

# Sphere Encapsulated Monte Carlo: Obtaining minimum energy configurations of large aromatic systems

Kimberly Bowal<sup>1</sup>, Peter Grančič<sup>2</sup>, Jacob W. Martin<sup>1,4</sup>, Markus Kraft<sup>1,3,4</sup>

released: 21 May 2019

<sup>1</sup> Department of Chemical Engineering  
and Biotechnology  
University of Cambridge  
West Site, Philippa Fawcett Drive  
Cambridge, CB3 0AS  
United Kingdom

<sup>3</sup> School of Chemical and  
Biomedical Engineering  
Nanyang Technological University  
62 Nanyang Drive  
637459  
Singapore  
E-mail: [mk306@cam.ac.uk](mailto:mk306@cam.ac.uk)

<sup>2</sup> Institute of Soil Research  
University of Natural Resources  
and Life Sciences  
Peter Jordan-Straße 82  
A-1190 Vienna  
Austria

<sup>4</sup> Cambridge Centre for Advanced Research  
and Education in Singapore (CARES)  
CREATE Tower, 1 Create Way  
138602  
Singapore

Preprint No. 229



---

*Keywords:* Monte Carlo, basin-hopping method, polycyclic aromatic hydrocarbon, heterogeneous cluster, nanostructure

**Edited by**

Computational Modelling Group  
Department of Chemical Engineering and Biotechnology  
University of Cambridge  
West Site, Philippa Fawcett Drive  
Cambridge, CB3 0AS  
United Kingdom

**Fax:** + 44 (0)1223 334796

**E-Mail:** [c4e@cam.ac.uk](mailto:c4e@cam.ac.uk)

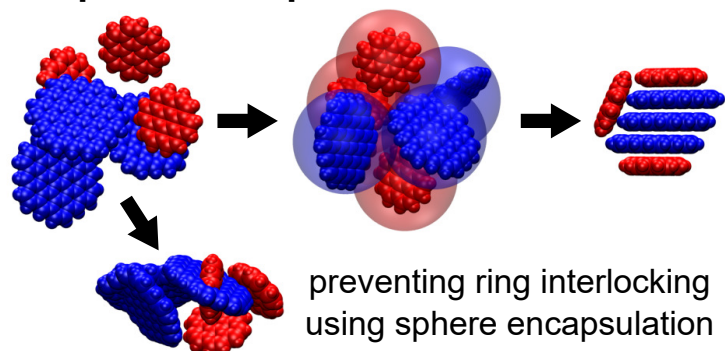
**World Wide Web:** <http://como.cheng.cam.ac.uk/>



## Abstract

We introduce a simple global minimisation approach that is able to find minimum energy configurations of clusters containing aromatic molecules. The translational and rotational perturbations required in Monte Carlo-based methods often lead to unrealistic configurations within which two or more molecular rings intersect, causing many of the computational steps to be rejected and the optimisation process to be inefficient. Here we develop a modification of the basin-hopping global optimisation procedure tailored to tackle problems with intersecting molecular rings. Termed the Sphere Encapsulated Monte Carlo (SEMC) method, this method introduces sphere-based rearrangement and minimisation steps at each iteration and its performance is shown through the exploration of potential energy landscapes of polycyclic aromatic hydrocarbon (PAH) clusters, systems of interest in combustion and astrophysics research. The SEMC method provides clusters that are accurate to 5% mean difference of the minimum energy at a 10-fold speed up compared to previous work using advanced molecular dynamics simulations. Importantly, the SEMC method captures key structural characteristics and molecule size partitioning trends as measured by the molecular radial distances and coordination numbers. The advantages of the SEMC method are further highlighted in its application to previously unstudied heterogeneous PAH clusters.

### Sphere Encapsulated Monte Carlo



#### Highlights:

- Sphere Encapsulated Monte Carlo method is developed to overcome ring interlocking
- Minimum energy configurations are determined at low computational expense
- Applied to clusters beyond the scope of existing methods

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Methods</b>	<b>5</b>
<b>3</b>	<b>Results and discussion</b>	<b>7</b>
3.1	SEMC method development . . . . .	8
3.1.1	Coronene cluster benchmark . . . . .	8
3.1.2	Temperature parameter selection . . . . .	8
3.1.3	Process development for heterogeneous clusters . . . . .	9
3.2	Heterogeneous PAH clusters . . . . .	10
3.2.1	Detailed analysis of representative heterogeneous cluster . . . . .	11
3.2.2	Large PAH clusters compared to REMD . . . . .	12
3.3	Evaluation of complex PAH clusters . . . . .	15
<b>4</b>	<b>Conclusions</b>	<b>17</b>
<b>A</b>	<b>Supplementary Information</b>	<b>19</b>
A.1	Intermolecular potentials . . . . .	19
	<b>References</b>	<b>20</b>

# 1 Introduction

Stable arrangements of nanoclusters may provide structural and thermodynamic properties and their determination is an area of great interest in the physical sciences. For example, stable arrangements can help to identify native states of a protein [47], the experimental structure of an atomic cluster [1], or the density of a molecular cluster [53]. The geometric ground state of a system represents the global minimum of its corresponding potential energy surface (PES). Global minimum determination is not a trivial task and strongly depends on the properties of the PES function, which are usually unknown. In addition, there is often more than one stable configuration, which results in an energy that is a non-convex function of the nuclear coordinates. Scanning across the entire PES to ensure the global minimum is reached is a challenging task at best, and computationally unfeasible for most systems of interest. For example, a cluster containing 13 atoms has at least 988 minima [37], and a cluster containing 147 atoms possesses something on the order of  $10^{60} - 10^{259}$  minima [57].

Many global optimisation methods have been developed to tackle this problem for a variety of systems. One popular method uses consecutive stochastic displacements and local minimisation steps to efficiently sample a complex PES. This Monte Carlo minimisation method was first used to examine protein folding by Li and Scheraga [33] in 1987 and subsequently generalised and popularised by Wales and Doye [60] as the basin-hopping method. The basin-hopping minimisation method simplifies the search of a complex PES by transforming it to a series of low energy plateaus, called basins [60]. This is done by taking the energy of each sampled configuration as the energy of its nearest local minimum, consequently removing energy barriers associated with transition states while retaining all minima. In order for the searching method to move between basins, all components within the system studied are randomly translated and rotated between each minimisation step. Implementing this large stochastic rearrangement prevents the system from being trapped in a local minima and promotes scanning of the entire PES. The basin-hopping method has been shown to be an effective global optimisation technique that allows for the determination of energy minima through exploration of a small fraction of the existing local minima, and has successfully determined the minimum energy configurations of many atomic clusters [2, 15, 16, 18, 35]. Due to its simple implementation and low computational cost, the basin-hopping method has been applied to many fields [61].

Since its development several decades ago there have been many tools implemented with the basin-hopping algorithm to improve the efficiency of global optimisation for specific systems, including parallel tempering [38], path memory to avoid revisiting sites [20], order parameters [23, 45], trial operators [44], additional jump moves [29], reseeding, symmetrising, and taboo lists [58]. More details can be found in a recent review on the developments in basin-hopping global optimisation by Wales [59]. Despite these advances the basin-hopping method remains primarily limited to atomic systems or those in which molecules can be represented as spherical or lumped superatomic components [17, 26, 32, 49]. This is because the required stochastic rearrangement steps make it poorly suited to evaluate complex molecular systems, such as clusters containing aromatic molecules.

Clusters of polycyclic aromatic hydrocarbons (PAHs) are systems of intense interest within

the fields of combustion [5, 42], aerosols [14], and astrophysics [4]. However they present a challenging system for global optimisation methods due to their strong molecular anisotropy. The large random rotation and translation steps required for stochastic basin-hopping sampling cause frequent intersections and ring interlocking between aromatics, leading to many unrealistic overlapping geometries and thus causing the process to become inefficient [43]. Minimum energy structures of PAH clusters have been explored using basin-hopping methods [43, 53], evolutionary algorithms [3], and other algorithms involving iterative random movements and local minimisations [21, 50], however due to computational inefficiency these studies are limited to small PAH clusters. Recently, a coarse-grained potential was developed that represents each PAH as an ellipsoid [27]. This approach smooths the PES and avoids this sampling problem, although at the expense of interaction accuracy and transferable parameters. This coarse-grained potential was used with the basin-hopping method to evaluate low energy structures of clusters containing up to 20 molecules [27] and provide thermodynamic and aggregation behaviour of clusters containing up to eight PAHs using dynamic modelling methods [25, 28].

