Meta-Code for Systematic Analysis of Chemical Addition (SACHA): Application to Fluorination of C_{70} and Carbon Nanostructure Growth

Christopher P. Ewels,*[,†,‡,⊥] Gregory Van Lier,[§,||,⊥] Paul Geerlings,[§] and Jean-Christophe Charlier[||]

Institut des Matériaux Jean Rouxel, CNRS UMR6502, 2 rue de la Houssinière, 44322 Nantes, France, Laboratoire de Physique des Solides, Université Paris Sud, CNRS UMR8502, Batîment 510, 91405 Orsay, France, Research Group of General Chemistry (ALGC), Free University of Brussels (VUB), Pleinlaan 2, B-1050 Brussels, Belgium, and Unité de Physico-Chimie et de Physique des Matériaux (P.C.P.M.), Université Catholique de Louvain (UCL), Place Croix du Sud, 1 (Boltzmann), B-1348 Louvain-la-Neuve, Belgium

Received April 5, 2007

We present a new computer program able to systematically study chemical addition to and growth or evolution of carbon nanostructures. SACHA is a meta-code able to exploit a wide variety of pre-existing molecular structure codes, automating the otherwise onerous task of constructing, running, and analyzing the large number of input files that are required when exploring structural isomers and addition paths. By way of examples we consider fluorination of the fullerene cage C_{70} and carbon nanostructure growth through C_2 addition. We discuss the possibilities for extension of this technique to rapidly and efficiently explore structural energy landscapes and application to other areas of chemical and materials research.

INTRODUCTION

Whereas there are many codes available for postprocessing of electronic structure calculations (for example, to determine infrared or EELS spectra), preprocessing techniques are less common. With continuing increases in computing speed, the calculation time for ab initio structural optimization now sometimes approaches the time required to construct the initial input structures and indeed may no longer be the rate-limiting step. For problems such as the determination of most stable isomers or addition paths for atomic or molecular addends,[1−3] it is necessary to test a large number of similar structures; rather than the arduous task of manually constructing and testing these, it would be desirable to automate this process.

For this reason we have developed SACHA (Systematic Analysis of CHemical Addition), a modular Fortran code which constructs input data files for electronic structure calculations and then analyzes the resultant output, constructing further input files as a result. It does not perform the structural optimizations or property calculations itself (there is no need to reinvent the wheel) but instead is able to call a variety of available codes, from simple interatomic potentials such as the Brenner potential,[4] through to ab initio Hartree Fock (HF) calculations using Gaussian03.[5] It is also able to combine methods, for example, performing a wide range of isomer sampling using lower level semiempirical calculations and then selecting promising candidates for further testing using ab initio treatment. By classifying isomers and selecting promising candidates it can then use these as new candidates for further functionalization or growth, allowing automatic determination of functional group addition pathways, nanostructure growth, etc.

The methodology was already successfully tested for analyzing fluorine addition to C_{60}, where it enabled the determination of the different addition routes for experimentally identified intermediates, up to full $C_{60}F_{60}$ addition.[2] In this particular study, about 10 000 structures were considered, which makes it an unfeasible task to perform manually.

In this paper we describe the functionality of the approach and consider two different test examples: fluorination of the C_{70} fullerene using a combined semiempirical and ab initio approach and carbon nanostructure growth through C_2 addition using structural scoring coupled with density functional tight binding calculations.[6,7] Although we have developed SACHA for analysis of addition patterns, there is no reason its use should be limited to this class of problems. In the analysis and conclusions we discuss limitations and possible future developments for the method, notably in the context of electronic structure calculation data mining.[8,9]

METHODOLOGY

There are many situations where neither Monte Carlo nor Molecular Dynamics are appropriate techniques for resolving routes through complex phase spaces. For example, Molecular Dynamics is of only limited use where reaction barriers linking configurations are high and thus rarely sampled (although we note recent attempts to overcome this using minima biasing[10] and/or genetic algorithms). For the problems we are considering, molecular dynamics also introduces unnecessary complexity when there is a set of clearly defined minima to sample. Monte Carlo techniques are closer to those proposed here. Indeed SACHA could be considered a "highly directed" Monte Carlo methodology, where a table of

* Corresponding author e-mail: chris@ewels.info. Also include any requests for material and for the source code.
† Institut des Matériaux Jean Rouxel.
‡ Université Paris Sud.
§ Free University of Brussels.
|| Université Catholique de Louvain.
⊥ These two authors contributed equally to this paper.

