Journal of Solid State Chemistry 184 (2011) 1484–1491

Contents lists available at ScienceDirect

Journal of Solid State Chemistry

journal homepage: www.elsevier.com/locate/jssc

Probing the structure of complex solids using a distributed computing approach—Applications in zeolite science

Samuel A. French [1], Rosie Coates, Dewi W. Lewis *, C. Richard A. Catlow

Department of Chemistry, University College London, 20 Gordon Street, London, WC1H 0AJ, UK

A R T I C L E I N F O

Article history:
Received 13 April 2010
Received in revised form
4 March 2011
Accepted 4 April 2011
Available online 12 April 2011

Keywords:
E-science
Zeolites
Modelling

A B S T R A C T

We demonstrate the viability of distributed computing techniques employing idle desktop computers in investigating complex structural problems in solids. Through the use of a combined Monte Carlo and energy minimisation method, we show how a large parameter space can be effectively scanned. By controlling the generation and running of different configurations through a database engine, we are able to not only analyse the data "on the fly" but also direct the running of jobs and the algorithms for generating further structures.

As an exemplar case, we probe the distribution of Al and extra-framework cations in the structure of the zeolite Mordenite. We compare our computed unit cells with experiment and find that whilst there is excellent correlation between computed and experimentally derived unit cell volumes, cation positioning and short-range Al ordering (i.e. near neighbour environment), there remains some discrepancy in the distribution of Al throughout the framework. We also show that stability–structure correlations only become apparent once a sufficiently large sample is used.

1. Introduction

Many of the physicochemical properties of solids that can be technologically exploited, such as catalytic, conductivity and optical properties, depend on the details of the structure, both local and long range. Many such properties arise from the presence and distribution of defects within the material. Thus, in the field of catalysis, many active sites are known to be defect centres such as low coordinated surface sites, vacancies or interstitials. More generally, key properties may be determined by the nature of the disorder of specific atoms within the host lattice. Thus, the distribution of metal cations at different lattice sites in a perovskite-structured material can influence, for example, the ion conductivity and the magnetic properties of the material. Such materials, containing only two cation sites, can be considered a simple case. More generally, the ability to characterise accurately the distribution of different configurations is often limited, in the first instance by experimental techniques and secondly by the very large number of potential configurations that would need to be calculated. Whilst computational methods, particularly Monte Carlo (MC) methods can be successful in modelling such complex systems, they are often limited by the number of different configurations that can be considered and by the fact that the critical local relaxations within the structure are often neglected. However, we do note the substantial progress made by Allan and co-workers who have developed a hybrid Monte Carlo method that is effective in modelling disorder in dense oxide materials [1]. Similarly, we note methods that allow the constriction of all possible, crystallographically distinct solid solution configurations [2].

The focus of this paper will be cation and aluminium distributions in zeolites—a key class of materials, whose application in catalysis, gas separation and ion-exchange is directly related to the distribution of atoms within a crystalline lattice. These crystalline microporous aluminosilicates comprise networks of channels and cavities, commensurating in size (4–10 Å) with small molecules that give rise both to size and shape selectivity in their use as catalysts and to their molecular sieving properties. The active sites in such materials arise from the presence of aluminium within what otherwise can be considered a silica framework: each Al, substituting for Si, requires the presence of a cationic species to preserve charge neutrality. In natural and many synthetic materials, the charge compensation is initially provided by alkali metal cations, typically sodium, present as loosely bound species within the pores of the material. For catalytic applications, these cations are often exchanged and replaced by protons, which bind to a bridging oxygen (between a Si and Al) forming a very strong Brønsted acid centre. But regardless of the method of preparation, or the end use, the precise distribution of the Al within the framework and hence

---

* Corresponding author.
E-mail address: d.w.lewis@ucl.ac.uk (D.W. Lewis).
[1] Current address: JMTC, Sonning Common, Reading, RG4 9NH, UK.

0022-4596/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jssc.2011.04.010

the extra-framework species is established during the crystallisation process. But what controls the atomic distribution and is it possible to characterise accurately this distribution? Furthermore, can we identify those parameters that may allow us to control the distribution to such an extent as to allow the optimisation of (for example) catalytic activity within a zeolitic material?