We have previously studied PAH clusters containing pyrene, coronene, ovalene, hexabenzocoronene, or circumcoronene using molecular dynamics methods [9–12, 22]. Our most recent work exploring heterogeneous PAH clusters [7] used replica exchange molecular dynamics (REMD), an advanced molecular dynamics implementation that determines minimum energy stable cluster configurations through the exchange of information between parallel systems running across a large range of temperatures. In this way, high temperature replicas overcome energetic barriers and provide information necessary for low energy replicas to settle in minimum energy configurations. We found that REMD was able to provide minimum energy stable configurations for clusters containing up to 100 PAHs, but at a high computational cost due to the large number of parallel simulations required. Interestingly, a partitioning of PAHs was observed within these heterogeneous clusters, with larger PAHs residing near the cluster centre and smaller PAHs remaining around the cluster surface [7]. These REMD simulations required a radial position potential to pull evaporated PAHs back to the cluster. Although this was shown to play a role in the high temperature replicas without significantly impacting the low temperature replicas of interest, we thought it important to verify the molecule-size partitioning using an independent Monte Carlo approach which requires no such position potential. In addition, an improved sampling method would allow for the arrangements of aromatic clusters to be robustly sampled with significantly less computational expense than REMD, allowing for larger and more complex systems to be explored.

This work presents a novel method of implementing the basin-hopping minimisation method for aromatic clusters at a reasonable computational expense, the Sphere Encapsulated Monte Carlo (SEMC) method. Through additional steps incorporating sphere-based rearrangements and minimisations, low energy configurations of homogeneous and heterogeneous PAH clusters are identified and compared with our recent work using advanced and costly molecular dynamics simulations.

## 2 Methods

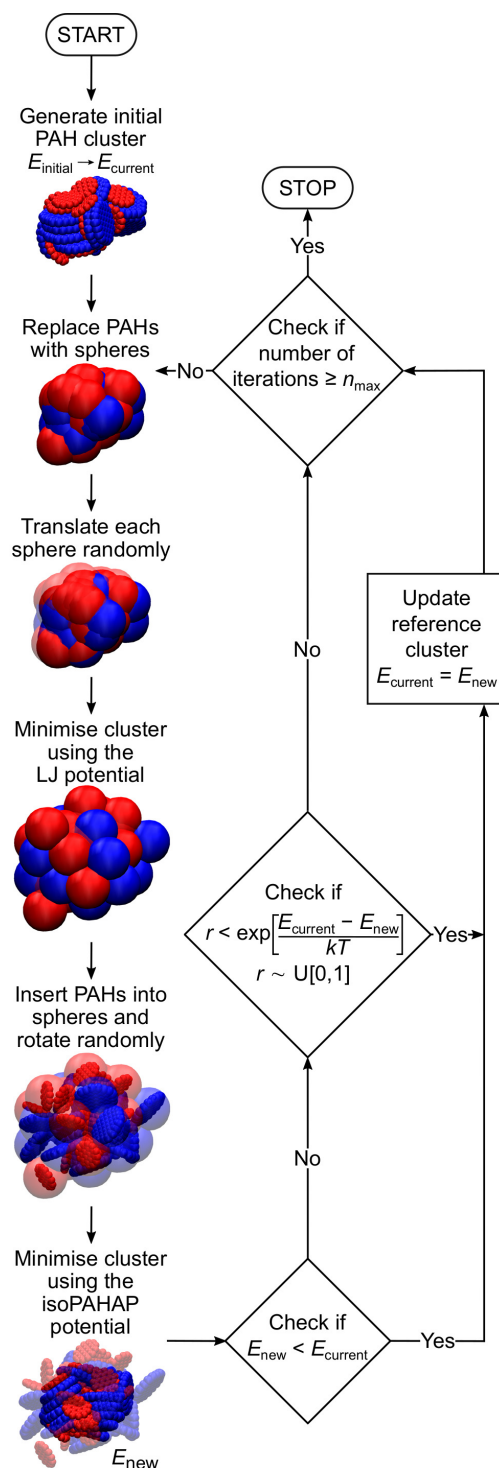
The SEMC method provides a computationally efficient means to explore the complex PES of PAH clusters in order to evaluate their stable molecular arrangements. Following the basin-hopping framework, the SEMC method consists of a series of random perturbations and minimisations in order to locate minimum energy structures. The key novelty of the SEMC method lies in the use of spherical particles to represent PAH molecules in order to improve the efficiency of the rearrangement steps.

The procedure of the SEMC method is as follows, also shown as a flow chart in Figure 1. An initial PAH cluster is generated by a random distribution of PAH molecules within a large spherical volume with a radius of 25 Å, with an energy of  $E_{\text{initial}}$ . Each PAH is then represented by a spherical particle with a diameter corresponding to the diameter of the largest PAH in the cluster (the importance of having spheres of equal size is discussed later). The position of each spherical particle is randomly translated by a three-dimensional vector generated from a Gaussian distribution with a mean of zero and a standard deviation of 2.5 Å. This rearranged system is minimised using a computationally efficient Lennard-Jones (LJ) potential to provide a minimum energy configuration of spherical particles. Minimisation is conducted using the conjugate gradient method for a maximum of  $10^4$  iterations or until the convergence criterion is fulfilled. This step chiefly serves to adjust the randomly positioned spheres so that they form a compact cluster in which the spheres do not overlap.

The PAH molecules are then reinserted into their corresponding spherical particles and are randomly rotated using Euler angles that produce uniformly distributed orientations [31] with a standard deviation of 1 radian. Due to the use of spherical particles previously, the PAHs are allowed to freely rotate without intersecting with neighbouring molecules. The molecular cluster is then minimised using the isoPAHAP potential to generate a minimum energy molecular structure. The isoPAHAP potential is a PAH-specific potential developed from high accuracy symmetry-adapted perturbation theory calculations [54] and has been used successfully in a number of previous studies on PAH systems [9, 21, 55]. This molecular energy minimisation uses the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm [8] for a maximum  $10^5$  iterations or until the maximum force acting on any atom in the system converges to 0.01 eV/Å.

As in the basin-hopping method, the energy of this arrangement is compared with the most recent cluster minimum energy and moves are accepted or rejected based on the Metropolis criterion [24, 34]. Hence if the new cluster energy,  $E_{\text{new}}$ , is lower than the previous cluster minimum energy,  $E_{\text{current}}$ , then the corresponding new configuration is accepted as the updated current reference configuration and a new iteration begins from this new system. Otherwise, the new configuration is accepted with the probability  $\exp[(E_{\text{current}} - E_{\text{new}})/(kT)]$ . Here,  $k$  refers to the Boltzmann constant and  $T$  to the simulation temperature. A temperature of 10 K is used for the smallest clusters made up of fewer than 5 molecules, 500 K for the larger clusters containing up to 32 molecules, and 1000 K for the largest clusters containing 100+ molecules. The selection of these temperatures will be discussed further.

The SEMC method is implemented in purpose-made Python scripts, calling on the freely available software package EON [13, 40] for minimisation steps. EON software performs



**Figure 1:** Flow chart of the SEMC method. The cluster images show a  $CIR_{16}COR_{16}$  cluster, where CIR molecules are shown in blue and COR molecules are shown in red.

atomic scale simulations and allows the user to define interatomic descriptions, such as the LJ and isoPAHAP potentials used in this work. All clusters undergo at least three



independent repetitions each with  $n_{\max} = 1000$  iterations.

Heterogeneous PAH clusters containing different combinations of pyrene (PYR), coronene (COR), ovalene (OVA), and circumcoronene (CIR) are simulated. These molecules are selected because they span the range of PAH sizes relevant to soot particles [5] and interstellar media [52]. The notation  $AAA_xBBB_y$  will be used to describe the clusters studied, where AAA and BBB refer to the molecule types and x and y give the number of molecules.

Average radial distances and coordination numbers are calculated to provide insight into molecular arrangements and cluster structure. The average radial distance of molecule type  $k$ ,  $r_k$ , is calculated using the following equation:

$$r_k = \frac{1}{N_k} \sum_{m \in I_k} R_m \quad (1)$$

where  $N_k$  is the number of molecules within  $k \in \{\text{small, large}\}$ , referring to the small and large molecule types,  $I_k$  is the set of molecule indices within  $k$ , and  $R_m$  is the radial distance between the cluster geometric centre and the geometric centre of molecule  $m$ .