10.1021/ci700121z CCC: $37.00
Published on Web 10/25/2007

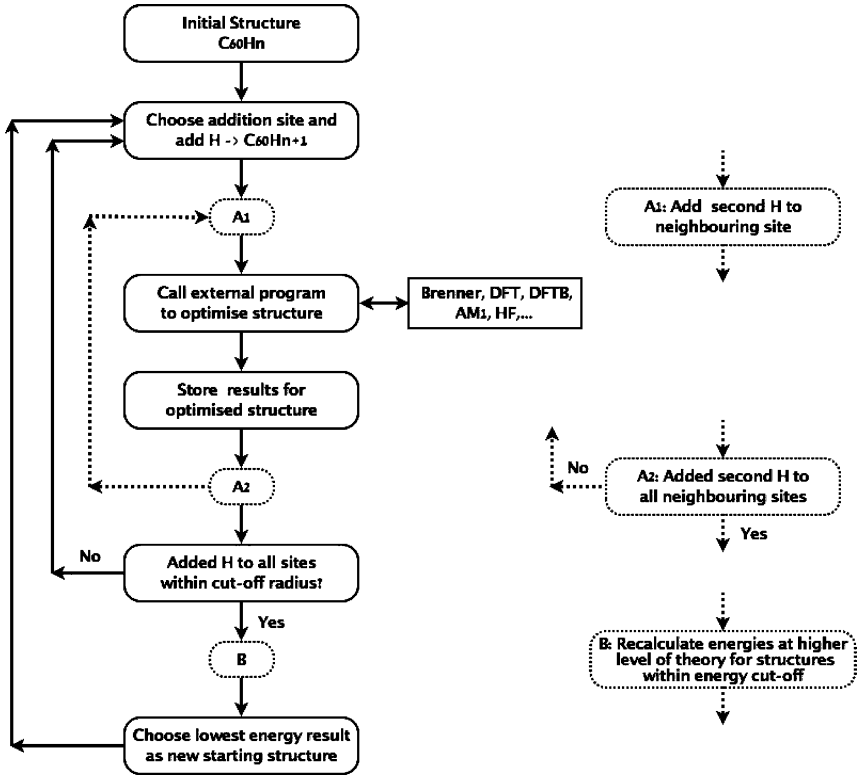

Figure 1. Flowchart showing operation of SACHA for determining addition patterns, in the example shown, hydrogen addition. Optional additional steps are marked in dashes.

potential moves is established, but instead of randomly selecting moves based on the Metropolis algorithm[11] the lowest energy step is consistently chosen. However the aim of Monte Carlo and SACHA calculations are different; whereas a Monte Carlo calculation may simulate system evolution over time using an exhaustive look-up table of potential random moves, SACHA aims to dynamically develop such a look-up table. And while, by removing the random element in system selection we may restrict our ability to fully explore the phase space, the resultant time savings allow us to tackle problems beyond the reach of conventional Monte Carlo.

SACHA was originally developed in order to automate the process of generating structural isomers and the resultant geometry optimizations. A generic map of the method is shown in Figure 1. Taking hydrogenation of C_{60} as an example, in the code's simplest mode of operation we first take C_{60} and seed it with a single H atom.

The code then iterates through all the possible addition sites for a second hydrogen atom. For each site the appropriate structure is generated, followed by an input file for the external code to be used (e.g., Gaussian03[5]). This code is called, and, once completed, the optimized structure and total energy are extracted from the output file of the external code. These results are sorted and tabulated, with the most stable structure taken as the base structure for the addition of a further hydrogen atom, as the addition process is repeated. In this way a sequence of $C_{60}H_n$ ($n = 1,2,...$) structures is generated, where each structure $C_{60}H_n$ is the lowest energy structure possible for H addition to $C_{60}H_{n-1}$. In cases where several isomers have very close energy it is possible to take the second, third, etc. most stable structure as a starting point for further addition, to examine possible divergence at that point.

To this basic setup we have added a number of optional refinements. Although above we have taken the most stable structure for H_n as the structure to hydrogenate to H_{n+1}, this may not always be the most stable addition route. For example, with pairwise atom addition such as H_2 gas addition, it is possible that with the use of single atom addition steps, a less stable addition site for the first atom still gives the more stable paired addition site after the second atom is added. With single atom addition steps the most stable addition route could thus be missed. In this case SACHA can test pairwise addition by trying all sites for a single H together with adding a second H to each of the three possible neighboring sites (addition '*A1*' and '*A2*' in Figure 1). More general pairwise testing, i.e., on non-neighboring sites, is possible but computationally very expensive and chemically less relevant. Such an approach can also be used to model, for example, bridging oxygen addition, i.e., where the addend is placed over a bond center rather than an atom.

In the case of large systems, it may be relevant to analyze the addition around only a given part of the system, for instance if there will be no interaction between addition sites if they are too far apart. For this, a second option has been introduced: rather than testing all possible structures we can instead choose to impose a *cutoff radius* for functionalization, defined as the number of carbon neighbors between the C atom being considered for the next addition and the initially functionalized carbon atom. If these two atoms are further apart than the provided *cutoff radius*, the site is not considered. As the choice of the *cutoff radius* can lead to different addition patterns, it can be used to further direct the addition route by restricting the addition to only part of the system (see below). The method requires structural optimization of a large number of isomers, and hence it is

Table 1. List of Molecular Structure Codes Which SACHA Is Currently Able To Utilize[a]

theory type	calculation type	external code used
interatomic pair potential optimization	geometry optimization	built into SACHA
Brenner potential	energy minimization/molecular dynamics steepest descent	Fungimol[13]
force-field molecular mechanics using parametrized UFF	single point energy/molecular dynamics steepest descent	DeMon 2004[14]
force-field molecular mechanics	molecular mechanics and dynamics/geometry optimization	Tinker[15]
semiempirical methods	single point energy/Geometry optimization	Gaussian03[5]
ab initio Hartree−Fock	single point energy/Geometry optimization	Gaussian03[5]
density functional tight binding	single point energy/geometry optimization/molecular dynamics steepest descent	DeMon 2004[14]
density functional theory	single point energy/geometry optimization	AIMPRO[16]/Gaussian03[5]
polygon identification within structural network	arbitrary scoring based on analysis of neighbor map locating number (and proximity) of polygons with n sides ($n = 3,12$), optionally penalizing neighboring pentagons	built into SACHA

[a] I.e., has the ability to construct an input data file, run the code, and extract the resultant structural coordinates, total energy, and property calculation results where applicable.

