, when representing molecular clusters, another set of 3 real values is added to each particle in order to encode the Euler angles that describe the orientation of the corresponding molecule in the 3D space [2]. All genes have a predetermined minimum and maximum bound: for the coordinates they define a 3D box in which the particles must be placed, whereas the orientation components range between 0 and 2π. The EA explores the search space defined by these settings, trying to identify the cluster with the global minimum energy.

To enhance search efficiency, all solutions are locally optimized before evaluation and therefore the solution pool is composed just by local optima. L-BFGS is the quasi-Newton standard local search procedure adopted to hybridize evolutionary cluster geometry optimization, as it cleverly combines modest memory requirements with fast convergence [63]. Several works describe modified local search procedures to speed up convergence and to increase the likelihood of discovering promising solutions. The so-called two-stage local search considers a modified biased potential that changes the potential energy surface and enhances the likelihood of discovering nonicosahedral global minima [83, 96]. This modified potential comprises a set of parameters that favor the appearance of specific shapes (e.g., ellipsoidal vs. spherical). As these settings must be specified prior to the optimization, this local search strategy cannot be used in situations where the shape of the global optimum is not known. In the algorithm applied in this work, only the standard L-BFGS is considered. Selection of parents is performed with the tournament operator, a method that allows for an effective adjustment of the selective pressure [12]. In a previous work we proposed a structural elite operator that was able to keep the best solutions in the population, together with the promotion of clusters with diverse structural shapes [106]. In our most recent publications, we have been proposing a more efficient steady-state approach: it relies on a fitness-based replacement strategy, in which the new solutions generated by the EA replace the worst individuals from the population, providing that they have better quality (i.e., lower potential) [16]. To prevent premature convergence, a diversity mechanism forbids the coexistence of similar solutions in the pool. Similarity between individuals can be accessed either by fitness or by structural properties. Results obtained with several chemical systems demonstrate the advantage of relying in structural distance measures [16].

Cut and Splice crossover operators are clearly the best option to combine sub-clusters of parents when creating novel solutions [92, 95, 97, 103, 98]. Their distinctive feature is the operation in the 3D space, aiming at preserving some semantic properties of the parents. Different flavors exist in the literature, depending on the constraints considered when applying the operator (e.g., selection based on a cutting plane vs. selection based on minimal Euclidean distance) and on the type of cluster being optimized (atomic/molecular; homogeneous/heterogeneous). As for mutation, two classes of operators are usually considered:

random mutation takes a particle and chooses a new arbitrary location for it inside the 3D box defining the search space; sigma mutation slightly perturbs the coordinates of a particle with a value sampled from a normal distribution [16].

3.3 Microsolvation of alkali-metal ions

The theoretical investigation on the microsolvation of alkali-metal ions by water and other molecules has been recently reviewed by us [Phil. Trans. R. Soc.]. The substitution of the solvent molecules by atoms reduces the complexity of the theoretical problem, both in modeling the potential energy interaction and during the search of the global minimum structure (which, now, does not depend on orientational degrees of freedom). Because of this, we have chosen the microsolvation study of alkali-metal ions with argon atoms to illustrate the application of our global optimization methodology.

The study of the microsolvation of alkali-metal ions with argon constitutes a good starting point to understanding the solvation phenomena involving these ionic species in the more complex biological environment. To establish the interaction potential model, we have employed two-body functions. While the Ar-Ar pair-potential is represented by the function Rydberg-London function of Cahill and Persegian [107], the interaction between the alkali-cation and argon was modeled as:

$$V(R) = aR^c \exp(-bR) - \chi_{pol}(d_{pol}, R)\frac{C_4}{R^4} - \chi_{disp}(d_{disp}, R)\frac{C_6}{R^6} \tag{3.1}$$

where a, b, c, C_4, C_6 are fitting parameters, while the damping functions for both polarization and dispersion components have the general form:

$$\chi(d, R) = \begin{cases} 1 & R > d; \\ \exp\left[-(d/R - 1)^2\right] & R \leq d. \end{cases} \tag{3.2}$$

We should note that d refers to d_{pol} (d_{disp}) in the calculation of polarization (dispersion) component. The parameters have been obtained by fitting the function to *ab initio* data

Table 3.1: Parameters of the Na^+-Ar and K^+-Ar potentials that were obtained by fitting Eqs. (3.1) and (3.2) to CCSD(T) *ab initio* energies. See the text.

Potential parameters	**Systems**	
	Na^+-Ar	K^+-Ar
a/eV	5040.47	4523.69
b/Å^{-1}	4.25	3.478
c	1.815	0.25
C_4/eV Å^{-4}	8.34	11.078
C_6/eV Å^{-6}	181.207	79.51
d_{pol}/Å	7.95	3.415
d_{disp}/Å	2.60	3.669

calculated at the CCSD(T) framework by using GAMESS package [108] and a basis set with quadruple-zeta quality for Li^+-Ar and Na^+-Ar (cc-pVQZ for lithium [53] and aug-cc-pVQZ for argon [109]) and with triple-zeta quality for K^+-Ar (ATZP basis set [110, 111]). The *ab initio* electronic energy were corrected for the basis-set superposition error (BSSE) with the counterpoise method [112], and, during the calculations, 7 and 10 orbitals were frozen for Na^+-Ar and K^+-Ar systems, respectively. The values of the parameters are given in Table 1, according to the fit performed using 25 and 26 ab initio points distributed in the intervals of 2.3 up to 10.0 Å and 2.5 up to 10.0 Å for Na^+-Ar and K^+-Ar, respectively, while the corresponding ones for Li^+-Ar have been published in Ref. [113]. All ab initio points are available from the authors upon request.

This pair-potential approach appears to be reasonable, since we are mainly concerned with the structural features of the microsolvation clusters and the three-body interactions are not expected to affect too much the geometries arising from models based on two-body functions. Actually, we have recently investigated the microsolvation of Li^+ by argon [113] and we showed that the inclusion of three-body interactions to model the $Li^+(Ar)_N$ clusters is relevant for describing energetic features, but it is less important to establish the global minimum geometry. In this reference [113], we have showed that, in general, the global minima of the potential with three-body interactions have always higher energy and, usually, lower symmetry than the corresponding ones for the PES that includes only two-body terms. Indeed, the $Li^+(Ar)_N$ clusters obtained from a potential based on pair-wise interactions failed the main energetic features up to N = 10, and the $Li^+(Ar)_2$, $Li^+(Ar)_3$ and

$Li^+(Ar)_{10}$ structures become wrong when compared with the ones from the MP2 optimization and the single-point CCSD(T) calculation for the MP2 geometries. For the other hand, the structures obtained using the potential energy surface (PES) including three-body terms had a good agreement with the corresponding ones optimized at the *ab initio* level up to N=8. For larger clusters, additionally, the comparison between potentials with and without three-body forces showed significant energetic and some structural differences for various of the cluster sizes. However, the results obtained with both surfaces were able to explain the high stability peaks for structures with "magic numbers" N= 4, 6, 14, 16 and 34 that arises from the experimental mass spectral intensities [114].

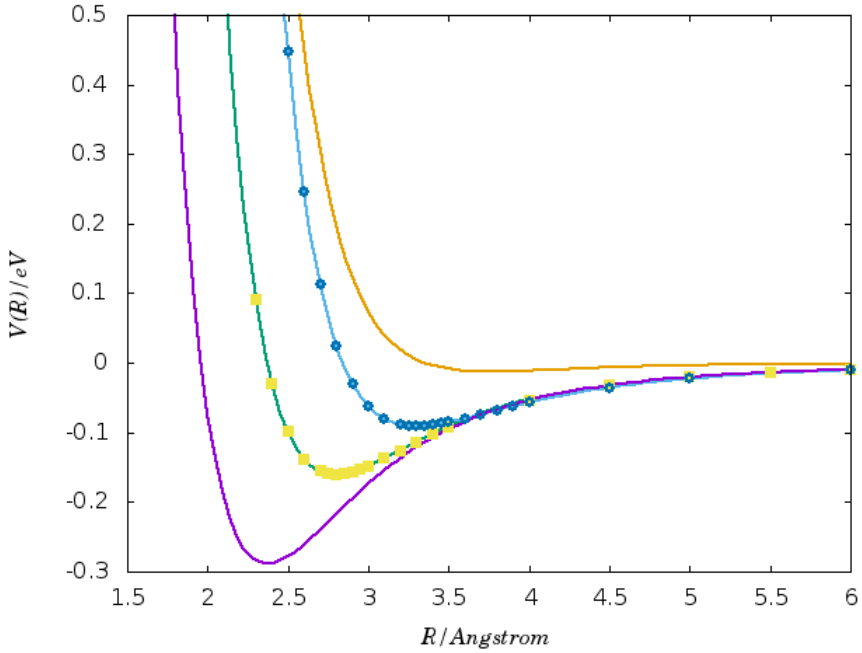


Figure 3.1: Potential energy curves for Ar-Li^+ (magenta), Ar-Na^+ (green), Ar-K^+ (blue) and Ar-Ar (orange) interactions. The Ar-Ar curve is from Ref. [Cahil], while the ion-Ar ones correspond to fits of Eq. (3.1) to ab initio points (also represented for Ar-Na^+ and Ar-K^+).

In the remaining of this section, we report new results for $Na^+(Ar)_N$ and $K^+(Ar)_N$ (with N=2-40) and compare them with previous studies on similar systems. In Figure 3.2, we show the putative global minimum structures for $Na^+(Ar)_N$. This system has been previously studied by Rhouma et al. [9] by employing two potential models that include 3-body interactions. In spite of some discrepancies for small-size clusters, where 3-body effects are expected to be more apparent, the present results give similar global minimum structures

up to N=10; exceptions arise only for $Na^+(Ar)_2$ and $Na^+(Ar)_3$. Like for $Li^+(Ar)_N$, the clusters grow up by surrounding the Na^+ by an increasing number of argon atoms. A similar pattern is observed in Figure 3.3 for $K^+(Ar)_N$. Nonetheless, several structural differences arise for the solvation of Li^+, Na^+ and K^+ with argon. Whereas the maximum coordination number is 6 for the $Li^+(Ar)_N$ clusters [113], it increases to 8 (12) for the solvation of Na^+ (K^+), as one can observe from Figure 3.2 (Figure 3.3). This may be attributed to the increasing size of the ions, which is apparent from the diatomic curves displayed in Figure 3.1: the equilibrium geometries are 2.37 Å, 2.85 Å and 3.33 Å for Li^+-Ar, Na^+-Ar and K^+-Ar, respectively. Although such a behavior with the increasing size of alkali-metal ion may be expected for other solvents, exceptions have been observed when there is a certain degree of competition between ion-solvent and solvent-solvent interactions (see, e.g., Refs. [115, 116, 117, 118] for the microhydration case). We show in Figure 3.4 that, for most cases, the ion is displaced off the center of the cluster. In fact, during the completion of the first solvation shell the ion occupies the center of the structure only for the clusters $Li^+(Ar)_4$, $Li^+(Ar)_6$, $Na^+(Ar)_6$, $Na^+(Ar)_8$, $K^+(Ar)_9$ and $K^+(Ar)_{12}$. It is worth noting that such clusters correspond to maxima of the second energy difference, i.e.,

$$\Delta_2 E = -2E_N + E_{N-1} + E_{N+1} \tag{3.3}$$

where E_N, E_{N-1} and E_{N+1} are the energies of clusters with *N*, *N-1* and *N+1* argon atoms, respectively; the $\Delta_2 E$ curves are represented in Figure 3.5 and the maxima are usually designated as "magic numbers". For the three alkali-ions, the strongest "magic number" is associated with the closure of the corresponding first solvation shell. Whereas the $Li^+(Ar)_6$ cluster has O_h symmetry, the $K^+(Ar)_{12}$ structure is an icosahedron. In the case of Na^+, though the first solvation shell closes at N=8 (in agreement with the work of Rhouma et al. [9], which uses a potential model with explicit three-body terms), another strong peak occurs for the $Na^+(Ar)_{10}$ cluster; we note from Figure 3.2 that both $Na^+(Ar)_8$ and $Na^+(Ar)_{10}$ structures belong to the D_{4d} point group of symmetry, but two of the argon atoms of the latter are already in the second solvation. Concerning larger clusters, the $K^+(Ar)_N$ global minimum structures tend to be more symmetric than the $Na^+(Ar)_N$ ones. Like for Ar_N

clusters, the $K^+(Ar)_N$ global minima are icosahedric-type structures, which is against the global optimization result achieved for a pair-potential model by Hernández-Rojas and Wales [119], and that predicts icosahedral packing only beyond N=49. Indeed, the equilibrium distance and the well-depth of the K^+-Ar bond are both more similar to the corresponding values in the Ar-Ar interaction than the Li^+-Ar and Na^+-Ar ones. It is also clear from Figure 3.4 that, in the case of large clusters, the Li^+ ion tends to occupy a more central position than Na^+ or K^+, which may be explained by the fact that the Li^+-Ar bond is strongest one (see Figure 3.1). In turn, only very small peaks of the $\Delta_2 E$ curves are observed for the larger clusters in Figure 3.5, which emphasizes the small contribution of each Ar-Ar interaction in the second solvation shell for the total energy of the aggregate. Nonetheless, it is worth noting that we can reproduce the mass-spectra "magic numbers" observed in time-of-flight experiments [120] for $K^+(Ar)_{12}$, $K^+(Ar)_{18}$ and $K^+(Ar)_{22}$, while a previous global optimization study by Hernández-Rojas and Wales [119] estimated a completely different series of clusters with enhanced stability.

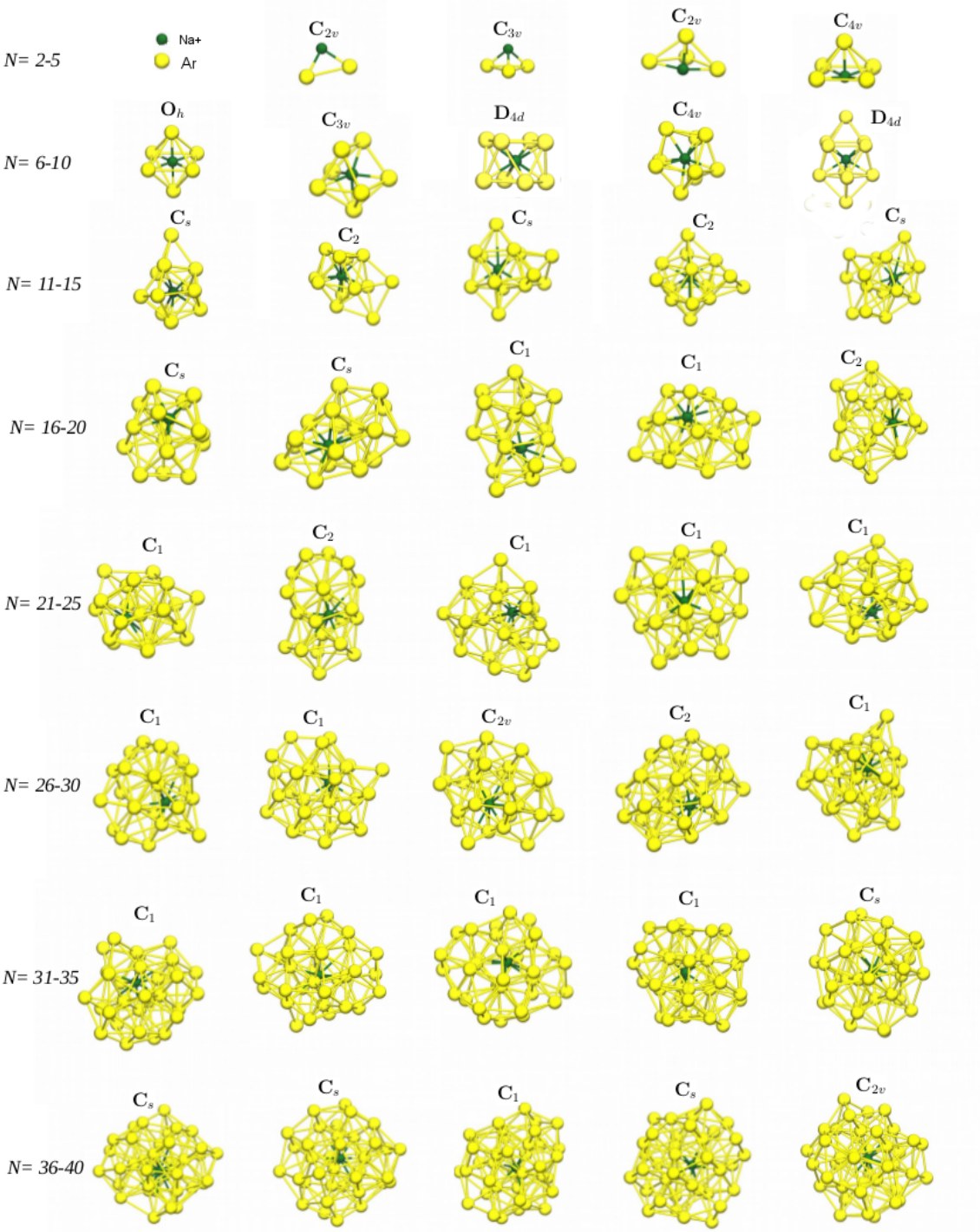


Figure 3.2: Putative global minimum structures for the $Na^+(Ar)_N$ clusters. The Cartesian coordinates are available from https://apps.uc.pt/mypage/faculty/qtmarque/en/clusters.

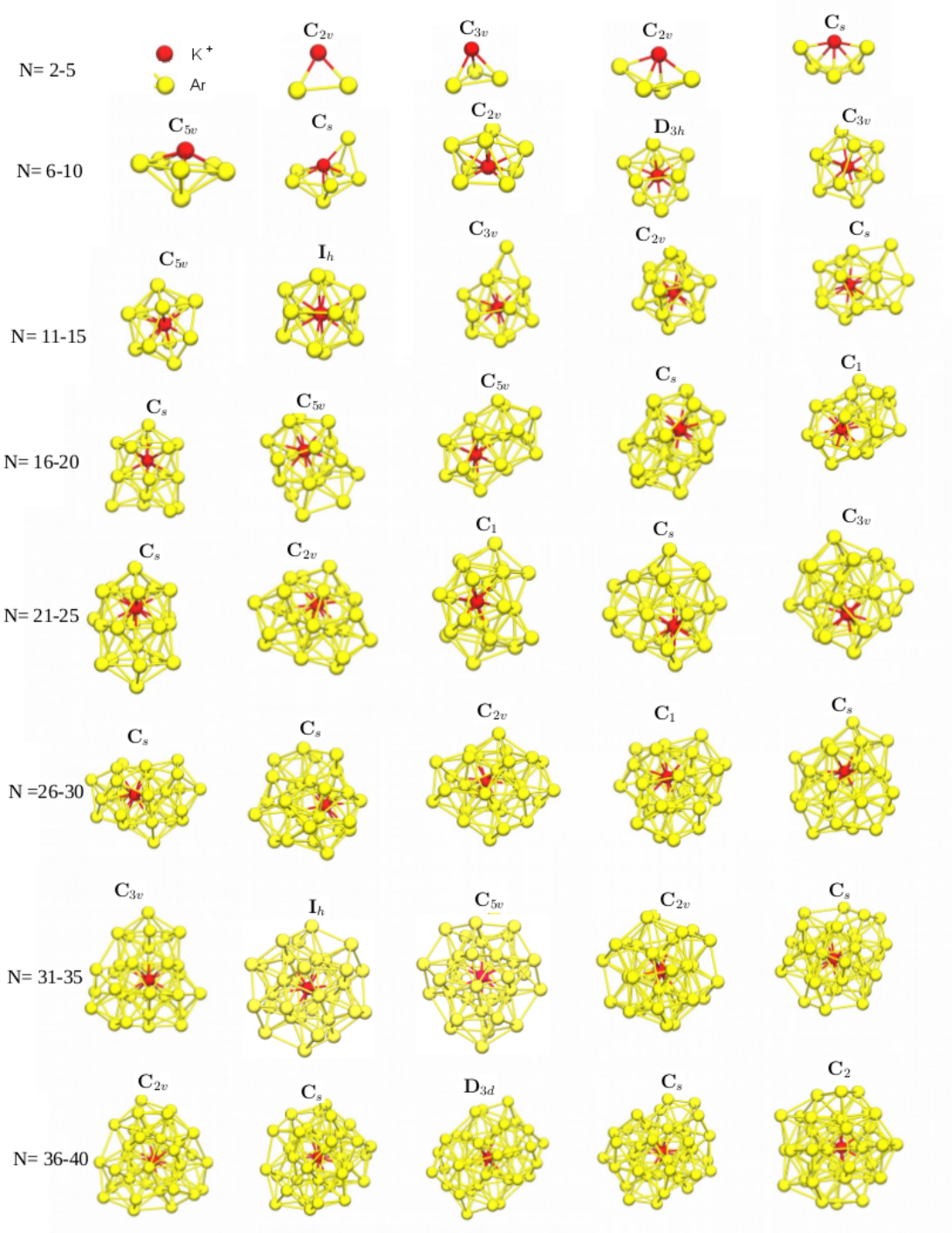


Figure 3.3: Putative global minimum structures for the $K^+(Ar)_N$ clusters. The Cartesian coordinates are available from https://apps.uc.pt/mypage/faculty/qtmarque/en/clusters.

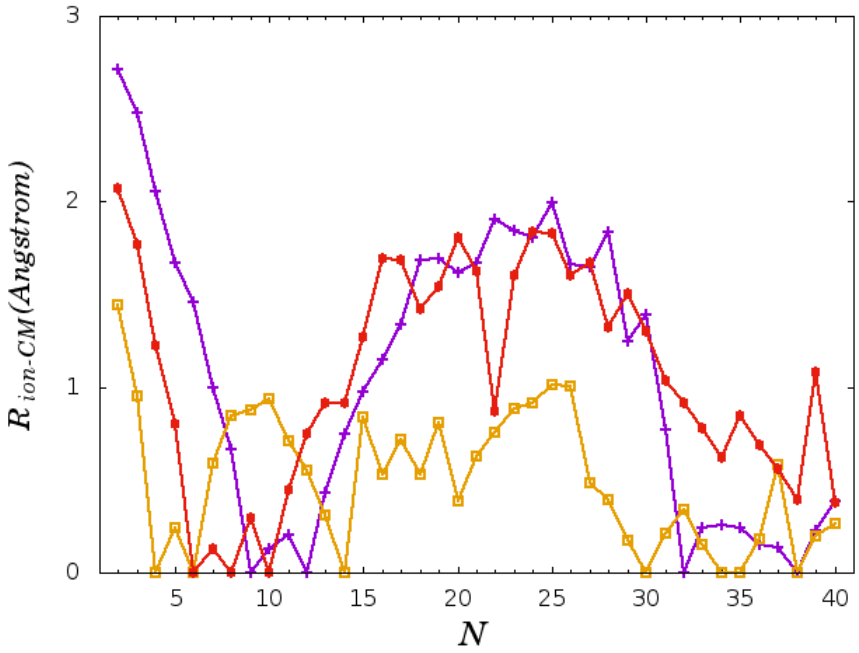


Figure 3.4: Distance separating the ion from the center of mass of the cluster: solid line (orange), $Li^+(Ar)_N$; dashed line (red), $Na^+(Ar)_N$; dotted line (magenta), $K^+(Ar)_N$.

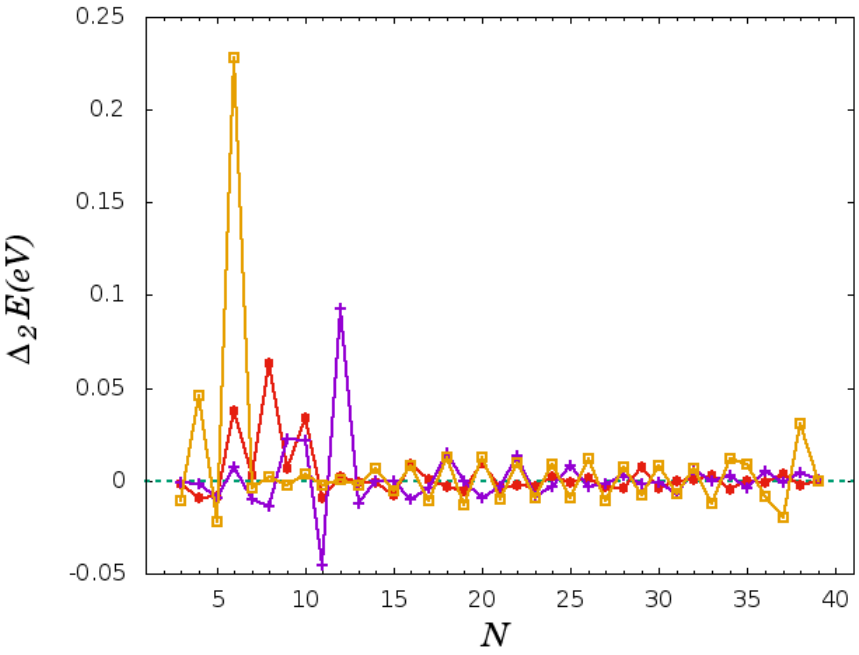


Figure 3.5: Second energy difference of the global minimum structures: solid line (orange), $Li^+(Ar)_N$; dashed line (red), $Na^+(Ar)_N$; dotted line (magenta), $K^+(Ar)_N$.

3.4 Binary transition-metal clusters

Metal alloys constitute a stringent test for global optimization algorithms, since there are several minima resulting from swapping different types of atoms within the same structure (i.e., the so-called homotops [121]). Our EA has been applied to study several two-component systems, including binary Lennard-Jones clusters [101], mixtures of rare-gas atoms [104], Zn-Cd alloy [122] and colloidal aggregates [123, 105]. In this section, we analyze the global minimum structures of binary clusters of the Cu-Rh alloy that have been searched with our EA. Figure 3.6 shows the putative global minimum structures of the $(RhCu)_N$ (N=2-21) clusters. In spite of having atoms of two different species, there are some symmetric structures especially for small-size clusters, and even the $(RhCu)_{13}$ global minimum has C_{3v} symmetry. Due to the difference in the cohesion energies of rhodium and copper (they are [124] 5.75 eV/atom and 3.49 eV/atom, respectively), the tendency of the growing clusters is to keep the Rh atoms next to each other, while increasing the Rh-Cu nearest-neighbor interactions as much as possible. Thus, it is apparent for larger clusters that the Cu atoms segregate on the surface of the structure, whereas rhodium occupy preferentially more central positions. Since the number of atoms of each type is the same and the radius of Rh is larger than the Cu one, we cannot observe a perfect core-shell structure. Nonetheless, we may say that the global minimum structures display a clear core-shell type of segregation with the copper atoms on the surface. In contrast, we recall here our recent work on $(ZnCd)_N$ clusters [122], where a perfect icosahedral core-shell structure is observed for most sizes.

It is interesting to observe in Figure 3.7 that most of $(RhCu)_N$ (N=2-21) clusters are not stable in comparison with the neighbor-size ones. Indeed, the strongest "magic number" arises for N=19, though two other small peaks occur for N=13 and N=17. In particular, the corresponding N=19 homogeneous cluster resembles a very organized structure that essentially has Oh symmetry, i.e., it displays a truncated octahedron motif (cf. Figure 3.6).

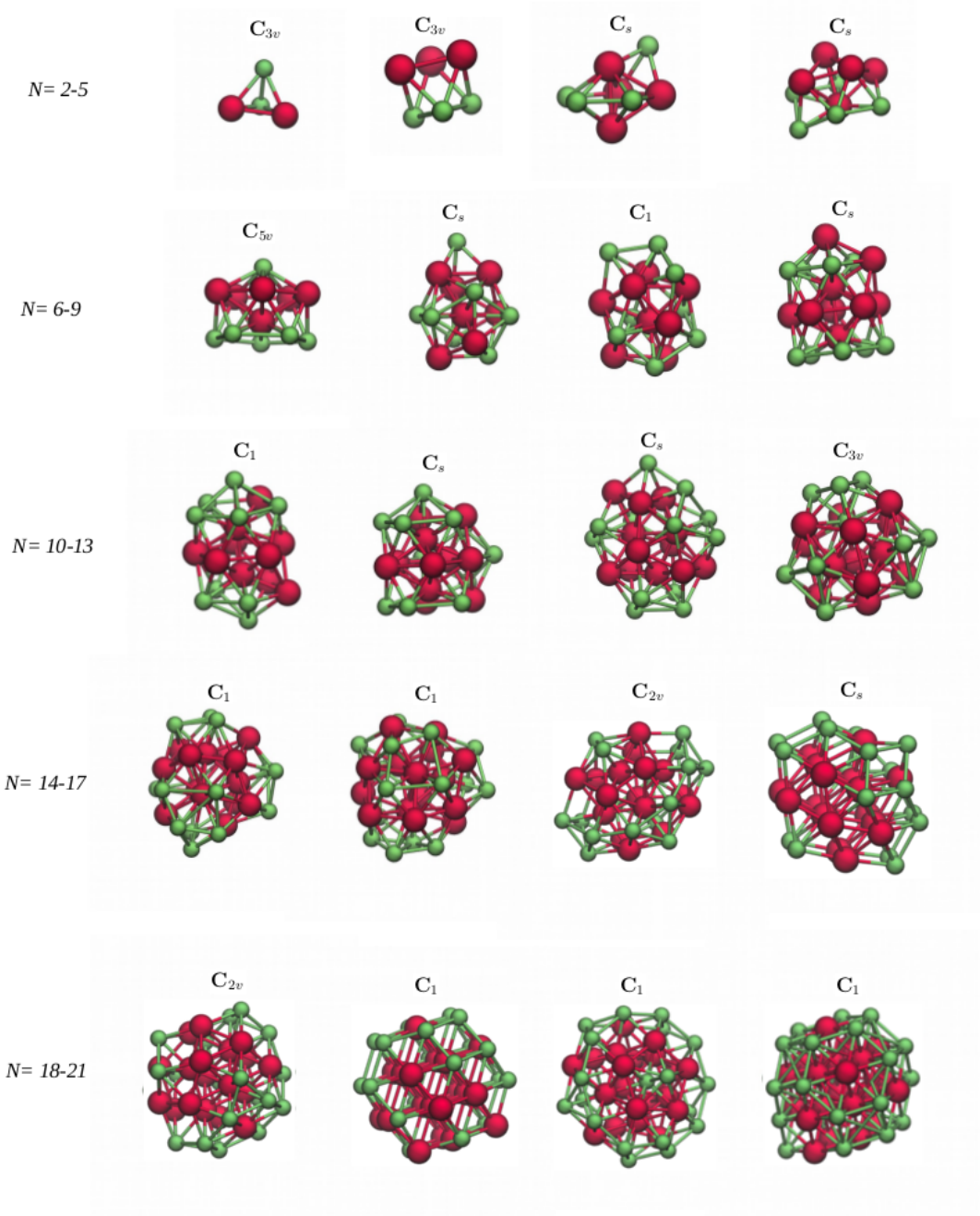


Figure 3.6: Putative global minimum structures for the $(RhCu)_N$ clusters: red (green) spheres represent Rh (Cu) atoms. Also shown are the corresponding point groups of symmetry.

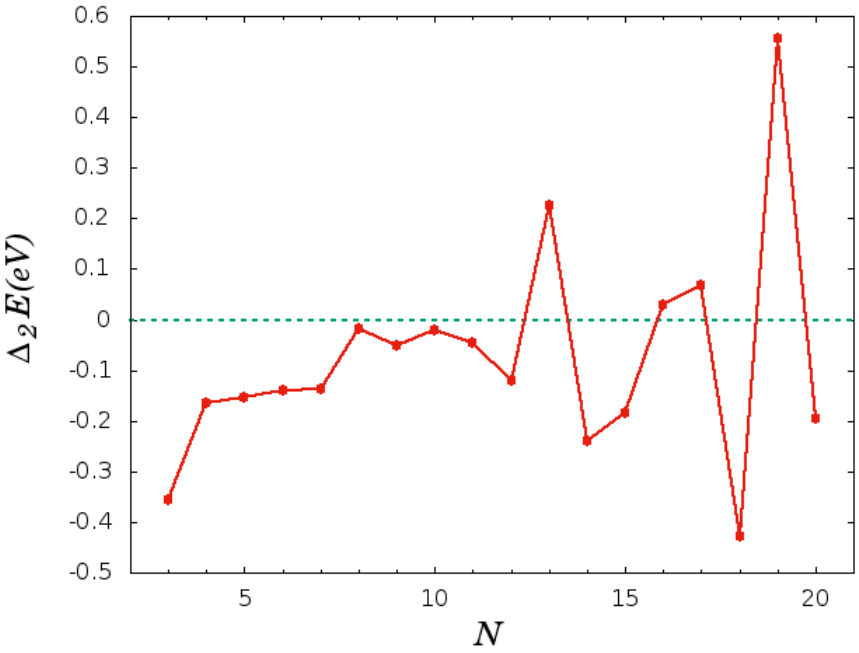


Figure 3.7: Second energy difference of the global minimum structures for the $(RhCu)_N$ clusters.

3.5 New self-strategies for global optimization: application to Morse clusters

Current research in the evolutionary cluster area mainly focuses on enhancing specific components aiming at a better efficiency/effectiveness tradeoff. Remarkable examples are the definition of simple recipes for maintaining diversity or the proposal of novel genetic operators [125].

Such efforts correspond to an exploitation of existing frameworks and we believe that these proposals are not enough to foster the appearance of innovative evolutionary approaches. On the contrary, we consider that truly original ideas should be tested, in order to allow for a relevant enhancement of the behavior of EAs for cluster geometry optimization. Das and Wales recently proposed one interesting idea [126]. They presented a machine learning algorithm to estimate the potential energy level of a given configuration. Results are reported for a simple triatomic molecule and the system is able to make a reliable prediction

about the final local minimum taking a sequence on configurations as input. Despite being tested only in simple systems, the accuracy of the classification algorithm suggests that it might be used to advise a hybrid EA to only perform the costly local minimization step in promising departing configurations.

In this work, we propose a novel self-adaptive EA for cluster geometry optimization. The proposal is directly related to the area of Multimeme Memetic Algorithms [127], which comprises self-adaptive memetic methods that simultaneously co-evolve solutions and settings encoded in each individual. It also draws inspiration from hyper-heuristics [88, 128, 89], a field that deals with the development of computational frameworks whose goal is to automatize the design of methods to solve hard global optimization problems. In concrete, our approach allows individuals to autonomously choose several settings and algorithmic components, thus defining their own search strategy.

A brief perusal of the cluster evolutionary optimization literature confirms that many different components can be selected to create a hybrid EA for this task, e.g., in what concerns the distance measure used to estimate similarity or the transformation operators that are applied to selected parents. Also, a considerable number of settings need to be defined. These are important decisions, as they will clearly impact the behavior of the algorithm. In addition, particularly when dealing with chemical systems with unknown properties, it might be difficult to correctly identify which components are better for that specific task. Finally, an optimization run comprises several stages and different components/settings might be better adapted to specific periods.

The EA proposed in this section has a general behavior similar to the one presented in Section 3.2. However, each solution from the population pool will comprise two components: a regular representation of the coordinates of the particles that compose the aggregate and individual settings that specify how this solution will be processed by the algorithm. We consider the following individual information pertaining the definition of the self-optimization strategy: Crossover type: individuals can adopt either Generalized C&S or standard one-point crossover. Mutation type: individuals can adopt either Sigma mutation or Random mutation. If sigma mutation is active, then the individual also encodes its own

standard deviation, which can take one of the following values: {0.01, 0.05, 0.1, 0.2}.

Distance measure: individuals can estimate similarity to other solutions using Fitness distance, Structural distance or Center of mass distance [16];

Figure 3.8 highlights an example of an individual chromosome. The first part of the solution encodes the Cartesian coordinates of the atoms that belong to the cluster being optimized. The second part contains the selection of components and settings that help to define its own search strategy.