The coordination number of molecule type  $k$ ,  $CN_k$ , provides the average number of molecules within a distance  $R^{(\text{cutoff})}$ . This is calculated as the average of the cumulative radial distribution function integrated over the centre of the cluster (set as the origin at 0) to a maximum cutoff radial distance,  $R^{(\text{cutoff})}$ :

$$CN_k(R^{(\text{cutoff})}) = \frac{1}{N_k} \sum_{m \in I_k} \int_0^{R^{(\text{cutoff})}} \sum_{j=1}^{N_{\text{total}}} \delta(D_{mj} - R) dR \quad (2)$$

$k \in \{\text{small, large, total}\}$  refers to the small molecules, large molecules, and all molecules within the cluster, respectively.  $D_{mj}$  is the distance between the geometric centres of molecules  $m$  and  $j$ .  $R^{(\text{cutoff})}$  corresponds to the radius of the molecule type of interest, so  $R_{\text{PYR}}^{(\text{cutoff})} = 0.46$  nm,  $R_{\text{COR}}^{(\text{cutoff})} = 0.47$  nm,  $R_{\text{OVA}}^{(\text{cutoff})} = 0.59$  nm, and  $R_{\text{CIR}}^{(\text{cutoff})} = 0.71$  nm.  $CN_{\text{total}}$  values use the radius of the larger molecule in the system considered. These provide  $CN$  values from zero to two, corresponding to an isolated molecule and a molecule sandwiched between two others, respectively.

In this work, homogeneous clusters containing 2 – 32 coronene molecules are simulated to provide an assessment of the SEMC method in a system previously studied using several optimisation methods. Following this, investigations into the low energy cluster structures of heterogeneous clusters containing up to 100 molecules are performed. Relevant results are compared to recent simulations of PAH clusters using REMD [7] and the mean absolute differences between the methods are reported, illustrating the accuracy and efficiency of the SEMC method.

### 3 Results and discussion

Key components of the SEMC method development, including evaluation of fundamental interactions with benchmark tests of COR clusters, selection of the temperature parameter, and important considerations for heterogeneous cluster systems are discussed. This

is followed by qualitative and quantitative results for heterogeneous PAH cluster systems, presented first in detail for a representative cluster and followed by a summary of results from all clusters. A direct comparison is made with previous molecular dynamics calculations. Finally, the results from several other systems are given, illustrating the capability of the SEMC method to predict the properties of larger and more complex clusters than was previously feasible using other computational methods.

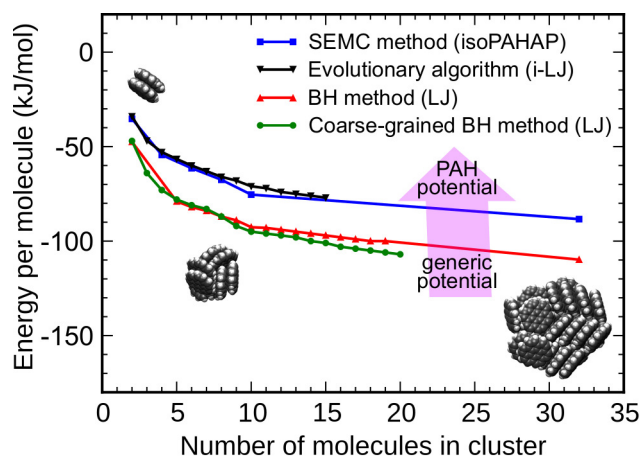
## 3.1 SEMC method development

### 3.1.1 Coronene cluster benchmark

Coronene is a seven-ringed PAH often studied as a model representation of graphene, since it is the smallest PAH that presents a central carbon ring chemically bound only to other carbons [41, 46, 64]. COR clusters also provide benchmark systems for studying  $\pi$ - $\pi$  interactions and are useful for testing the reliability of novel optimisation methods. The SEMC method is applied to homogeneous COR clusters containing 2–32 molecules. This allows the minimum structures produced using this novel method to be compared with those determined by other established global minimisation methods: the basin-hopping method using all-atom [43] or coarse-grained potentials [27] and an evolutionary algorithm method [3], as shown in Figure 2. The lower energies presented by the two basin-hopping methods are not due to the optimisation method but are a consequence of the use of generic LJ parameters, which are known to overestimate the stacking energy of PAHs and influence cluster morphology [39]. This has been corrected by potentials such as the isoPAHAP potential [54] and the improved LJ potential [3] which are parameterised for PAH interactions. The energy shift caused by an accurate PAH potential is highlighted with a purple arrow in Figure 2. With this consideration, the minimum configurations and energies of COR clusters obtained using the SEMC method agree well with literature results.

### 3.1.2 Temperature parameter selection

Temperature is a key free parameter in the SEMC method, controlling the degree to which higher energy configurations are accepted and thus determining the movement between minimum energy basins as the simulation progresses. For relatively small values of the temperature parameter, the system may exhibit the tendency to remain stuck in a current local minimum. On the other hand, too high values of the temperature parameter are known to promote many higher energy configurations to be accepted, thus promoting significant movement across energy basins without searching within minimum energy wells. Selection of the temperature parameter is dependent on system size, and therefore an assessment is conducted to evaluate appropriate values for the clusters of interest. The number of steps required to obtain the minimum energy for different cluster sizes is assessed for at least three independent simulations starting from different initial configurations and the minimum energy results for each case are shown in Table 1. It is seen that for a small homogeneous cluster, a low temperature of 10 K is favourable, while the higher temperatures of 500 K and 1000 K are able to locate the minimum energy values



**Figure 2:** *Energies of homogeneous PAH clusters containing 2–32 coronene molecules obtained with the SEMC method using the isoPAHAP potential and published global optimisation methods. The values obtained by the evolutionary algorithm method using an improved Lennard-Jones potential come from Bartolomei et al. [3], the basin-hopping method using a Lennard-Jones potential come from Rapacioli et al. [43], and the coarse-grained basin-hopping method using a Lennard-Jones potential from Hernández-Rojas et al. [27]. The purple arrow highlights the energy shift caused by using PAH-specific interaction parameters compared to generic LJ parameters which are known to cause enhanced binding. Cluster snapshots shown are from the SEMC method.*

more quickly in the 32 molecule and 100 molecule clusters, respectively. Therefore a temperature of 1000 K is used for the largest clusters, 10 K is used for the smallest clusters, and 500 K is used for the intermediate clusters containing more than 5 but fewer than 100 molecules. These selected temperatures provide satisfactory movement across the PES and convergence toward a minimum energy configuration for the clusters investigated. In most cases, the minimum energy is determined within 1000 iterations at all temperatures considered. While a more rigorous temperature selection process is possible and may result in a more efficient programme for specific systems of interest, this is not considered necessary for this work.

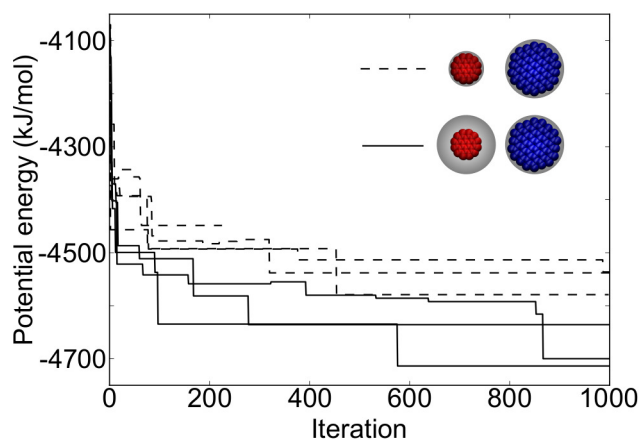
### 3.1.3 Process development for heterogeneous clusters

Initial SEMC simulations used spherical particles with diameters corresponding to the size of the molecule they were replacing. For heterogeneous PAH clusters containing more than one type of PAH, this caused the LJ minimisation step to act on a cluster containing spheres of two different sizes. It was observed that these heterogeneous LJ clusters favoured an arrangement in which the smaller spheres were located in the cluster centre in order to equalise surface and bulk bonds. The favouring of a core-shell configuration was reported in other heterogeneous cluster studies containing spherical components of different sizes [18, 36]. In the SEMC process, a size-dependent partitioning of spheres is undesired since the purpose of the sphere encapsulated steps is to randomly rearrange

**Table 1:** Number of iterations of the SEMC method required to obtain the minimum energy (kJ/mol) using different values of the temperature parameter (K).