often necessary to compromise on the level of theory used. Since it is a modular code, SACHA can use different levels of theory for different tasks. For example, a two-step mechanism can be used where all structural optimizations are performed at a lower theoretical level and ordered according to the lowest energy, followed by single-point (or further optimization) calculations at a higher level of theory for all structures within a given *energy cutoff* of the lowest energy system, in order to determine the final energy ordering (addition 'B' in Figure 1). This has successfully been applied when studying fluorination addition routes of C_{60} buckminsterfullerene, where a two-step mechanism was used with all structural optimizations at the semiempirical AM1 level of theory,[12] followed by single point energy calculations for all structures within a 1 eV *energy cutoff* of the best energy at the HF/3-21G level of theory for final energy ordering.[2] This methodology is especially useful when the first, lower level of theory used gives good geometry optimizations, but where isomer energy ordering might differ from, e.g., ab initio calculations. Since SACHA becomes especially useful when a large number of systems need to be considered, the structural optimization of these becomes the time limiting step. For this, the structures generated can be 'preoptimized' by adapting the atomic coordinates around the addition site rather than just adding the addend's coordinates. When, for example, a functionalized carbon atom shifts from sp^2 to sp^3 coordination, the pyramidalization angle can be automatically increased with an adjustable amount in order to save on the number of required steps to optimize the structure.

Note that the general approach of SACHA contains a number of restrictions, although none of these are inherent:

First, since chemical addition typically occurs on the outer surfaces of a system, addition is forced in our case to occur on the side opposite to the system's center of mass. This restriction can easily be lifted or modified as required.

Second, we do not take explicit advantage of system or site symmetry at present. Nonetheless symmetry duplication can be minimized by the use of a two-step mechanism, since the choice of the systems for the second, higher theoretical level step can be restricted not only by the *energy cutoff* but also by only considering the systems within a given *isomer energy cutoff*. This means that for isomers with a total energy within, e.g., 1×10^{-6} au from each other, only one is considered for higher-level treatment, with the assumption that such close energies must indicate symmetrically identical systems. This cutoff value has been used for the AM1/Hartree−Fock 3-21G treatment described below but can be adjusted according to the level of theory considered.

Table 2. List of Structural Tasks SACHA Can Currently Perform

task	notes
H addition	either isolated H, or H_2 on neighboring sites
F addition	either isolated F, or F_2 on neighboring sites
O addition	in a bond bridging configuration (epoxide addition)
C_2 addition	bridging a bond pair from the same hexagon, heptagon, etc. in an sp^2-carbon network, trying all possible bond pair combinations for each polygon
C−C bond rotation	Stone−Wales rotations
C atom substitution	substitution by N, B, etc.; can also be used to *remove* C atoms systematically, i.e., substituting atoms by vacancies

Finally, as a more fundamental limitation, we do not consider surface migration of species after addition, which is discussed further below. In more general terms, we are assuming that energy is the key factor for determination of common isomeric structures. In practice other factors could also play an important role in the decision process such as kinetic processes. Some inclusion of such effects can be incorporated, for example, restricting allowed addition sites to lie within a certain radius of current addition sites by adjusting the cutoff radius or forcing first neighbor pairwise addition.

The codes with which SACHA can currently interact are given in Table 1, along with various options within these codes. It also has routines for simple interaction with job queuing systems, allowing these calculations to be queued and monitored, with SACHA 'hibernating' until the job has passed to the front of the queue and executed. In this way it has successfully been used to run on both serial and parallel architectures. Modification to use another code or queuing system would simply require adding a new interface module.

Although in the discussion above we have used the example of hydrogen addition to demonstrate the formalism of the code, SACHA is capable of performing a range of different chemical 'operations' to any given input structure, as listed in Table 2. Again the modular nature of the code means it is relatively simple to add further tasks.

RESULTS FOR EXAMPLE PROBLEMS

We now present two different example problems in order to further demonstrate the possible application of SACHA: namely fluorination of C_{70} as an example of pairwise chemical addition, and analysis of growth mechanisms, in this case of carbon nanotubes through C_2 carbon dimer ingestion. Reaction enthalpies are approximated as the energy difference

$$E_R = E(C_{70}F_n) - E(C_{70}F_{n-2}) - E(F_2)$$

providing a measure of the exothermicity of fluorination.

TEST CASE 1: FLUORINATION OF C_{70}

As an initial test case, we analyze the fluorination of the fullerene molecule C_{70}. We have previously used SACHA to study the fluorination of C_{60},[2] where the technique successfully identified a number of fluorine addition routes and key stable species ('magic numbers' such as $C_{60}F_{18}$, $C_{60}F_{20}$, and $C_{60}F_{36}$). Fluorination has been applied very early in the functionalization of fullerenes, as it provides for a fast and easy addition method, improving solubility and thus procesability.[17] In recent years, fluorination is also emerging as an important process for functionalizing and chemically activating carbon nanotubes.[18]

There are no systematic theoretical studies of full addition paths for fluorinated C_{70} species in the literature to date. Experimental studies using MnF_3 as fluorinating agent did not find any stable fluorinated species with less than 34 fluorine atoms, with a broad range of species being produced within compositions between $C_{70}F_{36}$ and $C_{70}F_{44}$.[19] One of the few species to be analyzed theoretically in detail is $C_{70}F_{36}$,[20] in which various isolated benzenoid rings and double bonds were identified. Thus from the available experimental data it is clear that C_{70} behaves quite differently under fluorination to C_{60}. Chemical addition to C_{70} is governed by the presence of 5 benzenoid rings at its equator, with low reactivity toward addition due to the extended electronic delocalization.[21]