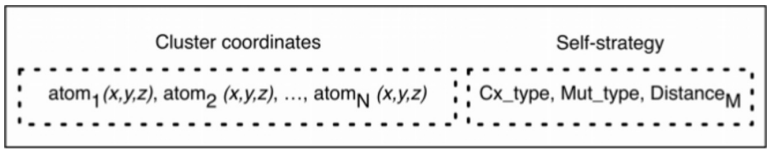


Figure 3.8: General organization of an individual chromosome: each individual comprises the coordinates of the atoms that belong to the cluster and its self-strategy.

The individual settings encoded in a solution are inherited by its descendants. When generating a descendant, encoded self-strategies are mutated at a predetermined rate (0.05 in this study) to prevent stagnation of the optimization behavior. All encoded settings are specific of a single solution, with the exception of the crossover type. When two parents engaged in a crossover operation have different crossover types, one of them is randomly selected.

The self-strategy encoded in solutions is straightforward and several other components could be considered. Also, evolution and cooperation/competition between existing strategies might be performed using other, enhanced, rules. Here, our goal is to propose a simple approach that can act as a proof of concept for the advantage of relying on self-adapted EAs for cluster geometry optimization. In accordance we applied the self-adaptive EA to the 14 hard Morse instances identified in [16], corresponding to the following cluster sizes: 41, 43, 46, 47, 53, 59, 60, 61, 62, 68, 70, 73, 74, 75. We do not seek for an enhanced effectiveness in this specific situation, as the EA success rate reported in the above-mentioned reference results from a carefully tuned algorithm to the potential function being addressed. Following our research hypothesis, our main goal is to verify whether the self-adaptive framework is able to successfully identify the components/settings that foster the discovery of the global

minima for the instances considered in this study. The EA global settings are the same as reported in the above mentioned study: Number of runs: 30; Population size: 100; Tournament Size: 5; Crossover rate: 0.7; Mutation rate: 0.05; Evaluations: $1x10^7$; Maximum Local Search Length: 1000.

As a rule, we verified that the self-adaptive EA was able to find the global minimum in several runs, for all the instances considered in the study. The success rate ranged between 3% and 33% and, for each instance, tended to be slightly below that reported for the fined tuned EA [16]. As mentioned before, this was an expected result, as the optimization algorithm is solving two problems simultaneously: it seeks for the best optimization strategy, at the same time it is trying to find optimal configurations for the clusters being optimized. Given the limited computational budget, we could anticipate that the final absolute optimization performance would be inferior when compared to carefully tuned algorithms that only need to seek for cluster with minimum potential. The most important outcome here is to verify if self-adaptation is identifying the most promising components and results confirm that this is indeed happening. The 3 panels from Figure 3.9 show the prevalence of C&S crossover and Sigma mutation throughout the optimization run: panel a) refers to the Morse cluster with 43 atoms, panel b) for the instance with 68 atoms and panel c) for the instance with 74 atoms. The same trend is visible for other instances. Since the beginning of the run, the percentage of individuals in the solution pool encoding C&S crossover and sigma mutation steadily increases and reaches over 90% for this mutation operator and around 80% for the crossover. The self-adaptive EA is thus able to identify and promote the application of the suite of operators that the literature in this area recognizes as the most effective ones. A detailed inspection of the strategy configurations encoded in global optimal solutions discovered by the hybrid EA reveals that not a single one was obtained with random mutation. This outcome confirms that sigma mutation is mandatory to perform the final adjustments that lead to the discovery of clusters with minimum potential energy. As for crossover, around 75% of global optimal solutions encode the application of C&S crossover, corroborating that semantic aware operators are better equipped to explore the search space. It is interesting to notice that, when taking the largest instances considered in this study ($N \geq 70$), the advantage of C&S vanishes. For these instances, the percentage of global min-

ima encoding this operator drops to about 50%, the same percentage of standard one-point crossover. This suggests that C&S like crossovers might have some difficulties when dealing with clusters above a given number of particles. This is an effect that was not previously reported in the literature and that might be another factor that helps to justify the decrease in the EA success rate, when clusters grow in size. We will return to this topic in a future work.

To complete the analysis of the self-adaptive EA, in Figure 3.10 we present the evolution of the 3 distance measures encoded in the solutions belonging to the solution pool. The 3 panels report results obtained with the 3 instances from the previous figure. The outcomes are similar for instances of different size and confirm that fitness based distances are not suitable for accessing the similarity between atomic clusters. The two structural based distances are predominant in the solution pool and, among the two, the center of mass distance, obtains the higher percentage. This is in line with previous studies that suggest this distance measure as the best option to estimate cluster similarity [62, 16] and supports the hypothesis that a straightforward self-adaptive strategy is able to correctly identify components that enhance the optimization ability of an EA for cluster optimization.

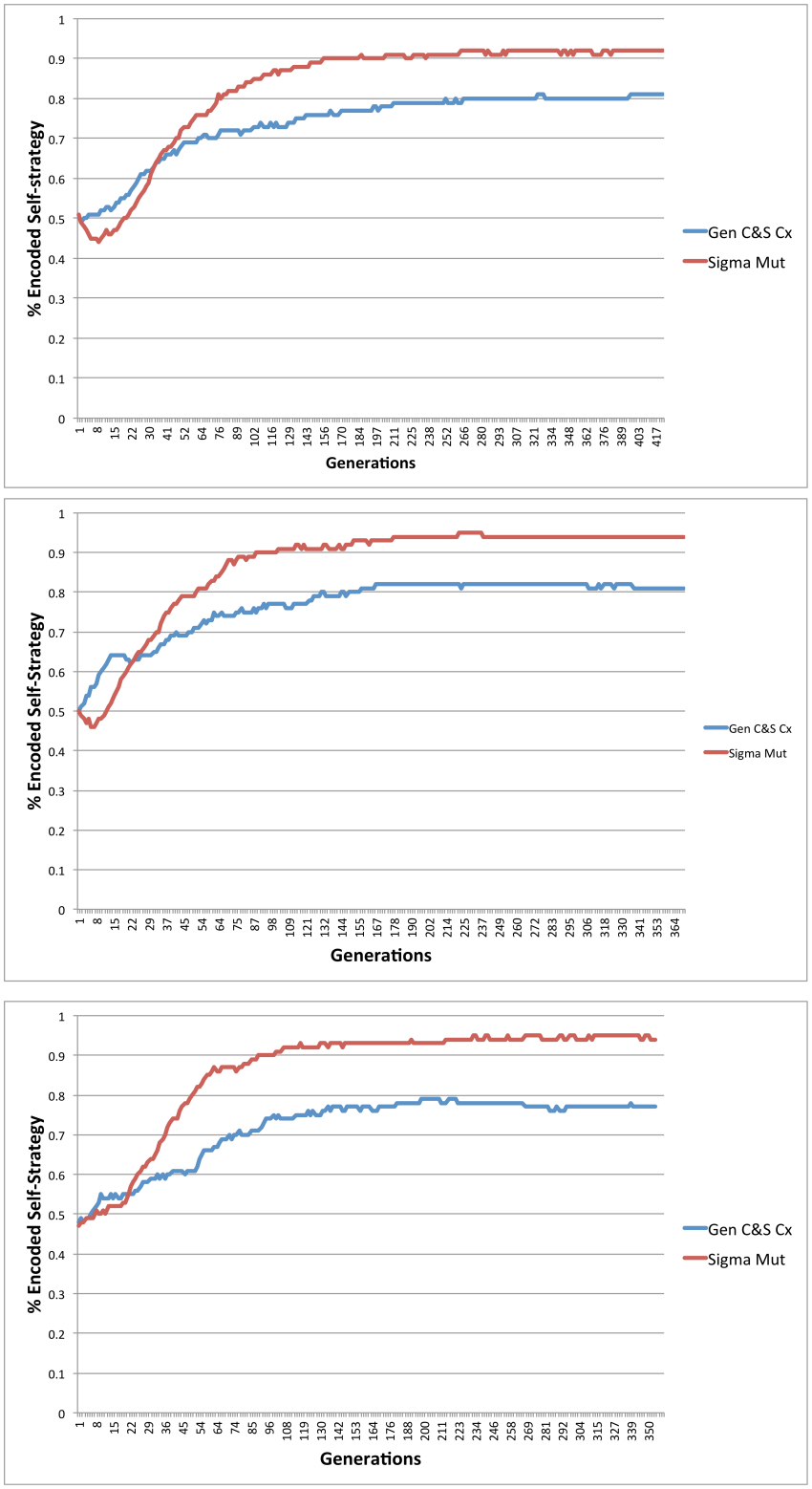


Figure 3.9: Evolution of the percentage of individuals encoding Gen C&S crossover and sigma mutation in the optimization of three Morse instances: a) 43 atoms; b) 68 atoms; c) 74 atoms.

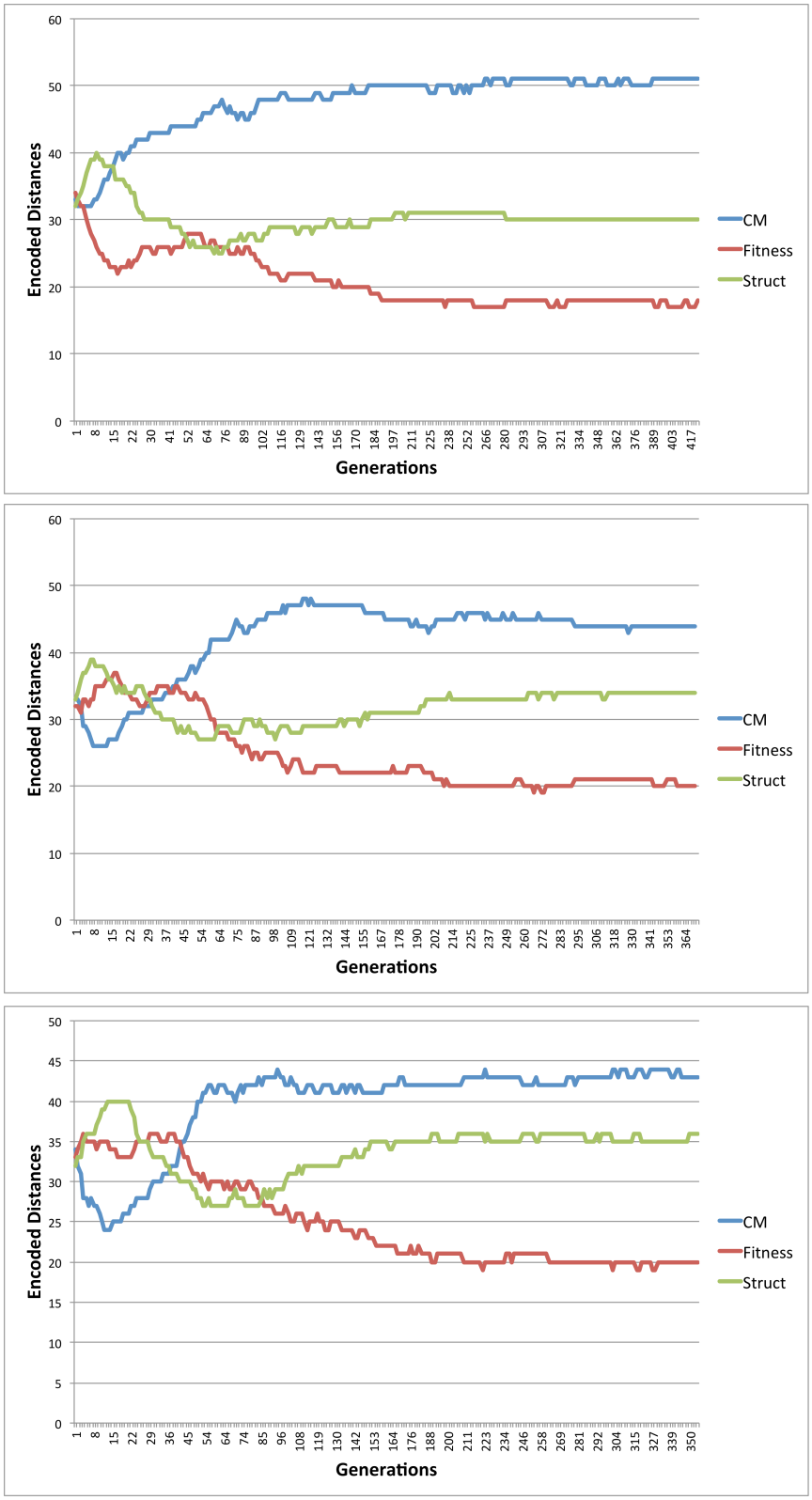


Figure 3.10: Evolution of the percentage of individuals encoding each of the distance measures in the optimization of three Morse instances: a) 43 atoms; b) 68 atoms; c) 74 atoms.

3.6 Conclusions

We reviewed the work of our group on the development of EAs for searching low-energy structures of atomic clusters. To illustrate the potentialities of this methodology, we have applied our EAs to the study of alkali-ion microsolvation with argon and for discovering global minimum structures of Rh-Cu binary clusters. The results for $Li^+(Ar)_N$, $Na^+(Ar)_N$ and $K^+(Ar)_N$ show that strong "magic numbers" are associated with the closure of the first solvation shell. Among these three systems, the global minimum structures for $K^+(Ar)_N$ clusters are, in general, the most symmetric ones. It is also worth noting that simple two-body potential employed for $K^+(Ar)_N$ clusters allowed to reproduce the experimental "magic numbers" observed [Luder] for $K^+(Ar)_{12}$, $K^+(Ar)_{18}$ and $K^+(Ar)_{22}$. Concerning the bimetallic clusters, we have shown that the global minimum structures present segregation of Cu atoms on the surface of the aggregates, which is compatible with the experimental cohesion energies of rhodium and copper. Finally, we should emphasize that EAs are state-of-the-art methods for geometry optimization and, in the past twenty years, they have been successfully applied to chemical systems with distinct properties. In addition, the preliminary results presented in this work suggest that self-adaptation is a promising avenue for future research. We proposed a novel self-adaptive framework that is able to simultaneously discover good quality solutions for a given cluster optimization problem, while it crafts a method specifically adapted to the search space being explored. We hypothesize that this framework is particularly advantageous in situations where knowledge about the structural properties of the system being optimized is not available and our future research will address this important endeavor.

Chapter 4

Exploring the first-shell and second-shell structures arising in the microsolvation of Li^+ by rare gases

An evolutionary algorithm (EA) has been employed to search for the low-energy structures of Li^+Ar_n and Li^+Kr_n clusters up to $n = 14$. Two analytical model-functions have been used to describe the interaction potential: one is based only on a sum of all pair-potentials, while the other includes also three-body interactions. Specifically, a new potential energy surface (PES) for Li^+Kr_n that includes three-body interactions has been modeled by calculating *ab initio* aug-cc-pVQZ/CCSD(T) energies for the Li^+Kr and Li^+Kr_2 fragments; for Li^+Ar_n, we have used our recently developed PES (Phys Chem Chem Phys 2017; 19: 25707-25716). Except for a limited number of cluster sizes from the second solvation shell, the global minimum structures are similar for both Li^+Ar_n and Li^+Kr_n. Modifications in the octahedral structure of the first solvation shell lead to a high-energy penalty, whereas the second solvation shell shows a panoply of minima with similar energies that are likely to be inter-converted, since the relevant PES is apparently flat. The reliability of both analytical potential models to describe low-energy structures was further investigated by performing post-optimization at the MP2 level of theory. For $n = 2$ and $n = 3$, we confirmed that it is essential to include three-body terms in the PES to reproduce the low-energy structures.

In addition, the MP2 calculations indicate that the energy re-order of the global-minimum structure observed for Li^+Kr_8 is related to the Kr_3 Axilrod-Teller-Muto term included in the PES.

4.1 Introduction

Solvation is a complex phenomena with great relevance for the broad area of chemical physics. It is intimately related with the interactions that can be established among solute and solvent molecules, while the contribution from entropic effects can only be correctly described when incorporating all degrees of freedom of the system. Two main theoretical approaches are followed in the study of solvation: (i) continuum solvation models that do not treat solvent molecules as individual entities; (ii) explicit microsolvation methodology, where the solvent molecules are successively added to the system. Although the continuum models are able to incorporate the polarization of the solvent, they cannot explicitly describe interactions with a strong directional character and, in this way, environments involving various solvents, ionic solutes and biological systems are difficult to be accounted for. To somehow reduce these drawbacks, continuum-solvent methods [129] have been improved to approximately incorporate the structure of the solvent [130], as well as the formation of the cavity for solute and the dispersion interaction by including additional empirical-based terms [131, 132]. In turn, the microsolvation approach leads to a higher computational cost, because it has to calculate the contribution to solvation of all possible isomers of the relevant clusters. Nonetheless, microsolvation allows for a detailed understanding of the phenomena at the molecular level, since solute and solvent species are explicitly represented and, thus, all interactions can be incorporated. Particularly important in the study of microsolvation is the exploration of the low-energy structures of the clusters that are representative of the first solvation shells. This may be accomplished by employing state-of-the-art global optimization algorithms, like the ones developed in our group and other groups over the last decade [133], complemented with electronic structure calculations.

For this, one may carry out the global optimization search directly at the *ab initio*

or DFT level of theory [90, 134], but this has the disadvantage of being very time-consuming. As an alternative, it is usual to construct an analytical potential energy surface, which is then used with the global optimization method to generate a relevant set of low-energy structures [135]; another tested approach for generating a promising pool of structures has also been achieved [136, 137] through optimization with standard optimization algorithms coupled to semi-empirical Hamiltonians [138, 139] or the Hartree-Fock method. These are subsequently re-optimized at a higher level of theory by employing *ab initio* or DFT methods (see, *e.g.*, Refs. [140, 141]).

The theoretical and experimental study of the microsolvatation of ions is an important issue for diverse areas of knowledge [142, 143, 144, 145]. In particular, the solvation of alkali-metal ions with rare-gas atoms is theoretically appealing since it leads to an electronic closed-shell system that can be treated by a single-determinant wave function. Moreover, the interaction between the alkali-metal ion and the rare-gas is dominated by electrostatic forces. Hence, it mainly depends on the polarizability of the rare-gas, and on both the charge and radius of the ion. Accordingly, cluster systems involving an alkali-metal ion surrounded by rare-gas atoms have been already addressed in literature [120, 9, 114, 113, 146]. Nonetheless, there is still room to improve the quality of the information about the energy and structural motifs of the clusters that grow during such microsolvation phenomena.

The main motivation for the present work is the identification of similarities/ discrepancies in low-energy minimum structures arising from two types of potential functions used to model the microsolvation clusters of alkali-metal ions with rare-gas atoms. One is based only on a sum of all pair-potentials, while the other includes also three-body interactions between the constituents of the cluster. As test cases we study the solvation of Li^+ in argon and krypton aggregates, and have employed an evolutionary algorithm (EA) to search for the low-energy structures of these clusters up to 14 rare-gas atoms. This is important to visualize the low-energy landscape of the clusters arising in the formation of the first solvation shell and initial steps of the second one. The interaction between the Li^+ and krypton atoms is described by a new potential energy surface (PES) modeled in the present paper. In what concerns the solvation with argon, we use the global minima data obtained in our previous

study [113, 146], but we had to search for other low-energy structures of Li^+Ar_n clusters. Specifically, we have performed a detailed comparison between the low-energy structures of both aggregates obtained by using the two different potential energy functions. For similar structures obtained from both potential energy surfaces, we also focus on the energy order of those minima and, specifically, a post-optimization is carried out by employing electronic structure calculations with the MP2. In the case of global minima, we have also verified whether the Li^+Ar_n structures are distinct from the Li^+Kr_n ones.

The paper is organized as follows. In Section 4.2, we describe the methodology employed in the present investigation. The results are presented and discussed in Section 4.3. The main conclusions are gathered in Section 4.4.

4.2 Methodology

4.2.1 Analytical potential energy surfaces

We have employed two types of potential energy surface (PESs) to model the interactions for both Li^+Ar_n and Li^+Kr_n clusters: one includes up to three-body components in a many-body expansion approach (hereafter designated as PES I), while the other is written as a sum of all pair-potentials (hereafter designated as PES II). Thus, PES I is described by the expression:

$$\begin{aligned} V(\mathbf{R}) &= \sum_j V_{\mathrm{Li^+X}}(R_{ij}) + \sum_j \sum_{k>j} V_{\mathrm{X_2}}(R_{jk}) + \\ &\sum_j \sum_k V_{\mathrm{Li^+X_2}}(R_{ij}, R_{ik}, R_{jk}) \\ &+ \sum_j \sum_k \sum_m V_{\mathrm{X_3}}(R_{jk}, R_{km}, R_{jm}) \end{aligned} \tag{4.1}$$

where the i-index refers to the Li^+ ion and indices j, k, and m label three distinct rare-gas atoms of a cluster system; X refers to either Ar or Kr. In Eq. (4.1), the $V_{\mathrm{Li^+X_2}}$ terms describe the three-body contributions due to the interaction between the dipoles on rare-gas atoms

induced by the lithium ion, while V_{X_3} is given by the well-known Axilrod-Teller-Muto (ATM) long-range potential. Conversely, the three-body interactions are missing in PES II, since it incorporates only the two-body terms of Eq. (4.1). The analytical functions of both two- and three-body terms have been described in our previous work [113] and are, now, detailed in the Supplementary Information.

The Li^+Ar_n potential was modeled in our recent work [113], while the Li^+Kr_n one is built in the present study by following the steps described in that paper. The Li^+X and Li^+X_2 terms of the PESs were fitted to *ab initio* points calculated with the GAMESS program [147] by employing the coupled cluster method that includes single and double excitations plus triple excitations obtained through perturbation theory [*i.e.*, CCSD(T)]. As for the three-body term of both Ar_3 and Kr_3, we have employed the Axilrod-Teller-Muto (ATM) potential [148, 149, 150] whose C_9 parameters were taken from the literature [151]. In addition, the two-body potentials for Ar_2 and Kr_2 were described by the simple Rydberg-London functions reported by Cahill and Parsegian [107].

Specifically for the determination of Li^+Kr_n potential, the CCSD(T) method with the augmented correlation consistent polarization valence quadruple zeta (aug-cc-pVQZ) basis set for the lithium ion [152] and for the krypton atoms [56] has been employed to calculate *ab initio* energies for 29 (209) geometries of the Li^+Kr (Li^+Kr_2) system; 14 electrons for each Kr were frozen. Then, these energies were corrected for the basis-set superposition error (BSSE) by applying the counterpoise method [112]. Finally, such corrected energies were used to fit the corresponding analytical functions. For the most difficult case of fitting the Li^+Kr_2 three-body terms, we have employed the GAFit program [21], which is based on a genetic algorithm [153, 154] and has been designed to significantly facilitate the task of developing potential energy surfaces. All details and the values of the parameters are given in the Supplementary Information.

4.2.2 Geometry optimization procedure

We have employed an EA to search for the low-energy structures of PES I and PES II for the title systems. For each cluster size, the EA starts with an initial pool of

random solutions and iteratively applies state-of-the-art genetic operators to discover the global minimum structures. Each solution encodes the real-valued Cartesian coordinates of the particles composing the aggregate (either Li^+ and n Ar atoms or Li^+ and n Kr atoms). The coordinate values range between 0 and $8.5 \times (n+1)^{1/3}$, where $n+1$ represents the size of the cluster. A minimum distance constraint enforces that the distance between every pair of particles is never below $3.5\,a_0$ ($2.4\,a_0$) for Li^+Ar_n (Li^+Kr_n) clusters.

The iterative processing of the solutions comprises a set of sequential operations. First, tournament selection probabilistically chooses the best solutions of the current pool to be the parents of a new set of clusters. Then, heterogeneous generalized cut and splice crossover [92, 155], a phenotypic genetic operator specifically designed to work with mixed clusters [156], is applied to pairs of parents and obtains two new solutions. The resulting offspring are slightly modified with sigma mutation, an operator that randomly perturbs the $3D$ location of atoms [16]. The new solutions are locally optimized with the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS) [157, 63], a quasi-Newton method that efficiently drives the solution to the local optimum at the bottom of the current basin of attraction on the energy landscape. After local relaxation, the final solution is evaluated with the appropriate function (PES I and PES II). The EA includes a swapping-type procedure in the evaluation step. This operation selects two atoms from different types (*i.e.*, Li^+ and a rare-gas atom), exchanges their positions ands accepts the change if it leads to a decrease in the energy of the resulting cluster. The EA applied in this study adopts an enhanced steady-state replacement strategy. After generating a new set of descendants, they replace the current solutions providing that they have lower potential energy and that no two similar cluster coexist in the pool. Structural similarity between solutions is estimated with the center of mass distance, a measure originally proposed in Ref. [62] and widely used in cluster geometry optimization studies [16]. A steady-state EA model that maintains structural diversity is better equipped to postpone premature convergence and can also maintain a diverse set of good quality local minima in the pool of solutions. The EA iterative cycle is repeated until a predefined number of evaluations of the interaction potential is achieved. The settings of the EA are the following: Number of runs: 30; Population size: 100; Evaluations: 500000 (1000000) for Li^+Ar_n (Li^+Kr_n); Tourney size: 5; Crossover rate: 0.7; Mutation rate:

0.05; Standard deviation for sigma mutation: 0.1.

Since the EA evolves a pool of different solutions, at any point of the optimization we can choose a given number of distinct low energy structures to be printed out. On the other side, additional structures have been searched by carrying out local optimizations departing from geometries obtained with different PESs (see also Sections 4.3.2 and 4.3.3). In order to distinguish between minima and saddle-points, we have performed a normal-mode analysis by diagonalizing the mass-weighted Hessian matrix for each structure. In turn, the comparison between two minimum geometries has been carried out by employing the SAICS program [101], which calculates the root-mean-square distance (RMSD) for the best overlap between the structures and, furthermore, is able to identify enantiomers.

4.3 Results and Discussion

4.3.1 Li^+Kr_2 PES: main features and comparison with Li^+Ar_2

The first step in the establishment of PES I for a given cluster comprises the construction of the PES for Li^+Kr_2 by fitting an adequate analytic function to *ab initio* data that has been corrected for the BSSE with the counterpoise method; a similar procedure has been adopted for Li^+Ar_2 in our previous study [113]. We compare in Figure 4.1(a) the Li^+Kr curve obtained in this work with the Li^+Ar one from Ref. [113]. We note that the Li^+-Kr interaction leads to a deeper potential well, in comparison with the corresponding one for Li^+-Ar; the equilibrium geometry is $4.764\,a_0$ ($4.484\,a_0$) for Li^+Kr (Li^+Ar), where the potential has the minimum value of $-13.1\,mE_h$ ($-10.6\,mE_h$). In turn, the vibrational frequency of Li^+Kr is slightly red-shifted in relation to the value of $267.5\,cm^{-1}$ obtained by us [113] for the lighter Li^+Ar molecule. As shown in Table 4.1 for Li^+Kr and in Ref. [113] for Li^+Ar, our data is in good agreement with the values given in literature by other authors [158, 159, 160, 161, 162, 163].

In addition, we represent for completeness in Figure 4.1(b) the Ar_2 and Kr_2 potential curves [107] used in the present study. The minimum of the Kr_2 (Ar_2) potential energy has

the value of $-0.64\,mE_h$ ($-0.46\,mE_h$) and it arises at an equilibrium geometry of $7.585\,a_0$ ($7.109\,a_0$). It should be mentioned that, despite the simplicity of the analytic form, both Ar_2 and Kr_2 curves are accurate for the purposes of the present work, since they can reliably describe the long-range tail of the interactions as well as the spectroscopically relevant region around the corresponding minimum [107].

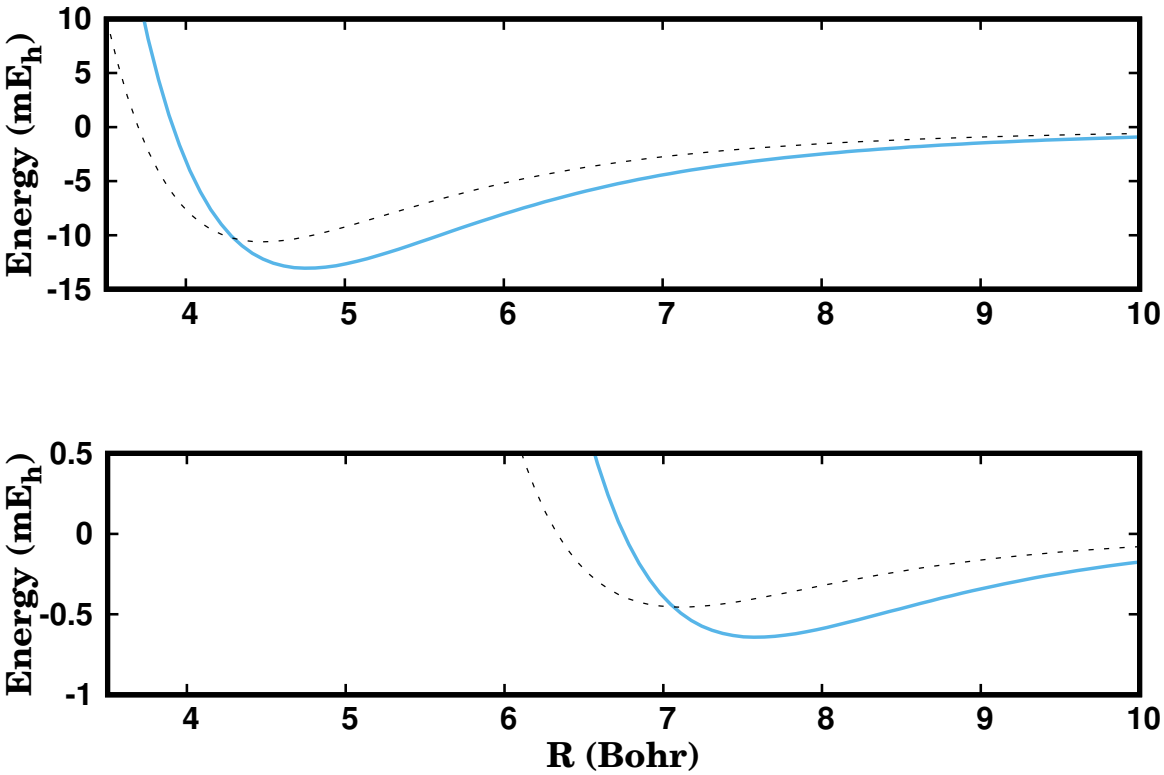


Figure 4.1: Two-body potential energy curves used for PES I and PES II: (a) Li^+Ar and Li^+Kr; (b) Ar_2 and Kr_2. The Li^+Kr curve has been fitted in this work to *ab initio* data, while the potential function for Li^+Ar (Ar_2 and Kr_2) has been taken from Ref. [113] (Ref. [107]). Solid (dashed) lines are used for systems with krypton (argon) atoms.

Regarding the Li^+Kr_2 potential, we represent the geometry and energy parameters from PES I and PES II in Table 4.1. In comparison to both PES I and available theoretical data, PES II tends to overestimate the Li^+Kr_2 dissociation energy; a similar finding has been previously reported [113] for Li^+Ar_2. Furthermore, the geometry given in Ref. [160] for

Li^+Kr_2 has θ_e= 149.9º (*i.e.*, about 14º smaller than the corresponding value for Li^+Ar_2 [160]), but those authors state that such bent minimum is very shallow and, hence, it should be regarded as quasilinear. Also in line with the results for Li^+Ar_2, the PES I for Li^+Kr_2 underestimates the dissociation energy obtained by Andrejeva *et al.* [160], which has been attributed to the BSSE correction of our *ab initio* data [113]. It is worth noting that the addition of three-body interactions leads to an equilibrium bond angle of 180º in PES I for both Li^+Ar_2 and Li^+Kr_2. Such angle becomes 104.9º (105.5º) for Li^+Ar_2 (Li^+Kr_2) in PES II. The increase of the bonding angle on going from Li^+Ar_2 to Li^+Kr_2 may be due to a longer Kr-Kr equilibrium distance in comparison to the Ar-Ar one.

4.3.2 Structures from PES I: Li^+Ar_n *vs.* Li^+Kr_n

We represent in Figure 4.2 the global minimum structures obtained with PES I of both Li^+Ar_n and Li^+Kr_n (with $n = 2 - 14$). It is worth noting that both systems lead to the same structures for most of the cluster sizes. In a limited number of cases, that is, n = 8, 10 and 12, the global minimum structures of Li^+Ar_n and Li^+Kr_n are different. For those sizes, the lowest-energy structures of Li^+Ar_n correspond to local minima in PES I of Li^+Kr_n and *vice versa*. Although not shown in Figure 4.2, PES II leads to different global minimum structures of the two systems only for n = 13. It appears, then, that the inclusion of three-body terms in the interaction potentials leads to increase the difference between the global-minimum structures of Li^+Ar_n and Li^+Kr_n.

The first solvation shell of both Li^+Ar_n and Li^+Kr_n is build up in the same manner. A linear structure arise for n = 2, while the rare-gas atoms occupy the vertices of an equilateral triangle with Li^+ in the center for n = 3. In turn, a tetrahedral structure is the global minimum for n = 4 and the addition of the fifth rare-gas atom leads to a quadrangular pyramid where Li^+ is slightly above the base plane. Finally, the first solvation shell is completed with six rare-gas atoms forming an octahedron with Li^+ in the center. Since the Li^+-rare-gas interaction is stronger than that involving rare-gas atoms alone, the growing clusters tend to retain the octahedral structure of the first solvation shell.

Table 4.1: Geometry and spectroscopy parameters[(a)] for Li^+Kr and Li^+Kr_2. Energies in cm^{-1} and distances in a_0.

	Li^+Kr		
Source	R_e	D_e	ω_e
PES I, II	4.764	2867	263.4
[159]	4.762	2859	262
[160]	4.743	2899	
[164, 165], experimental	4.58	3204	
	Li^+Kr_2		
Source	R_e	θ_e	D_e
PES I	4.793	180.0	2591.0
PES II	4.764	105.5	3008.0
[160]	4.762	149.9	2710

[(a)] The parameters R_e, D_e, ω_e and θ_e are, respectively, the equilibrium distance, the dissociation energy of only one atom of krypton in both Li^+Kr and Li^+Kr_2, vibrational frequency, and the equilibrium bonding angle.

Actually, the above mentioned differences arising in the global minimum structures of Li^+Ar_n and Li^+Kr_n occur in the second solvation shell. Indeed, we observe in Figure 4.2 that, for $n = 8$, 10 and 12, the Li^+Kr_n global minima are more elongated than the Li^+Ar_n ones; in general, the latter are more symmetric structures. Nonetheless, the two types of structures (*i.e.*, elongated or more spherical) are quasi-degenerated for both microsolvation systems. For instance, the energy of the elongated structure for Li^+Kr_8 (Li^+Ar_8) is 57.52 mEh (51.03 mEh), while the corresponding value for the more spherical one is 57.42 mEh (51.05 mEh). We should mention that the existence of quasi-degenerated structures is an indication of a very flat PES, which is dominated by the weak Ar-Ar (or Kr-Kr) interactions that lead to small potential barriers connecting the different minima. Thus, one expects

a "fluid-like" second solvation shell even at low temperatures (*i.e.*, well-below room temperature), where the Li^+Ar_n and Li^+Kr_n clusters may easily assume different low-energy morphologies.