Cluster	Temperature	Minimum energy	Iteration
COR <sub>4</sub>	10	-218	398
	500	-202	995
OVA <sub>16</sub> PYR <sub>16</sub>	10	-2860	897
	50	-2860	743
	100	-2860	812
	200	-2860	701
	500	-2860	629
OVA <sub>50</sub> PYR <sub>50</sub>	500	-9910	594
	1000	-9910	6

the PAH cluster. Thus the SEMC method uses equally sized spheres, corresponding to the size of the largest PAH in the cluster, eliminating any partitioning behaviour within the rearrangement steps. This modification promotes lower energy cluster configurations, as shown in Figure 3 for a CIR<sub>16</sub>COR<sub>16</sub> cluster. This indicates that a cluster containing smaller PAHs in the core is less energetically favoured, in contrast to the results seen for binary LJ sphere systems.



**Figure 3:** Accepted energies as a function of iteration for the CIR<sub>16</sub>COR<sub>16</sub> cluster. Three cases initialised with different configurations are given for both simulations using spheres of different sizes in the rearrangement step (shown with dashed lines) and simulations using a uniform sphere size (shown by solid lines).

### 3.2 Heterogeneous PAH clusters

The SEMC method is used to determine the low energy structures of a number of heterogeneous PAH clusters: CIR<sub>8</sub>COR<sub>8</sub>, OVA<sub>8</sub>PYR<sub>8</sub>, CIR<sub>16</sub>COR<sub>16</sub>, OVA<sub>16</sub>PYR<sub>16</sub>, CIR<sub>50</sub>COR<sub>50</sub>,

and OVA<sub>50</sub>PYR<sub>50</sub>.

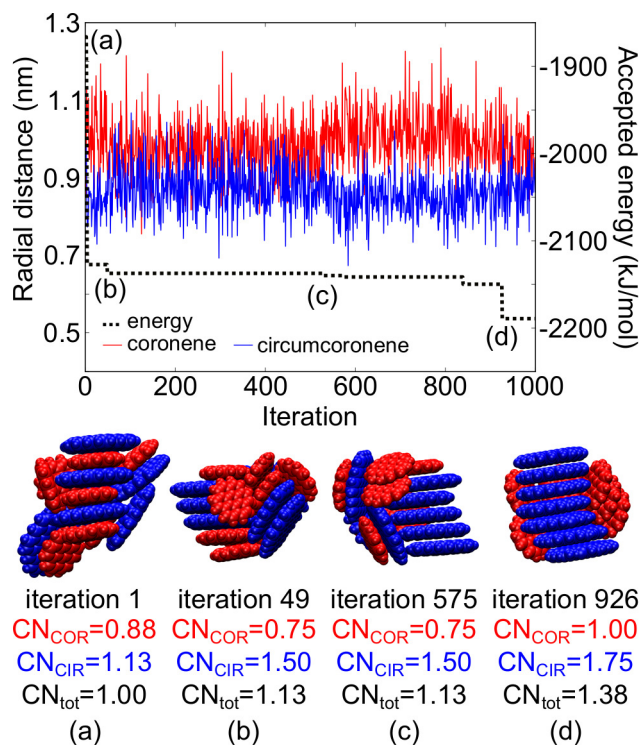
We are interested in the minimum energy arrangements of these molecular clusters, in particular to evaluate if the SEMC method is able to capture the interactions between planar aromatics and the effect of two molecule types within a single cluster. To do this, we have assessed the cluster configurations qualitatively and using several quantitative measures. To aid in understanding this analysis, the results are first reported and discussed in detail for an intermediate-sized representative PAH cluster, CIR<sub>8</sub>COR<sub>8</sub>, followed by results from all clusters studied.

### 3.2.1 Detailed analysis of representative heterogeneous cluster

Figure 4 shows the accepted cluster energies over the course of an SEMC simulation evaluating a CIR<sub>8</sub>COR<sub>8</sub> cluster, as well as images of the cluster configurations at low energy points. In this case, the accepted cluster energies decrease steadily as the simulation progresses and the lowest energy configuration is found near to the end, providing a clear example of progressive cluster arrangements at decreasing energies. At iteration 1, the cluster has a potential energy of -1841 kJ/mol and the cluster shows some stacking of the molecules although it is not compact and the two molecule types are mixed. As the simulation progresses through iterations 49 (-2137 kJ/mol) and 575 (-2142 kJ/mol), these lower energy clusters show increasingly well-stacked structures with molecules interacting more strongly with other molecules of the same type such that the two molecule types separate. The final lowest energy cluster at iteration 926 (-2189 kJ/mol) contains a single central stack of CIR molecules surrounded by smaller stacks of COR molecules. This is the expected low energy cluster arrangement since the strength of intermolecular interactions increases with the number of carbon nearest neighbours [43].

A partitioning of molecule types with the cluster is also seen quantitatively in the movement of the two molecule types, expressed as radial distances in Figure 4. Radial distances, measured as the average distance to the cluster centre for each molecule type during the course of the simulation, provide an indication of the separation of different molecule sizes within the cluster. The translational and rotational rearrangement steps in the optimisation procedure allow the molecule radial distances to vary considerably, but since the current lowest energy state is always used as the starting configuration for the rearrangement and minimisation of the next iteration, the divergence of the molecule radial distances highlights that the low energy structures have a configuration in which the larger molecules are closer to the cluster centre than the smaller molecules. This molecule size partitioning agrees with recent experimental work on the molecule size distribution within carbonaceous soot particles collected from a flame [6].

To provide an indication of the extent of stacking within each cluster, coordination numbers ( $CN$ ) are calculated for the total system and each molecule type. The  $CN$  provides a simple indication of the number of nearest neighbours for each molecule type within the cluster. For the CIR<sub>8</sub>COR<sub>8</sub> cluster, the degree of ordering of the cluster, as measured by  $CN_{\text{total}}$ , increases as the energy decreases. This is dominated by an increase of the close interactions and stacking by the CIR molecules, seen by an increase in  $CN_{\text{CIR}}$ . This is clearly indicated by the formation of a CIR column, moving from stacks mixed with COR at higher energies to CIR dimers and trimers to a central homogeneous CIR stack at the



**Figure 4:** Accepted energies and molecular radial distances as a function of iteration for the  $CIR_8COR_8$  cluster. Cluster snapshots and corresponding molecular coordination numbers are shown for key minimum energy configurations. CIR molecules are shown in blue and COR molecules are shown in red.

lowest energy. The interactions of the smaller COR molecules are secondary, showing an increase in stacking to form dimers and trimers as the energy decreases but only around the dominant CIR stack. In all simulations leading to the lowest energy configurations, low energy clusters are well stacked and a partitioning of the molecules is observed as the iterations progress, with the larger molecules found closer to the cluster centre than the smaller molecules, as seen here. In this way, a decrease in energy is seen to increase the stacking order of the molecules by causing more compact stacks to form and the two molecule types to partition. These trends are also seen within the  $OVA_8PYR_8$  and larger heterogeneous clusters, and the results are recorded in Table 2.