We have used SACHA in a two-stage selection process. Starting from C_{70}, all possible $C_{70}F_2$ species are produced where the two fluorine atoms are added pairwise to neighboring sites. These structures are then fully optimized at the AM1 level of theory,[12] and the results are sorted by energy. All structures with an energy up to 1 eV difference with the most stable system, and an isomer energy cutoff radius of 1×10^{-6}, are then considered for the second stage where their single point energies are recalculated at the HF/3-21G level of theory. Of these, the structure with the lowest HF/3-21G calculated energy is then taken as the new starting point for further F_2 addition. All calculations are performed by calling the external Gaussian03 program[5] from within SACHA. Since we only aim at further showing the possibilities of the SACHA program, extensive analysis of all addition routes for C_{70} fluorination is clearly beyond the scope of this article. We have thus chosen to analyze only the path resulting from addition to the most stable isomers at each $C_{70}F_{2n}$ stage, i.e., the thermodynamic addition path. Further addition to less stable isomers leads to bifurcations and less obvious addition patterns that need closer analysis and will be presented elsewhere.

Figure 2 shows the distribution of lowest energy isomer energies (up to 25 kcal/mol difference from the most stable isomer) obtained in this way for the fluorine addition sequence. It can be seen that at certain addition steps (e.g., $C_{70}F_6$ or $C_{70}F_8$) there are other isomers with energies very close to the ground-state structure; these are points where bifurcation of the addition pathway may be possible (other isomers have no feasible alternative addition sites; this often coincides with a more exothermic addition). As found previously for C_{60} which shows such a bifurcation point at $C_{60}F_6$ (between the 'S' and 'T' addition patterns with an energy difference of about 4 kcal/mol),[2] the energy differences for C_{70} are of the same order at certain stages, which also suggest the presence of different addition paths.

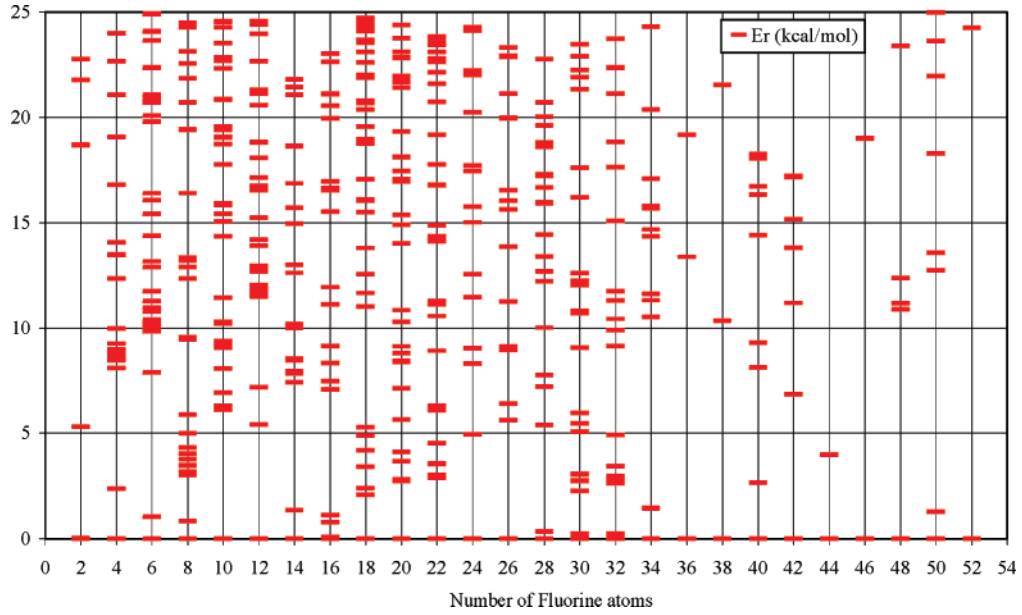

Figure 2. Distribution of lowest energy isomer energies obtained by AM1 structural optimization followed by HF/3-21G single point energy calculations. $C_{70}F_n$ isomers are generated through F_2 addition to the most stable $C_{70}F_{n-2}$ isomer. When multiple isomers have energies close to the ground-state structure energy, bifurcation of the addition pathway becomes a possibility although not considered here.

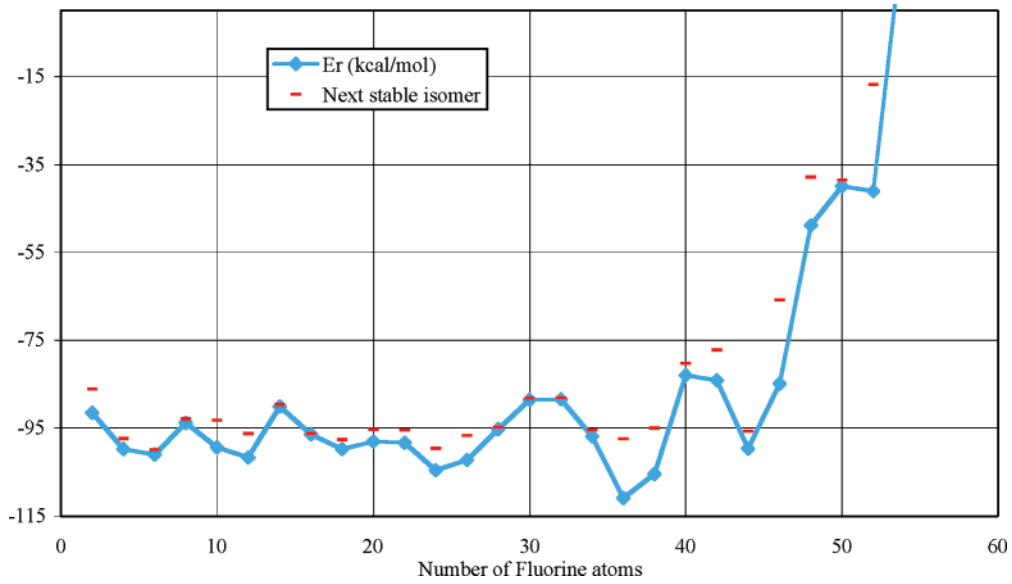

Figure 3. E_R for $C_{70}F_{n-2}$ and F_2 to give $C_{70}F_n$ (kcal/mol) versus the number of fluorine atoms added.