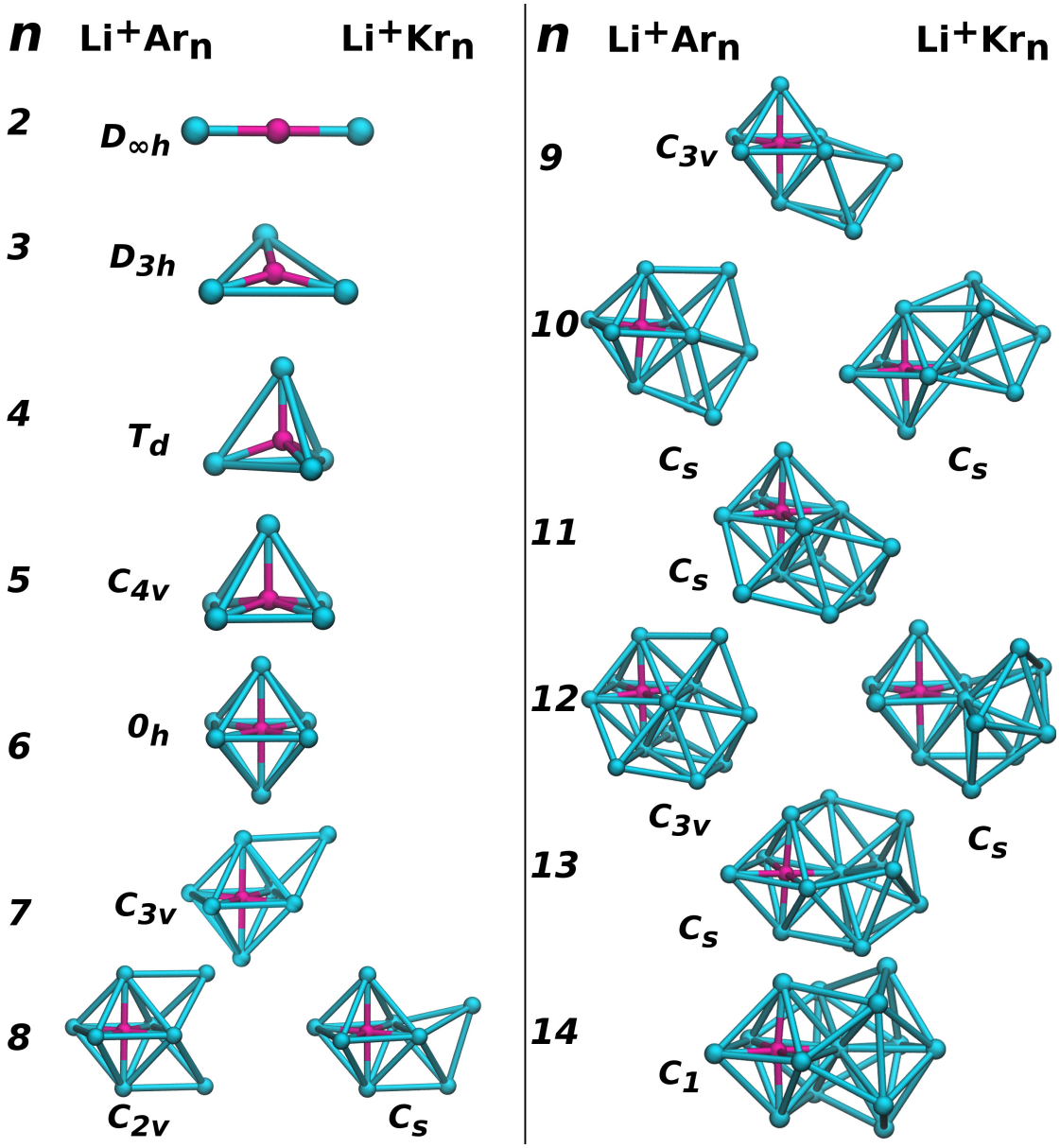


Figure 4.2: Global minimum structures for PES I of both Li^+Ar_n and Li^+Kr_n (with $n = 2 - 14$). In the cases where the two cluster systems lead to different structures, they are shown by separated representations. For each structure, the symmetry point group is also displayed.

A clearly distinct situation arises for the first solvation shell, where the ion-rare-gas interaction is the dominant one. In this case, the rare-gas atoms are very attached to the lithium ion and, hence, putting one of them in any other position leads to a high energy penalty. Because of this, the low-energy landscape essentially comprises the global minimum structure, as it is apparent from Figure 4.3. We observe in this figure that, except structure (i) for $n = 5$, all the other configurations (either minima or saddle points) fall high in energy and differ significantly from the global minimum. We should also emphasize that minima structures are associated to very low (though positive) normal-mode frequencies, which is a characteristic of quite flat regions in the PES. In turn, structure (iii) for Li^+Ar_6 and structures (i) for Li^+Ar_4 and Li^+Kr_4 (Figure 4.3) have one low imaginary-frequency and, hence, they may be assigned as very flat saddle-point configurations.

The analysis of the local minima in the second solvation shell of both Li^+Ar_n and Li^+Kr_n clusters is based on Figure 4.4. In this figure, we represent only minima (in solid) and saddle-points (in dashed) that, simultaneously, are separated among each other by more than $0.1\,mE_h$ and show an RMSD larger than $0.5\,a_0$ when compared to the corresponding global minimum structure; We represent in this figure the energy of low-lying local minima (in solid) and saddle-points (in dashed) as a function of the cluster size (for $n \geq 7$); also indicated through red lines for each cluster size are the lowest-energy structures that have only five rare-gas atoms in the first solvation shell (*i.e.*, without an octahedral core). As a general trend from Figure 4.4, we observe that the Li^+Kr_n clusters have more low-energy minima than the Li^+Ar_n ones. Also, most of the latter tends to be concentrated in a narrow energy-window slightly above the corresponding global minimum (*i.e.*, less than $1\,mE_h$), while the former are usually more separated and spread out over a larger energy-region. In turn, for the smaller clusters it is possible to identify two sets of local minima that are distributed by distinct energy-regions (*i.e.*, essentially below and above $1.5\,mE_h$), which leads to a gap that is larger for Li^+Ar_n. Such energy gap between the two sets of minima suggests significant structural differences. In particular, the lowest-energy structure with five rare-gas atoms in the first solvation shell tend to be among the upper set of minima displayed in Figure 4.4 for each cluster size. In the case of Li^+Kr_n clusters with $n \geq 11$, however, it becomes difficult to make a clear distinction between the two sets of minima. In accordance with this pattern,

the lowest-energy structure with five rare-gas atoms in the first solvation shell for $n = 11$, 12 and 13 arises at around 1.5 mE_h (or even below), which is clearly distinct from the pattern observed for smaller cluster sizes.

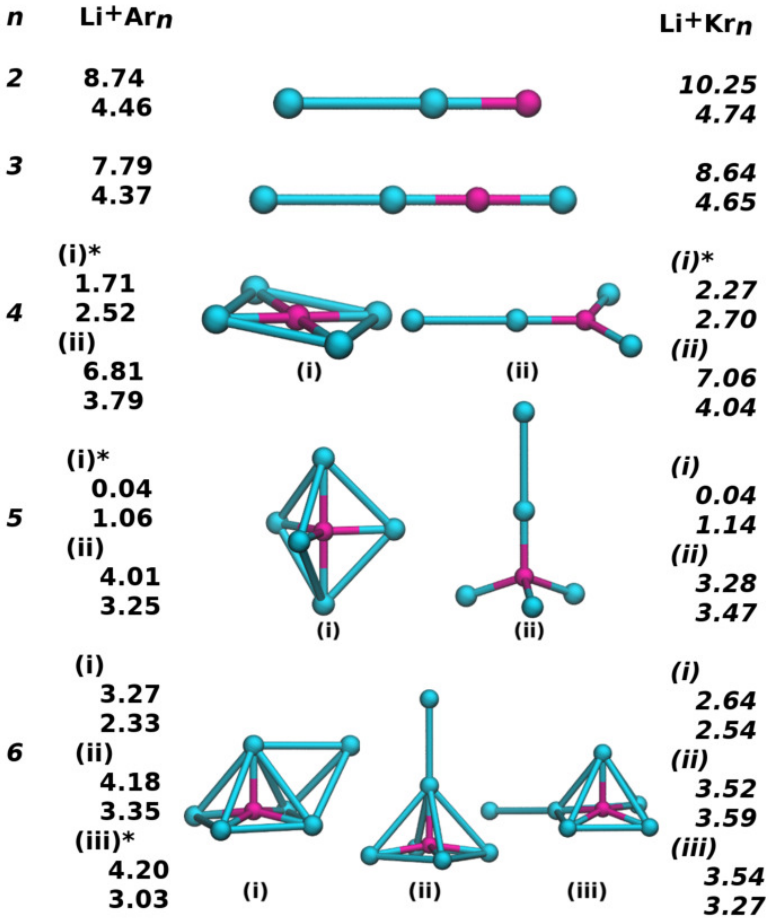


Figure 4.3: Local minimum structures discovered for PES I of both Li^+Ar_n and Li^+Kr_n (with $n = 2 - 6$). An asterisk indicates that the structure is actually a saddle point rather than a minimum. For each structure, the first and the second values are, respectively, the energy (in mE_h) and the RMSD (in a_0), both calculated in relation to the corresponding global minimum.

In Figure 4.5, we represent the lowest-energy structures with only five atoms in the first solvation shell for both Li^+Ar_n and Li^+Kr_n; these structures correspond to the red lines in Figure 4.4. We observe in Figure 4.5 that these clusters show essentially the same

structure for the first solvation shell, *i.e.*, a trigonal bipyramid with the Li^+ and three rare-gas atoms forming the common base. The main difference between the Li^+Ar_n and Li^+Kr_n clusters arises in the second shell for cluster sizes between $n = 8$ and $n = 11$. Whereas the external krypton atoms are always close together (forming a kind of subcluster), the argon atoms are more spread around the first shell, thus, having a greater interaction with the species of the core (mainly with the ion). This may be due to the fact that the interaction among krypton atoms is stronger than for argon. Related to this, we notice that the energy penalty for forming such kind of structure (in spite of the global minimum) is always larger for the Li^+Ar_n clusters. Nonetheless, five out of eight Li^+Kr_n structures show RMSD values in comparison to the corresponding global minimum larger than those for Li^+Ar_n.

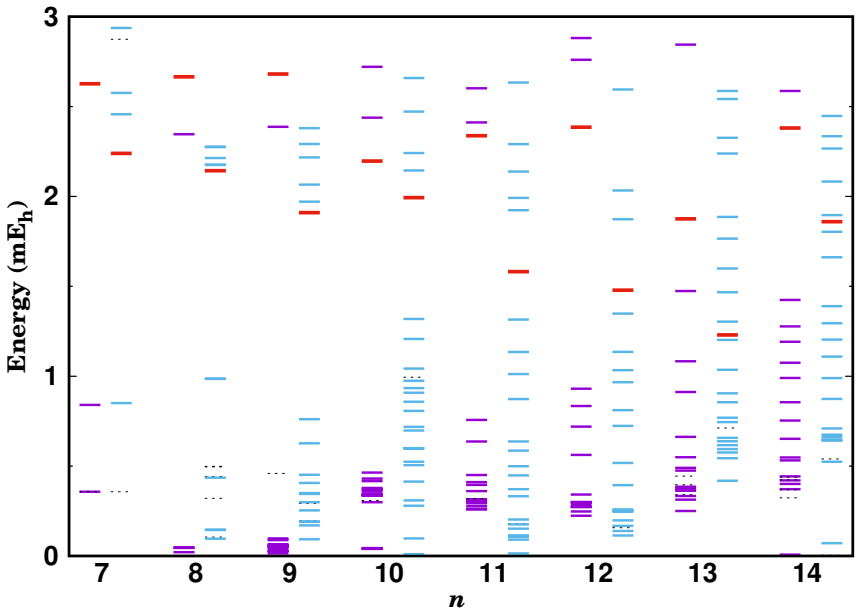


Figure 4.4: Local minima (solid lines) and saddle points (dashed lines) discovered in PES I for Li^+Ar_n and Li^+Kr_n (with $n = 7 - 14$). For each cluster size, the set of energy-minima on the left (right) are for Li^+Ar_n (Li^+Kr_n). Also shown by the red lines are the lowest-energy minima with only five rare-gas atoms in the first solvation shell. Energies are relative to the corresponding global minimum. This representation is not intended to be exhaustive, but it aims to give a correct qualitative picture about the distribution of the minima across the energy-range considered.

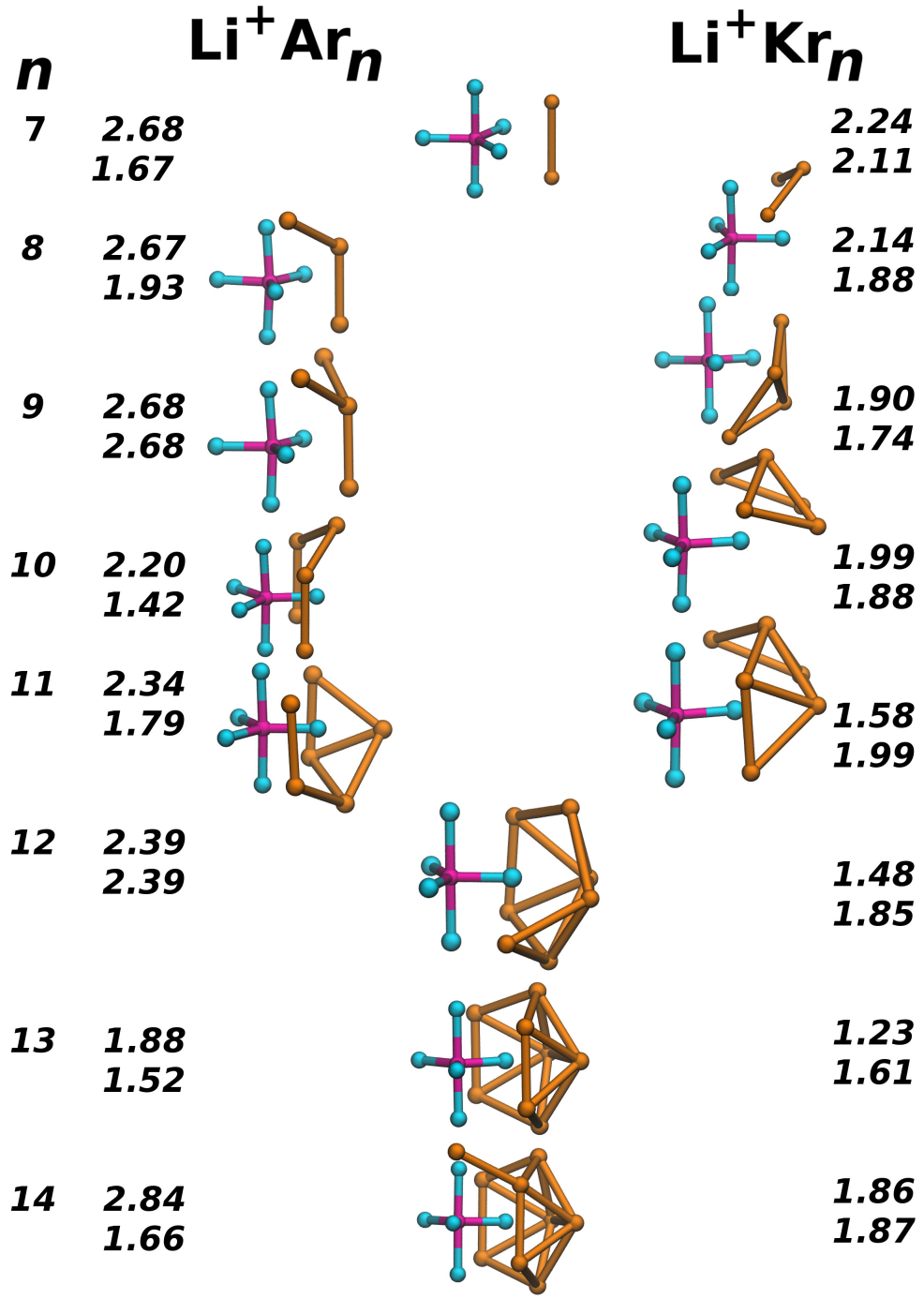


Figure 4.5: First low-lying local-minimum structures with 5 rare-gas atoms in the first solvation shell that were discovered in PES I of Li^+Ar_n and Li^+Kr_n (with $n = 7 - 14$). Such minima correspond to red lines in Fig. 4.4. For clearness, rare-gas atoms from the first (second) solvation shell are displayed in cyan (brown), while Li^+ is in magenta. For each structure, the first and the second values are, respectively, the energy (in mE_h) and the RMSD (in a_0), both calculated in relation to the corresponding global minimum.

4.3.3 Mapping the low-energy minima of PES I into PES II

The influence of three-body interactions in the global minimum of Li^+Ar_n clusters has been the aim of our previous study [113, 146]. It has been shown that significant differences between global minima arising from PES I and PES II are apparent for several sizes of the Li^+Ar_n clusters. In the present work, we have performed a detailed comparison between the global minimum structures obtained with PES I and PES II for both Li^+Ar_n and Li^+Kr_n (with $n = 2 - 14$). Also, we have looked further at other low-energy minima, besides the global one, for the smaller clusters (*i.e.*, for $n = 2 - 7$). Specifically, we have carried out local optimization to identify the structures in PES II that are obtained from the low-energy minima and saddle-points of PES I.

Figure 4.6 shows the mapping of low-energy minima and saddle-point structures of PES I into PES II; the average energy per rare-gas atom as a function of the size of the cluster is represented for Li^+Ar_n and Li^+Kr_n (with $n = 2 - 7$). One may observe that, for both systems, the energy per rare-gas atom of the global minima in PES I increases with the cluster size, while it shows a non-monotonous behavior in PES II (*i.e.*, it decreases up to $n = 4$ and, then, increases or $n \geq 5$). Such energetic pattern arising for both microsolvation systems has been previously indicated as one of the most relevant outcomes revealing the importance of incorporating three-body interactions in the PES of Li^+Ar_n clusters [113, 146].

It is worth noting from Figure 4.6 that, by local optimization from various minimum and saddle-point structures of PES I, we can only produce one or two minima in PES II for both Li^+Ar_n and Li^+Kr_n. Thus, PES I appears to be more complex than PES II for the two systems, in the sense that it comprises a larger number of low-energy minima. In the case of Li^+Ar_n, however, there are additional saddle-point structures in the PES II that map to minima or saddle-points in PES I. It is particularly interesting to observe in the top panel of Figure 4.6 that no structure of Li^+Ar_2 and Li^+Ar_3 maps to the corresponding global minimum structures in PES II. Conversely, it is possible to obtain Li^+Kr_2 and Li^+Kr_3 global minima in PES II departing from the lowest-energy structures in PES I. Indeed, Figure 4.6 reveals the existence of several saddle-point structures in the Li^+Ar_n PES II that are not present in the case of the Li^+Kr_n system, even though the former is expected to be, generally, more flat

than the latter one.

As it has been already noted for the Li^+Ar_n clusters [113, 146], the main influence of three-body interactions in the structures from the first solvation shell occurs for Li^+Kr_2 and Li^+Kr_3; one recalls that the Li^+Kr_2 (Li^+Kr_3) global minimum in PES I has a linear (planar) geometry, while it becomes a bent (pyramidal) structure in PES II. In turn, global minimum structures for $n = 4-6$ (and also for Li^+Kr_7, Li^+Ar_7 and Li^+Ar_8 that already have atoms in the second solvation shell) are those presented in Figure 4.2 and, by contrast, they are the same in PES I and PES II; see also Refs. [113, 146].

However, second-shell global minimum structures from PES I and PES II show significant differences for $n \geq 9$ ($n \geq 8$) for Li^+Ar_n (Li^+Kr_n). For these larger clusters, PES I tends to generate global minimum structures with the ion in a more peripheral position than that from PES II. To make easier the discussion that follows, we designate the former (latter) as A (B); both types of structures for $n = 9$ and $n = 14$ are illustrated in Figure 4.7. Note that these structures are similar for Li^+Ar_n and Li^+Kr_n clusters. In turn, Table 4.2 represents the total energy of structures A and B for both PES I and PES II of Li^+Kr_9 and Li^+Kr_{14}, as well as the corresponding contributions from the first and the second solvation shells, and the interaction between first- and second-shell species. The energy difference between structures A and B tends to increase for both PESs with the size of the cluster. In particular, such difference even reaches 3.8 mEh for $n = 14$ with the PES I. One may observe from Table 4.2 that, as expected, the energy contribution from the first solvation shell is identical from both structures. In turn, the second-shell configuration of structure A is always energetically favored, while the contribution from the interaction between the first- and second-shell particles benefits structure B. Thus, the adoption of structure A or B results from a subtle balance between these two contributions for both PES I and PES II.

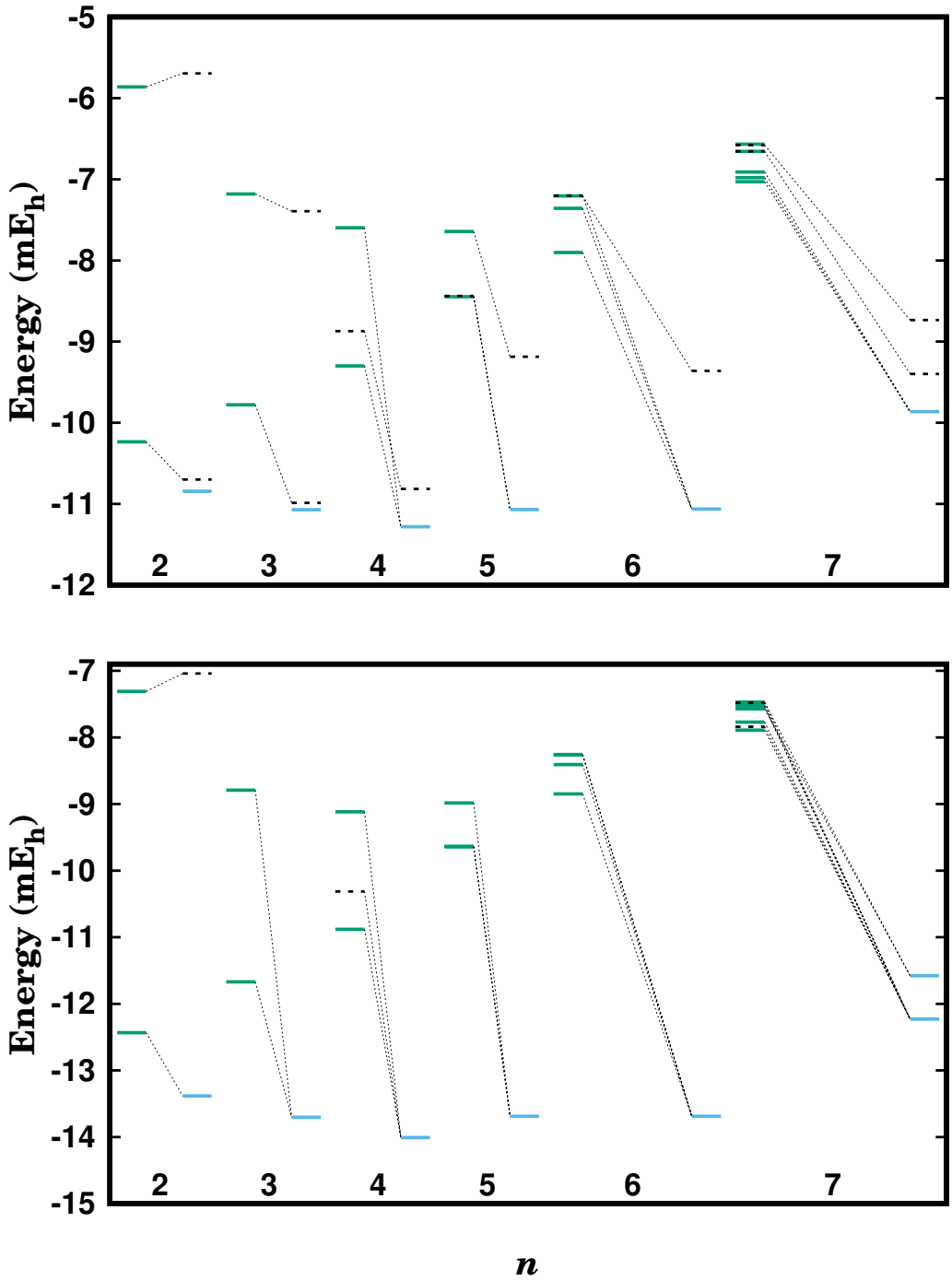


Figure 4.6: Minima (solid lines) and saddle-points (dashed lines) of PES II obtained by local optimization from the low-energy structures of PES I (as shown by the dotted lines), including the global one; the y-axis refers to the average energy per rare-gas atom. Top panel is for Li^+Ar_n $(n = 2-7)$ clusters, while bottom panel corresponds to the Li^+Kr_n $(n = 2-7)$ ones. Minima from PES I (PES II) are in green (light blue).

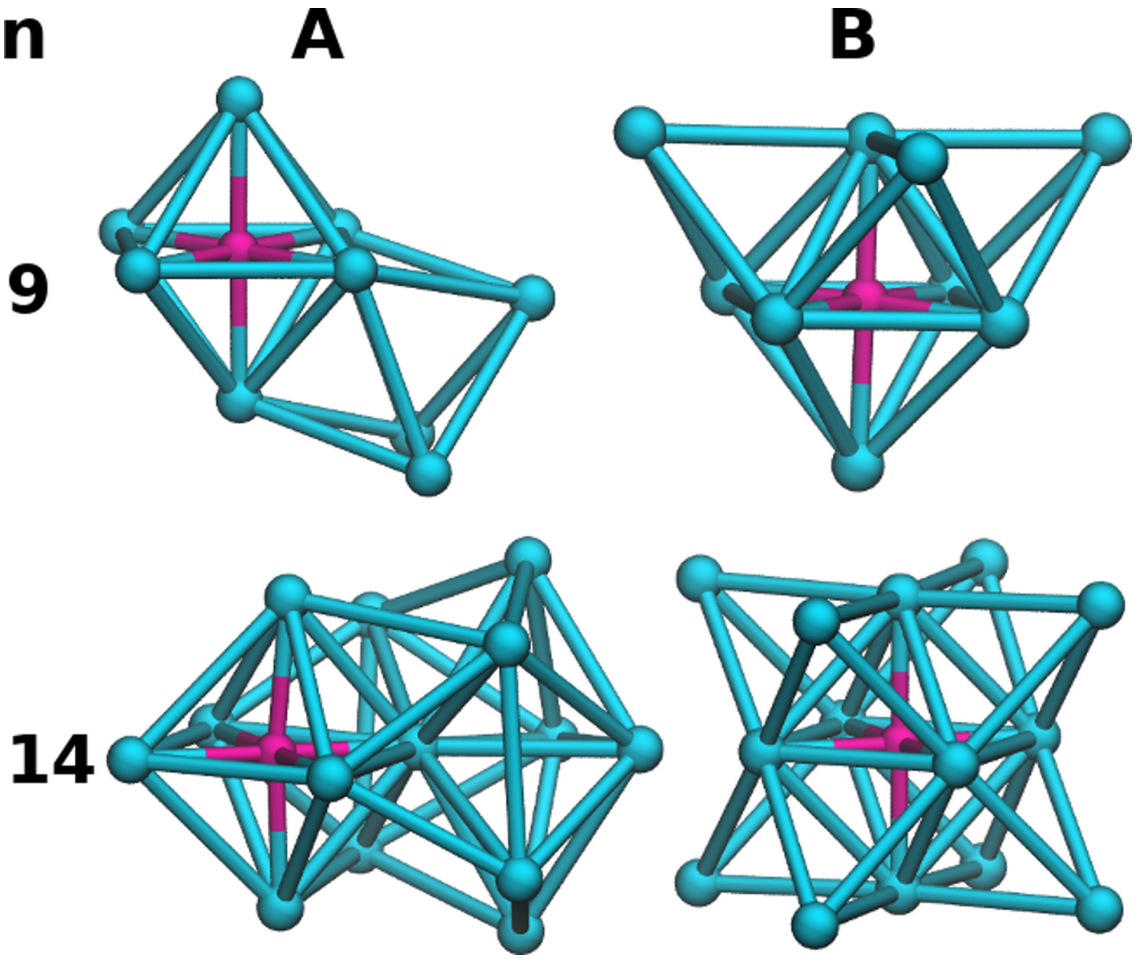


Figure 4.7: A- and B-type structures for $n = 9$ and $n = 14$ of Li^+-rare-gas clusters; Li^+ is displayed in magenta, while rare-gas atoms are in cyan.

When going from structure A to structure B, the energy contribution from the second-shell configuration varies by $1.6\,mE_h$ ($8.3\,mE_h$) for Li^+Kr_9 (Li^+Kr_{14}) with the PES I, while the counterpart from the first shell – second shell interaction is only $-1.3\,mE_h$ ($-4.2\,mE_h$); accordingly, structure A is the most stable in PES I. The reverse arises in the case of PES II, where the corresponding values for Li^+Kr_9 (Li^+Kr_{14}) are $1.5\,mE_h$ ($7.1\,mE_h$) and $-3.2\,mE_h$ ($-8.0\,mE_h$); hence, structure B is the most stable in PES II. Although not shown, a similar result is obtained for the analogous Li^+Ar_n structures.

Table 4.2: Energy analysis of the A-type and B-type structures of Fig. 4.7 in PES I and PES II. In each case, first entry corresponds to structure A, while second entry is for structure B.

	Li^+Kr_9		Li^+Kr_{14}	
energy contribution (mE_h)	PES I	PES II	PES I	PES II
total	-60.0	-91.4	-74.1	-110.3
	-59.6	-92.9	-70.3	-111.4
first shell	-53.1	-82.0	-52.8	-81.9
	-53.1	-81.9	-53.1	-82.1
second shell	-1.90	-1.92	-9.7	-9.0
	-0.25	-0.37	-1.4	-1.9
first shell – second shell	-5.0	-7.4	-11.6	-19.4
	-6.3	-10.6	-15.8	-27.4

4.3.4 Re-optimization of low-energy structures at the MP2 level

In order to evaluate the reliability of the low-energy structures given by PES I and PES II for both Li^+Ar_n and Li^+Kr_n, we have performed the re-optimization at the MP2/aug-cc-pVTZ level of all minima and saddle-points obtained for $n = 2$ and 3, in a total of 17 geometries (*cf.* Figure 4.6). From the results displayed in Figure 4.8, we may conclude that all (minima) structures considered from PES I, independently of the system, are reproduced and keep the same energetic order at the MP2 level; also, the geometry only slightly changes during the optimization procedure. Conversely, a more complex situation occurs for the structures from PES II, since the geometry optimization produces either different structures or changes the type of the stationary point (from a saddle-point to a minimum).

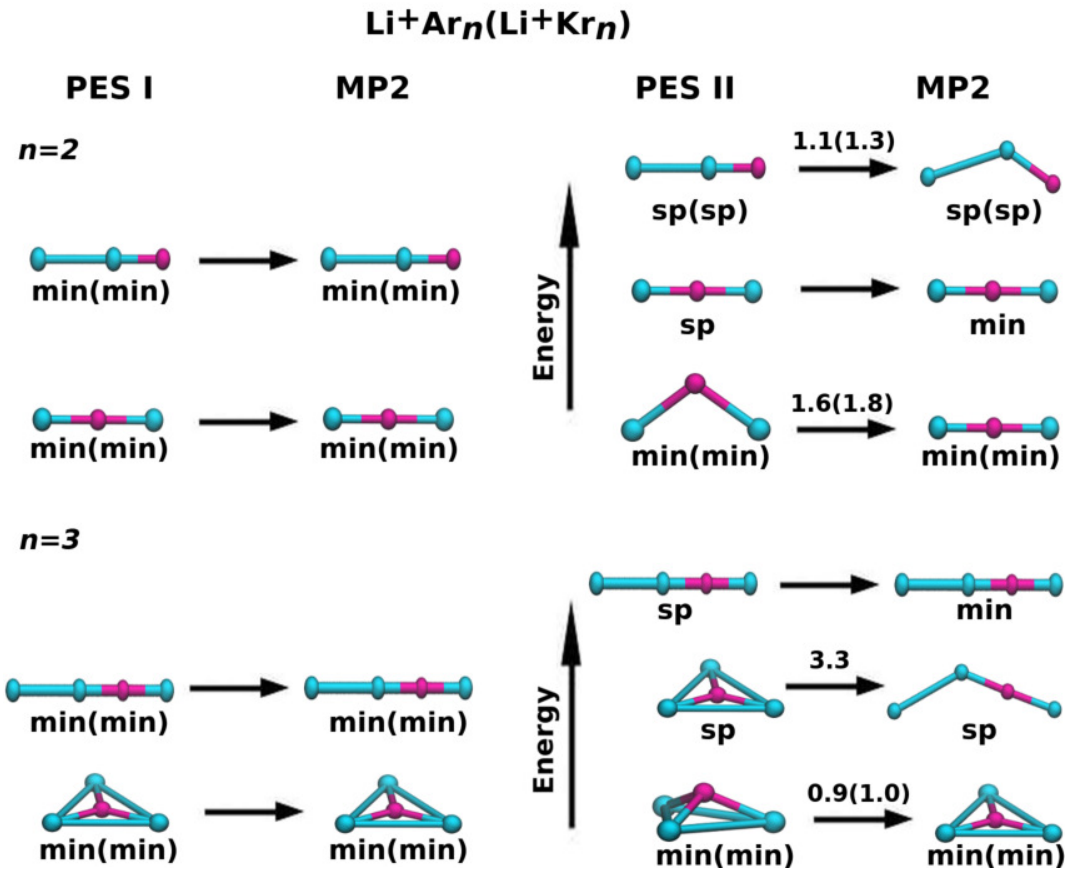


Figure 4.8: Low-energy stationary points obtained at the MP2/aug-cc-pVTZ level of theory by local optimization from PES I and PES II structures for both Li^+Ar_n and Li^+Kr_n (with n = 2 and 3). Each structure is labeled as minimum (min) or saddle-point (sp), and the numerical values over the arrows are for the corresponding RMSDs in a_0 (no value appears over the arrow when the RMSD is essentially zero). Values and labels in parenthesis are for Li^+Kr_n, while the remaining ones are for Li^+Ar_n. The energy arrows refer to the energetic order of the structures in PES I and PES II. The MP2 energies (in E_h) for the left-side structures, from bottom up, are -1061.320180 and -1061.310773 (-1588.355893 and -1588.346910) for Li^+Ar_2 (Li^+Ar_3), and -5511.730154 and -5511.719004 (-8263.970194 and -8263.959900) for Li^+Kr_2 (Li^+Kr_3). The corresponding values for the right-side structures are -1061.320180 and -1061.310696 (or -1588.355893, -1588.346835 and -1588.346910) for both minima and the saddle-point of Li^+Ar_2, respectively (for Li^+Ar_3), and -5511.730154 and -5511.718913 (or -8263.970194) for Li^+Kr_2 (Li^+Kr_3). Colors for rare-gas atoms and Li^+ are as in Fig. 4.7.