### 3.2.2 Large PAH clusters compared to REMD

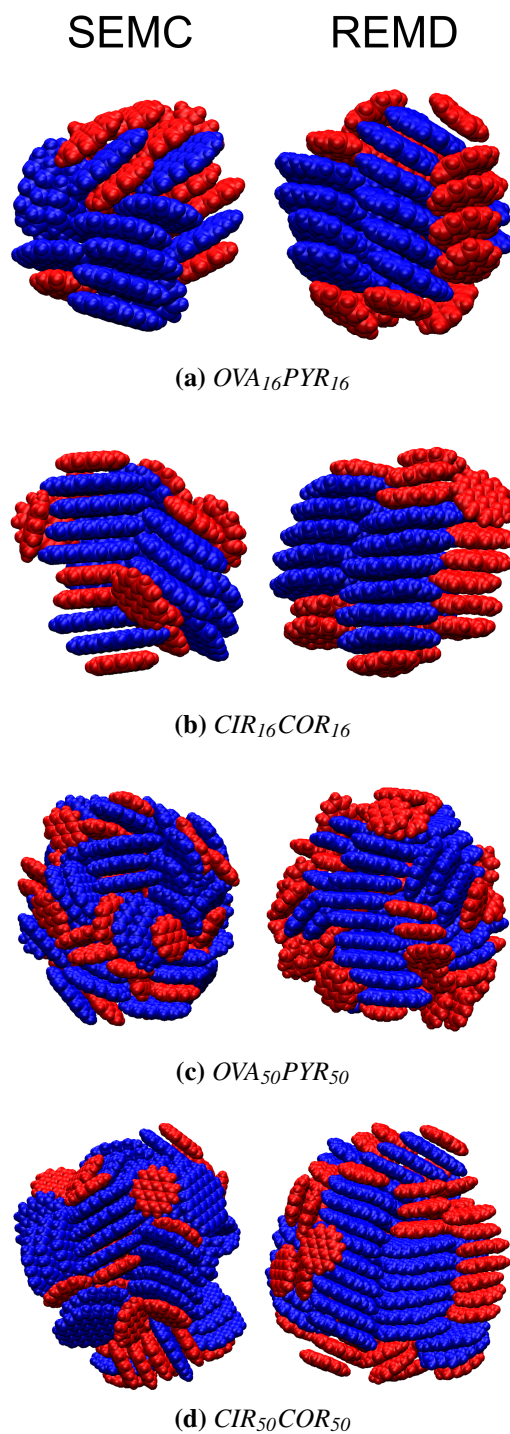
Minimum energy configurations of the  $CIR_{16}COR_{16}$ ,  $OVA_{16}PYR_{16}$ ,  $CIR_{50}COR_{50}$ , and  $OVA_{50}PYR_{50}$  clusters are evaluated since they can be directly compared to previous REMD simulation results [7]. Figure 5 shows the lowest energy structures for these large heterogeneous clusters, alongside the clusters determined by REMD [7]. As with the 16 molecule system described above, the SEMC method finds low energy clusters in which the PAHs form stacks with the larger molecules within the cluster core. The smaller molecules are predominantly situated on the ends and sides of these stacks, although there is some mixing within the larger stacks. The clusters exhibit parallel stacks in a

herringbone-like motif, which has been identified as the most stable configuration for large homogeneous clusters [43] and is similar to the bulk PAH crystal structure [30]. A parallel molecular columnar structure has also been observed experimentally in similarly sized PAH clusters [63].

Compared to the  $\text{CIR}_x\text{COR}_x$  clusters, the  $\text{OVA}_x\text{PYR}_x$  clusters present less structured motifs in which smaller stacks cause a weaker partitioning between the molecule types. This may be due to the smaller size and more elongated shape of the OVA and PYR molecules compared to the CIR and COR molecules, which promotes electrostatically-stabilised T-shaped molecular configurations over the sandwich-like  $\pi$ - $\pi$  interactions that form stacks. This is in agreement with previous work on homogeneous clusters that found large PAHs tend to adopt a herringbone-like structure more readily than small or linear PAHs [50, 53]. The smaller degree of stacking seen in clusters determined using the SEMC method may also be in part due to the use of a spherical volume in the encapsulating step required to prevent molecular overlap in the SEMC method. Sphere-containing clusters do not form stacked structures and thus this rearrangement step may introduce a bias unfavourable for planar molecules due to a misalignment between molecules. It is also seen that large 100 molecule clusters are less well-stacked than smaller clusters using both the SEMC and REMD methods, which suggests that the simulation length scales are not able to determine the global minimum energy configurations. However these results are still useful since the same cluster size trends are seen in experimental images of combustion-produced particles [19], indicating that processes creating these PAH clusters may not produce particles in a global minimum energy arrangement. It is well-known that many systems are strongly influenced by local minima and other features of the PES beyond the global minimum itself [62]. Thus low energy arrangements determined by the SEMC method for clusters containing a large number of PAHs may not correspond to the global minimum configuration but still provide useful information to relevant low energy systems.

A quantitative evaluation of these clusters, including the energies, radial distances, and  $CN$  values of the minimum energy configurations, shows good agreement between the SEMC and REMD methods, as recorded in Table 2. The SEMC method obtains minimum energy clusters with a 5% mean difference in intermolecular energy compared to REMD results for all clusters considered. Note that the intermolecular energies of heterogeneous PAH clusters are determined from an energy minimisation after the simulations in order to allow comparison between the two modelling methods. Similarly, the radial distances determined by the SEMC method are generally accurate to within 10% of the REMD results, ranging from 3–15% for the different cluster compositions and sizes. The molecular  $CN$  values and system  $CN_{\text{total}}$  values determined by the SEMC method present 20% and 10% mean differences, respectively. The largest differences, found in the  $CN_{\text{small}}$  values, can be identified by the tendency for the REMD produced clusters to form outer stacks made up of the smaller molecule type compared to the positioning of smaller molecules more independently around the cluster surface by the SEMC method.  $CN$  values provide an indication of molecular interactions and therefore the relative magnitudes of molecule types are more useful than a direct comparison.

Most importantly, the key structural motifs and molecule size partitioning trends observed from REMD results are evident in the minimum energy configurations determined by the



**Figure 5:** Snapshots of final cluster configurations from SEMC simulations (shown on the left for each cluster pair) compared with REMD simulation results [7] (shown on the right) for heterogeneous PAH clusters of different sizes. Larger molecules (CIR, OVA) are coloured blue and smaller molecules (COR, PYR) are red.



SEMC method. Both the SEMC and REMD methods produce minimum energy clusters in which the molecules are well-stacked, as indicated by the high ( $> 1$ )  $CN_{\text{total}}$  values consistent across all cluster sizes and compositions. In all cases, the larger molecules (CIR and OVA) have higher  $CN$  values than the corresponding smaller molecules (COR and PYR), suggesting that they are less likely to be found on the surface of the cluster. The  $CN_{\text{large}}$  values are above one, corresponding to each molecule experiencing close interaction with two other molecules, indicating a stacked structure. The  $CN_{\text{small}}$  values are smaller, corresponding to surface molecules only located near one other molecule. Likewise, the  $r_{\text{small}}$  values are nearly always larger than the  $r_{\text{large}}$  values, describing clusters in which the smaller molecules are found further away from the cluster centre than large molecules. This agreement between the two methods provides an independent confirmation of the REMD-derived conclusion that low energy heterogeneous PAH clusters favour a partitioning of molecule sizes with the larger PAHs in the cluster centre and thus also confirms that the position potential implemented in the REMD simulations [7] to prevent evaporation in the high energy replicas did not considerably affect the results observed in the low energy replicas.

Computational expense is reported in Table 2 to provide an approximate cost comparison between the SEMC and REMD simulation methods. This highlights the chief strength of the SEMC method since it is able to produce low energy structures using significantly less computational power than REMD, requiring only 8–13% of the computer time for the clusters studied. The SEMC method provides a means of determining the low energy configurations of homogeneous and heterogeneous aromatic clusters using significantly fewer computational resources than previous methods. This allows the simulation of aromatic clusters without the use of extensive supercomputing facilities, as required with REMD.

### 3.3 Evaluation of complex PAH clusters

Two additional systems are considered to further illustrate the potential of the SEMC method: clusters containing nine different molecule types and a cluster containing 150 molecules of three different molecule types.

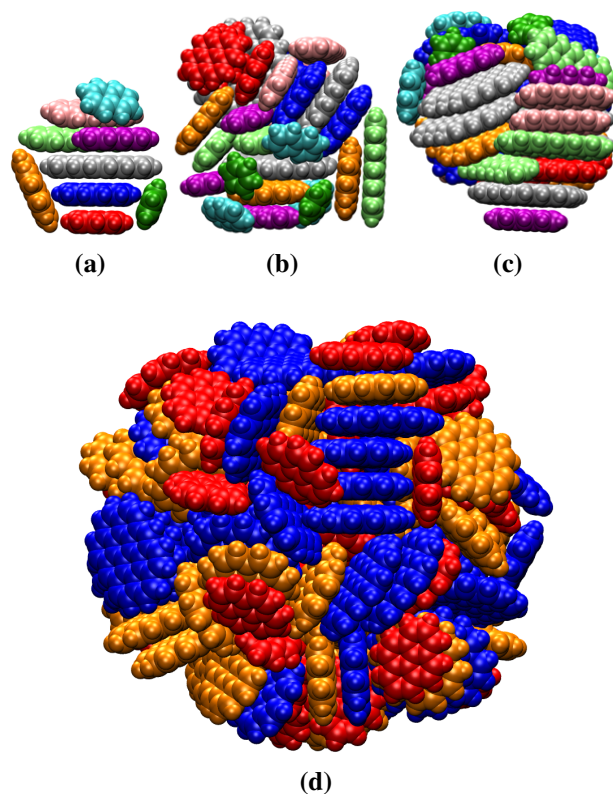
REMD requires significant computational resources, especially when considering heterogeneous clusters containing molecules of dissimilar sizes. This is because the difference in melting point of the constituent molecules demands a larger temperature range and thus an increase in the number of replicas required within the REMD method. SEMC simulations have no such dependency on the molecular melting points, allowing this method to consider a diverse set of heterogeneous clusters at a significantly lower cost than REMD. Thus there is great potential to use the SEMC method for highly inhomogeneous systems, which are believed to be representative of particles found in real environments [48, 51]. As an example, Figure 6(a-c) shows the minimum energy arrangements of clusters containing 9 molecule types at three different cluster sizes. These results show similar behaviour to the binary heterogeneous clusters discussed previously, with stacked columnar structures dominating. The larger molecules also reside in the centre of these stacks and the smaller molecules reside around the outside, engaging in T-shaped interactions.