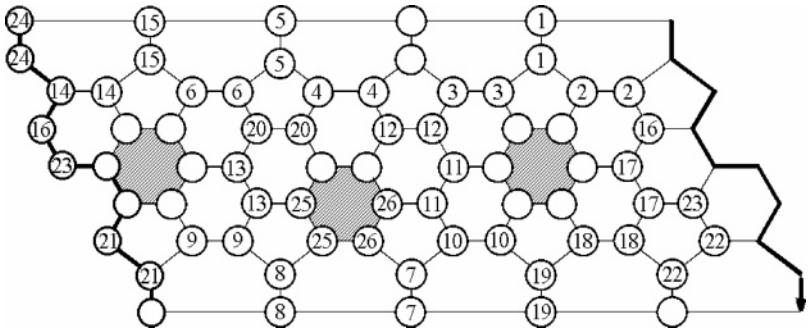

Figure 4. Modified Schlegel diagram, where C_{70} is 'cut open' following the C_5 axis, showing the fluorine pair addition order. The shaded rings show the benzenoid rings (see text).

Figure 3 shows the reaction energy for F_2 addition for each fluorinated C_{70}. The addition sequence is much smoother than for C_{60} fluorination,[2] showing that the addition path does not lead to structures that are exceptionally stable or unstable (i.e., with significantly higher exothermicity) up to $C_{70}F_{36}$. This is consistent with the experimental observation that there are no significant 'magic number' $C_{70}F_x$ fullerenes for $x <$ 36.[19]

The most stable addition route identified is given in Figure 4, where a Schlegel diagram depicts the addition order of F_2 pairs to C_{70}. Addition starts with a pattern analogous to the 'T' addition pattern previously observed for C_{60} fluorination.[2] This addition pattern is repeated on the same C_{70} hemisphere twice to form $C_{70}F_{12}$. The next addition to the same hemisphere is 7.4 kcal/mol less exothermic as compared to addition to the other hemisphere, and thus further addition occurs there. Conversely, the addition to the second hemisphere follows an "S" addition pattern, which becomes more stable as compared to the "T" pattern due to previous structural distortion caused by fluorine addition. In this way, $C_{70}F_{20}$ is formed, after which fluorine atoms are added for the first time to the equatorial belt. For $C_{70}F_{30}$, one unfluorinated benzenoid and one corranulene subunit persists, showing the tendency of the fullerene fluorination to be driven by increased aromaticity, in accordance with observed C_{60} fluorination.[2] Further addition leads to $C_{70}F_{36}$, which has a significantly higher exothermicity. This coincides with the experimental observation that $C_{70}F_{36}$ and $C_{70}F_{38}$ can be obtained as stable structures. Nonetheless, the structure we identified for $C_{70}F_{36}$ has two unfluorinated benzenoid rings and one naphtalenoid ring and differs from the one identified theoretically by Clare and Kepert[20] as being the most stable $C_{70}F_{36}$ having four benzenoid rings. It is indeed possible that via another addition path the latter structure can be obtained, since as discussed above we only consider the most stable addition here, although no experimental identification of the structure for $C_{70}F_{36}$ has been performed which could provide final proof.

Further addition leads to $C_{70}F_{44}$ before the exothermicity greatly drops. The important role of aromaticity can still be observed upon further fluorination, where 3 remaining benzenoid rings remain unfluorinated for $C_{70}F_{48}$, after which two more fluorine pairs are added to one of these before the reaction becomes endothermic for $C_{70}F_{54}$ and above. This coincides with the experimental observation that isomers between $C_{70}F_{36}$ and $C_{70}F_{44}$ have been observed together, but no higher isomers were obtained upon fluorination with MnF_3.[19]

Note that the method as implemented does not take advantage of site symmetry at present and does not consider surface migration of species after addition, which is discussed further below. Nonetheless, although out of a total number of systems of 2485 that would need to be considered for a single atom addition for the full analysis of only one reaction path from C_{70} up to $C_{70}F_{70}$, only about 1700 systems are considered for paired addition. (The actual number depends

on the actual addition path, i.e., at which stage added pairs are adjacent or not). But this number is further reduced by using a two-step mechanism where only symmetry-independent isomers are considered by using an *isomer energy cutoff* of 1×10^{-6} au. Although this still means that in the first calculation step systems will be considered that are symmetrically equivalent, this excess of calculations is mainly important for the first few additions giving products with higher symmetry. In consequence, for the remaining additions, less than 10% of excess calculations are typically performed at the first, lower theory level calculation step, which are being filtered out by the *isomer energy cutoff* before the higher level calculations. This introduces a workaround for symmetry treatment, which further reduces the total amount of systems treated at higher theory level to less than 1000, resulting in an important computing time reduction. Despite these potential limitations the technique successfully gives a complete addition pathway, which is in excellent agreement with the limited available experimental information, providing stable structural models for every $C_{70}F_{2n}$, up to $C_{70}F_{52}$.