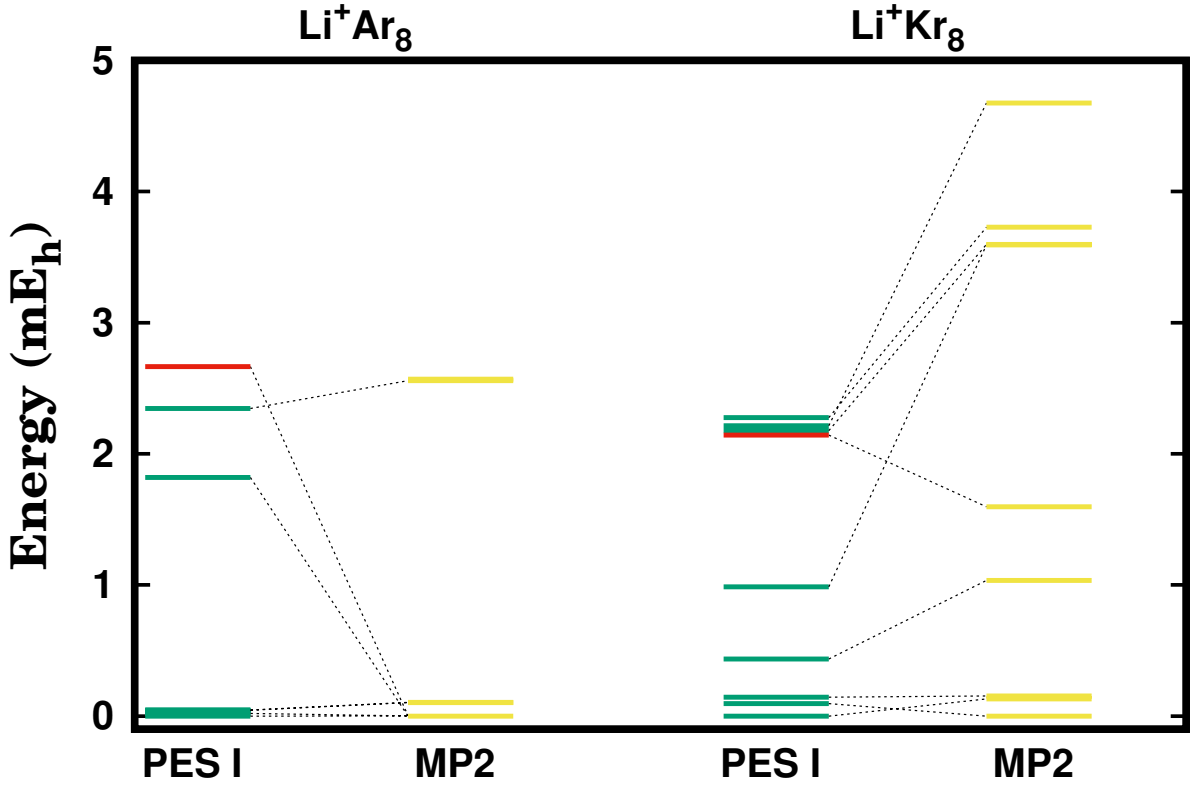


Figure 4.9: Low-energy minima obtained at the MP2/aug-cc-pVTZ level of theory by local optimization from PES I minima for Li^+Ar_8 and Li^+Kr_8. In each case, the energies of the minima are referred to the lowest-energy structure. As in Fig. 4.4, the red lines indicate structures from PES I that have only 5 rare-gas atoms in the first solvation shell.

For Li^+Ar_2, the highest-energy structure changes the geometric motif during the optimization though keeping as a saddle-point, while the two lowest-energy structures (*i.e.*, the bent minimum and the linear saddle-point) lead both to the MP2 global minimum (with a linear geometry). In the case of Li^+Ar_3, the highest-energy saddle-point is transformed into a minimum at the MP2 level (with essentially the same geometry as in PES II, but changing the energetic order with respect to the other saddle-point), and both the lowest-energy minimum and saddle-point change their structures upon re-optimization. In turn, the PES II structures for both Li^+Kr_2 and Li^+Kr_3 keep the same type of stationary point, but varies the geometry by the re-optimization at the MP2 level of theory.

Furthermore, we have investigated how the low-energy landscapes of Li^+Ar_8 and Li^+Kr_8 given by PES I map to the MP2/aug-cc-pVTZ level of theory; this is the smallest cluster size where differences can be found in global-minimum structures of both systems. Thus, structures corresponding to the energy-levels displayed in Figure 4.4 have been re-optimized at that level of theory. The use of triple-zeta basis set in spite of the quadruple-zeta one is due to the high computational cost to treat these aggregates. Figure 4.9 shows a diagram where the low-energy minima of PES I are connected through dotted lines with the corresponding re-optimized MP2 configurations.

A first glance from Figure 4.9 shows that the optimization at the MP2 level leads to a better energy discrimination among different structures of both Li^+Ar_8 and Li^+Kr_8 that are quasi-degenerated in PES I. Because more than one structure in PES I leads to the same minimum after re-optimization at the MP2 level, it is likely that PES I overestimates the number of minima, especially for Li^+Ar_n clusters. In particular, the global minimum energy structure of the MP2 optimization is obtained from the global and three local minima of the PES I, including the highest energy that have only five argon atoms in the first solvatation shell.

It is also interesting to note in Figure 4.9 that the energy range of the minima resulting from the MP2 optimization is very different between Li^+Ar_8 and Li^+Kr_8. While the MP2 Li^+Ar_8 minima spawn essentially the same energy-range of the corresponding ones from PES I, there are minima in the case of Li^+Kr_8 that become very high in energy. Moreover, the structures of Li^+Kr_8 found by the MP2 optimization exhibit energy differences for the ground state larger than the ones observed from PES I. For example, the local minima in PES I (MP2) with the highest energy shown in Fig. 4.9 is 2.3 mEh ($4.7\,mE_h$) above corresponding ground state.

Additionally, we observe in Fig. 4.9 an energy inversion between the global minimum and the second-lowest energy minimum of Li^+Kr_8 upon MP2 optimization. A similar behavior has been apparent for the corresponding minimum structures of PES I and PES II, as discussed in Section 4.3.3 for Li^+Ar_n and Li^+Kr_n clusters from the second solvation shell. In turn, we show in Table 4.3 that, for Li^+Kr_8, the global minimum structure from PES I has

a lower MP2/aug-cc-pVTZ energy than the one from PES II. However, the MP2 optimization departing from those structures leads to an inversion in the energy order of the two minima; the two optimized structures are quasi-degenerated (the energy difference is $\sim 0.1\,\mathrm{mE_h}$), but they are structurally distinct since the corresponding RMSD value for the best overlap is $0.58\,a_0$.

Table 4.3: Ab initio total energies (in E_h) obtained from single-point calculations and after geometry optimization for the global minimum structures of Li^+Kr_8 in PES I, PES II and PES I without including the Kr_3 ATM term; all calculations were carried out at the MP2/aug-cc-pVTZ level of theory.

	Li^+Kr_8 global minimum structures		
	PES I	PES II	PES I (without ATM)
Single-point	-22025.142005	-22025.139829	-22025.142073
After optimization[(a)]	-22025.142397	-22025.142529	-22025.142530

[(a)] By using a convergence criterion of $1.0\times10^{-6}\,E_h$, the number of optimization steps carried out from the global minimum structures of PES I, PES II and PES I (without ATM) is 18, 10 and 45, respectively.

It is reasonable to expect that the difference in the global minimum structures of PES I and PES II arise due to the Kr_3 ATM interaction term. Indeed, when excluding the ATM term from PES I, one obtains a global minimum structure which leads to an energy similar to the one from PES II after MP2 optimization, and the two geometries are actually identical (*i.e.*, RMSD=$0.04\,a_0$). Although not shown in Table 4.3, the re-optimization at the MP2 level of the Li^+Ar_8 structures corresponding to the global minimum of PES I, PES II and PES I without the ATM contribution leads always to the same geometry.

4.4 Conclusions

We have carried out a detailed global optimization study in order to figure out the low-energy landscape of the clusters resulting from the microsolvation of Li^+ by either argon or krypton atoms. A thorough comparison resorting to the effect of the solvent species on the structure and energetics has been performed for clusters with up to 14 rare-gas atoms. All calculations for the Li^+Ar_n clusters employed previously proposed [113, 146] PESs that include either both two- and three-body interactions (PES I) or just the two-body terms (PES II). As for the Li^+Kr_n clusters, we have developed analogous PES I and PES II based on *ab initio* energies that were calculated at the CCSD(T)/aug-cc-pVQZ level of theory and corrected for the BSSE with the CP methods.

The global minimum structures obtained from PES I are, in general, similar for both Li^+Ar_n and Li^+Kr_n. However, most of the Li^+Ar_n low-lying minima tends to be concentrated in a narrow energy-window close to the corresponding global minimum, while the Li^+Kr_n ones are in larger number and spread out over the energy range.

We have also shown that, independently of the microsolvation systems, distinct energetic patterns are apparent when comparing structures from first and second solvation shells. Indeed, the energy gap between local minima is larger for clusters with the uncompleted first solvation shell, which can be attributed to the stronger Li^+-rare-gas interaction. Accordingly, the octahedral structure of the first solvation shell is maintained for larger clusters, since there is a high-energy penalty when one rare-gas atom is removed from the neighborhood of Li^+ to a more external position. In contrast, clusters from the second solvation shell show a panoply of energetically similar low-lying minima that, in apparently flat PESs like the present ones, are expected to inter-convert among each other, even at low temperatures. Thus, we believe that the first solvation shell has a rigid structure, while the second solvation shell presents a "fluid-like" behavior.

The comparison between the lowest-energy minimum structures from PES I and PES II shows that significant differences arise for Li^+Ar_n (Li^+Kr_n) at $n = 2$, 3 and $n = 9-14$ ($n = 2$, 3 and $n = 8-14$). The latter structures contain rare-gas atoms in the second solvation

shell with the ion essentially in the center (off-center) for PES II (PES I). The stability of such type of structures result from a subtle balance between the interaction involving the second-shell atoms alone and that from the inter-shell species energy contribution. In turn, re-optimization of the n=2 and 3 low-energy minima from both PES I and PES II at the MP2/aug-cc-pVTZ level of theory have confirmed that the incorporation of three-body terms in the PES is essential to reproduce the main structural motifs. Furthermore, we have also re-optimized Li^+Ar_8 and Li^+Kr_8 low-energy structures at the MP2 level. It is worth noting that the MP2 Li^+Ar_8 (Li^+Kr_8) minima spawn essentially the same (larger) energy-range of (than) the corresponding ones from PES I. In turn, the energy inversion between the global minimum and the second-lowest energy minimum in PES I of Li^+Kr_8 upon MP2 optimization has been attributed to the Kr_3 ATM interaction term. Due to this kind of energy re-order, we advocate the use of several minima in the post-optimization procedure so that one increases the probability to get the global minimum structure at a higher level of theory.

SUPPORTING INFORMATION

Additional supporting information may be found online in: https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002

Chapter 5

Microsolvation of Li^+ in a mixture of argon and krypton: unveiling the most stable structures of the clusters

The microsolvation of Li^+ by both argon and krypton atoms has been studied based on a new potential energy surface that includes two- and three-body interactions; the potential terms involving the lithium ion were calibrated with CCSD(T)/aug-cc-pVQZ energies after corrected for the basis-set superposition error. The structures of the $Li^+Ar_nKr_m$ ($n + m \leq 20$) clusters arising from global optimization show a first solvation shell preferentially occupied by krypton atoms. This binary-solvent microsolvation clusters are most stable when the total number of krypton (argon) atoms occupy the first (second) solvation shell.

5.1 Introduction

Solvation is a ubiquitous phenomenon in chemistry, which has been treated under several perspectives from both the experimental and theoretical sides. Perhaps the most detailed way to look at such a physical process is by using the microsolvation approach

where the solvent entities (atoms or molecules) are added stepwise to the solute [145]. This leads to the formation of clusters that grow, and whose structural and energetic properties explain about the way the solvent surrounds the solute. From the theoretical side, the major challenges to set up the study of microsolvation are twofold: (i) accurately describe the interaction potential, and (ii) efficiently optimize the clusters to discover the lowest-energy structure (and, eventually, other low-energy minima). In general, the former requires the use of state-of-the-art methods to perform electronic-structure calculations that are, then, employed in a least-squares fit [21] to an adequate analytical potential function (see, *e.g.*, Refs. [135, 140]). In turn, the latter is a very active area of research whose endeavor has given fruitful results, with the development of state-of-the-art global-optimization algorithms [59, 92, 81, 103, 166] that have been already applied in the study of microsolvation clusters [140, 167, 117, 113, 146, 18].

The most used approach comprises using state-of-the-art techniques to carry out global geometry optimization of the microsolvation clusters modelled with an adequate analytical interaction potential that has been previously fitted to *ab initio* data (see, *e.g.*, Refs. [135, 140]); the obtained minima structures are, then, re-optimized at a higher level of theory to get more accurate energy values. Although alternative approaches have been proposed [90, 134], they are computationally more expensive.

Among the great number of publications related to the study of solvation, those involving the formation and characterization of small ionic clusters have attracted much interest [168, 169, 170, 142, 171, 172]. Indeed, a great variety of clusters with ionic species and atomic and molecular systems, respectively, as solutes and solvents have been the subject of recent studies (see, *e.g.*, Refs. [173, 174, 175, 176, 177, 178] and references cited therein). These charged clusters can be used to gain detailed insight into solvation phenomena at the molecular level by establishing the bridge connecting the isolated ion in the gas-phase and the solvated ion in solution.

In previous studies, we have assessed the importance of including up to three-body interactions for describing the microsolvation of Li^+ with either argon [113, 146, 13] or krypton [13] atoms. Here, we explore the ability of a similar potential model to reproduce the

main structural features of microsolvation clusters that contain Li^+ and atoms of both argon and krypton. To our knowledge, this is one of the first works on the solvation of an alkaline ion by mixtures of rare gases. In contrast, heterogeneous rare-gas clusters have received the attention of theoreticians and experimentalists [179, 180, 104, 181, 182, 183]. One of the central issues discussed in such studies is the preferential site occupancy of the different types of atoms in the rare-gas aggregates. In the particular case of large heterogeneous Ar-Kr clusters, though it can occur that some krypton atoms are present on the surface and some argon atoms in the bulk, the most favorable structure show preferentially krypton atoms in the bulk and argon atoms on the surface [184].

In this paper, we investigate solvent-selectivity effects in heterogeneous clusters resulting from the microsolvation of Li^+ with both argon and krypton. To achieve that purpose, we have employed an evolutionary algorithm to perform a global optimization study on the $Li^+Ar_nKr_m$ clusters. More specifically, we have investigated the preferential location of argon and krypton atoms around the lithium ion, and whether the two rare-gases are mixed together or, conversely, tend to be separated when forming the microsolvation cluster.

5.2 $Li^+Ar_nKr_m$ Potential

Based on our previous works on Li^+Ar_n and Li^+Kr_n [113, 146, 13], we have developed a new potential energy surface (PES) to model the mixed $Li^+Ar_nKr_m$ clusters. The analytical representation of the PES is given by the expression:

$$
\begin{aligned}
V(\mathbf{R}) &= \sum_j V_{\mathrm{Li^+Ar}}(R_{ij}) + \sum_m V_{\mathrm{Li^+Kr}}(R_{im}) \\
&+ \sum_j \sum_{k>j} V_{\mathrm{Ar_2}}(R_{jk}) + \sum_m \sum_{n>m} V_{\mathrm{Kr_2}}(R_{mn}) + \sum_j \sum_m V_{\mathrm{ArKr}}(R_{jm}) \\
&+ \sum_j \sum_{k>j} V_{\mathrm{Li^+Ar_2}}(R_{ij}, R_{ik}, R_{jk}) + \sum_m \sum_{n>m} V_{\mathrm{Li^+Kr_2}}(R_{im}, R_{in}, R_{mn}) \\
&+ \sum_j \sum_m V_{\mathrm{Li^+ArKr}}(R_{ij}, R_{im}, R_{jm}) + \sum_j \sum_{k>j} \sum_{l>k} V_{\mathrm{Ar_3}}(R_{jk}, R_{kl}, R_{jl}) \\
&+ \sum_m \sum_{n>m} \sum_{o>n} V_{\mathrm{Kr_3}}(R_{mn}, R_{no}, R_{mo}) + \sum_j \sum_m \sum_{n>m} V_{\mathrm{ArKr_2}}(R_{jm}, R_{jn}, R_{mn}) \\
&+ \sum_j \sum_{k>j} \sum_m V_{\mathrm{Ar_2Kr}}(R_{jm}, R_{km}, R_{jk}) \qquad (5.1)
\end{aligned}
$$

where the index i refers to the Li^+ ion and indices j, k, and l (m, n, and o) label three distinct argon (krypton) atoms of the cluster system. The $V_{\mathrm{Li^+ArKr}}$ term was obtained in the present work (see below), while the other ones were taken from Ref. [113] ($V_{\mathrm{Li^+Ar}}$ and $V_{\mathrm{Li^+Ar_2}}$), Ref. [13] ($V_{\mathrm{Li^+Kr}}$ and $V_{\mathrm{Li^+Kr_2}}$), Ref. [107] ($V_{\mathrm{Ar_2}}$ and $V_{\mathrm{Kr_2}}$), and Ref. [103] (V_{ArKr}). Although the improved Lennard-Jones function of Pirani and collaborators [185, 186, 187], which has three parameters with a precise physical meaning, could be preferred to model the pair-wise interactions in Eq.(5.1), we adopted for consistency the same type of potentials as in our previous works on Li^+–rare-gas clusters [113, 146, 13]. In addition, we have employed the Axilrod-Teller-Muto function [148, 149, 150] to describe all the three-body rare-gas interactions, whereas the corresponding C_9 parameters were obtained from Ref. [151] for $V_{\mathrm{Ar_3}}$ and $V_{\mathrm{Kr_3}}$, and Ref. [188] for $V_{\mathrm{ArKr_2}}$ and $V_{\mathrm{Ar_2Kr}}$.

The three-body Li^+ArKr interaction for different energies have been calculated at the CCSD(T)/aug-cc-pVQZ level of theory, followed by the correction for the basis-set superposition error (BSSE), which uses the counter-poise method [112]; all calculations were performed by using the GAMESS software package [189]. The aug-cc-pVQZ atomic basis sets for Li, Ar and Kr atoms are given in Refs. [53], [109] and [56], respectively. In the calculations, 5 (14) orbitals for each Ar (Kr) atom were frozen. Specifically, the Li^+ArKr three-body term is obtained by fitting the analytic function which establishes the interaction

between the dipoles on argon and krypton atoms induced by the lithium cation, that is,

$$V_{\mathrm{Li^+ArKr}}(R_{\mathrm{Li^+Ar}}, R_{\mathrm{Li^+Kr}}, R_{\mathrm{ArKr}}) = \left[\frac{\vec{\mu}_{\mathrm{Ar}} \cdot \vec{\mu}_{\mathrm{Kr}}}{R^3_{\mathrm{ArKr}}} - 3\frac{(\vec{\mu}_{\mathrm{Ar}} \cdot \vec{R}_{\mathrm{ArKr}})(\vec{\mu}_{\mathrm{Kr}} \cdot \vec{R}_{\mathrm{ArKr}})}{R^5_{\mathrm{ArKr}}}\right] \times \chi_{disp}(\beta_{disp}, \gamma_{disp}, R_{\mathrm{ArKr}}) \quad (5.2)$$

In Eq. (5.2), $\vec{\mu}_{\mathrm{Ar}}$ ($\vec{\mu}_{\mathrm{Kr}}$) is the dipole induced on the Ar (Kr) atom due to the Li^+ ion, which is defined as:

$$\vec{\mu}_{\mathrm{Ar(Kr)}} = \alpha_{\mathrm{Ar(Kr)}} \left[\chi_{pol}(\beta_{pol}, \gamma_{pol}, R_{\mathrm{Li^+Ar(Kr)}})\right]^{1/2} \frac{\vec{R}_{\mathrm{Li^+Ar(Kr)}}}{R^3_{\mathrm{Li^+Ar(Kr)}}} \quad (5.3)$$

where α_{Ar} (α_{Kr}) is the Ar (Kr) polarizability. In Eqs. (5.2) and (5.3) the χ damping functions have the form:

$$\chi(\beta, \gamma, R) = \begin{cases} 1 & R > \beta; \\ \exp\left[-\gamma(\beta/R - 1)^2\right] & R \leq \beta. \end{cases} \quad (5.4)$$

To maintain consistency with three-body $V_{\mathrm{Li^+Ar_2}}$ and $V_{\mathrm{Li^+Kr_2}}$ terms, we have used the parameters found in Refs. [113] and [13] to describe the $\vec{\mu}_{\mathrm{Ar}}$ and $\vec{\mu}_{\mathrm{Kr}}$ induced dipoles. Thus, only the χ_{disp} parameters were adjusted in the present paper. Within such conditions, the least-squares fit of the three-body $V_{\mathrm{Li^+ArKr}}$ term to 180 *ab initio* energies (that were calculated at different geometries of the Li^+ArKr species) led to the values $\beta_{disp} = 8.8598074$ a_0 and $\gamma_{disp} = 0.73526948$; the root-mean-square deviation obtained for this fitting was 8.2×10^{-5} E_h.

In Figure 5.1, the potential energy curves of the diatomic species Ar_2, Kr_2 and ArKr (top panel), and Li^+Ar and Li^+Kr (bottom panel) employed in the $Li^+Ar_nKr_m$ PES [Eq. (5.1)]. The interactions that involve Kr atoms have a deeper potential well and a larger equilibrium distance when compared with the Ar ones. In this sense, the ArKr interaction potential is a middle ground between the Ar_2 and Kr_2 potentials. The minimum of the ArKr potential energy has the value of $-0.53\,mE_h$ and it arises at an equilibrium geometry of $7.350\,a_0$, while to Kr_2 (Ar_2) the potential minimum is $-0.64\,mE_h$ ($-0.46\,mE_h$), and the equilibrium geometry is $7.585\,a_0$ ($7.109\,a_0$).

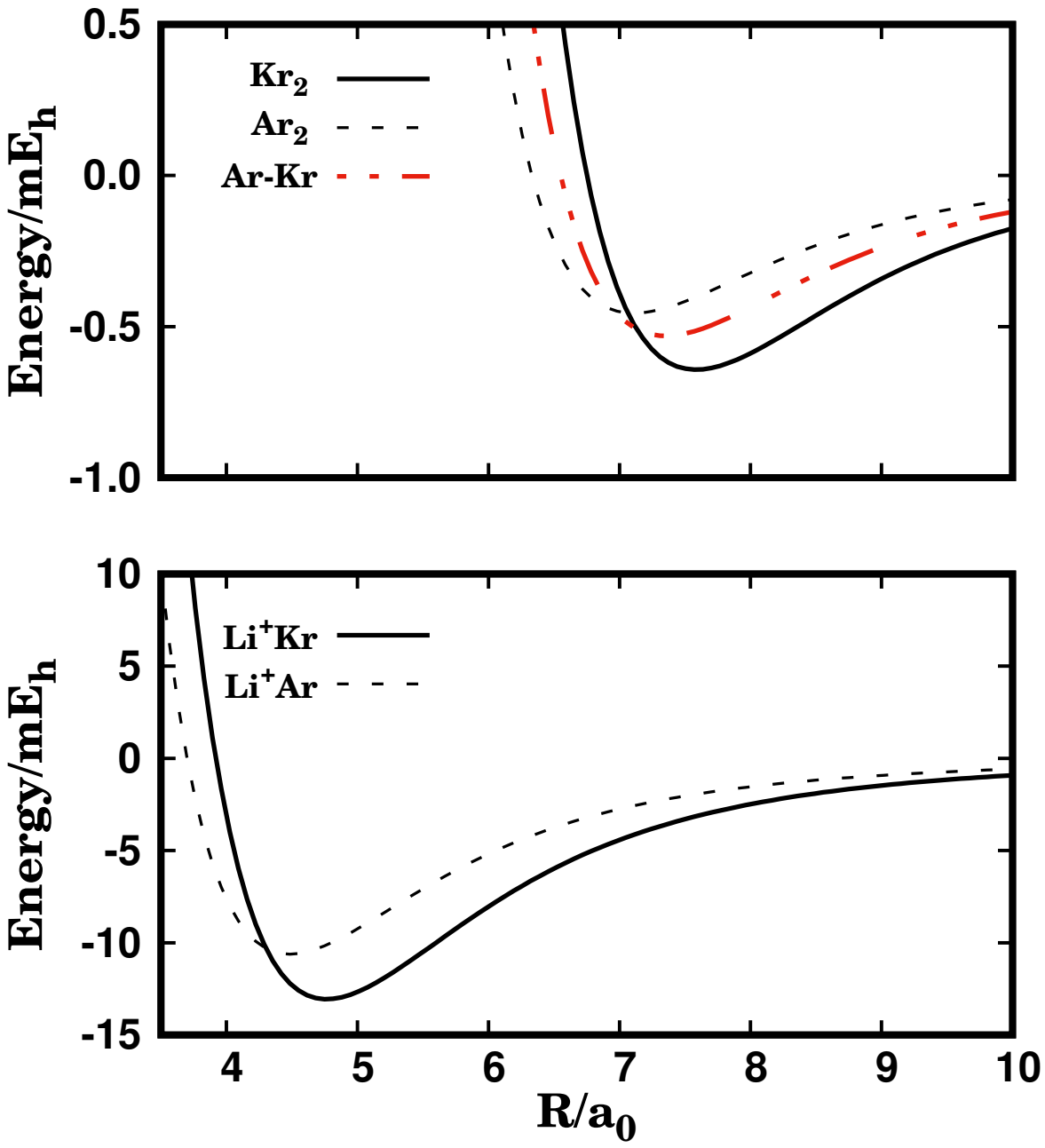


Figure 5.1: Potential energy curves for pair interactions involving the rare-gases (top panel) and for Li^+ with Ar and Kr atoms (bottom panel).

Similar behavior is observed for Li^+Ar_2, Li^+Kr_2 and Li^+ArKr triatomic species, where the latter has a minimum potential of -22.7 mE_h, while the other two are, respectively, $-20.5\,mE_h$ and $-24.9\,mE_h$. The energy differences between the minima of the systems Li^+Ar_2 and Li^+ArKr, and Li^+Kr_2 and Li^+ArKr, are slightly smaller than the energy difference

between the diatoms Li^+Ar and Li^+Kr (about $2.5\,mE_h$) [13], and is because of the influence of the three-body interaction terms. Another interesting aspect is that the energy required to remove the argon atom from Li^+ArKr ($D_e = 2106.0\ cm^{-1}$) is less than to remove the same atom from system Li^+Ar_2 ($D_e = 2162.3\,cm^{-1}$). On the other hand, it takes more energy to remove Kr from system Li^+ArKr ($D_e = 2643.1\,cm^{-1}$) than from system Li^+Kr_2 ($D_e = 2591.0\,cm^{-1}$). The reasoning for these observations may be found in the charge distribution for both argon and krypton atoms. From the analysis of Mulliken's population (see Table S1 in the Supporting Information), one concludes that the effective charge in the argon (krypton) atom is smaller (larger) in Li^+ArKr than in Li^+Ar_2 (Li^+Kr_2); hence, argon (krypton) shares less (more) electrons with Li^+ in Li^+ArKr than in Li^+Ar_2 (Li^+Kr_2).

5.3 Microsolvation Clusters: Energetics and Structure

We have searched for the global minimum structures of the $Li^+Ar_nKr_m$ clusters by applying an evolutionary algorithm (EA) developed in our group [103]. Basically, the EA evolves a pool of structures that are possible solutions for the global optimization problem. For that, it combines the generalized Cut&Splice crossover and sigma mutation operators [155] to explore different regions of the potential energy surface of the cluster with a local optimization procedure for the exploitation of a given basin of attraction. Since the exchange of positions between argon and krypton atoms within a structural motif may lead to different energy minima, the EA is equipped with a swap-mutation operator which attempts to explore the space of the homotop configurations [121]. Also, the strategy adopted for the replacement of structures in the pool, which is based on the comparison of both the energetic and structural properties of the clusters, is a key feature of the algorithm to reach good results [16]. Indeed, such a strategy aims at maintaining a certain degree of diversity in the pool of possible solutions over the global optimization procedure. Further details about the EA, including the best settings of the algorithm (also applied in this work), can be obtained from either the original paper [103] or a recent review [64].

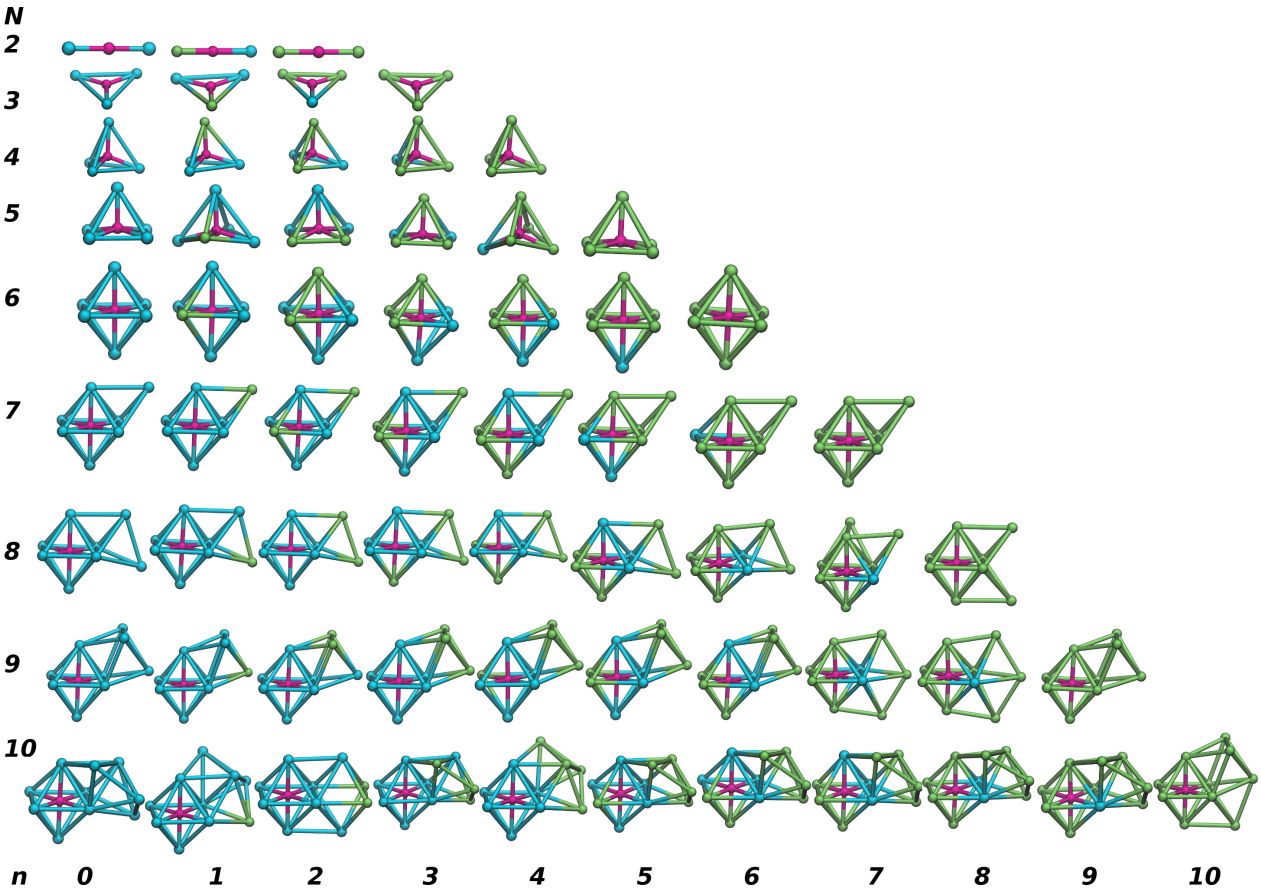


Figure 5.2: Putative global minimum for all compositions of the $Li^+Ar_nKr_m$ microsolvation clusters up to $N = 10$ (with $N = n + m$). All clusters are modeled by the analytical PES. Green (blue) spheres represent the argon (krypton) atoms.

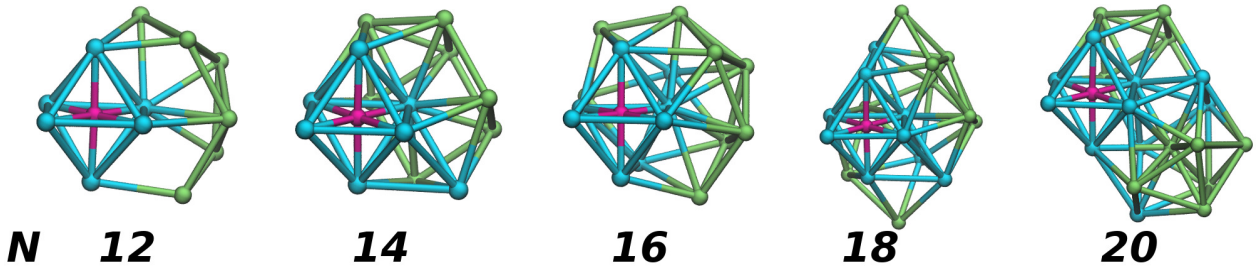


Figure 5.3: Putative global minimum structures of the $Li^+(ArKr)_{N/2}$ ($N = 12 - 20$) clusters that are modeled by the analytical PES. Green (blue) spheres represent the argon (krypton) atoms.

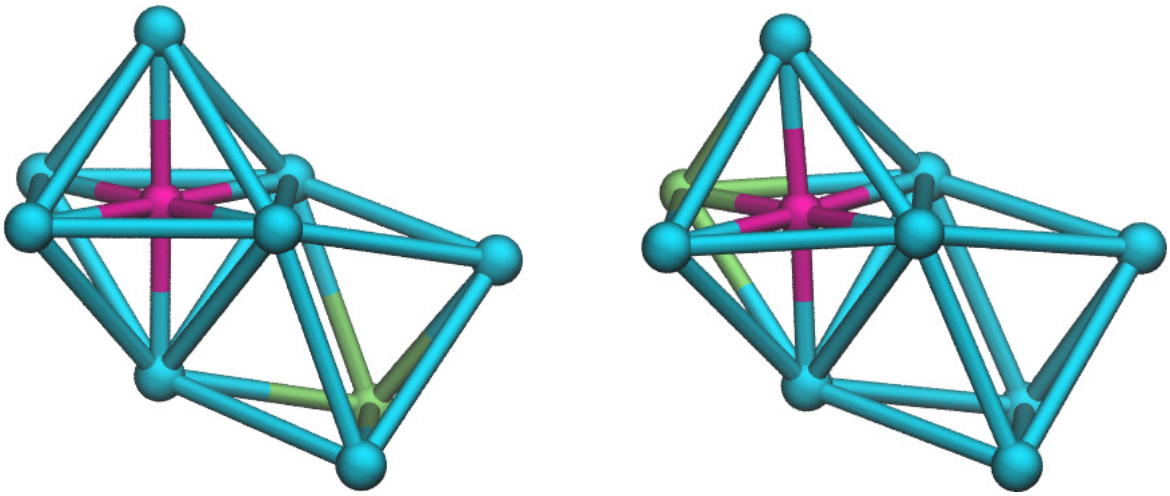
The putative global minimum structures of the analytical PES that were obtained

with the EA are represented in Figure 5.2 and Figure 5.3; the corresponding Cartesian coordinates and energies are given in the Supporting Information. We have studied the $Li^+Ar_nKr_m$ clusters up to $N = n+m = 10$ for all compositions of the rare-gases. In turn, we display in Figure 5.3 the lowest-energy structures for clusters with an equal number of argon and krypton atoms for $12 \leq N \leq 20$. One observes in Figure 5.2 that the shape of the global minimum is independent of the composition for almost all cluster sizes up to $N = 7$. The only exception to this trend occurs for $N = 5$, where the structures with either one argon or one krypton show a distorted pyramidal geometry. It is worth noting that the regular square pyramid observed for the other clusters with $N = 5$ does not appear even as a local minimum for $Li^+Ar_4Kr_1$ and $Li^+Ar_1Kr_4$.