**Table 2:** Final intermolecular energy ( $E_{\text{int}}$ , kJ/mol), average radial distance ( $r$ , nm), coordination number ( $CN$ ), and computation run time (CPU kilohours) of heterogeneous cluster systems from SEMC and REMD simulations. Values are obtained from post-simulation minimised clusters to allow comparison between the SEMC and REMD methods. The subscripts refer to the large (CIR, OVA) and small (COR, PYR) molecule types within each cluster as well as the total system. REMD simulation results are obtained from Bowal et al. [7].

Cluster property	Simulation method	CIR <sub>8</sub> COR <sub>8</sub>	CIR <sub>16</sub> COR <sub>16</sub>	CIR <sub>50</sub> COR <sub>50</sub>	OVA <sub>8</sub> PYR <sub>8</sub>	OVA <sub>16</sub> PYR <sub>16</sub>	OVA <sub>50</sub> PYR <sub>50</sub>
$E_{\text{int}}$	SEMC	-2220	-4650	-16100	-1320	-2920	-9420
	REMD	–	-4870	-15800	–	-2900	-10900
	% difference	–	5	2	–	1	15
$r_{\text{large}}$	SEMC	0.82	1.05	1.76	0.79	0.97	1.55
	REMD	–	1.02	1.57	–	0.89	1.38
	% difference	–	3	11	–	9	12
$r_{\text{small}}$	SEMC	1.02	1.35	1.86	0.86	1.07	1.49
	REMD	–	1.42	2.11	–	1.23	1.74
	% difference	–	5	13	–	14	15
$CN_{\text{large}}$	SEMC	2.0	1.9	1.7	1.6	1.6	1.4
	REMD	–	2.0	2.0	–	2.0	1.8
	% difference	–	5	16	–	22	25
$CN_{\text{small}}$	SEMC	1.0	0.81	0.68	0.88	0.56	0.52
	REMD	–	1.1	1.1	–	0.56	0.68
	% difference	–	30	47	–	0	27
$CN_{\text{total}}$	SEMC	1.6	1.5	1.5	1.4	1.4	1.2
	REMD	–	1.7	1.8	–	1.5	1.4
	% difference	–	13	18	–	7	15
CPU khours	SEMC	0.12	0.43	4.48	0.04	0.24	2.10
	REMD	–	5.32	56.2	–	1.80	23.0
	% difference	–	170	170	–	153	167

A large cluster containing 150 molecules of OVA, COR, and PYR is evaluated to show the potential of the SEMC method to study systems above the reasonable size limit using the REMD method. The minimum energy structure of this heterogeneous cluster is shown in Figure 6(d). The molecular  $CN$  values,  $CN_{\text{OVA}} = 1.22$ ,  $CN_{\text{COR}} = 0.92$ , and  $CN_{\text{PYR}} = 0.54$ , indicate that the larger molecules more readily form stacks within the centre of the cluster than the smaller molecules, a result consistent with previous cluster structures.

Further work should be done to benchmark and improve the SEMC method. For example, the degree of stacking of the molecules may be tuned using the convergence criteria, defined as the maximum force less than a specified limit, however this may increase the



**Figure 6:** Complex PAH clusters evaluated by the SEMC method. Minimum energy clusters containing 9 aromatic molecules of varying sizes: benzene (dark green), naphthalene (cyan), anthracene (purple), tetracene (pink), pentacene (light green), pyrene (red), coronene (orange), ovalene (blue), circumcoronene (silver), denoted as  $BEN_xNAP_xANT_xTET_xPEN_xPYR_xCOR_xOVA_xCIR_x$ , are shown containing (a) one of each molecule type,  $x = 1$ ; (b) three of each molecule type,  $x = 3$ ; and (c) five of each molecule type,  $x = 5$ . The large  $OVA_{50}COR_{50}PYR_{50}$  cluster is shown in (d).

required simulation time and one can also expect convergence problems due to accumulation of numerical errors. In addition, the molecular isoPAHAP minimisation is the slowest step in the SEMC process and it is observed that a significant amount of computational time is spent at the end for a small decrease in system energy, as an outcome of relatively negligible modification of the molecular configuration. There is therefore potential to increase the overall speed of the SEMC process by implementing a less strict convergence criteria for the molecular minimisation.

## 4 Conclusions

The Sphere Encapsulated Monte Carlo method is developed in order to efficiently determine minimum energy configurations of clusters containing aromatic molecules. The

SEMC method builds on the framework of the basin-hopping method, in which a system is evaluated through a series of random perturbations and minimisations, through the addition of an intermediate sphere encapsulation of the molecules. The SEMC method is able to provide low energy cluster configurations of homogeneous coronene clusters in agreement with existing global minimisation methods.

The arrangements of PAHs in heterogeneous clusters are explored using the SEMC method. Binary PAH clusters containing up to 100 molecules present a parallel stacks motif, similar to the herringbone structure seen in bulk PAHs. The maximisation of carbon-carbon interactions dominates the arrangements formed, leading to smaller PAHs capping the ends of otherwise homogeneous stacks of larger PAHs. This agrees with previous work on small heterogeneous clusters and replica exchange molecular dynamics simulations. Minimum cluster energies are accurate to within 5% of REMD results using only a tenth of the computing resources. The SEMC method is also used to explore several complex PAH clusters beyond the scope of existing computational methods. In addition to its ability to determine good system minima with a fraction of the computational expense used previously, the SEMC method does not require additional constraints or large temperature ranges as required within REMD, and could be useful for generating minima to determine further system information such as transition state paths. The SEMC method provides a novel method of implementing a stochastic method to determine low energy arrangements of aromatic clusters with significantly fewer resources than existing methods, useful for evaluating nanostructures present in PAH environments such as combustion particulates and interstellar dust.

## Acknowledgements

We thank Prof David Wales for his comments at the beginning of this project. This project was supported by the National Research Foundation (NRF), Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. KB is grateful to the Cambridge Trust and the Stanley Studentship at King's College, Cambridge for their financial support. MK gratefully acknowledges the support of the Alexander von Humboldt foundation.

# A Supplementary Information

## A.1 Intermolecular potentials

The form and values of the potential energy function are known to influence the stability and arrangement of PAH clusters [39], so care must be taken when selecting the intermolecular description. In this work, two intermolecular potentials are used to provide the total interaction energy of the system as a sum over all pairwise atomic interactions within the SEMC method.

The isoPAHAP potential consists of a Born–Mayer term to describe short-range interactions, and a damped dispersion term and point charge electrostatic term to describe long-range interactions, as shown in Equation A.1.

$$U_{ab} = G \exp[-\alpha_{ab}(R_{ab} - \rho_{ab})] - \left[ 1 - \exp(-\beta R_{ab}) \sum_{h=0}^n \frac{(\beta R_{ab})^h}{h!} \right] \frac{C_{6,ab}}{R_{ab}^6} + \frac{q_a q_b}{R_{ab}} \quad (\text{A.1})$$

$U_{ab}$  denotes the interaction energy where  $a$  and  $b$  denote atomic sites within a molecule,  $G = 0.001$  hartree,  $\alpha_{ab}$  is the hardness parameter in the Born–Mayer term,  $R_{ab}$  is the atom-atom separation,  $\rho_{ab}$  is the shape parameter,  $\beta$  is the damping coefficient,  $C_{6,ab}$  is the dispersion coefficient, and  $q_a$  and  $q_b$  are the atomic point charges. Parameter values are found in Table S1. Further details about the form and parametrisation of the isoPAHAP potential can be found in Totton et al. [54, 56].