TEST CASE 2: GROWTH OF CARBON NANOSTRUCTURES THROUGH C_2 ADDITION

For our second example we choose a very different system, namely the growth of carbon nanostructures through the ingestion of C_2 dimers. While improved growth techniques now mean high purity fullerene samples can be produced directly, the precise growth mechanisms for fullerenes, nanotubes, and other carbon nanostructures are still in question.

There are many theoretical studies of fullerene growth in the literature, for example, using density functional tight binding molecular dynamics, allowing random C_2 coalescence.[22] However fullerene structures are typically formed in timescales of milliseconds, unattainable in full molecular dynamics calculations. Performing theoretical studies incorporating the full kinetic processes and catalytic mechanisms would be a truly formidable task. Hence one of the key theoretical questions to tackle first is an elucidation of formation *mechanisms* rather than entire growth processes.

The energetically most favorable sp^2-carbon structures contain a maximum of hexagons, for aromatic and curvature reasons. In addition, for energetic stability the isolated pentagon rule (IPR) is favored,[23] namely for reasons of aromatic stability pentagons should not neighbor each other.

Fullerene growth models normally invoke C_2 addition from the gas phase to pre-existing smaller clusters.[24] Such models commonly then invoke bond rotation and reconstruction (such as the 'pentagon road' for fullerene growth[25,26]) as a way of rearranging the lattice to restore the isolated pentagon rule and minimize nonhexagonal structures. These bond rotations ("Stone−Wales" rotations[27]) involve the rotation of two carbon atoms through 90° about the center of their shared bond. This rotation modifies the four polygons neighboring the atom pair, decreasing the number of sides of one opposing pair and increasing the number of sides of the other opposing pair by one in each case. Aside from dislocation climb processes, such bond rotations are the only reported process for translation of pentagons through an sp^2-network and hence are assumed to be responsible for assuring sp^2-carbon networks obey the IPR.

However if such processes are to be responsible for reorganization within carbon nanostructure walls, high energetic barriers (~6 eV) must be overcome,[28] rendering them inaccessible at moderate annealing temperatures. These barriers can be somewhat decreased through catalysis by metastable interstitials,[29] but such catalytic species will be rare and short-lived.

We thus pose the hypothetical question: Are such rotations strictly necessary after C_2 addition in order to maintain isolated pentagons? Or can selective site C_2 addition in itself generate genuine IPR structures? It is important to note that we are not attempting to realistically simulate the growth of carbon nanostructures with this study; we do not include a myriad of other potentially relevant processes such as carbon dimer ejection.[30] Instead the intention is to determine, via a highly driven selection process coupled with quantum chemical calculations, whether there exists growth routes to IPR structures with no bond rotation "annealing" steps required.

We started with a polyaromatic hydrocarbon sheet centered on a single pentagon, $C_{45}H_{15}$. Using SACHA we then considered the breaking of two C−C bonds in the same polygon and insertion of C_2. We allow this for all polygons in all possible arrangements, with the exception of breaking C−H bonds, and then "score" the resultant structure with a series of weighted penalties for the number of nonhexagonal patches, with an extra penalty for neighboring pentagons. A given C_2 addition step can typically have up to 5 possible 'best' additions leading to the same number of pentagons and heptagons added, removed, paired, etc., and this scoring technique is unable to distinguish between them. Therefore we use a postprocessing option, namely these identically scoring winning structures were then geometrically optimized using the deMon density functional tight binding (DFTB) code, and the energetically most stable structure is taken as the start structure for further C_2 addition. This demonstrates that many levels of compromise are possible with such techniques between computational speed and theoretical rigor.

The resultant structures are shown in Figure 5, with the addition of up to 55 C_2 dimers. Although there are some considerable asymmetries in the structure (induced when there are multiple moves at one step with the same cost, at which point the program selects the first that it encounters, breaking the system symmetry), it is clear from the images that in all cases the majority of the pentagons are isolated, confirming our hypothesis that *with C_2 addition alone* it is possible to produce an extended carbon structure which largely obeys the IPR. Despite this there remain several structural defects in all cases, suggesting that although no bond rotations may be required for structural growth, for highly ordered completely graphitic nanostructures, some bond rotations will be necessary. These are typically performed experimentally during high temperature annealing after the actual synthesis, allowing for pentagon translation apart from the growth conditions.

When there are clusters of heptagons and pentagons they are often located alongside the terminating hydrogen atoms. This again is reasonable, since the system is limited in this region in the moves it can make. The calculation could be continued further to see whether the structure will develop into a tubular structure diameter constrained by the terminat-