A different situation is apparent in Figure 5.2 for $N \geq 8$. Indeed, the homogeneous clusters Li^+Ar_8 and Li^+Kr_8 (as well as Li^+Ar_{10} and Li^+Kr_{10}) show different structures, as already noticed in previous work [13]; hence, modifications in the shape of the global minima would be expected as the composition is changed from Ar-rich to Kr-rich clusters. Nonetheless, the global minimum structures for most of the $N = 8$ compositions are like that for the homogeneous Li^+Kr_8 cluster. Although such a trend is also apparent for $N = 10$, a larger diversity of structures arises among all the compositions of this cluster size. In turn, it is interesting to observe in Figure 5.2 that, despite showing equal structures for both $N = 9$ homogeneous clusters, the geometrical motif for $Li^+Ar_7Kr_2$ and $Li^+Ar_8Kr_1$ appears to result from growing the Li^+Ar_8 global minimum. We should mention that such structural diversity is consistent with the panoply of minima having similar energies that arises within the second solvation-shell of both Li^+Ar_n and Li^+Kr_n clusters [13].

We also observe in Figure 5.2 and Figure 5.3 that the krypton atoms occupy preferentially the first solvation shell, which always show an octahedral geometry, and the argon atoms go to the second solvation shell. This competition between krypton and argon to occupy positions closer to the ion can be attributed to fact that the Li^+-Kr interaction is stronger than the Li^+-Ar one (*cf.* Figure 5.1). Moreover, we have verified by re-optimization at the MP2/aug-cc-pVTZ level for the Li^+ArKr_8 cluster that the global minimum is the structure where the first solvation shell is totally filled with krypton atoms [Figure 5.4(a)]

rather than having the argon closer to Li^+ [Figure 5.4(b)].



(a) (b)

Figure 5.4: Geometries obtained by local optimization at the MP2/aug-cc-pVTZ level of theory for the Li^+ArKr_8 global minimum (a) and the lowest-energy minimum with the argon atom in the first solvation shell (b). Green (blue) spheres represent the argon (krypton) atoms.

Although we should note that MP2 describes poorly the dispersion interactions involving rare-gas atoms [190], the present calculations are reasonable since we just want to carry out a comparison between the most stable structures of the Li^+ArKr_8 cluster. The energy difference between these two structural motifs from the analytical PES and given by the MP2 calculations are $0.65\,mE_h$ and $0.82\,mE_h$, respectively. The consistency of these results indicates that the present analytical PES has the ability to realistically describe the heterogeneous $Li^+Ar_nKr_m$ aggregates.

We also emphasize that argon and krypton do not appear to mix with each other in the microsolvation cluster, because the Kr-Kr interaction is stronger than the Ar-Kr and Ar-Ar ones (see Figure 5.1). Thus, the global minima for $N \geq 12$ (with an equal number of the two rare-gases) tend to display structures where sub-clusters of either argon or krypton can be clearly identified. In particular, for $N = 18$ and $N = 20$, there are two external argon sub-clusters separated by a central sub-cluster that contains the ion and the krypton atoms

(see Figure 5.3).

We now look further at the energetics involved in the formation of heterogeneous microsolvation clusters for each size, N. Figure 5.5 shows that, at a given N, the energy per rare-gas atom increases (decreases in absolute value) as krypton is replaced by argon, and such behavior is almost linear with the number of Ar atoms. For small cluster sizes, the energy variation is quite marked due to the large difference between the Li^+-Ar and Li^+-Kr interactions, *i.e.*, the potential of the latter is more attractive than the former one (*cf.* Figure 5.1).

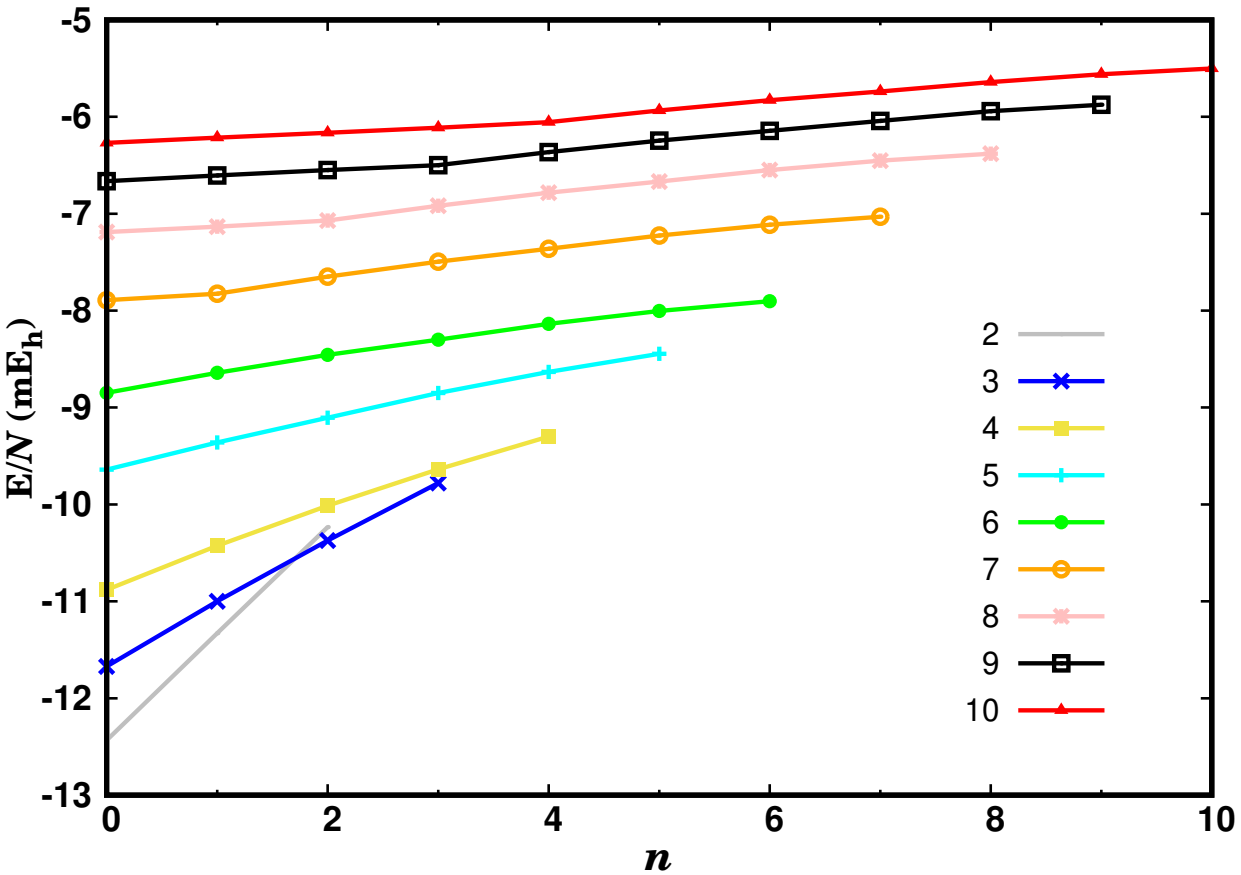


Figure 5.5: Energy per atom of rare-gas as a function of number (n) of Ar atoms in the composition of the microsolvation $Li^+Ar_nKr_{N-n}$ clusters. Each line corresponds to a given number N of rare-gas atoms (as inserted in the plot).

Furthermore, we may also observe in Figure 5.5 that, up to the closure of the first solvation shell, the energy variation with the composition is highly sensitive to the size of the

cluster, which may be because of the three-body interactions involving the ion. Conversely, the variation of energy for larger clusters is dominated by the weak interactions involving the rare-gas atoms (*cf.* Figure 5.1). Indeed, the energy lines in Figure 5.5 become almost parallel for $N \geq 6$, which also reflects the small difference in the rare-gas three-body interactions as N increases.

The relative stability of the mixed microsolvation cluster regarding the corresponding aggregates having only Ar or Kr atoms may be evaluated by computing the excess energy [191]:

$$E^*_{exc}(n,N) = E(\mathrm{Li^+Ar}_n\mathrm{Kr}_{N-n}) - n\frac{E(\mathrm{Li^+Ar}_N)}{N} - (N-n)\frac{E(\mathrm{Li^+Kr}_N)}{N} \tag{5.5}$$

where the first term is the global minimum energy of the mixed cluster, and $E(\mathrm{Li^+Ar}_N)/N$ ($E(\mathrm{Li^+Kr}_N)/N$) is the energy per rare-gas atom of $\mathrm{Li^+Ar}_N$ ($\mathrm{Li^+Kr}_N$). We should note that Eq. (5.5) allows to calculate the excess-energy curve for each N as function of n argon numbers. These curves are represented in Figure 5.6 for clusters up to $N = 10$.

Whereas the most stable clusters are associated with minima of the excess energy, values of E^*_{exc} greater than zero indicate unfavorable compositions. It is clear from Figure 5.6 that, up to $N = 6$, all the mixed compositions are unfavorable in relation to $\mathrm{Li^+Ar}_N$ and $\mathrm{Li^+Kr}_N$. As the aggregate increases beyond the first solvation shell, some compositions of the heterogeneous clusters become stable. We can see from Figure 5.6 that, independently of N, the most stable composition arises for clusters with 6 krypton atoms, which are expected to occupy the first solvation shell (see Figure 5.2).

Additionally, the second energy difference for each cluster size, *i.e.*,

$$\begin{aligned}\Delta_2 E(n,N) &= -2E(\mathrm{Li^+Ar}_n\mathrm{Kr}_{N-n}) \\ &\quad + E(\mathrm{Li^+Ar}_{n+1}\mathrm{Kr}_{N-n-1}) + E(\mathrm{Li^+Ar}_{n-1}\mathrm{Kr}_{N-n+1})\end{aligned} \tag{5.6}$$

establishes the stability of clusters in relation to the neighboring compositions.

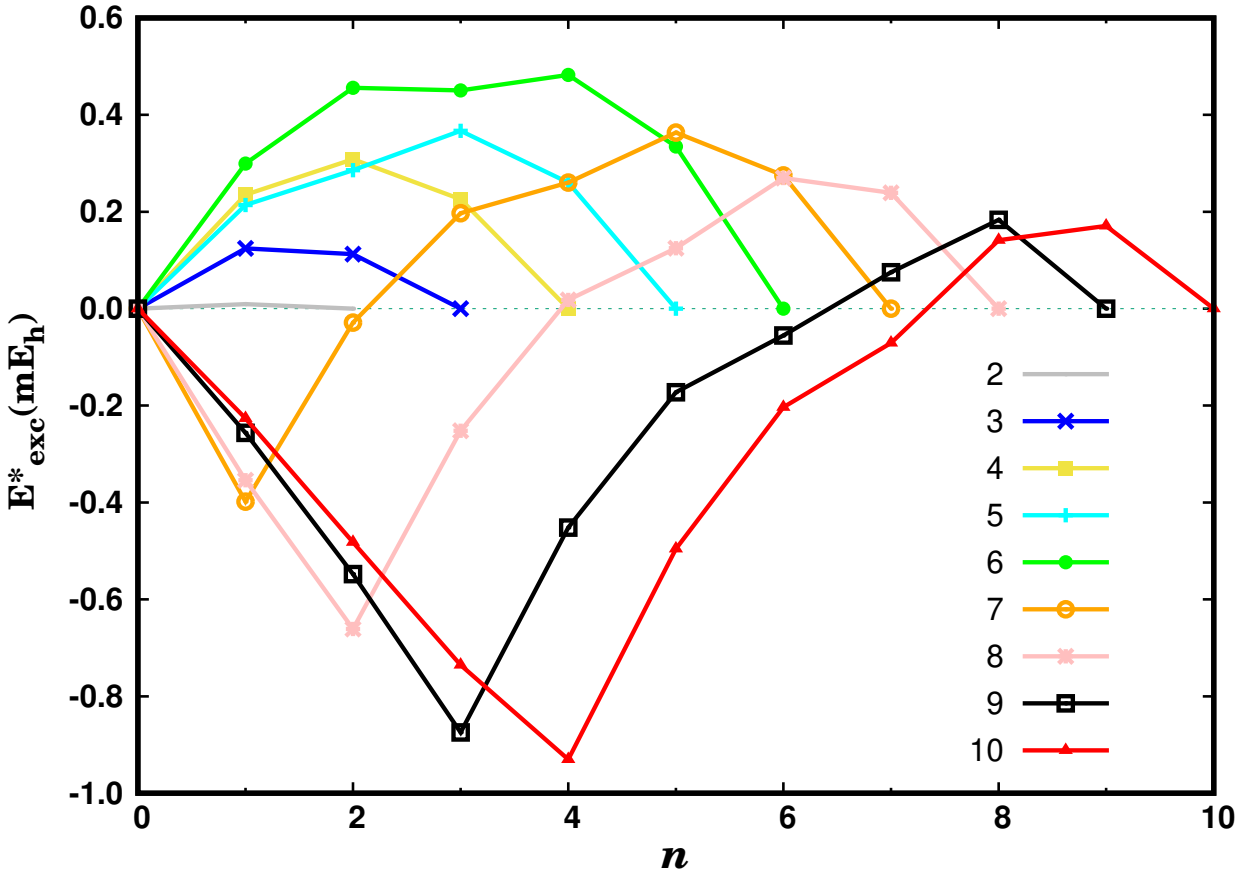


Figure 5.6: Excess energy as a function of number of Ar (n) in the composition of the microsolvation $Li^+Ar_nKr_{N-n}$ clusters. Each line corresponds to a given number N of rare-gas atoms (as inserted in the plot).

Thus, the inspection of Figure 5.7 shows that, for each curve with $N \geq 7$, the single prominent-peak corresponds to the cluster whose composition has 6 krypton atoms. This confirms the stability of structures with a first solvation shell totally filled with krypton atoms. Another interesting feature observed in Figure 5.7 is the small peak arising for each curve with $N \geq 6$, which appears to be the signature of a certain relative stability (though very weak) of clusters with the first solvation shell filled with 3 argon and 3 krypton atoms.

The energy of the putative global minimum structures for microsolvation clusters with an equal number of argon and krypton atoms is represented in Figure 5.8.

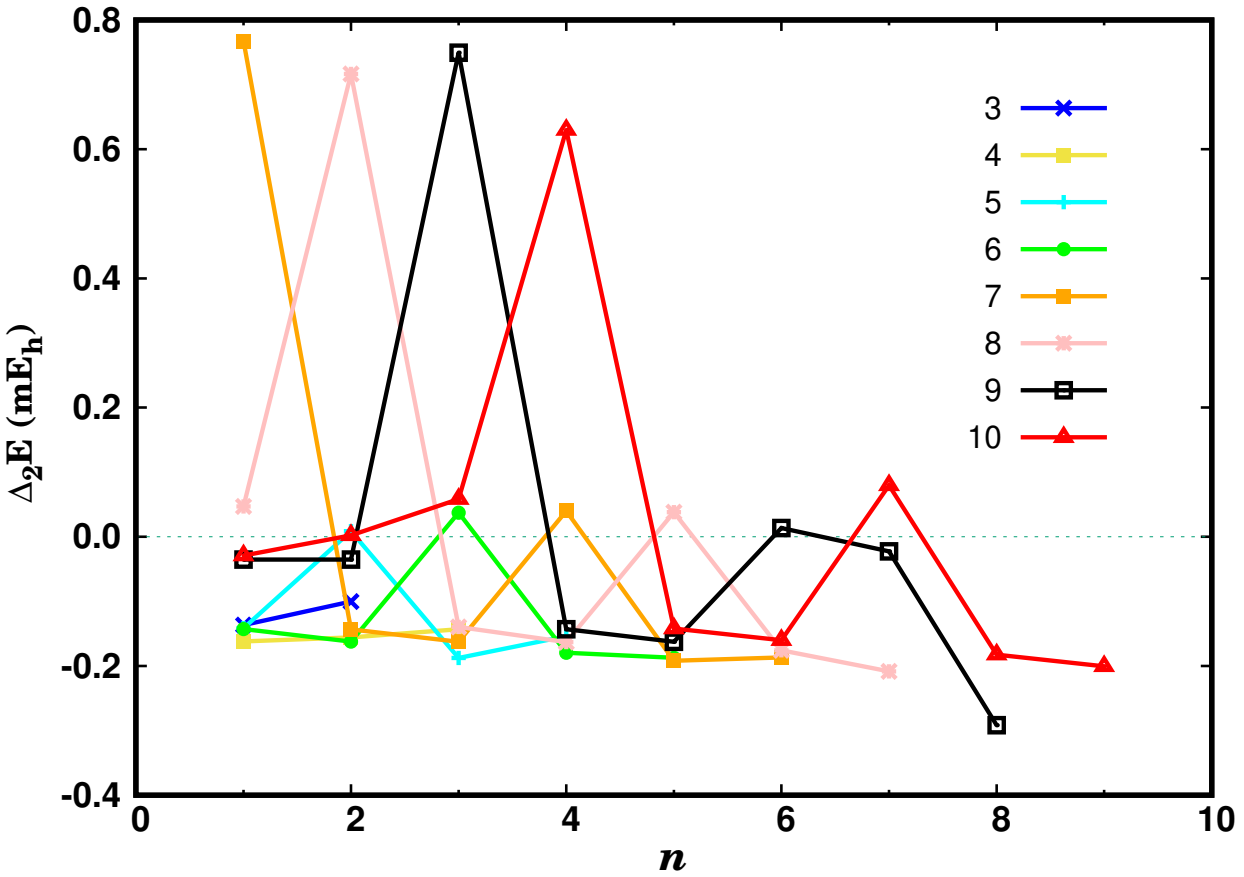


Figure 5.7: Second energy difference as a function of number of Ar (n) in the composition of the microsolvation $Li^+Ar_nKr_{N-n}$ clusters. Each line corresponds to a given number N of rare-gas atoms (as inserted in the plot).

It is shown in this figure that the energy per rare-gas atom strongly increases up to $N = 8$. The energy continues to increase for larger clusters, but at a lower rate since the rare-gas atoms become far apart from the ion. Moreover, we also show in Figure 5.8 for $N \geq 10$ ($N \geq 12$) the energy of local minima of clusters corresponding to structures with three (six) argon atoms in the first solvation shell.

As expected, the stability of the clusters decreases as more krypton atoms of the first solvation shell exchange positions with argon atoms from the second solvation shell.

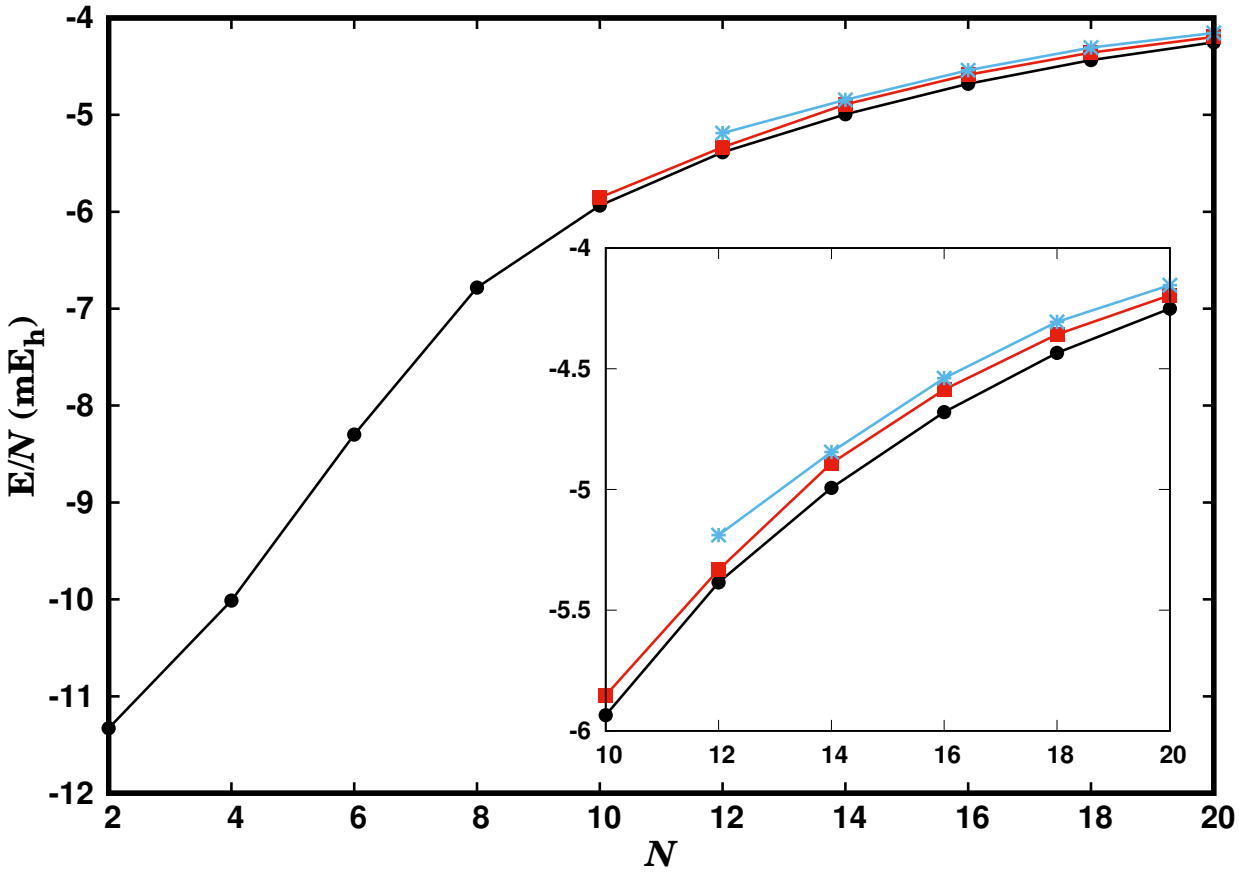


Figure 5.8: Energy of the $Li^+(Ar\,Kr)_{N/2}$ ($N = 2 - 20$) clusters. The energy per atom of rare gas is represented by the black lines (and dots) for the global minimum structures, while the red (blue) line are for the corresponding clusters where three (six) krypton atoms of the first solvation shell exchange their positions with the same number of argon atoms from the second solvation shell. The inset is an enlargement of the plot to detail the differences between the three lines.

It is worth noting in Figure 5.8 that the difference in the energy per rare-gas atom between the global minimum and the corresponding structures from the other two motifs keeps at a relatively small value as N increases. Nonetheless, one expects that high energy barriers must be overtaken when dynamically exchange a krypton atom from the first solvation shell by an external argon. In this sense, it becomes important to carry out a detailed study of the thermodynamic properties of these ionic aggregates. Note that a correct description of structural, energetic, dynamical and thermodinamical properties of the first solvation

shell of the heterogeneous $Li^+Ar_nKr_m$ clusters might be very sensitive to the accuracy of the potential energy.

5.4 Conclusions

We have developed an analytic PES to model the low-energy structures of the $Li^+Ar_nKr_m$ clusters. By using this PES and global optimization techniques, we have studied the energetics and the structure of clusters resulting from the microsolvation of Li^+ in a mixture of argon and krypton. The most stable structures of the microsolvation clusters show an octahedral first solvation shell totally filled with six krypton atoms, while argon is preferentially placed in the second solvation shell. This result was confirmed for Li^+ArKr_8 by *ab initio* calculations at the MP2/aug-cc-pVTZ level of theory, which is also an indication of the adequacy of the present PES to describe the energy landscape of the $Li^+Ar_nKr_m$ clusters. Additionally, our study shows that, even for larger clusters, the two types of rare-gas atoms are unlikely to be mixed. Indeed, the microsolvation of the ion leads to formation of clusters, where distinct argon and krypton sub-clusters are apparent. Finally, we should also note that the structures reported in this paper may be used in a future work as starting geometries for evaluating thermodynamic properties of the $Li^+Ar_nKr_m$ clusters through Monte-Carlo parallel-tempering technique.

SUPPORTING INFORMATION

Additional supporting information may be found online in: https://pubs.acs.org/doi/suppl/10.1021/acs.jpca.9b00960/suppl_file/jp9b00960_si_001.pdf

Chapter 6

Modeling microsolvation clusters with electronic-structure calculations guided by analytical potentials and machine learning techniques

We propose a new methodology to study, at the density functional theory (DFT) level, the clusters resulting from the microsolvation of alkali-metal ions with rare-gas atoms. The workflow begins with a global optimization search to generate a pool of low-energy minimum structures for different cluster sizes. This is achieved by employing an analytical potential energy surface (PES) and an evolutionary algorithm (EA). The next main stage of the methodology is devoted to establish the adequate DFT approach to treat the the microsolvation system, through a systematic benchmark study involving several combinations of functionals and basis sets, in order to characterize the global minimum structures of the smaller clusters. In the next stage, we apply machine learning classification algorithms to investigate how the low-energy minima of the analytical PES connect to the DFT ones. Unveiling this connection may provide useful insight to guide the choice of the most promising low-energy minima of large clusters to be optimized at the DFT level of theory. In this work, the methodology was applied to the Li^+Kr_n $(n = 2 - 14)$ microsolvation clusters for which

the most competitive DFT approach was found to be the B3LYP-D3/aug-pcseg-1.

6.1 Introduction

The study of microsolvation clusters is a relevant methodology explored by modern experimental techniques [192, 143, 193, 194], state-of-the-art computational approaches [117, 167, 195, 196, 197, 2] or both [198]. This allows to acquire insightful information about the energetics and structural properties of the clusters that have a role in the nucleation and solvation processes, thus, contributing to understand these phenomena at the molecular level.

The theoretical study of microsolvation may follow two different approaches. One method [140] searches for low-energy structures associated to an analytical potential energy surface (PES), which are then re-optimized at a higher level of theory, *i.e.*, by using *ab initio* methods or density functional theory (DFT). The other approach [199, 200, 201, 202, 90, 134] performs the global optimization search directly at the highest level of theory (usually DFT). Because the latter is very time-consuming and can be applied only for relatively small clusters, the former is the most employed one, despite being less accurate. In this approach, we may identify three main steps: (i) calculation and fitting the PES; (ii) global optimization of the cluster structure; (iii) re-optimization of the low-energy structures obtained in step (ii). To obtain a reasonable level of accuracy, the PES needs to be fitted to a large set of *ab initio* energies calculated with post-Hartree-Fock methods. Usually, the fitting procedure in step (i) applies only for a small fragment of the cluster (in general, involving two- and three-body interactions), which is then employed as the basic unit to build up the whole interaction. We should emphasize that such a modeling process requires a non-linear fitting, which benefits by the use of software recently developed by Rodríguez-Fernández *et al.* [21]. In turn, step (ii) is conducted by employing state-of-the-art global optimization methods that have been proposed in the literature over the last three decades [59, 92, 203, 204, 81, 82, 205, 95, 83, 84, 85, 206, 207, 208]. Our group has also developed evolutionary algorithms that have shown good performance in discovering low-energy structures of both atomic [133, 16, 103, 105] and

molecular [166] clusters. Finally, step (iii) is perhaps the most time-consuming, because the re-optimization is carried out at a high-level of theory and, in general, for many structures of each cluster size. Moreover, it is usual to observe a reordering of the energy minima from the analytical PES after re-optimization.

From a theoretical perspective, the microsolvation of alkali-metal ions with rare-gas atoms is appealing due to the electronic closed-shell character of these systems, which facilitates the calculation of the cluster interaction energy. Over the years, different groups have used those systems to study microsolvation on a theoretical basis [114, 119, 9, 209, 1, 18]. In particular, we have carried out global optimization studies [113, 146, 13, 19] on the Li^+Ar_n, Li^+Kr_n and $Li^+Ar_mKr_n$ clusters by employing potential energy surfaces that include up to three-body interaction terms and an evolutionary algorithm developed in our group [103]. Such studies were complemented in some specific cases by performing post-optimization *ab initio* calculations.

In this work, we focus on the post-optimization step and propose a methodology to be applied in the study of microsolvation clusters involving alkali-metal ions and rare-gas atoms. Having as reference MP2 and CCSD(T) calculations, one seeks for a DFT method that would be both theoretically accurate and computationally competitive for the study of large microsolvation clusters. In turn, we devise a strategy based on machine learning techniques to select a promising set of low-energy structures from those obtained on an analytical PES that will be, then, re-optimized at the DFT level of theory. The methodology is applied to the microsolvation of Li^+ by krypton atoms. Basically, it combines the information from both the analytical PES and DFT re-optimization of small clusters to extrapolate for the larger ones where the application of electronic-structure calculations are computationally expensive. We suggest that one may reach such achievement through the application of modern machine learning classification algorithms [210, 211, 212]. Special attention will be given to the selection of the most effective features that allow for an accurate identification of promising clusters.

The paper is organized as follows. We describe the methodology employed for the present study in Section 6.2, while the main results are presented and discussed in Section 6.3.

Finally, the conclusions are summarized in Section 6.4.

6.2 Methodology

The present methodology is described by the workflow represented in Figure 6.1. First, we employ an analytical PES and an EA to obtain a large set of low-energy minima for each size of the Li^+Kr_n clusters. Second, we re-optimize the global minimum structure of the smaller clusters at the MP2 level of theory by using aug-cc-pVQZ (or aug-cc-pVTZ) basis set up to $n = 6$ (or $n = 7$ and 8). Third, we perform single-point energy calculations at CCSD(T)/aug-cc-pVQZ (MP2/aug-cc-pVQZ) [MP2/aug-cc-pVTZ] level of theory for clusters up to $n = 5$ ($n = 6$) [$n = 8$]. The energies so obtained are taken as an accurate reference for subsequent comparison with the DFT calculations, up to $n = 8$, by using several combinations of functionals and basis sets. Fourth, a screening procedure is carried out to benchmark the most competitive DFT approach to treat Li^+Kr_n clusters in terms of accuracy and computational time consumed. Fifth, we investigate how the low-energy minima of the analytical PES connect to the corresponding ones obtained by local optimization with the established benchmark DFT approach. Since this is a time-consuming procedure due to the increasing number of local minima to be re-optimized, only small cluster sizes can be investigated. The main goal here is to investigate about the possibility of having an automatic procedure to select the local minima that are most suitable for a DFT post-optimization, in order to reach the the lowest-energy structure for each cluster size. Thus, the study is then carried out for clusters up to $n = 14$ by employing the machine learning approach to get insight about such purpose.

In the remaining of this section, we overview the analytical PES and the EA employed for the global optimization; the latter seeks for low-energy minimum structures described by the analytical PES. Additionaly, the machine learning predictor is also presented in the last part of current section.

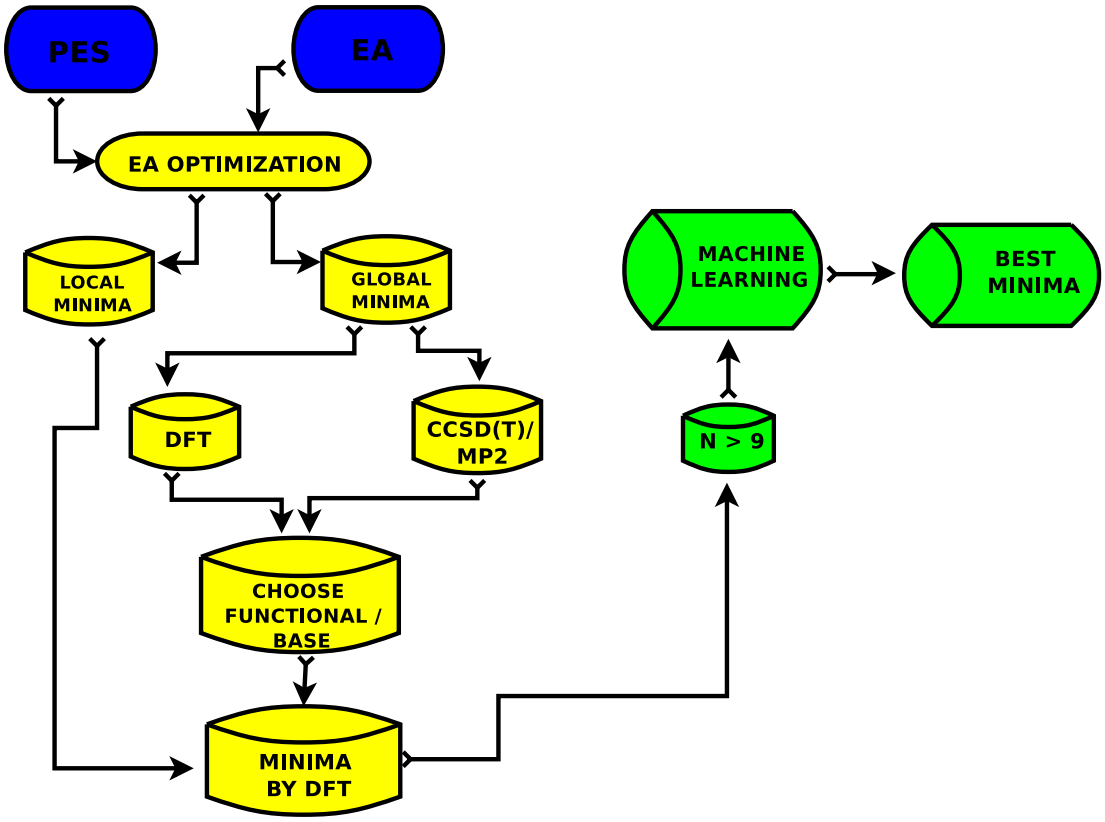


Figure 6.1: Diagram to include method DFT and Technique Machine Learning.

6.2.1 Analytical PES

The analytical PES employed in this work has been developed in our previous work [13]. It is described by the expression:

$$V(\mathbf{R}) = \sum_j V_{\mathrm{Li^+Kr}}(R_{ij}) + \sum_j \sum_{k>j} V_{\mathrm{Kr_2}}(R_{jk}) + \\ + \sum_j \sum_k V_{\mathrm{Li^+Kr_2}}(R_{ij}, R_{ik}, R_{jk}) + \sum_j \sum_k \sum_m V_{\mathrm{Kr_3}}(R_{jk}, R_{km}, R_{jm}) \quad (6.1)$$

where the summations run over all the two- and three-body contributions. We should mention that $V_{\mathrm{Li^+Kr}}$ ($V_{\mathrm{Kr_2}}$) is a pair-potential, which depends only on the corresponding Li^+-Kr (Kr-Kr) distance. In turn, $V_{\mathrm{Li^+Kr_2}}$ terms describe the three-body contributions due to the interaction between the dipoles on rare-gas atoms induced by the lithium ion and $V_{\mathrm{Kr_3}}$ is replaced by the well-known Axilrod-Teller-Muto (ATM) long-range potential for Kr_3. We

note that, in Eq. (6.1), the i-index is associated to the Li^+ ion, while the indices j, k, and m are used to label three distinct krypton atoms of a cluster system. The analytical functions of both two- and three-body terms as well as the corresponding fitting parameters have been presented in the original work[13] (see also Refs. [113, 146, 19]).

6.2.2 Global optimization method

We have employed a hybrid EA to obtain low-energy structures of the microsolvation clusters modelled by the analytical PES represented by Eq. (6.1). The EA has been developed in our group over the years [133, 155, 16, 103, 105] and, hence, only the main ingredients of the method are described below; for further details about the EA and other alternative approaches, see the recent review [64]. The algorithm follows a steady-state approach, enhanced with specific diversity measures aiming at postponing premature convergence. Each solution is represented by the 3D Cartesian coordinates of the atoms that compose the cluster. A minimum distance constraint is always enforced, preventing two atoms from getting too close to each other. Initial solutions are randomly generated. Afterwards, a quasi-Newton local optimization [63] method pushes every solution to the nearest local optimum. The energy at the local minimum corresponds to the quality of the solution.

From the set of current solutions, the sequential application of tournament selection, heterogeneous cut and splice crossover and sigma mutation obtains a pool of descendants [155, 103]. After local relaxation and evaluation, descendants replace old solutions, providing that have lower potential energy and that no two similar clusters co-exist in the solution pool. Structural dissimilarity between solutions is estimated with the center-of-mass distance, one of the most effective measures used in structural geometry optimization [62]. The optimization cycle is repeated until a pre-determined number of evaluation is reached. The most important setting of the EA used in this study are the following: number of runs: 30; population size: 100; evaluations: 1 000 000; tourney size: 5; crossover rate: 0.7; mutation rate: 0.05; standard deviation for sigma mutation: 0.1.