**Table S1:** Parameters of isoPAHAP in a.u.

Atom pair	$\rho$	$\alpha$	$C_6$
C C	6.0434	1.8783	30.282
C H	4.9562	1.7560	12.605
H H	4.1195	1.4043	5.2179

In this work, the Lennard-Jones potential is used for intermediate steps involving spherical particles. The Lennard-Jones potential takes the following form:

$$U_{ab} = 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \quad (\text{A.2})$$

where  $\varepsilon$  is the potential well depth,  $\sigma$  is the equilibrium distance, and  $r$  is the distance between particles. In this work,  $\sigma$  is the distance (in Angstroms) between the molecule centre to the most distant hydrogen centre and the van der Waals radius of hydrogen, and  $\varepsilon$  is an arbitrary value in eV that provides a sufficiently deep potential energy well. This potential does not provide the detailed PAH interactions of the isoPAHAP potential but allows for a computationally efficient determination of the minimum energy configuration of a cluster containing spherical particles, suitable for the rearrangement steps.

## References

- [1] A. Aguado, A. Vega, A. Lebon, and B. von Issendorff. Are zinc clusters really amorphous? A detailed protocol for locating global minimum structures of clusters. *Nanoscale*, 10(40):19162–19181, 2018. doi:10.1039/C8NR05517C.
- [2] F. Baletto and R. Ferrando. Structural properties of nanoclusters: Energetic, thermodynamic, and kinetic effects. *Reviews of Modern Physics*, 77(1):371–423, 2005. doi:10.1103/RevModPhys.77.371.
- [3] M. Bartolomei, F. Pirani, and J. M. C. Marques. Modeling coronene nanostructures: Analytical potential, stable configurations and ab initio energies. *The Journal of Physical Chemistry C*, 121(26):14330–14338, 2017. doi:10.1021/acs.jpcc.7b03691.
- [4] C. Boersma, J. D. Bregman, and L. J. Allamandola. Properties of polycyclic aromatic hydrocarbons in the northwest photon dominated region of NGC 7023. I. PAH size, charge, composition, and structure distribution. *Astrophysical Journal*, 769(2): 117–129, 2013. doi:10.1088/0004-637X/769/2/117.
- [5] M. L. Botero, E. M. Adkins, S. González-Calera, H. Miller, and M. Kraft. PAH structure analysis of soot in a non-premixed flame using high-resolution transmission electron microscopy and optical band gap analysis. *Combustion and Flame*, 164:250–258, 2016. doi:10.1016/j.combustflame.2015.11.022.
- [6] M. L. Botero, Y. Sheng, J. Akroyd, J. Martin, J. A. Dreyer, W. Yang, and M. Kraft. Internal structure of soot particles in a diffusion flame. *Carbon*, 141:635–642, 2019. doi:10.1016/j.carbon.2018.09.063.
- [7] K. Bowal, J. W. Martin, and M. Kraft. Partitioning of polycyclic aromatic hydrocarbons in heterogeneous clusters. *Carbon*, 143:247–256, 2019. doi:10.1016/j.carbon.2018.11.004.
- [8] R. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing*, 16:1190–1208, 1995. doi:10.1137/0916069.
- [9] D. Chen, T. S. Totton, J. Akroyd, S. Mosbach, and M. Kraft. Phase change of polycyclic aromatic hydrocarbon clusters by mass addition. *Carbon*, 77:25–35, 2014. doi:10.1016/j.carbon.2014.04.089.
- [10] D. Chen, T. S. Totton, J. Akroyd, S. Mosbach, and M. Kraft. Size-dependent melting of polycyclic aromatic hydrocarbon nano-clusters: A molecular dynamics study. *Carbon*, 67:79–91, 2014. doi:10.1016/j.carbon.2013.09.058.
- [11] D. Chen, J. Akroyd, S. Mosbach, and M. Kraft. Surface reactivity of polycyclic aromatic hydrocarbon clusters. *Proceedings of the Combustion Institute*, 35(2):1811–1818, 2015. doi:10.1016/j.proci.2014.06.140.

- [12] D. Chen, J. Akroyd, S. Mosbach, D. Opalka, and M. Kraft. Solid-liquid transitions in homogenous ovalene, hexabenzocoronene and circumcoronene clusters: A molecular dynamics study. *Combustion and Flame*, 162(2):486–495, 2015. doi:10.1016/j.combustflame.2014.07.025.
- [13] S. T. Chill, M. Welborn, R. Terrell, L. Zhang, J.-C. Berthet, A. Pedersen, H. Jonsson, and G. Henkelman. EON: Software for long time simulations of atomic scale systems. *Modelling and Simulation in Materials Science and Engineering*, 22(5):055002, 2014. doi:10.1088/0965-0393/22/5/055002.
- [14] N.-D. Dat and M. B. Chang. Review on characteristics of PAHs in atmosphere, anthropogenic sources and control technologies. *Science of the Total Environment*, 609:682–693, 2017. doi:10.1016/j.scitotenv.2017.07.204.
- [15] J. P. Doye. Lead clusters: Different potentials, different structures. *Computational Materials Science*, 35(3):227–231, 2006. doi:10.1016/j.commatsci.2004.07.009.
- [16] J. P. Doye and D. J. Wales. Polytetrahedral clusters. *Physical Review Letters*, 86(25):5719–5722, 2001. doi:10.1103/PhysRevLett.86.5719.
- [17] J. P. Doye, D. J. Wales, W. Branz, and F. Calvo. Modeling the structure of clusters of C<sub>60</sub> molecules. *Physical Review B*, 64(23):235409, 2001. doi:10.1103/PhysRevB.64.235409.
- [18] J. P. K. Doye and L. Meyer. Mapping the magic numbers in binary Lennard-Jones clusters. *Physical Review Letters*, 95(6):063401, 2005. doi:10.1103/PhysRevLett.95.063401.
- [19] V. Fernandez-Alos, J. K. Watson, R. vander Wal, and J. P. Mathews. Soot and char molecular representations generated directly from HRTEM lattice fringe images using Fringe3D. *Combustion and Flame*, 158(9):1807–1813, 2011. doi:10.1016/j.combustflame.2011.01.003.
- [20] S. Goedecker. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *The Journal of Chemical Physics*, 120(21):9911–9917, 2004. doi:10.1063/1.1724816.
- [21] P. Grančič, R. Bylsma, H. Meeke, and H. M. Cuppen. Evaluation of all-atom force fields for anthracene crystal growth. *Crystal Growth & Design*, 15(4):1625–1633, 2015. doi:10.1021/cg5013507.
- [22] P. Grančič, J. W. Martin, D. Chen, S. Mosbach, and M. Kraft. Can nascent soot particles burn from the inside? *Carbon*, 109:608–615, 2016. doi:10.1016/j.carbon.2016.08.025.
- [23] U. H. Hansmann and L. T. Wille. Global optimization by energy landscape paving. *Physical Review Letters*, 88(6):068105, 2002. doi:10.1103/PhysRevLett.88.068105.
- [24] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. doi:10.1093/biomet/57.1.97.