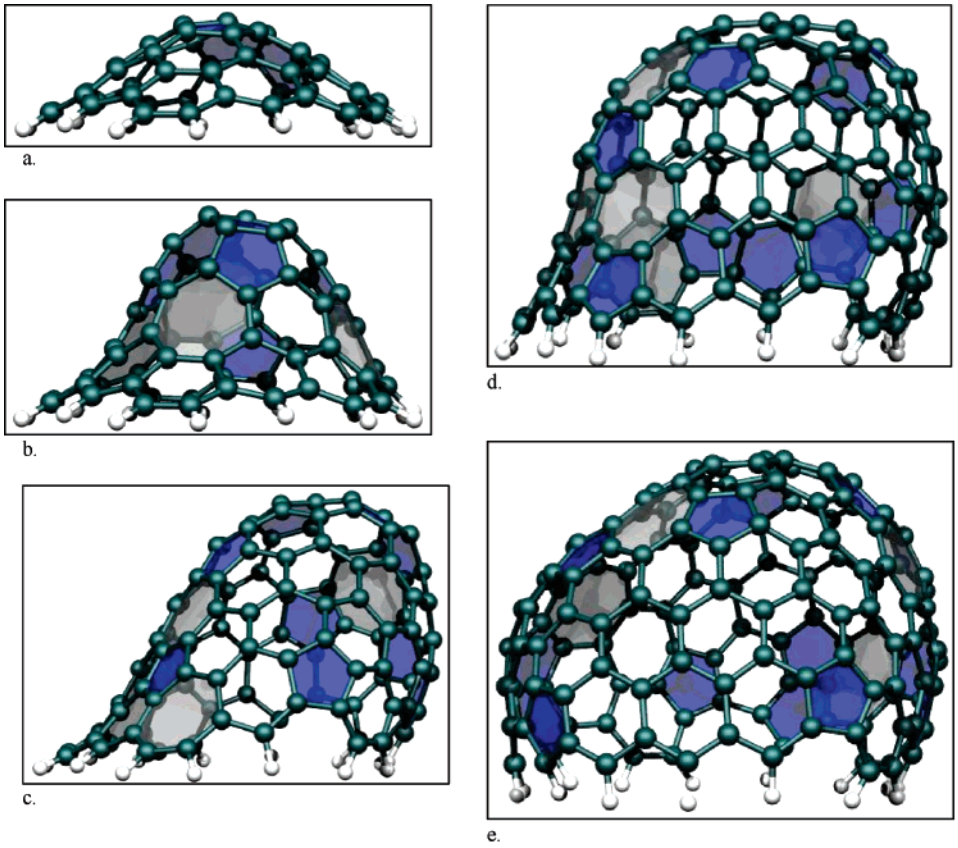

Figure 5. Simulated carbon nanostructure growth starting from $C_{45}H_{15}$, a polyaromatic hydrocarbon centered on a single pentagon. The sheet is 'grown' by SACHA through C_2 addition; in each case all possible sites are tested, and a site is chosen that minimizes the number of pentagons, heptagons, and paired pentagons. The figure shows the resultant hydrocarbon at different stages during growth, with (a) 47, (b) 67, (c) 105, (d) 123, and (e) 155 carbon atoms. Pentagons are shaded blue, heptagons are shaded gray, and hexagons are unshaded.

ing hydrogens or expand into a sphere. If indeed the terminating hydrogens are important in determining the resultant dimensions of the nanostructure, this may be reflected in experiment as some constraining nucleating particle such as a metal catalyst.

DISCUSSION AND FUTURE DEVELOPMENT

SACHA has the capability to operate with a variety of different structural optimization codes (given in Table 1), and it is a relatively simple matter to add others. Equally there are now a range of different tasks it can perform which have been coded in, listed in Table 2; again the modular structure means it is fairly simple to add others. An important additional function given its ability to generate large numbers of calculations is an option to submit calculations to a queuing system, waiting for each addition or growth step to complete before starting the next one.

Having developed the code this far it is a simple matter to extend it to study more complex reactions and systems. The code contains routines for rotating C−C bonds, allowing, in principle, studies of Stone−Wales[27] mediated structural evolution of carbon nanostructures, for example, the 'welding' of two polymerized C_{60} molecules to give a 'peanut' and ultimately nucleate a nanotube.[31] As another example application, with the addition of slightly more thorough bookkeeping of structures and energies it would be possible to 'grow' a library of carbon nanostructures, tabulating their ab initio properties. Although we have only discussed here the functionalization of carbon nanosystems (graphene, fullerenes, and nanotubes) there is no reason why it will not work with other materials such as boron-nitrogen nanostructures, or indeed surface functionalization, or growth of bulk crystal structures, zeolites, etc.

We note that by sampling many potential isomers we are gradually developing a 'look-up table' of structural configurations. Thus SACHA could be extended to a full Monte Carlo approach, evoking its isomer sampling routines whenever the Monte Carlo made a choice connecting it with structures not currently in the look-up table. This would ultimately result in a linked web of isomers and energies, but, unlike a systematic sampling of all possible isomers, would order the sampling starting with the most likely structures. Alternatively it would be possible to determine the lowest energy addition pattern route and then return to isomers closest in energy to the ground state and explore the lowest energy path from these, gradually developing a complete addition map.

Probably the major limitation of the current SACHA methodology is that we do not consider migration of addends between addition steps. This is not appropriate in many cases, but, for example, for wet chemistry fluorination of C_{60}, nuclear magnetic resonance (NMR) observations show conversion between $C_{60}F_{36}$ isomers over time, even at room temperature.[32] The SACHA results are applicable either where migration barriers are high or in regimes of rapid addition where migration is minimized. However it would be possible to extend the methodology to consider migration through inclusion of an extra step between additions if the

computing cost could be justified. For example, each surface addend could be allowed a single site hop, with the lowest energy isomer chosen for continuing addition. Such hopping could also be continued iteratively until an energy minimum is located. Alternatively, the addends could be alternately added and removed from the ball at each addition step, each time selecting the lowest energy variant. Ultimately the only restriction on such calculations is computational time.

In situations where the number and type of atoms remain the same (such as, for example, Stone−Wales bond rotations of a graphitic carbon nanostructure), each progressive step is a new generation of structure but can still be compared to those calculated from the generation before. This could allow for considerably more complex choice selection, for example, instead of employing a simple step-by-step progressive method, it should be possible to use instead a genetic algorithm or allow the code to intelligently 'rewind' to a previous generation should it find itself in an energetic 'cul-de-sac'.