6.2.3 Machine learning predictor

In order to obtain the global DFT minimum, not all local optima discovered by the EA need to be re-optimized at the DFT level of theory, as some of them will lead to clearly suboptimal solutions. Given the high computational burden of a DFT optimization, it is highly relevant to develop an automatic method that filters the solutions obtained by the EA, thus passing to the next optimization level only the most promising structures.

We created a Machine Learning (ML) classifier [210] trained with a dataset of low-energy PES local optima obtained by the EA, for clusters ranging between 7 and 14 Kr atoms. All these instances from the dataset were reoptimized at the DFT level and labelled according to the quality of the solution achieved after this step (*i.e.*, at a higher level of theory). They are considered either as promising, if the DFT method achieves a low energy cluster, or as unpromising, otherwise. This last class comprises low energy local optima obtained by the EA that should not be selected for reoptimization, as they do not lead to the lowest-energy structures obtained after reoptimization at the DFT level. Our goal is then to obtain a classifier that can accurately predict which local optima obtained by the EA should be reoptimized. If a model can be used to estimate how promising an EA local optima is, then the efficiency of the whole process can be highly enhanced, by reducing the time spent reoptimizing non-promising solutions at the higher level of theory.

The dataset has the following structure: the labelled examples belong to two classes, defined as promising *versus* unpromising clusters. Classification of the examples was performed manually, as previously explained. Each instance is defined by the following 6 numeric features: number of atoms, number of symmetry operations, hyperradius (ρ), deformation indices (ξ^- and ξ^+), potential energy of the cluster as obtained by the EA and the single point DFT energy at the EA geometry (E^{SP}_{DFT}). We should note that the first feature defines the size of the cluster, while the second one is related to the point group of symmetry of the structure. In turn, the characterization of the geometrical motif takes advantage of the next

three features that can be calculated by using the expressions [213, 214]:

$$\rho^2 = \frac{I_1 + I_2 + I_3}{2M_N} \tag{6.2}$$

$$\xi^+ = \frac{I_1 - I_2}{M_N \rho^2} \tag{6.3}$$

$$\xi^- = \frac{I_3 - I_2}{M_N \rho^2} \tag{6.4}$$

where M_N is the total mass of the cluster and I_1, I_2, and I_3 are the principal moments of inertia, that can be obtained by diagonalization of the corresponding tensor of inertia. In these equations, one has assumed that $I_1 \geq I_2 \geq I_3$, which leads to $\xi^- \leq 0$ and $\xi^+ \geq 0$. It is worth noting that ρ represents a measure of the compactness of the cluster, and the combined values of ξ^- and ξ^+ are an indication of the shape [213], which may be assigned as one of the tops: spherical, prolate, oblate or asymmetric.

The dataset comprises 159 instances, in which 64 are classified as promising structures and 95 are unpromising. A random forest classifier was applied to the data [211]. No special pre-processing steps were required, as the dataset does not contain any missing values and trees are scale-invariant, thus do not needing standardized data. As the classes are imbalanced, a simple random oversampling method was applied to the training set. This method creates a balanced set by random sampling the under-represented class.

The Python Scikit-learn package, version 0.20.3 [212] and the imbalanced learn package [215] were used in all the ML tests. We applied grid search to optimize the parameters of the random forest classifier and obtained the following settings: {forest size: 200; number of features considered at each split: square root of the total number of features; maximum depth of the tree: 5; minimum samples to split an internal node: 2; criterion: information gain entropy}.

6.3 Results and discussion

6.3.1 Low-energy structures of the analytical PES

As described in Section 6.2, the present methodology begins with a global optimization search on the analytical potential energy of the cluster system. These are low-cost structures that will be used as departing geometries for the post-optimization step at the higher level of theory that is adopted in the proposed methodology. In general, the global optimization study may be carried out by any method adequate for such purpose, including the EA described in Section 6.2.2. In particular, we employ here the low-energy structures of the Li^+Kr_n $(n = 1 - 14)$ clusters, as modeled with the PES described in Section 6.2.1, that were obtained with the EA in our previous work [13]. Thus, we avoid an exhaustive discussion about the structures obtained in the global optimization and the reader is addressed to the original work for details [13]. Nonetheless, it is important to emphasize that the Li^+Kr_n clusters close the first solvation shell at $n = 6$, whose global minimum structure is octahedral. Due to the strong ion-krypton interaction in comparison to the krypton-krypton one, the structure of the first solvation shell structure is maintained for large clusters. Conversely, the second solvation shell show a "fluid-like" behavior, with several energetically similar low-lying structures [13].

6.3.2 Benchmarking the DFT approach

We have calculated the ground state energy of the Li^+Kr_n $(n = 1 - 5)$ clusters by employing different methods. Initially, the global minimum structures of these clusters are generated performing a MP2 optimization from the respective EA geometries. Specifically, we have compared the CCSD(T)/aug-cc-pVQZ//MP2/aug-cc-pVQZ electronic energies calculated using the GAMESS package [216] and several DFT results obtained by employing the NWCHEM program [217] in the same geometries. The DFT calculations involve a combination of the B3LYP [45], Slater-VWN5 (SVWN) [46], xperdew91 (PW91) [38], PBE0 (PBE) [40] and m08-HX (M08) [37] functionals, and cc-pVDZ [52, 56], aug-cc-pVDZ [52, 56],

aug-pcseg-1 [218] and aug-pcseg-2 [218] basis sets. Additionally, the Grimme's DFT-D3 [29] dispersion correction is also employed for both B3LYP and PBE functionals, hereafter designated as B3LYP-D3 and PBE-D3, respectively.

For the comparison, we have considered an efficiency (η) measure, which is defined as the product of two other quantities: (i) the mean-square deviation of electronic energies (σ^E) between the values obtained by each functional/basis-set combination and the corresponding CCSD(T) ones, and (ii) the computational time (CT) cost of the DFT calculations:

$$\eta = \sigma^E \times \mathrm{CT} \tag{6.5}$$

The σ^E, CT and η values are shown in Table 6.1. In general, lower σ^E values are obtained when the DFT-D3 correction is employed, while the smallest computational cost is for the aug-pcseg-1 basis set.

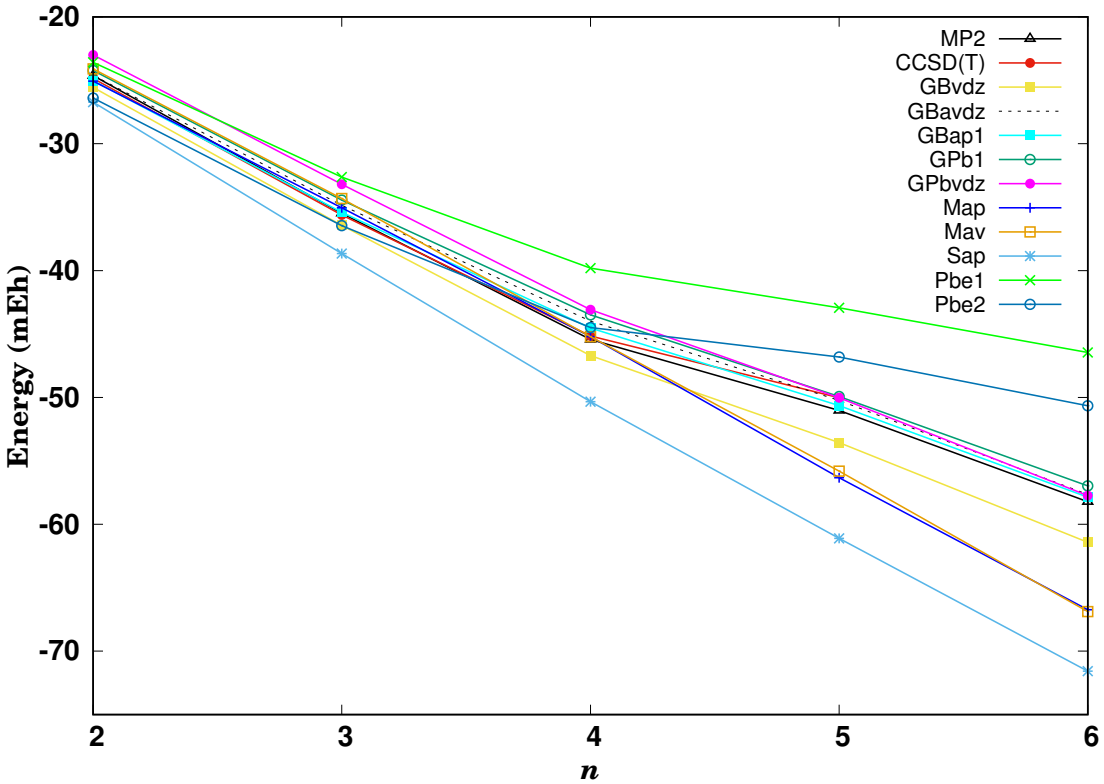


Figure 6.2: Graphic benchmarck DFT in relation the previous tabel. In comparation with calculating ab initio aug-cc-pVQZ/CCSD-T and aug-cc-pVQZ/MP2, until 5 and 6 atoms.

Table 6.1: Efficiency measure (η) of each combination of functional and basis set employed in the DFT calculations of the Li^+Kr_n ($n = 1 - 5$) clusters. Also represented are the two components of η, *i.e.*, σ^E and CT. See the text.

functional/basis-set	label	σ^E/mE$_h$	CT/s	η/mE$_h$ s	ranking
B3LYP/cc-pVDZ		135.7126	47.55	6453.13	
B3LYP/aug-cc-pVDZ		252.9000	82.45	20851.60	
B3LYP/aug-pcseg-1		234.0495	15.63	3657.02	
B3LYP/aug-pcseg-2		82.9882	95.05	7888.02	
PW91/cc-pVDZ		101.7733	49.85	5073.40	
PW91/aug-cc-pVDZ		423.4874	126.03	53370.00	
PW91/aug-pcseg-1		391.0059	14.40	5630.48	
PW91/aug-pcseg-2		224.0767	95.48	21393.72	
PBE/cc-pVDZ		126.8861	41.85	5310.18	
PBE/aug-cc-pVDZ		122.1766	64.25	7849.85	
PBE/aug-pcseg-1	Pbe1	102.5276	14.20	1455.89	8
PBE/aug-pcseg-2	Pbe2	14.9636	93.68	1401.72	7
SVWN/cc-pVDZ		265.2364	46.88	12432.96	
SVWN/aug-cc-pVDZ		137.7678	78.58	10825.11	
SVWN/aug-pcseg-1	Sap	133.3386	12.98	1730.07	9
SVWN/aug-pcseg-2		367.7625	86.20	31701.12	
M08/cc-pVDZ		45.1738	58.35	2635.89	
M08/aug-cc-pVDZ	Mav	29.0926	61.30	1783.37	10
M08/aug-pcseg-1	Map	29.6841	12.93	383.67	4
M08/aug-pcseg-2		105.3890	100.48	10588.96	
PBE-D3/cc-pVDZ	GPbvdz	20.2180	41.25	833.99	6
PBE-D3/aug-cc-pVDZ		75.5964	66.38	5017.71	
PBE-D3/aug-pcseg-1	GPb1	7.0031	13.40	93.84	2
PBE-D3/aug-pcseg-2		34.8064	89.78	3124.74	
B3LYP-D3/cc-pVDZ	GBvdz	9.6238	47.55	457.61	5
B3LYP-D3/aug-cc-pVDZ	GBavdz	3.7078	83.10	308.12	3
B3LYP-D3/aug-pcseg-1	GBap1	1.7199	14.78	25.41	1
B3LYP-D3/aug-pcseg-2		47.6105	93.43	4448.01	

Considering the top-ten DFT approaches (*i.e.*, those with the lowest values of η, explicitly indicated in Table 6.1), we then recalculated the ground-state energy from structures obtained performing a DFT re-optimization. These results are displayed in Figure 6.2, jointly with the corresponding MP2/aug-cc-pVQZ (and CCSD(T)/aug-cc-pVQZ) values for Li^+Kr_n aggregates up to $n = 6$ ($n = 5$). Although showing a small separation for $n = 5$, the CCSD(T) and MP2 results follow essentially the same trend. Basically, this behavior may be described by two straight lines: one extending up to $n = 4$, while the other, with a lower slope, is apparent for $n \geq 4$. In contrast to such behavior, some of the DFT results (Map, Mav and Sap) follow a sole straight-line behavior up to $n = 6$; in particular, Sap performs badly in comparison with the *ab initio* results, even for the smallest cluster sizes. Other DFT results (Pbe1 and Pbe2) follow two straight lines as above mentioned for MP2 and CCSD(T), but clearly fail to reproduce the *ab initio* results, especially for $n \geq 5$.

It is also apparent from Figure 6.2 that the best five DFT results in comparison to the *ab initio* energies occur when including the Grimme's D3 dispersion correction in the calculations (*i.e.*, GBap1, GPb1, GBavdz, GBvdz and GPbvdz). In general, this observation is in agreement with the ranking displayed in Table 6.1; due to a smaller value of the CT parameter, however, Map results are ranked as 4 in Table 6.1, while GPbvdz that performs much better in terms of energy appears in the sixth position. Accordingly, we have selected GBap1, GPb1, GBavdz, GBvdz and GPbvdz to perform DFT calculations for $n = 7$ and 8. The corresponding DFT energies are represented in Figure 6.3 for $Li^+Kr_n, (n = 2 - 8)$ clusters. In this figure, we also show the previous CCSD(T)/aug-cc-pVQZ and MP2/aug-cc-pVQZ results for aggregates up to $n = 5$ and $n = 6$, respectively. Since we could not perform MP2 calculations with a so large basis set for $n > 6$, we also carried out MP2/aug-cc-pVTZ calculations for all aggregates up to $n = 8$. Such curve is also shown in Figure 6.3.

In this step we perform an analysis with $Li^+Kr_n, (n = 2 - 8)$ clusters by evaluating the mean-square deviation of electronic energies between these five DFT levels of theory and the corresponding *ab initio* ones. Specifically, we considered as reference data the energies calculated using CCSD(T)/aug-cc-pVQZ (up to $n = 5$), and MP2/aug-cc-pVQZ ($n = 6$) and MP2/aug-cc-pVTZ ($n = 7, 8$).

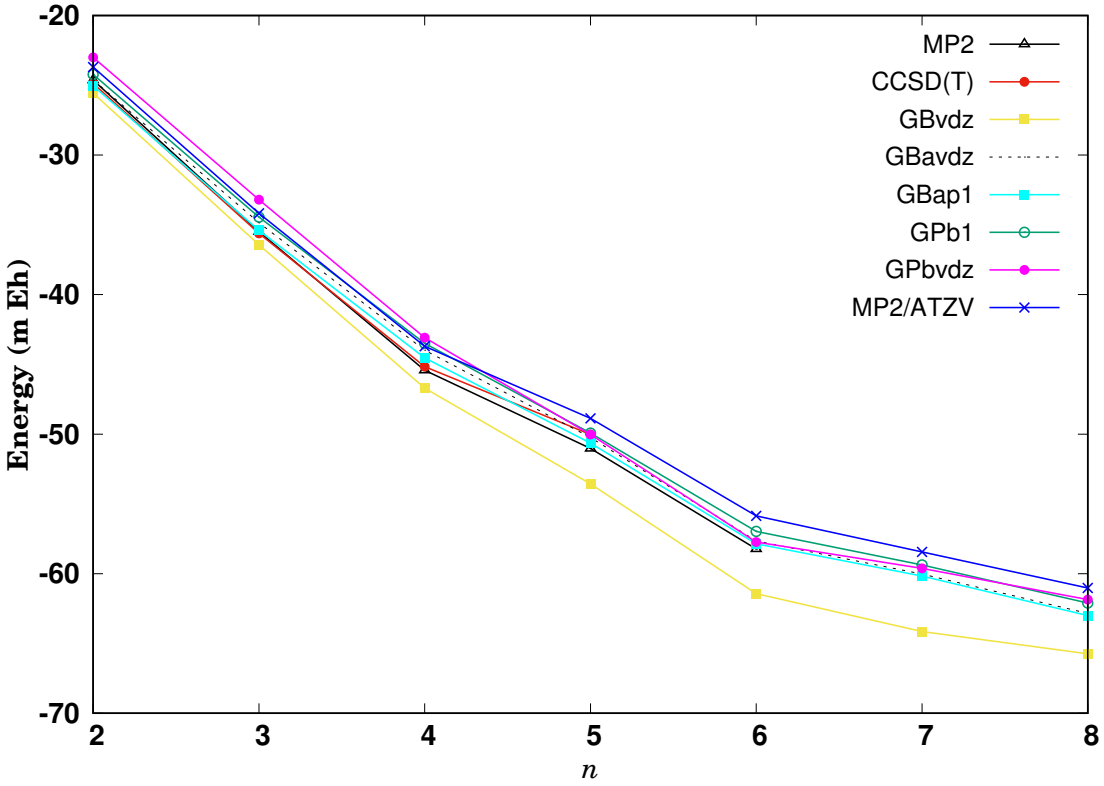


Figure 6.3: Graphic benchmarck DFT more refined in relation the previous graphic (Figure 6.2). In comparation with calculating ab initio aug-cc-pVQZ/CCSD-T and aug-cc-pVTZ/MP2, until 5 and 8 atoms.

These σ^E are presented in Table 6.2 and they confirm that the B3LYP-D3/aug-pcseg-1 level of theory presents a behavior more approximate with the CCSD(T) and MP2 ones. All subsequent DFT calculations were then performed with this combination of functional and basis set. Note that in Table 6.2 the PBE-D3/aug-pcseg-1 and B3LYP-D3/aug-pVDZ calculations had deviations close to B3LYP-D3/aug-pcseg-1 ones when are considered Li^+Kr_n aggregates up to $n = 8$, differently from what we have observed in Table 6.1, when B3LYP-D3/aug-pcseg-1 level of theory presents clearly the lowest σ^E. This occurs mainly due the MP2/aug-cc-pVTZ results, which are always larger than the *ab initio* results with quadruple-zeta basis-set, being closer to PBE-D3/aug-pcseg-1 and B3LYP-D3/aug-pVDZ ones.

Table 6.2: Mean-square deviation of electronic energies (σ^E) between the values obtained by each functional/basis-set combination and the corresponding CCSD(T)/aug-cc-pVQZ (up to $n = 5$), MP2/aug-cc-pVQZ ($n = 6$) and MP2/aug-cc-pVTZ ($n = 7, 8$).

functional/basis-set	σ^E/mE$_h$
B3LYP-D3/aug-pcseg-1	7.8878
PBE-D3/aug-pcseg-1	8.1969
B3LYP-D3/aug-cc-pVDZ	8.1973
PBE-D3/cc-pVDZ	15.9535
B3LYP-D3/cc-pVDZ	81.1473

6.3.3 DFT optimization departing from the minima of the analytical PES

By employing the DFT method with B3LYP-D3/aug-pcseg-1 (the best approach selected in the previous subsection), we have performed re-optimization of the global and local minima obtained with the EA for the Li^+Kr_n ($n = 7 - 14$) clusters; local minima with up to 3.0 mE_h above the global minimum have been selected for re-optimization. We should emphasize that the process of DFT re-optimization was carried out as follows: first, it is performed an optimization by using the default parameters of NWCHEM software; second, it is performed another minimization by employing thinner optimization-step parameters; then the energies of both optimizations are compared and, if the difference is greater than 0.1 mE_h, we perform another optimization with even thinner parameters. The DFT result considered is the energy obtained in the last optimization performed.

In Figure 6.4 we display the comparison between the optimized EA and re-optimized DFT energies (from the global ones), indicating the connection between them. We can observe in this figure that not always the global minimum obtained in the DFT re-optimization corresponds to the global minimum of the EA method. We can mention, for example, the

case of aggregate Li^+Kr_{10}, when the DFT global minimum is obtained from an EA local minimum located about 2 mE_h above of the EA global minimum, whereas the EA global minimum generates a DFT minimum with energy about 0.8 mE_h greater then the global DFT minimum.

To try to understand if there is any characteristic in the (global and local) EA minima that indicates which are the most likely to generate the global DFT minimum structure (after the re-optimization step), we have calculated some quantities of interest from each EA structure: the symmetry group, hyperradius and deformation indices, and the potential energy given by the analytical PES. These results are summarized in Table A.1 of Appendix, and an example of EA and structures of all minima considered for the Li^+Kr_7 is shown in Figure 6.5. In this figure, it is also shown also the EA and DFT energies in relation to the corresponding ground state. We can see that three different EA structures (with energy differences up to 2.9 mE_h) go into the minimum energy structure after DFT re-optimization (within optimization accuracy).

Analyzing the results in Table of the Apendix and Figure 6.5, we can hardly identify a simple pattern among the various quantities indicating whether a minimum (global or local) from the analytical PES will lead to the lowest-energy structure after the DFT post-optimization. In the next subsection, we have employed a modern ML approach to get any insight about this issue.

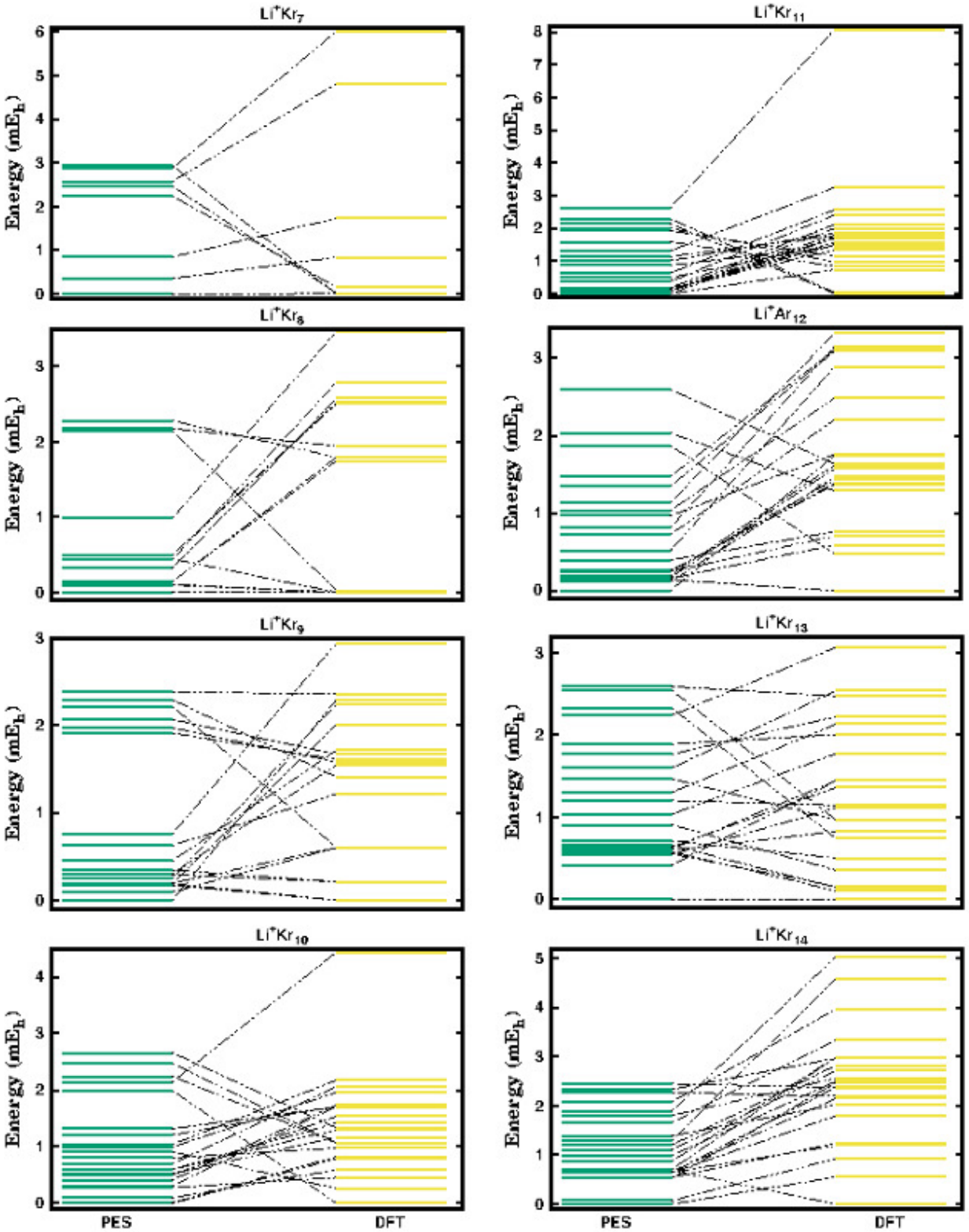


Figure 6.4: Comparation between optimization by GA and Post-optimization DFT.

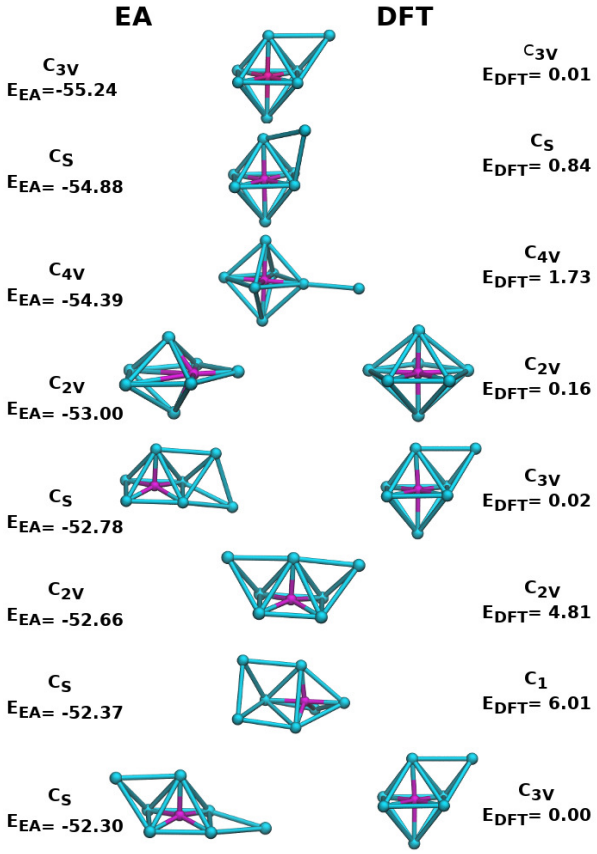


Figure 6.5: Structures of the Li^+Kr_7 minima obtained with the EA. Also indicated are the corresponding symmetry and energy in the analytical PES (second column) and the DFT energy in relation to the global minimum (third column). Other parameters used as geometrical descriptors in the ML method are also given in the second column. Energies are in mEh.

6.3.4 Training and validation of the ML classifier

A 4-fold cross validation model [210] was applied to access the effectiveness of the model. The global predictive accuracy obtained is around 61% and the detailed confusion matrix can be consulted in Figure 6.6.

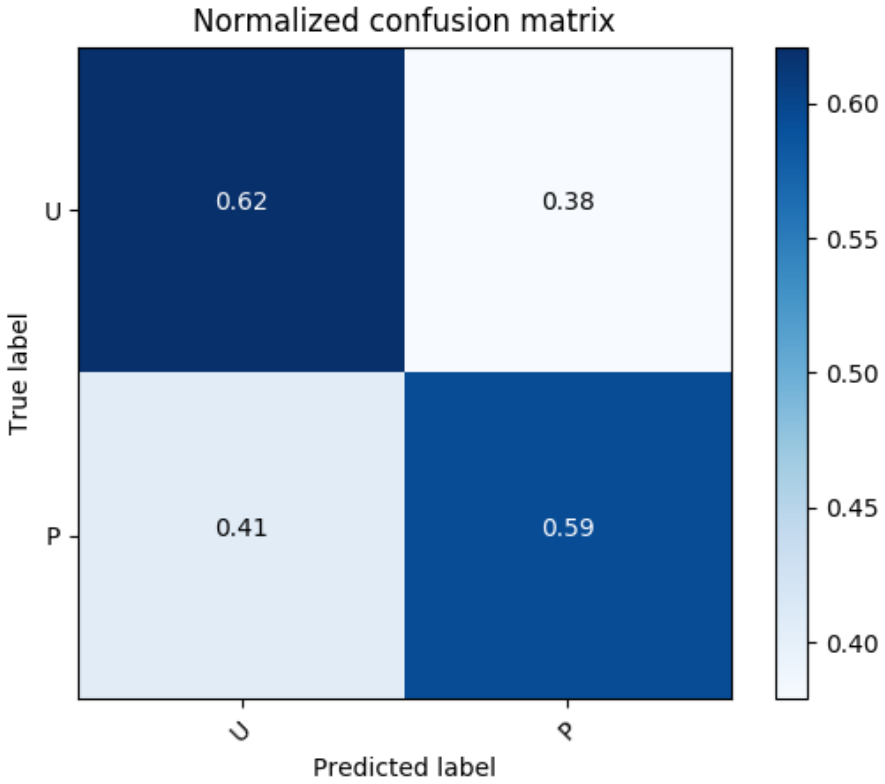


Figure 6.6: Confusion matrix obtained with the basic 6 features. The classes U and P represent, respectively, (U)npromising and (P)romising clusters.

Results reveal that there is a balanced classification of the two classes, but the accuracy is not impressive, as it misses the correct identification in approximately 4 out of each 10 clusters. As it stands, this simple model is still quite error-prone on the detection of promising geometries to optimize at the DFT level. Particularly undesirable is the rate of false negatives which is above 40%, as it corresponds to promising structures that are wrongly classified by the model.

Aiming at enhancing the predictive accuracy of the classifier, we added several new features that might help to describe important properties of the clusters. It is well-known that geometric characteristics of aggregates, as determined by local arrangements of particles, can impact the quality of clusters [16]. Therefore this information might help the classifier to enhance its effectiveness. For this study we consider the local connectivity information [219]. This measure creates a table with detailed information of how many connections each atom

has. Two atoms are connected if they stand at distance smaller than a given threshold. For the clusters on this study we consider a minimum distance of $8.0\,a_0$ and a maximum of 12 neighbors, leading to the addition of 12 descriptors.

The modified dataset has now 18 features to describe the 159 instances. We applied the classifier previously presented (with the same settings) and it achieved a global predictive accuracy of 65%. The detailed confusion matrix is presented in Figure 6.7. Results displayed in the matrix confirm that the geometric information of the clusters is relevant to predict how promising the structure is. The correct classifications increase in both classes, most notably in the class of unpromising clusters that raises to 67% of correct predictions. On the negative side, the rate of false negatives is 39%. This value is still too high, as the model will fail to highlight many promising clusters for DFT reoptimization.

To further enhance the classification accuracy of the ML model, we added an additional feature. The methodology proposed in this work comprises two sequential steps, in which promising clusters discovered by the EA are re-optimized at a DFT higher level of theory. This second step is an iterative process: first, local optima discovered by the EA are mapped to the DFT landscape and then the iterative DFT optimization proceeds. Accordingly, we take the first value obtained at the DFT level and use it as an additional feature. Note that this does not imply a high computational overhead, as only one DFT evaluation is performed. Before application, this feature was engineered in the following way: for each cluster size (between 7 and 14), we determined the minimum DFT value. Then, the value for each engineered feature was obtained by subtracting the corresponding minimum from the original DFT quantity for that feature.

The final dataset has now 19 features to describe the 159 instances. We applied the classifier previously presented (with the same settings) and it achieved a global predictive accuracy of 69%. The detailed confusion matrix is presented in Figure 6.8. There is a noteworthy increase in the classification accuracy of the promising clusters and the rate of false negatives drops to 30%. This confirms that the first energy obtained at the DFT level of theory is relevant and it provides an essential contribute to discriminate between promising and unpromising clusters.

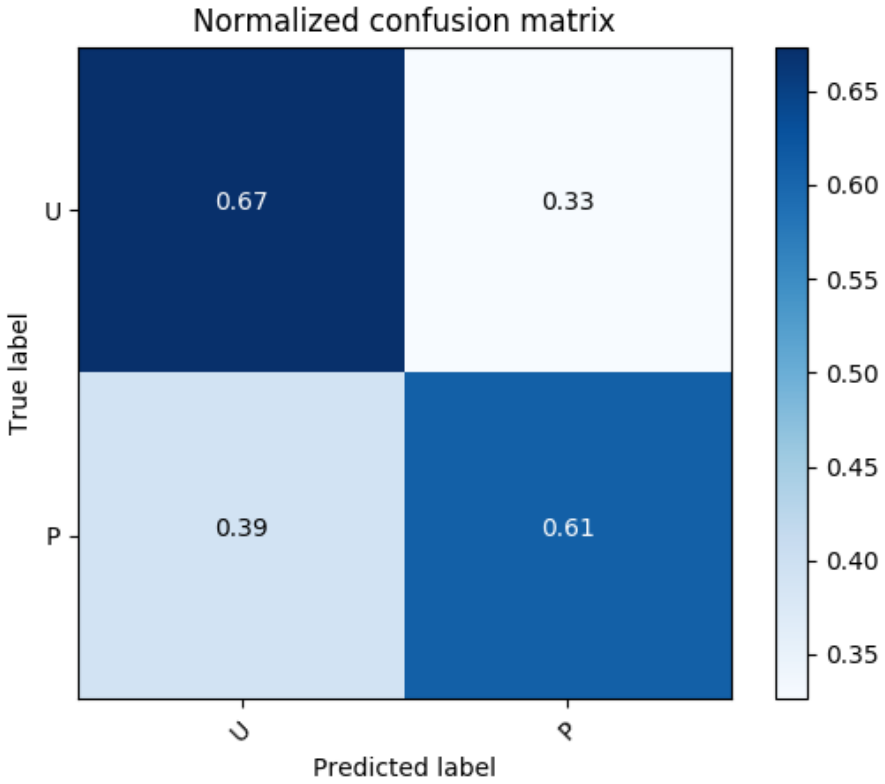


Figure 6.7: Confusion matrix obtained with the extended 18 features. The classes U and P represent, respectively, (U)npromising and (P)romising clusters.

To confirm that the new feature is indeed enhancing the predictive accuracy of the ML model we estimated the mutual information between each feature and the target. This measure is calculated using a nonparametric entropy estimation method based on the k-nearest neighbour approach and it estimates the dependency between each feature and the target [220]. It is 0 if the two variables are independent, whilst higher values correspond to higher dependencies. Values obtained are low (always below 0.14), confirming that single features are unable to accurately estimate the target. In any case, higher values are obtained by the features encoding the first DFT energy and the energy obtained by the EA. Even though weaker than the 2 previous features, a few connectivity values (namely, 3, 6, and 7 neighbors) also achieve a reasonable correlation with the target.