- [25] J. Hernández-Rojas and F. Calvo. Coarse-grained modeling of the nucleation of polycyclic aromatic hydrocarbons into soot precursors. *Physical Chemistry Chemical Physics*, 2019. doi:10.1039/C8CP07724J.
- [26] J. Hernández-Rojas, J. Bretón, J. Gomez Llorente, and D. Wales. Lowest-energy structures of  $(C_{60})_n X$   $X = Li^+, Na^+, K^+, Cl^-$  and  $(C_{60})_n Y Cl$  ( $Y = Li, Na, K$ ) clusters for  $n \leq 13$ . *The Journal of Chemical Physics*, 121(24):12315–12322, 2004. doi:10.1063/1.1819894.
- [27] J. Hernández-Rojas, F. Calvo, and D. J. Wales. Coarse-graining the structure of polycyclic aromatic hydrocarbons clusters. *Physical Chemistry Chemical Physics*, 18(20):13736–13740, 2016. doi:10.1039/c6cp00592f.
- [28] J. Hernández-Rojas, F. Calvo, S. Niblett, and D. Wales. Dynamics and thermodynamics of the coronene octamer described by coarse-grained potentials. *Physical Chemistry Chemical Physics*, 19(3):1884–1895, 2017. doi:10.1039/C6CP07671H.
- [29] M. Iwamatsu and Y. Okabe. Basin hopping with occasional jumping. *Chemical Physics Letters*, 399(4-6):396–400, 2004. doi:10.1016/j.cplett.2004.10.032.
- [30] R. Khanna, V. Sahajwalla, and R. H. Hurt. An atomistic technique for simulating non-covalent interactions in large ensembles of high-molecular-weight polyaromatics. *Carbon*, 43(1):67 – 77, 2005. doi:10.1016/j.carbon.2004.08.023.
- [31] J. J. Kuffner. Effective sampling and distance metrics for 3D rigid body path planning. In *Proceedings of the 2004 IEEE International Conference on Robotics and Automation*, pages 3993–3998, 2004. doi:10.1109/ROBOT.2004.1308895.
- [32] H. Kusumaatmaja, C. S. Whittleston, and D. J. Wales. A local rigid body framework for global optimization of biomolecules. *Journal of Chemical Theory and Computation*, 8(12):5159–5165, 2012. doi:10.1021/ct3004589.
- [33] Z. Li and H. A. Scheraga. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proceedings of the National Academy of Sciences*, 84(19):6611–6615, 1987. doi:10.1073/pnas.84.19.6611.
- [34] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi:10.1063/1.1699114.
- [35] T. F. Middleton, J. Hernández-Rojas, P. N. Mortenson, and D. J. Wales. Crystals of binary Lennard-Jones solids. *Physical Review B*, 64(18):184201, 2001. doi:10.1103/PhysRevB.64.184201.
- [36] M. Mravljak, T. Kister, T. Kraus, and T. Schilling. Structure diagram of binary Lennard-Jones clusters. *The Journal of Chemical Physics*, 145(2):024302, 2016. doi:10.1063/1.4954938.
- [37] J. Northby. Structure and binding of Lennard-Jones clusters:  $13 \leq N \leq 147$ . *The Journal of Chemical Physics*, 87(10):6166–6177, 1987. doi:10.1063/1.453492.



- [38] E. G. Noya and J. P. Doye. Structural transitions in the 309-atom magic number Lennard-Jones cluster. *The Journal of Chemical Physics*, 124(10):104503, 2006. doi:10.1063/1.2173260.
- [39] L. Pascazio, M. Sirignano, and A. D. Anna. Simulating the morphology of clusters of polycyclic aromatic hydrocarbons: The influence of the intermolecular potential. *Combustion and Flame*, 185:53–62, 2017. doi:10.1016/j.combustflame.2017.07.003.
- [40] A. Pedersen and H. Jónsson. Distributed implementation of the adaptive kinetic Monte Carlo method. *Mathematics and Computers in Simulation*, 80(7):1487–1498, 2010. doi:10.1016/j.matcom.2009.02.010.
- [41] R. Podeszwa. Interactions of graphene sheets deduced from properties of polycyclic aromatic hydrocarbons. *The Journal of Chemical Physics*, 132(4):044704, 2010. doi:10.1063/1.3300064.
- [42] A. Raj, M. Sander, V. Janardhanan, and M. Kraft. A study on the coagulation of polycyclic aromatic hydrocarbon clusters to determine their collision efficiency. *Combustion and Flame*, 157(3):523–534, 2010. doi:10.1016/j.combustflame.2009.10.003.
- [43] M. Rapacioli, F. Calvo, F. Spiegelman, C. Joblin, and D. J. Wales. Stacked clusters of polycyclic aromatic hydrocarbon molecules. *The Journal of Physical Chemistry A*, 109(11):2487–2497, 2005. doi:10.1021/jp046745z.
- [44] G. G. Rondina and J. L. Da Silva. Revised basin-hopping Monte Carlo algorithm for structure optimization of clusters and nanoparticles. *Journal of Chemical Information and Modeling*, 53(9):2282–2298, 2013. doi:10.1021/ci400224z.
- [45] G. Rossi and R. Ferrando. Global optimization by excitable walkers. *Chemical Physics Letters*, 423(1-3):17–22, 2006. doi:10.1016/j.cplett.2006.03.003.
- [46] H. Ruuska and T. A. Pakkanen. Ab initio study of interlayer interaction of graphite: Benzene-coronene and coronene dimer two-layer models. *The Journal of Physical Chemistry B*, 105(39):9541–9547, 2001. doi:10.1021/jp011512i.
- [47] O. Schueler-Furman, C. Wang, P. Bradley, K. Misura, and D. Baker. Progress in modeling of protein structures and interactions. *Science*, 310(5748):638–642, 2005. doi:10.1126/science.1112160.
- [48] W. Schutte, A. Tielens, and L. Allamandola. Theoretical modeling of the infrared fluorescence from interstellar polycyclic aromatic hydrocarbons. *The Astrophysical Journal*, 415:397–414, 1993. doi:10.1086/173173.
- [49] B. Strodel, J. W. Lee, C. S. Whittleston, and D. J. Wales. Transmembrane structures for alzheimer’s  $\alpha\beta_{1-42}$  oligomers. *Journal of the American Chemical Society*, 132(38):13300–13312, 2010. doi:10.1021/ja103725c.

- [50] H. Takeuchi. Structures, stability, and growth sequence patterns of small homoclusters of naphthalene, anthracene, phenanthrene, phenalene, naphthacene, and pyrene. *Computational and Theoretical Chemistry*, 1021:84–90, 2013. doi:10.1016/j.comptc.2013.06.029.
- [51] P. D. Teini, D. M. A. Karwat, and A. Atreya. Observations of nascent soot: Molecular deposition and particle morphology. *Combustion and Flame*, 158:2045–2055, 2011. doi:10.1016/j.combustflame.2011.03.005.
- [52] A. G. Tielens. Interstellar polycyclic aromatic hydrocarbon molecules. *Annual Review of Astronomy and Astrophysics*, 46:289–337, 2008. doi:10.1146/annurev.astro.46.060407.145211.
- [53] T. S. Totton, D. Chakrabarti, A. J. Misquitta, M. Sander, D. J. Wales, and M. Kraft. Modelling the internal structure of nascent soot particles. *Combustion and Flame*, 157(5):909–914, 2010. doi:10.1016/j.combustflame.2009.11.013.
- [54] T. S. Totton, A. J. Misquitta, and M. Kraft. A first principles development of a general anisotropic potential for polycyclic aromatic hydrocarbons. *Journal of Chemical Theory and Computation*, 6(3):683–695, 2010. doi:10.1021/ct9004883.
- [55] T. S. Totton, A. J. Misquitta, and M. Kraft. Assessing the polycyclic aromatic hydrocarbon anisotropic potential with application to the exfoliation energy of graphite. *The Journal of Physical Chemistry A*, 115(46):13684–13693, 2011.
- [56] T. S. Totton, A. J. Misquitta, and M. Kraft. A quantitative study of the clustering of polycyclic aromatic hydrocarbons at high temperatures. *Physical Chemistry Chemical Physics*, 14(12):4081–4094, 2012. doi:10.1039/C2CP23008A.
- [57] C. Tsai and K. Jordan. Use of an eigenmode method to locate the stationary points on the potential energy surfaces of selected argon and water clusters. *The Journal of Physical Chemistry*, 97(43):11227–11237, 1993. doi:10.1021/j100145a019.
- [58] D. J. Wales. Energy landscapes and structure prediction using basin-hopping. In A. R. Oganov, editor, *Modern Methods of Crystal Structure Prediction*, chapter 2, pages 29–54. Wiley-VCH, 2011. doi:10.1002/9783527632831.
- [59] D. J. Wales. Exploring energy landscapes. *Annual Review of Physical Chemistry*, 69:401–425, 2018. doi:10.1146/annurev-physchem-050317-021219.
- [60] 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. *The Journal of Physical Chemistry A*, 101(28):5111–5116, 1997. doi:10.1021/jp970984n.
- [61] D. J. Wales and H. A. Scheraga. Global optimization of clusters, crystals, and biomolecules. *Science*, 285(5432):1368–1372, 1999. doi:10.1126/science.285.5432.1368.
- [62] D. J. Wales, M. A. Miller, and T. R. Walsh. Archetypal energy landscapes. *Nature*, 394(6695):758, 1998. doi:10.1038/29487.

- [63] C. S. Wang, N. C. Bartelt, R. Ragan, and K. Thürmer. Revealing the molecular structure of soot precursors. *Carbon*, 129:537–542, 2018. doi:10.1016/j.carbon.2017.12.005.
- [64] Y. Zhao and D. G. Truhlar. A prototype for graphene material simulation: Structures and interaction potentials of coronene dimers. *The Journal of Physical Chemistry C*, 112(11):4061–4067, 2008. doi:10.1021/jp710918f.