For the calculation types discussed here, the structures are classified using their total energy. However there is no reason why another system parameter may not be used, for example, magnetizability, polarizability, or dielectric constant. In this way SACHA could be used to perform structural studies over landscapes other than the energetic. Toward this aim we have, for example, additional routines within SACHA for calling Gaussian03 to determine ring-centered NICS (nucleus independent chemical shifts) to study the aromaticity along with the addition path.

Hopefully from the discussion and examples given here it is clear that SACHA is a useful and productive preprocessing analysis tool for studying addition pathways and structural evolution in molecules and materials, both in its current incarnation and with the addition of the many possible expansions discussed above.

ACKNOWLEDGMENT

We thank Annick Loiseau for organizing the Ecole "Nanotubes: Science et Applications" held in Aussois (France), where the idea for SACHA was born. J.C.C. acknowledges the Fonds de la Recherche Scientifique (FNRS) as a Senior Research Associate. G.V.L. acknowledges the Research Foundation - Flanders (FWO) as a Postdoctoral Fellow. Parts of this work are also connected to the EU-STREP "Nano^{2}hybrids" (NMP-CT-2006-033311), to the Interuniversity Attraction Poles (PAI6) on "Quantum Effects in Clusters and Nanowires", and to the FAME European network of excellence.

REFERENCES AND NOTES

(1) Ewels, C. P.; Van Lier, G.; Heggie, M. I.; Charlier, J.-C. *Phys. Rev. Lett.* **2006**, *96*, 216103.
(2) Van Lier, G.; Cases Amat, M.; Ewels, C. P.; Taylor, R.; Geerlings, P. *J. Org. Chem.* **2005**, *70*, 1565−1579.
(3) Van Lier, G.; Ewels, C. P.; Zuliani, F.; De Vita, A.; Charlier, J.-C. *J. Phys. Chem. B* **2005**, *109*, 6153−6158.
(4) Brenner, D. W. *Phys. Rev. B* **1990**, *42*, 9458−9471.
(5) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 2003.
(6) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, 12947.
(7) Seifert, G.; Porezag, D.; Frauenheim, T. *Int. J. Quantum Chem.* **1996**, *58*, 185.
(8) Curtarolo, S.; Morgan, D.; Persson, K.; Rodgers, J.; Ceder, G. *Phys. Rev. Lett.* **2003**, *91*, 135503.
(9) Morgan, D.; Ceder, G.; Curtarolo, S. *Meas. Sci. Technol.* **2005**, *16*, 296−301.
(10) Laio, A.; Parrinello, M. *PNAS* **2002**, *99*, 12562−12566.
(11) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *1*, 1087−1092.
(12) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(13) Freeman, T. Fungimol 0.5.1. http://www.fungible.com/fungimol/ (accessed 15/7/2007).
(14) Köster, A. M.; Calaminici, P.; Casida, M. E.; Flores-Moreno, R.; Geudtner, G.; Goursot, A.; Heine, T.; Ipatov, A.; Janetzko, F.; del Campo, J. M.; Patchkovskii, S.; Reveles, J. U.; Salahub, D. R.; Vela, A. NRC: Ottawa, Canada, 2004 (The current authors would like to thank the DFTB research team at Dresden for allowing access to their code for this project.).
(15) *Tinker 4.2*; Jay Ponder Lab, Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine: St. Louis, MO 63110, U.S.A., 2004.
(16) Briddon, P. R.; Jones, R. *Phys. Stat. Sol. B* **2000**, *217*, 131−171.
(17) Taylor, R. *J. Fluor. Chem.* **2004**, *125*, 359−368.
(18) Glerup, M.; Krystic, V.; Ewels, C. P.; Holzinger, M.; Van Lier, G. In *Doped Nanomaterials and Nanodevices*; Chen, W., Ed.; American Scientific Publishers: 2007.
(19) Taylor, R.; Abdul-Sada, A. K.; Boltalina, O. V.; Street, J. M. *J. Chem. Soc., Perkin Trans.* **2000**, *2*, 1013−1021.
(20) Clare, B. W.; Kepert, D. L. *J. Mol. Struct. (Theochem)* **2002**, *583*, 45−62.
(21) Van Lier, G.; De Proft, F.; Geerlings, P. *Chem. Phys. Lett.* **2002**, *366*, 311−320.
(22) Zheng, G.; Irle, S.; Morokuma, K. *J. Chem. Phys.* **2005**, *122*, 014708.
(23) Kroto, H. W. *Nature* **1987**, *329*, 529.
(24) Endo, M.; Kroto, H. W. *J. Phys. Chem.* **1992**, *96*, 220.
(25) Kroto, H. W. *Science* **1988**, *242*, 1139.
(26) Smalley, R. E. *Acc. Chem. Res.* **1992**, *25*, 98.
(27) Stone, A. J.; Wales, D. J. *Chem. Phys. Lett.* **1986**, *128*, 501.
(28) Eggen, B. R.; Heggie, M. I.; Latham, C. D.; Jones, R.; Briddon, P. R. *Science* **1996**, *272*, 87.
(29) Ewels, C. P.; Heggie, M. I.; Briddon, P. R. *Chem. Phys. Lett.* **2002**, *351*, 178−182.
(30) Rohlfing, E. A.; Cox, D. M.; Kaklor, A. *J. Chem. Phys.* **1984**, *81*, 3322.
(31) Han, S.; Yoon, M.; Berber, S.; Park, N.; Osawa, E.; Ihm, J.; Tomanek, D. *Phys. Rev. B* **2004**, *70*, 113402.
(32) Avent, A. G.; Taylor, R. *Chem. Commun.* **2002**, 2726−2727.

CI700121Z