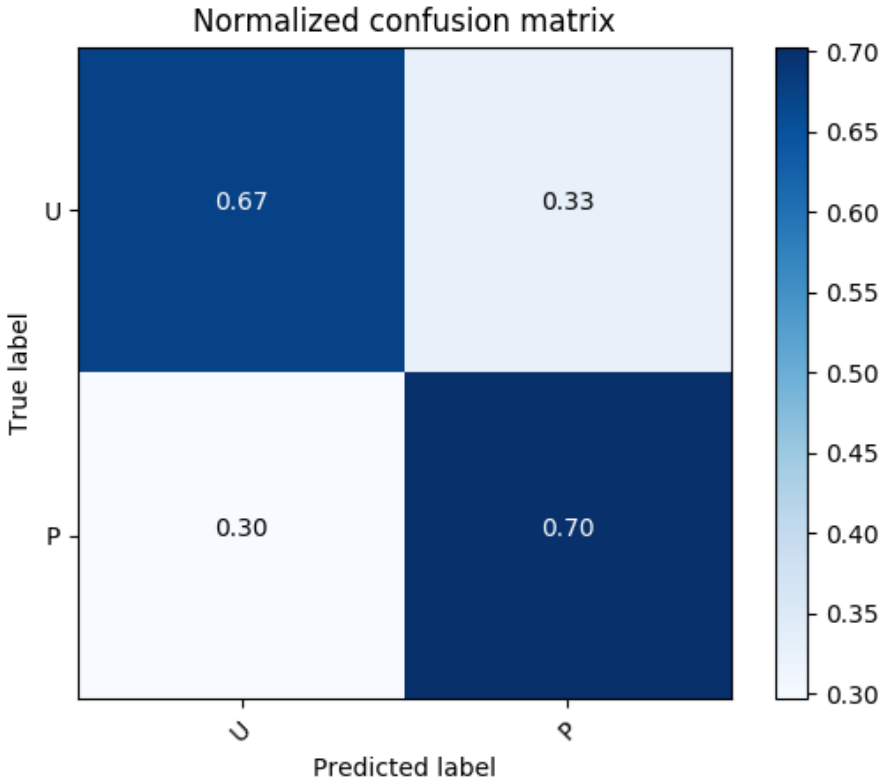


Figure 6.8: Confusion matrix obtained with the extended 19 features. The classes U and P represent, respectively, (U)npromising and (P)romising clusters.

6.4 Conclusions

We have devised a strategy based on ML techniques to reduce the computational burden related to the re-optimization at a high theoretical level of structures modeled by an analytical PES. The methodology relies on the re-optimization at the DFT level of theory and it has been applied to clusters resulting from the microsolvation of Li^+ by krypton atoms. The choice of the DFT approach resorted to a benchmark process involving several combinations of functionals and basis-sets, some of them including D3 dispersion corrections. By considering Li^+Kr_n clusters up to $n = 8$, we found out that B3LYP-D3/aug-pcseg-1 is the best choice when considering the computational time and the comparison with MP2 (or CCSD(T), for the smaller clusters) energies. The DFT/B3LYP-D3/aug-pcseg-1 method was

employed to re-optimize all the low-energy minima of the Li^+Kr_n $(n = 2 - 14)$, which were, then, used to test the ML classifier.

The ML classifier proposed in this work is effective in identifying promising structures for DFT reoptimization. The feature set comprising geometrical information and first DFT energetics was able to accurately detect the best local optima that were obtained by the analytical potential. This is a relevant result, as it promotes a massive efficiency enhancement in the DFT optimization step. Current results show that the ML model is accurate to detect promising clusters with the same size as those used in training. Our next research step will address a generalization that allows for the accurate detection of promising clusters with a higher number of particles.

SUPPORTING INFORMATION

The Supporting Information is available in Apendix.

Chapter 7

Conclusion and Remarks

In this thesi we have related a review of the tools developed for the construct and study of clusters. We begin the study on the development of EAs for searching low-energy structures of atomic clusters with the two-bodys potential. It were alkali-ion (Li, Na and K) microsolvation with argon, where the results show that strong "magic numbers" are associated with the closure of the first solvation and the global minimum structures for K^+Ar_n clusters are, in general, the most symsetric ones. For discovering global minimum structures of Rh-Cu transiton-metal clusters, the global minimum structures present segregation of Cu atoms on the surface of the aggregates, presenting compatibility with the experimental cohesion energies. And the self-adaptation show that is a promising avenue for future research.

In a second moment we perfom a comparision study of low-energy landscape of the clusters resulting from the microsolvation of Li^+Ar_n and Li^+Kr_n for clusters with up to 14 rare-gas atoms and that include either two and three-body interactions. Where the Li^+Ar_n low-lying minima tend to be concentrated in a narrow energy window close to the corresponding global minimum, whereas the Li^+Kr_n ones are in larger number and spread out over the energy range. We perform also a analysis of the local minima in the first and second solvation shell. A other curiosity in this study is the fact of the structures contain rare-gas atoms in the second solvation shell with the ion essentially in the center in PES II and off-center in PES I. The stability of such type of structures results from a subtle balance

between the interaction involving the second-shell atoms alone and that from the inter-shell species energy contribution. But in this stage confirmed that the incorporation of three-body terms in the PES is essential to reproduce the main structural motifs.

With the goal of discover the binary-solvent clusters most stable, it were proposed perform a study of solvent-binary of the $Li^+Ar_nKr_m$, and as answer, we obtain six krypton atoms occupied the first and second sovatation shell. Indeed, the microsolvation of the ion leads to formation of clusters, where distinct argon and krypton subclusters are apparent.

For last, to guide the choice of the most promising low-energy minima of Li^+Kr_n(n = 2 - 14) to be optimized at the DFT level of theory, it was proposed this work the Machine Learning classifer. The results show that the Machine Learing model is accurate to detect promising clusters with the same size as those used in training. The next step is analyse parameters that will contribute for allows an accurate detection of promising clusters with a higher number of particles.

Future prospects in this research involves performing a study of building molecular clusters. Until now we have only investigated the atoms clusters, since we are interested in performing a study of clusters of atomic ions by carbon gas molecules and heterocyclic organic molecules. We aim to analyze the thermodynamic properties using a code developed in our research group. To obtain better results we will optimize the use of Machine Learning technique. Moreover, to obtain better results in the fitting of two and three-bodies interation potential functions, where will be built potential functions that allow it to consider possible contributions of the four or five bodies.

Bibliography

[1] M. Slama, K. Issa, F. B. Mohamed, M. B. E. H. Rhouma, and F. Spiegelman, "Structure and stability of na^+ xe_n clusters," *Eur. Phys. J. D*, vol. 70, p. 242, 2016.

[2] J. M. C. Marques, F. B. Pereira, J. L. Llanio-Trujillo, P. E. Abreu, M. Albertí, A. Aguilar, F. Pirani, and M. Bartolomei, "A global optimization perspective on molecular clusters," *Phil. Trans. R. Soc. A*, vol. 375, p. 20160198, 2017.

[3] F. E. B. Mohamed, M. Slama, H. Hammami, M. B. E. H. Rhouma, and M. Hochlaf, "Microsolvation of no^+ in ar_n clusters: A theoretical treatment," *J. Chem. Phys.*, vol. 142, p. 20160198, 2017.

[4] E. Florez, N. Acelas, F. Ramirez, C. Hadad, and A. Restrepo, "Microsolvation of f^-," *Phys. Chem. Chem. Phys.*, vol. 20, pp. 8909–8916, 2018.

[5] W. K. Tang, M. Chau, and C. Siu, "Role of water in molecular oxygen activation in hydrated chromium(i) cluster ions: A theoretical insight," *Int. J. Mass Spectrom.*, vol. 436, pp. 118–126, 2019.

[6] M. Rastogi, C. Leidlmair, L. A. der Lan, J. O. Zárate, R. P. Tudela, M. Bartolomei, M. I. Hernández, J. Campos-Martínez, T. González-Lezana, J. Hernández-Rojas, J. Bretón, P. Scheier, and M. Gatchell, "Lithium ions solvated in helium," *Phys. Chem. Chem. Phys.*, vol. 20, pp. 25569–25576, 2018.

[7] M. Slama, K. Issa, M. Zbidi, and M. B. E. H. Rhouma, "Microsolvation of k^+ in xenon clusters: a three-body approximation and structural transition," *Mol. Phys.*, vol. 115, pp. 757–770, 2017.

[8] E. Yurtsever, E. Yildirim, M. Yurtsever, E. Bodo, and F. A. Gianturco, "Solvation of k^+ in helium droplets," *Eur. Phys. J. D*, vol. 43, pp. 105–108, 2007.

[9] M. B. E. H. Rhouma, F. Calvo, and F. Spiegelman, "Solvation of na^+ in argon clusters," *J. Phys. Chem. A*, vol. 110, pp. 5010–5016, 2006.

[10] E. Gomma, M. A. Tahoon, and A. Negm, "Aqueous micro-solvation of li^+ ions: Thermodynamics and energetic studies of li^+-$(h_2o)_n$ (n = 1–6) structures," *J. Mol. Liq.*, vol. 241, pp. 595–602, 2017.

[11] E. Bodo, E. Yurtsever, M. Yurtsever, and F. A. Gianturco, "Ionic dimers in he droplets: Interaction potentials for li_2^+-he, na_2^+-he and k_2^+-he and stability of the smaller clusters," *J. Chem. Phys.*, vol. 124, p. 074320, 2006.

[12] A. Eiben and J. Smith, *Introduction to Evolutionary Computation*. Springer, Berlin, Heidelberg, 2 ed., 2015.

[13] W. S. Jesus, J. M. C. Marques, F. V. Prudente, and F. B. Pereira, "Exploring the first-shell and second-shell structures arising in the microsolvation of li^+ by rare gases," *Int. J. Quantum Chem.*, vol. 119, p. e25860, 2019.

[14] Y. S. Doronin, V. N. Samovarov, and E. A. Bondarenko, "Heterogeneous cluster formation in a supersonic argon-krypton jet according to cathodoluminescence data in the vacuum ultraviolet," *Low Temp. Phys.*, vol. 32, pp. 251–255, 2006.

[15] A. G. Danil'chenko, S. I. Kovalenko, and V. N. Samovarov, "Phase separation into pure components in mixed ar-xe clusters," *Pis'ma v Zh. Èksper. Teoret. Fiz.*, vol. 84, pp. 385–389, 2006.

[16] F. B. Pereira and J. M. C. Marques, "A study on diversity for cluster geometry optimization," *Evol. Intel.*, vol. 2, pp. 121–140, 2009.

[17] J. M. C. Marques, A. A. C. C. Pais, and P. E. Abreu, "Generation and characterization of low-energy structures in atomic clusters," *J. Comput. Chem.*, vol. 31, pp. 1495–1503, 2010.

[18] J. M. C. Marques, W. S. Jesus, F. V. Prudente, F. B. Pereira, and N. Lourenço, "Revealing energy landscapes of atomic clusters by applying adaptive bio-inspired algorithms," in *Physical Chemistry for Chemists and Chemical Engineers: Multidisciplinary Research Perspectives* (A. V. Vakhrushev and J. V. J. R. Haghi, eds.), pp. 47–74, Oakville: Apple Academic Press, 2018.

[19] W. S. Jesus, J. M. C. Marques, and F. V. Prudente, "Microsolvation of li^+ in a mixture of argon and krypton: Unveiling the most stable structures of the clusters," *J. Phys. Chem. A*, vol. 123, pp. 2867–2873, 2019.

[20] W. S. Jesus, J. M. C. Marques, F. V. Prudente, F. B. Pereira, and N. Lourenço, "Modeling microsolvation clusters with electronic-structure calculations guided by analytical potentials and machine learning techniques," *J. Chem. Theory Comput.*, to be submitted (2019).

[21] R. Rodríguez-Fernández, F. B. Pereira, J. M. C. Marques, E. Martínez-Núñez, and S. A. Vázquez, "Gafit: a general-purpose, user-friendly program for fitting potential energy surfaces," *Comput. Phys. Commun.*, vol. 217, pp. 89–98, 2017.

[22] M. A. Castro and S. Canuto, "Métodos pertubativos para a obtenção de correlação eletrônica," in *Métodos de Química Teórica e Modelagem Molecular* (N. H. Morgon and K. Coutinho, eds.), pp. 113–144, São Paulo: Livraria da Física, 2007.

[23] I. Shavitt and R. J. Barlett, *Many-Body Methods in Chemistry and Physics*. Cambridge University Press, New York, Cambridge, 2 ed., 2009.

[24] G. E. Scuseria, "Analytic evaluation of energy gradients for the singles and doubles coupled cluster method including perturbative triple excitations: Theory and applications to foof and cr_2," *J. Chem. Phys.*, vol. 94, pp. 442–447, 1991.

[25] K. Raghavachari and G. Trucks, "Complete basis set limit second-order moller-plesset calculations for the fcc lattices of neon, argon, krypton, and xenon," *Chern. Phys. Lett.*, vol. 165, p. 511, 1989.

[26] P. Hoenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.*, vol. 136, p. B864, 1964.

[27] S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," *J. Comput. Chem.*, vol. 32, pp. 1456–1465, 2011.

[28] D. Sharapa, J. T. Margraf, A. Hesselmann, and T. Clark, "Accurate intermolecular potential for the c_60 dimer: The performance of different levels of quantum theory," *J. Chem. Theory Comput.*, vol. 13, pp. 274–285, 2017.

[29] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu," *J. Chem. Phys.*, vol. 132, p. 154104, 2010.

[30] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, and E. Kaxiras, "Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment," *J. Chem. Phys.*, vol. 114, p. 5149, 2001.

[31] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "Accurate description of van der waals complexes by density functional theory including empirical corrections," *J. Comput. Chem.*, vol. 25, p. 1463, 2004.

[32] P. Jurečka, J. Černý, P. Hobza, and D. R. Salahub, "Density functional theory augmented with an empirical dispersion term. interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations," *J. Comput. Chem.*, vol. 28, p. 555, 2007.

[33] N. Marom, A. Tkatchenko, M. Scheffler, and L. Kronik, "Describing both dispersion interactions and electronic structure using density functional theory: The case of metal−phthalocyanine dimers," *J. Chem. Theory Comput.*, vol. 6, pp. 81–90, 2010.

[34] S. Grimme, "Semiempirical hybrid density functional with perturbative second-order correlation," *J. Chem. Phys.*, vol. 124, p. 034108, 2006.

[35] A. Tkatchenko and M. Scheffler, "Accurate molecular van der waals interactions from ground-state electron density and free-atom reference data," *Phys. Rev. Lett.*, vol. 102, p. 073005, 2009.

[36] S. Steinmann, G. Csonka, and C. Corminboeuf, "Unified inter- and intramolecular dispersion correction formula for generalized gradient approximation density functional theory," *J. Chem. Theory Comput.*, vol. 5, pp. 2950–2958, 2009.

[37] Y. Zhao and D. Truhlar, "Exploring the limit of accuracy of the global hybrid meta density functional for main-group thermochemistry, kinetics, and noncovalent interactions," *J. Chem. Theory Comput.*, vol. 4, pp. 1849–1868, 2008.

[38] J. Perdew, J. Chevary, S. Vosko, K. Jackson, M. Pederson, D. Singh, and C. Fiolhais, "Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation," *Phys. Rev. B*, vol. 46, p. 6671, 1992.

[39] J. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.*, vol. 78, p. 1396, 1997.

[40] C. Adamo and V. Barone, "Toward reliable density functional methods without adjustable parameters: The pbe0 model," *J. Chem. Phys.*, vol. 110, p. 6158, 1999.

[41] D. Novikov, A. Freeman, N. Christensen, A. Svane, and C.O.Rodriguez, "Lda simulations of pressure-induced anomalies in c/a and electric-field gradients for zn and cd," *Phys. Rev. B*, vol. 56, p. 7206, 1997.

[42] Y. Zhang, W. Pan, and W. Yang, "Describing van der waals interaction in diatomic molecules with generalized gradient approximations: The role of the exchange functional," *J. Chem. Phys.*, vol. 107, p. 7921, 1997.

[43] B. Montanari, P. Ballone, and R. Jones, "Density functional study of molecular crystals: Polyethylene and a crystalline analog of bisphenol-a polycarbonate," *J. Chem. Phys.*, vol. 108, p. 6947, 1998.

[44] A. Becke, "Correlation energy of an inhomogeneous electron gas: A coordinate[U+2010]space model," *J. Chem. Phys.*, vol. 88, p. 1053, 1988.

[45] A. Becke, "Density-functional thermochemistry. iii. the role of exact exchange," *J. Chem. Phys.*, vol. 98, p. 5648, 1993.

[46] J. Slater and K. Jhonson, "Self-consistent-field x_α cluster method for polyatomic molecules and solids," *Phys. Rev. B*, vol. 5, p. 844, 1971.

[47] S. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis," *J. Phys.*, vol. 58, p. 1200, 1980.

[48] L. Viegas, "Assessment of model chemistries for hydrofluoropolyethers: A dft/m08-hx benchmark study," *Int J Quantum Chem.*, vol. 117, p. e25381, 2017.

[49] L. Viegas, "Multiconformer transition state theory rate constants for the reaction between oh and α, ω-dimethoxyfluoropolyethers," *Int J Chem Kinet.*, vol. 51, pp. 358–366, 2019.

[50] J.-W. Song, S. Tokura, T. Sato, M. Watson, and K. Hirao, "Long-range corrected density functional calculations of chemical reactions: Redetermination of parameter," *J. Chem. Phys.*, vol. 127, p. 154109, 2007.

[51] O. Vydrov and G. Scuseria, "Assessment of a long-range corrected hybrid functional," *J Chem. Phys.*, vol. 125, p. 234109, 2009.

[52] B. Prascher, D. Woon, K. Peterson, T. D. Jr, and A. Wilson, "Gaussian basis sets for use in correlated molecular calculations. vii. valence, core-valence, and scalar relativistic basis sets for li, be, na, and mg," *Theor. Chem. Acc.*, vol. 128, pp. 69–82, 2011.

[53] T. H. Dunning, Jr., "Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen," *J. Chem. Phys.*, vol. 90, pp. 1007–1023, 1989.

[54] R. Kendall, T. Dunning, and R. Harrison, "Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions," *J. Chem. Phys.*, vol. 96, p. 6796, 1992.

[55] R.C.Raffenetti, "General contraction of gaussian atomic orbitals: Core, valence, polarization, and diffuse basis sets; molecular integral evaluation," *J. Chem. Phys.*, vol. 58, p. 4452, 1973.

[56] A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning, Jr., "Gaussian basis sets for use in correlated molecular calculations. ix. the atoms gallium through krypton," *J. Chem. Phys.*, vol. 96, pp. 7667–7676, 1999.

[57] F. Jensen, "Polarization consistent basis sets. 4: The elements he, li, be, b, ne, na, mg, al, and ar," *J. Chem. Phys. A*, vol. 111, pp. 11198–11204, 2007.

[58] F. Jensen, "Polarization consistent basis sets. vii. the elements k, ca, ga, ge, as, se, br, and kr," *J. Chem. Phys.*, vol. 136, p. 114107, 2012.

[59] B. Hartke, "Global geometry optimization of clusters using genetic algorithms," *J. Phys. Chem.*, vol. 97, pp. 9973–9976, 1993.

[60] Y. Xiao and D. E. Williams, "Genetic algorithms: a new approach to the prediction of the structure of molecular clusters," *Chem. Phys. Lett.*, vol. 215, pp. 17–24, 1993.

[61] A. Cassioli, M. Locatelli, and F. Schoen, "Global optimization of binary lennard-jones clusters," *Optim. Methods Softw.*, vol. 24, pp. 819–835, 2009.

[62] A. Grosso, M. Locatelli, and F. Schoen, "A population-based approach for hard global optimization problems based on dissimilarity measures," *Math. Program. Ser. A*, vol. 110, pp. 373–404, 2007.

[63] D. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Math. Program. B*, vol. 45, pp. 503–528, 1989.

[64] J. M. C. Marques and F. B. Pereira, "Colloidal clusters from a global optimization perspective," *J. Mol. Liq.*, vol. 210, pp. 51–63, 2015.

[65] H. M. Chen, R. S. Liu, L.-Y. Jang, J.-F. Lee, and S. F. Hu, "Characterization of core-shell type and alloy ag/ au bimetallic clusters by using extended x-ray absorption fine structure spectroscopy," *Chem. Phys. Lett.*, vol. 421, p. 118, 2006.

[66] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, pp. 223–230, 1973.

[67] A. Schug, P. C. Whitford, Y. Levy, and J. N. Onuchic, "Mutations as trapdoors to two competing native conformations of the rop-dimer," *Proc. Natl. Acad. Sci. USA*, vol. 104, pp. 17674–17679, 2007.

[68] Y. Levy, S. S. Cho, T. Shen, J. N. Onuchic, and P. G. Wolynes, "Symmetry and frustration in protein energy landscapes: A near degeneracy resolves the rop dimer-folding mystery," *Proc. Natl. Acad. Sci. USA*, vol. 102, pp. 2373–2378, 2005.

[69] A. Schug and J. N. Onuchic, "From protein folding to protein function and biomolecular binding by energy landscape theory," *Curr. Opin. Pharmacol.*, vol. 10, pp. 709–714, 2010.

[70] M. A. Miller and D. J. Wales, "Energy landscape of a model protein," *J. Chem. Phys.*, vol. 111, pp. 6610–6616, 1999.

[71] G. A. Cox, T. V. Mortimer-Jones, R. P. Taylor, and R. L. Johnston, "Development and optimisation of a novel genetic algorithm for studying model protein folding," *Theor. Chem. Acc.*, vol. 112, pp. 163–178, 2004.

[72] G. A. Cox and R. L. Johnston, "Analyzing energy landscapes for folding model proteins," *J. Chem. Phys.*, vol. 124, p. 204714, 2006.

[73] A. J. Bennett, R. L. Johnston, E. Turpin, and J. Q. He, "Analysis of an immune algorithm for protein structure prediction," *Inform.*, vol. 32, pp. 245–251, 2008.

[74] M. C. Prentiss, D. J. Wales, and P. G. Wolynes, "Protein structure prediction using basin-hopping," *J. Chem. Phys.*, vol. 128, p. 225106, 2008.

[75] M. C. Prentiss, D. J. Wales, and P. G. Wolynes, "The energy landscape, folding pathways and the kinetics of a knotted protein," *PLoS Comput. Biol.*, vol. 6, p. e1000835, 2010.

[76] K. Klenin, B. Strodel, D. J. Wales, and W. Wenzel, "Modelling proteins: Conformational sampling and reconstruction of folding kinetics," *Biochim. Biophys. Acta.*, vol. 1814, pp. 977–1000, 2011.

[77] M. T. Oakley, D. J. Wales, and R. L. Johnston, "Energy landscape and global optimisation for a frustrated model protein," *J. Phys. Chem. B*, vol. 115, pp. 11525–11529, 2011.

[78] M. T. Oakley, D. J. Wales, and R. L. Johnston, "The effect of non-native interactions on the energy landscapes of frustrated model proteins.," *J. Atom. Molec. Opt. Phys.*, vol. 2012, p. 192613, 2012.

[79] D. J. Wales and T. Head-Gordon, "Evolution of the potential energy landscape with static pulling force for two model proteins," *J. Phys. Chem. B*, vol. 136, pp. 8394–8411, 2012.

[80] M. T. Oakley, E. G. Richardson, H. Carr, and R. L. Johnston, "Protein structure optimisation with a "lamarckian" ant colony algorithm," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 10, pp. 1548–1552, 2013.

[81] D. J. Wales and J. P. K. 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, pp. 5111–5116, 1997.

[82] D. J. Wales and H. A. Scheraga, "Global optimization of clusters, crystals, and biomolecules," *Science*, vol. 285, pp. 1368–1372, 1999.

[83] M. Locatelli and F. Schoen, "Efficient algorithms for large scale global optimization: Lennard-jones clusters," *Comput. Optim. Appl.*, vol. 26, pp. 173–190, 2003.

[84] X. Shao, L. Cheng, and W. Cai, "A dynamic lattice searching method for fast optimization of lennard-jones clusters," *J. Comput. Chem.*, vol. 25, p. 1693, 2004.

[85] H. Takeuchi, "Clever and efficient method for searching optimal geometries of lennard-jones clusters," *J. Chem. Inf. Model.*, vol. 46, pp. 2066–2070, 2006.

[86] S. Meyer-Nieberg and H. G. Beyer, "Self-adaptation in evolutionary algorithms," in *Parameter Setting in Evolutionary Algorithms* (F. G. . Lobo, C. F. Lima, and Z. Michalewicz, eds.), pp. 47–75, Berlin Heidelberg: Springer, 2007.

[87] H. G. Beyer and H. P. Schwefel, "Evolution strategies–a comprehensive introduction.," *Nat. Comput.*, vol. 1, pp. 3–52, 2002.

[88] E. K. Burke, M. Hyde, G. Kendall, G. Ochoa, E. Özcan, and J. R. Woodward, "A classification of hyper-heuristic approaches.," in *Handbook of Metaheuristics* (M. Gendreau and J. Y. Potvin, eds.), (US), pp. 449–468, Springer, 2010.

[89] G. Pappa, G. Ochoa, M. Hyde, A. Freitas, J. Woodward, and J. Swan, "Contrasting meta-learning and hyper-heuristic research: the role of evolutionary algorithms," *Genet. Program. Evol. M.*, vol. 15, pp. 3–35, 2014.

[90] S. Heiles and R. L. Johnston, "Global optimization of clusters using electronic structure methods," *Int. J. Quantum Chem.*, vol. 113, pp. 2091–2109, 2013.

[91] F. Cleri and V. Rosato, "Tight-binding potentials for transition metals and alloys," *Phys. Rev. B*, vol. 48, pp. 22–33, 1993.

[92] D. M. Deaven and K. M. Ho, "Molecular geometry optimization with a genetic algorithm," *Phys. Rev. Lett.*, vol. 75, pp. 288–291, 1995.

[93] Y. Zeiri, "Prediction of the lowest energy structure of clusters using a genetic algorithm," *Phys. Rev. E*, vol. 51, pp. R2769–R2772, 1995.

[94] B. Hartke, "Application of evolutionary algorithms to global cluster geometry optimization," in *"Structure and Bonding" series* (R. L. Johnston, ed.), vol. 110, pp. 33–53, Heidelberg: Springer, 2004.

[95] C. Roberts, R. L. Johnston, and N. T. Wilson, "A genetic algorithm for the structural optimization of morse clusters," *Theor. Chem. Acc.*, vol. 104, pp. 123–130, 2000.

[96] F. B. Pereira and J. M. C. Marques, "A self-adaptive evolutionary algorithm for cluster geometry optimization," in *Proceedings of the Eighth International Conference on*

Hybrid Intelligent Systems (HIS 2008), (Los Alamitos, CA, USA), pp. 678–683, IEEE Computer Society, 2008.

[97] F. B. Pereira and J. M. C. Marques, "Analysis of crossover operators for cluster geometry optimization," in *Computational Intelligence for Engineering Systems: Emergent Applications, Intelligent Systems, Control and Automation, Science and Engineering Series* (A. Madureira, J. Ferreira, and Z. Vale, eds.), (Verlag), pp. 77–89, Springer, 2010.

[98] Z. Chen, J. Xiangwei, J. Li, and L. Shushen, "A sphere-cut-splice crossover for the evolution of cluster structures," *J. Chem. Phys.*, vol. 138, p. 214303, 2013.

[99] L. Cheng, W. Cai, and X. A. Shao, "Connectivity table for cluster similarity checking in the evolutionary optimization method," *Chem. Phys. Lett.*, vol. 389, pp. 309–314, 2004.

[100] A. Cassioli, M. Locatelli, and F. Schoen, "Dissimilarity measures for population-based global optimization algorithms," *Comput. Optim. Appl.*, vol. 45, pp. 257–281, 2010.

[101] J. M. C. Marques, J. L. Llanio-Trujillo, P. E. Abreu, and F. B. Pereira, "How different are two chemical structures?," *J. Chem. Inf. Model.*, vol. 50, pp. 2129–2140, 2010.

[102] J. M. C. Marques, F. B. Pereira, and T. Leitão, "On the use of different potential energy functions in rare-gas cluster optimization by genetic algorithms: application to argon," *J. Phys. Chem. A*, vol. 112, p. 6079, 2008.

[103] J. M. C. Marques and F. B. Pereira, "An evolutionary algorithm for global minimum search of binary atomic clusters," *Chem. Phys. Lett.*, vol. 485, pp. 211–216, 2010.

[104] J. M. C. Marques and F. B. Pereira, "A detailed investigation on the global minimum structures of mixed rare-gas clusters: Geometry, energetics and site occupancy," *J. Comput. Chem.*, vol. 34, pp. 505–522, 2013.

[105] S. M. A. Cruz, J. M. C. Marques, and F. B. Pereira, "Improved evolutionary algorithm for the global optiization of clusters with competing attractive and repulsive interactions," *J. Chem. Phys.*, vol. 145, p. 154109, 2016.

[106] F. B. Pereira and J. M. C. Marques, "A hybrid evolutionary algorithm for cluster geometry optimization: The importance of structural elitism," in *Proceedings of the Eighth International Conference on Hybrid Intelligent Systems (HIS 2008)*, (Los Alamitos, CA, USA), pp. 911–914, IEEE Computer Society, 2008.

[107] K. Cahill and V. A. Parsegian, "Rydberg–london potential for diatomic molecules and unbonded atom pairs," *J. Chem. Phys.*, vol. 121, p. 10839, 2004.

[108] M. S. Gordon and M. W. Schmidt, "Advances in electronic structure theory: Gamess a decade later," in *Theory and Applications of Computational Chemistry: the first forty years* (C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria, eds.), pp. 1167–1189, Amsterdam: Elsevier, 2005.

[109] D. E. Woon and T. H. Dunning, Jr., "Gaussian basis sets for use in correlated molecular calculations. iii. the atoms aluminum through argon," *J. Chem. Phys.*, vol. 98, pp. 1358–1371, 1993.

[110] P. L. Barbieri, P. A. Fantin, and F. E. Jorge, "Gaussian basis sets of triple and quadruple zeta valence quality for correlated wave functions," *Mol. Phys.*, vol. 104, pp. 2945–2954, 2006.

[111] S. F. Machado, G. G. Camiletti, A. C. Neto, F. E. Jorge, and R. S. Jorge, "Gaussian basis set of triple zeta valence quality for the atoms from k to kr: Application and dft and ccsd(t) calculations of molecular properties," *Mol. Phys.*, vol. 107, p. 1713, 2009.

[112] S. F. Boys and F. Bernardi, "The calculation of small molecular interactions by the diferences of separate total energies. some procedures with reduced errors," *Mol. Phys.*, vol. 19, pp. 553–566, 1970.

[113] F. V. Prudente, J. M. C. Marques, and F. B. Pereira, "Solvation of li^+ by argon: How important are three-body forces?," *Phys. Chem. Chem. Phys.*, vol. 19, pp. 25707–25716, 2017.

[114] G. E. Froudakis, S. C. Farantos, and M. Velegrakis, "Mass spectra and theoretical modeling of li^+ne_n, li^+ar_n and li^+kr_n clusters," *Chem. Phys.*, vol. 258, pp. 13–20, 2000.

[115] F. Schulz and B. Hartke, "Dodecahedral clathrate structures and magic numbers in alkali cation microhydration clusters," *ChemPhysChem*, vol. 3, pp. 91–106, 2002.

[116] F. Schulz and B. Hartke, "A new proposal for the reason of magic numbers in alkali cation microhydration clusters," *Theor. Chem. Acc.*, vol. 114, pp. 357–379, 2005.

[117] B. S. González, J. Hernández-Rojas, and D. J. Wales, "Global minima and energetics of li^{+}(h$_2$o), and ca^{2+}(h$_2$o)$_{(n)}$ clusters for n ¡= 20," *Chem. Phys. Lett.*, vol. 412, pp. 23–28, 2005.

[118] J. L. Llanio-Trujillo, J. M. C. Marques, and F. B. Pereira, "New insights on lithium-cation microsolvation by solvents forming hydrogen-bonds: Water versus methanol," *Comput. Theor. Chem.*, vol. 1021, pp. 124–134, 2013.

[119] J. Hernández-Rojas and D. J. Wales, "Global minima for rare gas clusters containing one alkali metal ion.," *J. Chem. Phys.*, vol. 119, pp. 7800–7804, 2003.

[120] C. Lüder, D. Prekas, and M. Velagrakis, "Ion-size effects in the growth sequences of metal-ion-doped noble gas clusters.," *Laser Chem.*, vol. 17, p. 109, 1997.

[121] J. Jellinek and E. B. Krissinel, "Ni$_n$al$_m$ alloy clusters: Analysis of structural forms and their energy ordering," *Chem. Phys. Lett.*, vol. 258, pp. 283–292, 1996.

[122] C. M. A. Zanvettor and J. M. C. Marques, "On the lowest-energy structure of binary zn-cd nanoparticles: Size and composition," *Chem. Phys. Lett.*, vol. 608, pp. 373–379, 2014.

[123] S. M. A. Cruz and J. M. C. Marques, "Low-energy structures of clusters modeled with competing repulsive and either long- or moderate short-range attractive interactions," *Comput. Theor. Chem.*, vol. 1107, pp. 82–93, 2017.

[124] C. Kittel, *Introduction to Solid State Physics*. Hoboken, NJ: John Wiley & Sons, Inc, 8 ed., 2005.

[125] M. Dittner and B. Hartke, "Conquering the hard cases of lennard-jones clusters with simple recipes," *Comput. Theor. Chem.*, vol. 1107, pp. 7–13, 2017.

[126] R. Das and D. Wales, "Machine learning prediction for classification of outcomes in local minimisation," *Chem. Phys. Lett.*, vol. 667, pp. 158–164, 2017.

[127] N. Krasnogor and J. Smith, "Emergence of profitable search strategies based on a simple inheritance mechanism," in *Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation* (M. Kaufmann, ed.), pp. 432–439, Inc, 2001.

[128] E. K. Burke, M. Hyde, G. Kendall, and J. Woodward, "A genetic programming hyper-heuristic approach for evolving 2-d strip packing heuristics," *IEEE Trans. Evolut. Comput.*, vol. 14, pp. 942–958, 2010.

[129] A. Klamt and G. Schuurmann, "Cosmo: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," *J. Chem. Soc., Perkin Trans. 2*, pp. 799–805, 1993.

[130] A. Klamt, "Conductor-like screening model for real solvents: A new approach to the quantitative calculations of solvation phenomena," *J. Phys. Chem.*, vol. 99, pp. 2224–2235, 1995.

[131] S. Miertus, E. Scrocco, and J. Tomasi, "Electrostatic interaction of a solute with a continuum. a direct utilization of ab initio molecular potentials for the prevision of solvent effects," *Chem. Phys.*, vol. 55, pp. 117–129, 1981.

[132] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, "Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions," *J. Phys. Chem. B*, vol. 113, pp. 6378–6396, 2009.

[133] F. B. Pereira, J. M. C. Marques, T. Leitão, and J. Tavares, "Analysis of locality in hybrid evolutionary cluster optimization," in *Proceedings of the 2006 IEEE Congress on Evolutionary Computation, Vols. 1-6*, (Vancouver), pp. 2270–2277, CEC, 2006.

[134] F. Silva, B. Galvão, G. Voga, M. Silva, D. Rodrigues, and J. Belchior, "Exploring the mp2 energy surface of nanoalloy clusters with a genetic algorithm: Application to sodium-potassium," *Chemical Physics Letters*, vol. 639, pp. 135–141, 2015.

[135] R. A. Lordeiro, F. F. Guimarães, J. C. Belchior, and R. L. Johnston, "Determination of main structural compositions of nanoalloy clusters of $cu_x au_y$ $(x + y \leq 30)$ using a genetic algorithm approach," *Int. J. Quantum Chem.*, vol. 95, p. 112, 2003.

[136] J. F. Pérez, C. Z. Hadad, and A. Restrepo, "Structural studies of the water tetramer," *Int. J. Quantum Chem.*, vol. 108, pp. 1653–1659, 2008.

[137] J. F. Pérez, E. Florez, C. Z. Hadad, P. Fuentealba, and A. Restrepo, "Stochastic search of the quantum conformational space of small lithium and bimetallic lithium-sodium clusters," *J. Phys. Chem. A*, vol. 112, pp. 5749–5755, 2008.

[138] J. J. P. Stewart, "Optimization of parameters for semiempirical methods i. method," *J. Comput. Chem.*, vol. 10, no. 2, pp. 209–220, 1989.

[139] J. J. P. Stewart, "Optimization of parameters for semiempirical methods ii. applications," *J. Comput. Chem.*, vol. 10, no. 2, pp. 221–264, 1989.

[140] J. L. Llanio-Trujillo, J. M. C. Marques, and F. B. Pereira, "New insights on lithium-cation microsolvation by solvents forming hydrogen-bonds: Water versus methanol," *Comput. Theor. Chem.*, vol. 1021, pp. 124–134, 2013.

[141] M. X. Silva, B. R. L. Galvão, and J. C. Belchior, "Theoretical study of small sodium-potassium alloy clusters through genetic algorithm and quantum chemical calculations," *Phys. Chem. Chem. Phys.*, vol. 16, pp. 8895–8904, 2014.

[142] S. Roszak and J. Leszczynski, "Ab initio studies of the microsolvation of ions," *J. Chem. Phys. A*, vol. 107, pp. 949–955, 2003.

[143] M. K. Beyer, "Hydrated metal ions in the gas phase," *Mass. Spectrom. Rev.*, vol. 26, p. 517, 2007.

[144] Y. Marcus, "Effect of ions on the structure of water: Structure making and breaking," *Chem. Rev.*, vol. 109, pp. 1346–1370, 2009.

[145] C. Hadad, E. Florez, N. Acelas, G. Merino, and A. Restrepo, "Microsolvation of small cations and anions," *Int. J. Quantum Chem.*, vol. 119, p. e25766, 2019.

[146] F. V. Prudente, J. M. C. Marques, and F. B. Pereira, "Correction: Solvation of li^+ by argon: how important are three-body forces?," *Phys. Chem. Chem. Phys.*, vol. 20, pp. 16877–16882, 2018.

[147] M. W. Schmidt, K. K. Baldridge, J. A. Boats, S. T. Elbert, M. S. Gorgon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. Montgomery, Jr., "General atomic and molecular electronic structure system," *J. Comput. Chem.*, vol. 14, pp. 1347–1363, 1993.

[148] B. M. Axilrod and E. Teller, "Interaction of the van der waals type between three atoms," *J. Chem. Phys.*, vol. 11, p. 299, 1943.

[149] Y. Muto, "Force between nonpolar molecules," *Proc. Phys. Math. Soc. Japan*, vol. 17, p. 629, 1943.

[150] B. M. Axilrod, "Triple[U+2010]dipole interaction. i. theory," *J. Chem. Phys.*, vol. 19, p. 719, 1951.

[151] J. Standard and P. C. Certain, "Bounds to two- and three-body long-range interaction coefficients for s-state atoms," *J. Chem. Phys.*, vol. 83, p. 3002, 1985.

[152] R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, "Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions," *J. Chem. Phys.*, vol. 96, pp. 6796–6806, 1992.

[153] J. M. C. Marques, F. V. Prudente, F. B. Pereira, M. M. Almeida, A. M. Maniero, and C. E. Fellows, "A new genetic algorithm to be used in the direct fit of potential energy curves to ab initio and spectroscopic data," *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 41, no. 8, p. 085103 (15pp), 2008.

[154] M. M. Almeida, F. V. Prudente, C. E. Fellows, J. M. C. Marques, and F. B. Pereira, "Direct fit of spectroscopic data of diatomic molecules by using genetic algorithms: Ii. the ground state of rbcs," *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 44, p. 225102 (8pp), 2011.

[155] F. B. Pereira, J. M. C. Marques, T. Leitão, and J. Tavares, "Designing Efficient Evolutionary Algorithms for Cluster Optimization: A Study on Locality," in *Advances in Metaheuristics for Hard Optimization, Springer Natural Computing Series* (P. Siarry and Z. Michalewicz, eds.), (Berlin), pp. 223–250, Springer, 2008.

[156] F. B. Pereira and J. M. C. Marques, "Towards an effective evolutionary approach for binary lennard-jones clusters," in *Proceedings of the 2010 IEEE Congress on Evolutionary Computation*, (Barcelona), pp. 1–7, CEC, 2010.

[157] J. Nocedal, "Updating quasi-newton matrices with limited storage," *Math. Comp.*, vol. 35, pp. 773–782, 1980.

[158] J. Lozeille, E. Winata, P. Soldan, E. P. F. Lee, L. A. Viehland, and T. G. Wright, "Spectroscopy of li^+·rg and li^+- rg transport coefficients (rg = he-rn)," *Phys. Chem. Chem. Phys.*, vol. 4, pp. 3601–3610, 2002.

[159] D. Bellert and W. H. Breckenridge, "Bonding in ground-state and excited-state a^+·rg van der waals ions (a = atom, rg = rare-gas atom): A model-potential analysis," *Chem. Rev.*, vol. 102, pp. 1595–1622, 2002.

[160] A. Andrejeva, A. M. Gardner, J. B. Graneek, R. J. Plowright, W. H. Breckenridge, and T. G. Wright, "Theoretical study of m^+-rg_2 (m^+ = li, na, be, mg; rg = he-rn)," *J. Phys. Chem. A*, vol. 117, pp. 13578–13590, 2013.

[161] G. R. Ahmadi and I. Roeggen, "The interatomic potential for the $x^1\sigma$ state of $arli^+$," *J. Phys. B: At. Mol. Opt. Phys.*, vol. 27, pp. 5603–5620, 1994.

[162] C. W. Bauschlicher Jr., H. Partridge, and S. R. Langhoff, "Comparison of the bonding between ml^+ and ml_2^+ (m=metal, l=noble gas)," *Chem. Phys. Lett.*, vol. 165, pp. 272–276, 1990.

[163] J. J. Szymczak, K. T. Giju, S. Roszak, and J. Leszczynski, "The li^+ cations − the descendant of h^+ or an ancestor of na^+? the properties of li^+ar_n ($n = 1 - 6$) clusters," *J. Phys. Chem. A*, vol. 108, pp. 6570–6574, 2004.

[164] L. A. Viehland, "Interaction potentials for li^+−rare-gas systems," *Chem. Phys.*, vol. 78, pp. 279–294, 1983.

[165] L. A. Viehland, "Interaction potentials for the alkali ion-rare-gas systems," *Chem. Phys.*, vol. 85, pp. 291–305, 1984.

[166] J. L. Llanio-Trujillo, J. M. C. Marques, and F. B. Pereira, "An evolutionary algorithm for the global optimization of molecular clusters: Application to water, benzene, and benzene cation," *J. Phys. Chem. A*, vol. 115, pp. 2130–2138, 2011.

[167] F. Schulz and B. Hartke, "Dodecahedral clathrate structures and magic numbers in alkali cation microhydration clusters," *ChemPhysChem*, vol. 3, pp. 98–106, 2002.

[168] P. Jena and Q. Sun, "Super atomic clusters: Design rules and potential for building blocks of materials," *Chem. Rev.*, vol. 118, pp. 5755–5870, 2018.

[169] Z. Luo, A. W. Castleman, Jr., and S. N. Khanna, "Reactivity of metal clusters," *Chem. Rev.*, vol. 116, pp. 14456–14492, 2016.

[170] S. R. Gadre, S. D. Yeole, and N. Sahu, "Quantum chemical investigations on molecular clusters," *Chem. Rev.*, vol. 114, pp. 12132–12173, 2014.

[171] E. J. Bieske and O. Dopfer, "High-resolution spectroscopy of cluster ions," *Chem. Rev.*, vol. 100, pp. 3963–3998, 2000.

[172] A. W. Castleman, Jr. and R. G. Keesee, "Ionic clusters," *Chem. Rev.*, vol. 86, pp. 589–618, 1986.

[173] O. Dopfer, "Spectroscopic and theoretical studies of ch_3^+-rg_n clusters (rg=he, ne, ar): From weak intermolecular forces to chemical reaction mechanisms," *Int. Rev. Phys. Chem.*, vol. 22, p. 437, 2003.

[174] E. D. Pillai, T. D. Jaeger, and M. A. Duncan, "Ir spectroscopy of $Nb^+(N2)_n$ complexes: Coordination, structures, and spin states," *J. Am. Chem. Soc.*, vol. 129, pp. 2297–2307, 2007.

[175] K. Butterworth, C. T. Chiang, B. Cunningham, M. Freindorf, T. R. Furlani, R. L. DeLeon, and J. F. Garvey, "Reactions within fluorobenzene-ammonia heterocluster ions: Experiment and theory," *J. Phys. Chem. A*, vol. 116, pp. 18771–883, 2012.

[176] J. Cui, X. Zhou, G. Wang, C. Chi, Z. Liu, and M. Zhou, "Infrared photodissociation spectroscopy of mass selected homoleptic copper carbonyl cluster cations in the gas phase," *J. Phys. Chem. A*, vol. 117, pp. 7810–7817, 2013.

[177] A. D. Brathwaite, H. L. Abbott-Lyon, and M. A. Duncan, "Distinctive coordination of co vs n_2 to rhodium cations: An infrared and computational study," *J. Phys. Chem. A*, vol. 120, pp. 7659–7670, 2016.

[178] K. Ding, H. Xu, Y. Yang, T. Li, Z. Chen, Z. Ge, W. Zhu, and W. Zheng, "Mass spectrometry and theoretical investigation of VN_n^+ (n = 8, 9, and 10) clusters," *J. Phys. Chem. A*, vol. 122, pp. 4687–4695, 2018.

[179] E. Fasshauer, M. Förstel, M. Mucke, T. Arion, and U. Hergenhahn, "Theoretical and experimental investigation of electron transfer mediated decay in arkr clusters," *Chem. Phys.*, vol. 482, pp. 226–238, 2017.

[180] M. Förstel, M. Mucke, T. Arion, T. Lischke, M. Pernpointner, U. Hergenhahn, and E. Fasshauer, "Long-range interatomic coulombic decay in arxe clusters: Experiment and theory," *J. Phys. Chem. C*, vol. 120, pp. 22957–22971, 2016.

[181] M. Förstel, M. Mucke, T. Arion, A. M. Bradshaw, and U. Hergenhahn, "Autoionization mediated by electron transfer," *Phys. Rev. Lett.*, vol. 106, p. 033402, 2011.

[182] M. Lundwall, M. Tchaplyguine, G. Öhrwall, R. Feifel, A. Lindblad, A. Lindgren, S. L. Sörensen, S. Svensson, and O. Björneholm, "Radial surface segregation in free heterogeneous argon/krypton clusters," *Chem. Phys. Lett.*, vol. 392, pp. 433–438, 2004.

[183] D. Piscitelli, A. V. Phelps, J. de Urquijo, E. Basurto, and L. C. Pitchford, "Ion mobilities in xe-ne and other rare-gas mixtures," *Phys. Rev. E*, vol. 68, p. 046408, 2003.

[184] M. Lundwall, H. Bergersen, A. Lindblad, G. Öhrwall, M. Tchaplyguine, S. Svensson, and O. Björneholm, "Preferential site occupancy observed in coexpanded argon-krypton clusters," *Phys. Rev. A*, vol. 74, p. 043206, 2006.

[185] D. Cappelletti, G. Liuti, and F. Pirani, "Generalization to ion-neutral systems of the polarizability correlations for interaction potential parameters," *Chem. Phys. Lett.*, vol. 183, pp. 297–303, 1991.

[186] F. Pirani, G. S. Maciel, D. Cappelletti, and V. Aquilanti, "Experimental benchmarks and phenomenology of interatomic forces: open-shell and electronic anisotropy effects," *Int. Rev. Phys. Chem.*, vol. 25, pp. 165–199, 2006.

[187] F. Pirani, S. Brizi, L. F. Roncaratti, P. Casavecchia, D. Cappelletti, and F. Vecchiocattivi, "Beyond the lennard-jones model: a simple and accurate potential function probed by high resolution scattering data," *Phys. Chem. Chem. Phys.*, vol. 10, p. 5489, 2008.

[188] A. Kumar and W. J. Meath, "Pseudo-spectral dipole oscillator strengths and dipole-dipole and triple-dipole dispersion energy coefficients for hf, hcl, hbr, he, ne, ar, kr and xe," *Mol. Phys.*, vol. 54, pp. 823–833, 1985.

[189] M. W. Schmidt, K. K. Baldridge, J. A. Boats, S. T. Elbert, M. S. Gorgon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. Montgomery, Jr., "General atomic and molecular electronic structure system," *J. Comput. Chem.*, vol. 14, p. 1347, 1993.

[190] A. Hermann and P. Schwerdtfeger, "Complete basis set limit second-order moller-plesset calculations for the fcc lattices of neon, argon, krypton, and xenon," *J. Chem. Phys.*, vol. 131, p. 244508, 2009.

[191] R. Ferrando, J. Jellinek, and R. L. Johnston, "Nanoalloys: from theory to applications of alloy clusters and nanoparticles," *Chem. Rev.*, vol. 108, pp. 845–910, 2008.

[192] M. Miyazaki, A. Fujii, T. Ebata, and N. Mikami, "Infrared spectroscopy of size-selected benzene−water cluster cations $[c_6h_6 - (h_2o)_n] + (n = 1 - 23)$: Hydrogen bond net-

work evolution and microscopic hydrophobicity," *The Journal of Physical Chemistry A*, vol. 108, no. 48, pp. 10656–10660, 2004.

[193] T. Premke, E.-M. Wirths, D. Pentlehner, R. Riechers, R. Lehnig, A. Vdovin, and A. Slenczka, "Microsolvation of molecules in superfluid helium nanodroplets revealed by means of electronic spectroscopy," *Frontiers in Chemistry*, vol. 2, p. 51, 2014.

[194] G. Schwaab, F. Sebastiani, and M. Havenith, "Ion hydration and ion pairing as probed by thz spectroscopy," *Angewandte Chemie International Edition*, vol. 58, no. 10, pp. 3000–3013, 2019.

[195] F. Schulz and B. Hartke, "Structural information on alkali cation microhydration clusters from infrared spectra," *Phys. Chem. Chem. Phys.*, vol. 5, p. 5021, 2003.

[196] B. Viswanathan, C. J. Barden, F. Ban, and R. J. Boyd, "Calibration of a computational scheme for solvation: Group i and ii metal ions bound to water, formaldehyde and ammonia," *Molecular Physics*, vol. 103, no. 2-3, pp. 337–344, 2005.

[197] J. Srinivasa Rao, T. Dinadayalane, J. Leszczynski, and G. N. Sastry, "Comprehensive study on the solvation of mono- and divalent metal cations: Li+, na+, k+ be2+, mg2+ and ca2+," *The Journal of Physical Chemistry A*, vol. 112, pp. 12944–12953, 10 2008.

[198] B. Hartke, A. Charvat, M. Reich, and B. Abel, "Experimental and theoretical investigation of microsolvation of na^+-ions in the gas phase by high resolution mass spectrometry and global cluster geometry optimization," *J. Chem. Phys.*, vol. 116, pp. 3588–3600, 2002.

[199] A. N. Alexandrova, "H·(h2o)n clusters: Microsolvation of the hydrogen atom via molecular ab initio gradient embedded genetic algorithm (gega)," *J. Phys. Chem. A*, vol. 114, pp. 12591–12599, 2010.

[200] R. M. Forck, J. M. Dieterich, C. C. Pradzynski, A. L. Huchting, R. A. Mata, and T. Zeuch, "Structural diversity in sodium doped water trimers," *Phys. Chem. Chem. Phys.*, vol. 14, pp. 9054–9057, 2012.

[201] S. Heiles, A. J. Logsdail, R. Schäfer, and R. L. Johnston, "Dopant-induced 2d–3d transition in small au-containing clusters: Dft-global optimisation of 8-atom au–ag nanoalloys," *Nanoscale*, vol. 4, pp. 1109–1115, 2012.

[202] C. Heard and R. L. Johnston, "A density functional global optimisation study of neutral 8-atom cu-ag and cu-au clusters," *Eur. Phys. J. D*, vol. 67, p. 34, 2013.

[203] S. K. Gregurick, M. H. Alexander, and B. Hartke, "Global geometry optimization of $(ar)_n$ and $b(ar)_n$ clusters using a modified genetic algorithm," *J. Chem. Phys.*, vol. 104, pp. 2684–2691, 1996.

[204] J. A. Niesse and H. R. Mayne, "Global optimization of atomic and molecular clusters using the space-fixed modified genetic algorithm method," *J. Comput. Chem.*, vol. 18, pp. 1233–1244, 1997.

[205] R. H. Leary, "Global optimization on funneling landscapes," *J. Global Optim.*, vol. 18, pp. 367–383, 2000.

[206] H. Takeuchi, "Novel method for geometry optimization of molecular clusters: Application to benzene clusters," *J. Chem. Inf. Model.*, vol. 47, pp. 104–109, 2007.

[207] L. Cheng, Y. Feng, J. Yang, and J. Yang, "Funnel hopping: Searching the cluster potential energy surface over the funnels," *J. Chem. Phys.*, vol. 130, p. 214112, 2009.

[208] D. Schebarchov and D. J. Wales, "Communication: A new paradigm for structure prediction in multicomponent systems," *J. Chem. Phys.*, vol. 139, p. 221101, 2013.

[209] M. Al-Ahmari, S. Saidi, J. Dhiflaoui, F. Hassen, and H. Berriche, "Structure and stability of the li^+xe_n and $lixe_n$ clusters," *J. Clust. Sci.*, vol. 26, pp. 913–924, 2015.

[210] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer Series in Statistics, New York, NY, USA: Springer New York Inc., 2001.

[211] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, Oct 2001.

[212] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[213] V. Aquilanti, A. Lombardi, and E. Yurtsever, "Global view of classical clusters: the hyperspherical approach to structure and dynamics," *Phys. Chem. Chem. Phys.*, vol. 4, p. 5040, 2002.

[214] V. Aquilanti, A. Lombardi, and M. B. Sevryuk, "Phase-space invariants for aggregates of particles: hyperangular momenta and partitions of the classical kinetic energy," *J. Chem. Phys.*, vol. 121, p. 5579, 2004.

[215] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, no. 17, pp. 1–5, 2017.

[216] M. Schmidt, J. L. Calvé, and M. Mons, "Structural transitions in benzene–argon clusters: Size and temperature effects," *J. Chem. Phys.*, vol. 98, pp. 6102–6120, 1993.

[217] M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. van Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus, and W. de Jong, "Nwchem: A comprehensive and scalable open-source solution for large scale molecular simulations," *Comput. Phys. Commun.*, vol. 181, p. 1477, 2011.

[218] F. Jensen, "Unifying general and segmented contracted basis sets. segmented polarization consistent basis sets," *J. Chem. Theory Comput.*, vol. 10, pp. 1074–1085, 2014.

[219] L. Cheng, W. Cai, and X. Shao, "A connectivity table for cluster similarity checking in the evolutionary optimization method," *Chemical Physics Letters*, vol. 389, no. 4, pp. 309 – 314, 2004.

[220] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, p. 066138, Jun 2004.

Appendix A

Table of parameters

Table of parameters used in Machine Learning. Energy(EA) and E. DFT are in mEh. And E. First is the energy first obtained by DFT optimization, in Eh.

Min	Oper. Sym.	Energy(EA)	ρ^2	ξ^+	ξ^-	E. DFT	E. First
N=7							
0	6	-0.055239	34.05	5.19E-07	-3.15E-01	0.010100	-19281.683591
1	2	-0.054882	36.21	2.22E-03	-3.54E-01	0.839012	-19281.682855
2	8	-0.054388	42.49	2.19E-06	-4.48E-01	1.734086	-19281.681994
3	4	-0.052999	34.11	1.91E-01	-1.44E-01	0.156677	-19281.679778
4	2	-0.052782	38.32	8.00E-03	-4.14E-01	0.020084	-19281.678141
5	4	-0.052663	38.43	2.55E-03	-4.23E-01	4.812058	-19281.678594
6	2	-0.052365	39.51	4.95E-02	-3.86E-01	6.012252	-19281.677257
8	2	-0.052301	44.04	6.26E-02	-5.66E-01	0.000000	-19281.677529

Min	Oper. Sym.	Energy(EA)	ρ^2	ξ^+	ξ^-	E. DFT	E. First
N=8							
0	2	-0.057515	39.32	9.31E-02	-2.98E-01	0.001289	-22035.163414
1	4	-0.057419	38.50	1.76E-01	-1.18E-01	0.000000	-22035.163674

2	4	-0.057410	38.70	1.02E-01	-2.66E-01	0.003185	-22035.163515
3	12	-0.057371	42.30	5.04E-07	-5.19E-01	1.775595	-22035.163467
4	4	-0.057370	40.43	1.37E-01	-2.22E-01	1.747003	-22035.163501
5	2	-0.057195	40.84	8.90E-02	-3.24E-01	2.577242	-22035.162668
6	2	-0.057080	43.54	6.71E-02	-3.96E-01	2.793591	-22035.162488
7	1	-0.057073	42.94	7.08E-02	-3.83E-01	0.001164	-22035.162482
8	1	-0.057019	41.48	2.10E-01	-8.80E-02	2.502404	-22035.162767
9	2	-0.057017	43.91	4.07E-02	-4.57E-01	2.474407	-22035.162714
10	1	-0.056530	49.22	8.84E-02	-4.07E-01	3.462294	-22035.161825
11	2	-0.055372	39.92	2.56E-01	-1.05E-01	0.024429	-22035.159147
12	2	-0.055339	40.49	8.31E-02	-3.06E-01	1.942022	-22035.159160

Min	Oper. Sym.	Energy(EA)	ρ^2	ξ^+	ξ^-	E. DFT	E. First
				N=9			
0	6	-0.059983	43.61	9.19E-07	-3.65E-01	2.294202	-24788.643088
1	1	-0.059890	45.65	1.76E-02	-4.04E-01	0.603371	-24788.642920
2	1	-0.059813	43.75	2.09E-01	-1.42E-01	0.000131	-24788.643188
3	2	-0.059793	43.51	1.73E-01	-1.50E-01	0.000000	-24788.643195
4	2	-0.059792	44.41	1.16E-01	-2.93E-01	2.010347	-24788.643207
5	2	-0.059787	45.25	2.69E-02	-3.90E-01	0.603813	-24788.643003
6	4	-0.059730	43.68	2.10E-01	-1.70E-01	2.242815	-24788.643066
7	1	-0.059691	43.39	1.92E-01	-1.26E-01	0.211421	-24788.643276
8	1	-0.059684	43.45	1.81E-01	-9.94E-02	1.549979	-24788.643520
9	1	-0.059638	44.48	1.56E-01	-1.59E-01	0.211224	-24788.643363
10	2	-0.059633	47.92	8.46E-02	-4.55E-01	1.705346	-24788.643298
11	2	-0.059633	46.94	1.16E-01	-3.85E-01	1.705043	-24788.643248
12	2	-0.059587	47.21	1.16E-02	-4.21E-01	3.264154	-24788.642184
13	2	-0.059577	43.14	1.52E-01	-7.69E-02	1.403675	-24788.643708
14	2	-0.059532	46.13	1.48E-01	-3.11E-01	1.714972	-24788.643552

15	6	-0.059482	46.15	2.33E-01	-1.80E-06	1.962000	-24788.643411
16	2	-0.059357	47.71	2.41E-01	-1.39E-01	1.217279	-24788.642273
17	2	-0.059222	47.66	2.36E-01	-1.47E-01	2.937057	-24788.642402
18	2	-0.059205	52.37	6.91E-02	-5.15E-01	3.132715	-24788.642255
19	1	-0.059203	50.23	5.71E-02	-3.97E-01	3.060980	-24788.642342
20	1	-0.058073	43.73	1.59E-01	-2.23E-01	1.611102	-24788.639346
21	1	-0.058012	45.67	1.41E-01	-2.86E-01	1.579006	-24788.638208
22	2	-0.057958	44.02	1.58E-01	-1.87E-01	0.010810	-24788.638399
23	1	-0.057917	43.16	2.20E-01	-1.24E-01	1.669842	-24788.640195
24	1	-0.057766	45.35	1.61E-01	-2.40E-01	0.603498	-24788.638708
25	1	-0.057691	45.03	2.18E-01	-7.43E-02	1.406398	-24788.638853
27	1	-0.057646	45.83	2.18E-01	-6.63E-02	2.526725	-24788.638769
28	2	-0.057603	45.20	1.46E-01	-1.08E-01	2.353670	-24788.639035

Min	Oper. Sym.	Energy(EA)	ρ^2	ξ^+	ξ^-	E. DFT	E. First
				N=10			
0	2	-0.062692	47.65	1.12E-01	-2.83E-01	0.788837	-27542.123417
1	4	-0.062682	47.73	6.83E-02	-3.43E-01	0.825916	-27542.123413
2	2	-0.062594	46.36	1.53E-01	-1.58E-01	0.593658	-27542.123611
3	2	-0.062413	47.07	2.59E-01	-8.12E-03	0.454166	-27542.123697
4	2	-0.062383	49.17	1.05E-01	-2.81E-01	1.713518	-27542.122618
5	2	-0.062279	50.44	1.02E-01	-2.95E-01	1.739629	-27542.122478
6	1	-0.062187	49.34	1.77E-01	-1.76E-01	1.537975	-27542.122677
7	1	-0.062168	49.36	1.60E-01	-1.82E-01	1.168878	-27542.122785
8	2	-0.062094	49.60	1.05E-01	-2.89E-01	1.302600	-27542.123056
9	6	-0.062094	53.21	5.83E-07	-5.32E-01	1.423041	-27542.122942
10	1	-0.061994	54.60	2.46E-03	-5.38E-01	2.062690	-27542.122021
11	4	-0.061885	54.78	1.07E-01	-4.92E-01	0.975133	-27542.123057
12	1	-0.061784	50.76	1.13E-01	-3.23E-01	0.259062	-27542.123447

13	2	-0.061698	49.96	7.91E-03	-3.10E-01	1.954582	-27542.122305
14	1	-0.061650	52.70	2.50E-01	-1.52E-01	2.172992	-27542.122449
15	2	-0.061485	53.06	2.96E-01	-1.03E-01	1.713553	-27542.122255
16	2	-0.061374	54.27	1.87E-01	-3.24E-01	1.703321	-27542.122517
17	2	-0.060699	47.75	1.34E-01	-1.80E-01	0.000000	-27542.119017
18	2	-0.060548	46.88	2.48E-01	-3.27E-03	4.438359	-27542.119422
19	1	-0.060451	49.00	2.02E-01	-1.46E-01	1.065697	-27542.118824
20	1	-0.060220	50.04	1.47E-01	-2.90E-01	1.065977	-27542.118550
21	1	-0.060033	51.90	1.29E-01	-2.66E-01	1.280031	-27542.117118

Min	Oper. Sym.	Energy(EA)	ρ^2	ξ^+	ξ^-	E. DFT	E. First
				N=11			
0	2	-0.065188	49.36	1.56E-01	-5.04E-02	0.852905	-30295.604167
1	1	-0.065173	50.21	1.24E-01	-1.99E-01	1.856407	-30295.603310
2	1	-0.065097	50.94	9.57E-02	-2.30E-01	1.809134	-30295.603229
3	1	-0.065080	52.43	1.59E-01	-1.69E-01	2.116458	-30295.602923
4	1	-0.065077	51.20	1.20E-01	-2.24E-01	1.912930	-30295.603209
5	1	-0.065074	50.27	1.49E-01	-1.52E-01	1.609036	-30295.603480
6	1	-0.065041	50.16	1.20E-01	-1.83E-01	1.678947	-30295.603341
7	1	-0.065036	50.36	1.83E-01	-1.37E-01	1.460625	-30295.603615
8	2	-0.065012	51.46	1.20E-01	-2.45E-01	2.232196	-30295.602946
9	1	-0.065008	51.36	1.24E-01	-2.33E-01	2.234682	-30295.602957
11	1	-0.064816	55.38	9.55E-02	-4.00E-01	2.524715	-30295.602537
13	2	-0.064688	53.05	1.91E-01	-1.92E-01	1.615972	-30295.603598
14	2	-0.064550	53.97	2.02E-01	-6.90E-02	2.679298	-30295.602383
15	2	-0.064500	54.48	2.06E-01	-2.04E-01	1.571840	-30295.603691
16	2	-0.064315	52.41	1.20E-01	-3.63E-02	1.986634	-30295.602994
18	1	-0.064176	52.36	2.39E-02	-1.60E-01	1.906555	-30295.603133
20	1	-0.063873	60.42	8.96E-02	-3.21E-01	3.367624	-30295.601795

21	1	-0.063606	48.92	1.74E-01	-7.34E-02	0.955061	-30295.600443
23	1	-0.063264	51.54	1.69E-01	-8.88E-02	0.000000	-30295.599162
24	2	-0.063195	52.37	2.07E-01	-1.04E-01	1.080317	-30295.599025
25	2	-0.063049	50.56	1.69E-01	-1.73E-02	0.118584	-30295.598557
26	1	-0.062896	53.45	1.60E-01	-2.66E-01	0.146338	-30295.597675
27	2	-0.062554	55.92	4.17E-02	-9.13E-02	8.204090	-30295.596699

Min	Oper. Sym.	Energy(EA)	ρ^2	ξ^+	ξ^-	E. DFT	E. First
				N=12			
0	2	-0.068011	53.02	1.17E-02	-2.49E-01	1.746926	-33049.083773
1	1	-0.067897	53.43	1.65E-01	-1.46E-01	1.461569	-33049.083984
2	2	-0.067872	54.14	1.23E-01	-1.06E-01	0.000000	-33049.083520
3	2	-0.067853	52.29	1.77E-02	-2.23E-01	0.592425	-33049.083832
4	1	-0.067844	53.40	2.31E-01	-2.88E-02	1.573507	-33049.083930
5	2	-0.067813	53.50	1.90E-01	-1.09E-01	1.621839	-33049.083978
6	1	-0.067763	52.38	5.61E-02	-1.30E-01	1.432103	-33049.084058
7	2	-0.067763	52.40	6.12E-02	-1.24E-01	1.386949	-33049.084212
8	6	-0.067762	52.40	1.24E-05	-4.44E-02	0.705481	-33049.084825
9	1	-0.067752	52.89	1.16E-01	-1.27E-01	1.369925	-33049.084035
10	1	-0.067617	55.47	2.56E-01	-1.92E-02	0.764734	-33049.083077
11	1	-0.067493	56.00	1.02E-01	-3.05E-01	2.891435	-33049.082785
12	2	-0.067287	56.91	1.24E-01	-2.44E-01	2.206907	-33049.083364
13	1	-0.067199	59.02	9.07E-02	-3.57E-01	3.153291	-33049.082338
14	1	-0.067044	60.41	1.30E-01	-3.40E-01	1.760021	-33049.082448
15	1	-0.066977	55.49	6.34E-02	-1.43E-01	2.487384	-33049.082885
16	2	-0.066876	65.25	1.02E-01	-4.80E-01	3.321041	-33049.082371
18	2	-0.066663	62.82	6.28E-02	-4.23E-01	3.091813	-33049.082412
19	1	-0.066533	50.54	1.00E-01	-4.67E-02	3.107956	-33049.082412
20	2	-0.066138	54.48	4.59E-02	-6.37E-02	0.472771	-33049.078556

22	4	-0.065977	53.53	1.05E-02	-2.49E-02	1.288849	-33049.078766
23	1	-0.065416	54.83	9.13E-02	-1.94E-01	1.637127	-33049.080836

Min	Oper. Sym.	Energy(EA)	ρ^2	ξ^+	ξ^-	E. DFT	E. First
				N=13			
0	2	-0.071183	54.10	8.66E-02	-1.37E-01	0.000000	-35802.565254
1	6	-0.070765	56.84	1.09E-05	-1.19E-01	1.446655	-35802.563707
2	2	-0.070639	57.27	2.47E-01	-1.29E-02	1.112691	-35802.564121
3	1	-0.070632	54.94	1.52E-01	-7.78E-02	0.157119	-35802.564755
4	2	-0.070607	54.86	1.52E-01	-8.08E-02	0.101550	-35802.564751
5	2	-0.070588	55.56	5.85E-02	-9.79E-02	0.826453	-35802.564479
6	2	-0.070566	58.01	1.56E-01	-2.30E-01	1.444054	-35802.563917
7	1	-0.070545	56.29	1.82E-01	-3.54E-02	1.363202	-35802.564042
8	2	-0.070526	54.02	2.68E-02	-7.12E-02	0.136745	-35802.565070
9	4	-0.070471	53.80	7.84E-02	-6.29E-02	0.495371	-35802.564676
11	1	-0.070279	58.92	2.47E-01	-3.93E-02	0.335665	-35802.563564
12	1	-0.070148	58.72	1.10E-01	-1.27E-01	1.773203	-35802.563436
14	1	-0.069981	59.69	1.33E-01	-2.60E-01	1.143799	-35802.563790
15	1	-0.069879	60.84	1.49E-01	-2.64E-01	2.133464	-35802.563044
16	1	-0.069716	58.82	8.07E-02	-1.95E-01	0.964231	-35802.564174
18	2	-0.069584	60.56	1.61E-02	-1.62E-01	2.541237	-35802.562727
19	1	-0.069418	66.20	8.38E-02	-4.17E-01	2.223576	-35802.562932
21	4	-0.069297	64.35	1.37E-01	-3.32E-01	2.006718	-35802.561746
22	1	-0.068944	62.54	9.44E-02	-2.17E-01	3.067354	-35802.562303
23	1	-0.068856	60.15	4.99E-02	-1.59E-01	0.738364	-35802.563909
24	2	-0.068641	58.27	1.59E-01	-3.13E-02	0.962929	-35802.559207
25	1	-0.068596	66.20	1.90E-01	-2.48E-01	2.474741	-35802.562638

Min	Oper. Sym.	Energy(EA)	ρ^2	ξ^+	ξ^-	E. DFT	E. First
				N=14			
0	1	-0.074087	55.42	7.02E-02	-1.02E-01	0.553471	-38556.046371
1	4	-0.074016	55.60	6.91E-02	-6.08E-02	0.925911	-38556.046306
2	1	-0.073563	59.37	1.66E-01	-2.78E-02	2.478019	-38556.044830
3	2	-0.073547	56.01	2.50E-03	-1.16E-01	1.232636	-38556.045880
4	1	-0.073446	57.22	5.53E-02	-5.33E-02	1.800972	-38556.045398
5	1	-0.073441	58.88	1.05E-01	-1.85E-01	2.383094	-38556.044887
6	8	-0.073440	56.06	3.08E-06	-1.17E-01	1.196044	-38556.045753
7	2	-0.073426	59.13	1.42E-01	-1.32E-01	2.149698	-38556.045649
8	2	-0.073412	59.97	1.34E-01	-1.56E-01	0.000000	-38556.045012
9	1	-0.073377	58.55	1.07E-01	-3.82E-02	2.990700	-38556.044377
10	2	-0.073214	59.25	1.51E-01	-1.88E-02	2.991847	-38556.044355
12	2	-0.073097	63.54	1.27E-01	-2.56E-01	2.500665	-38556.043055
13	2	-0.072978	59.90	1.44E-01	-1.50E-01	2.722543	-38556.044651
14	1	-0.072883	65.93	8.10E-02	-3.79E-01	2.824549	-38556.042582
15	1	-0.072792	65.22	3.00E-02	-3.95E-01	4.592459	-38556.042777
16	1	-0.072697	61.72	1.21E-01	-2.10E-01	2.028528	-38556.044732
17	1	-0.072425	66.19	2.02E-01	-2.27E-01	3.343386	-38556.043517
19	1	-0.072283	67.47	1.88E-01	-2.67E-01	2.559150	-38556.042812
20	1	-0.072191	70.40	9.53E-02	-3.56E-01	5.037569	-38556.042440
21	1	-0.072004	70.04	6.86E-02	-4.54E-01	3.953463	-38556.041401
22	1	-0.071821	67.15	3.20E-02	-3.67E-01	2.195139	-38556.042678
23	1	-0.071752	65.87	2.22E-01	-1.31E-01	2.968349	-38556.044146
24	1	-0.071639	60.01	1.16E-01	-1.57E-01	2.371281	-38556.040646