Central to all these questions, is our ability to model such structures and calculate their relative stability. We therefore, require a methodology that can first of all construct chemically reasonable structures and then evaluate the relative stability of each of these configurations, the numbers of which can be very large for disordered systems. Monte Carlo methods are naturally suited to generating ensembles of configurations and have proven valuable in such problems. However, it may be difficult in some cases for such methods to allow for structural relaxation for all configurations. For example, when considering systems with cation disorder (such as zeolites), MC moves will involve swapping pairs of atoms but may not include considerations of how the lattice surrounding these atoms relaxes to accommodate such a change. In the case of flexible structures such as zeolites, where incorporation of Al into the framework is readily accommodated and where the open nature of the pores means that the charge-compensation extra-framework cations are mobile, such local relaxations will have a considerable influence on the stability of a particular configuration. Note, these changes are to some extent more complex than in dense systems where relaxations are usually highly symmetric [1]. The structural rearrangement becomes critical as the number of Al and cations increases. For example, the presence of a single Al in the framework will be compensated by a cation that will coordinate directly to the oxygen atoms adjacent to the Al. However, when multiple Al sites are close together, it becomes more difficult to envisage how to incorporate multiple cations within the pores of the material [3,4].

One possible solution is to perform a full energy minimisation calculation on each configuration. However, it soon becomes clear that such an approach involves excessive and possibly unfeasible computational expense, as the configurational space here is vast. For example, a typical loading of aluminium (Si:Al = 11) and extra-framework cations in a single crystallographic unit cell of Mordenite gives a theoretical number of unique configurations exceeding *ca.*10^{13}, although many of these will be symmetrically equivalent [5,6]. To evaluate simply the energetics of such a sample is costly; to perform optimisation requires a huge increase in resource. However, limiting the degrees of freedom and performing only a limited optimisation can considerably reduce the computational expense. Furthermore, if we are able to deduce, from experience, particular geometries that lead to either low or high energy configurations, we may be able to focus our search of parameter space only to those configurations of interest.

Recently, the advent of distributed computing strategies has brought about a potential step-change in approaching problems where large configurational space has to be considered. Such methods typically take advantage of ‘‘free’’ CPU cycles on desktop PCs: successful examples include climateprediction.net and Richards' drug discovery screensaver [7]. Thus, one can envisage a strategy whereby configurations are generated and then optimised and evaluated on a large number of distributed computers. However, such an approach does little to improve the methodology; we are simply exploiting a larger number of computer resources. A far more profitable approach is to develop a strategy where the results of optimised structures can be used to steer subsequent configuration generation or where optimisations can be modified according to progress. The large configurational space that can now be considered also leads to problems of data management. Relational databases (such as MySQL [8]) allow us not only to store data but also to retrieve and analyse selected configurations based on search criteria. Thus by developing analysis tools (e.g. geometry analysis) that store data in such databases, we increase our ability to extract trends within our results.

As a test case we consider the zeolite Mordenite, widely utilised as an acid catalyst, which can be formed with a wide range of Si:Al ratios, with a variety of extra-framework cations. Previous studies, by Gray et al. [5] and Gray [6], developed and applied the combined Monte Carlo/energy minimisation (MC/EM) strategy used here. However, the computational cost of such an approach, both at the time and even now, if a serial approach is taken, is limited severely by the number of configurations that need to be considered. Nevertheless, in general, excellent agreement with experimental data was found. Cation positions (compared with crystallographic data) and local Si–Al distributions (compared to NMR data) were very well reproduced, as was the general unit cell geometry. However, some discrepancies were found in the relative occupation of the various framework sites (the T-sites that are occupied by either Si or Al). Experimentally, such (partial) occupations are derived by a long established method, whereby consideration of the average T–O distance is an indicator of the Al occupation [9]: increased T–O distances indicate a higher proportion of Al, although we note more recent improvements by Alberti and Gottardi [10] and Alberti et al. [11].

We therefore present here the development of a database-driven methodology for the sampling of aluminium and cation distributions in zeolites that utilises distributed computing resources. With the properties in the database we can easily interrogate for patterns and clustering of properties in order to improve our knowledge of the zeolite structure and the factors that control it. Of course, several properties control Al and cation positions, but with this information we will be able to improve greatly the efficiency of the MC generation step by including more rules and more importantly we can pass this vital information to aid XRD structure refinement.

2. Computational detail

We now give details of the various components of our distributed computing strategy. Our methodology takes a Monte Carlo method for the generation of configurations and couples it to a job submission process that ensures accelerated and improved convergence of jobs whilst at the same time optimally loading the distributed computing resource. Job control and progress, together with the analysis tools (the final component of our strategy), all utilise database tables to store and manipulate results.

2.1. Configuration generation

We utilise a Monte Carlo (MC) method to generate the initial configuration by populating a siliceous framework with Al and then incorporating the charge compensating cations [5,6]. The initial framework is obtained by performing a full optimisation of a supercell, taken from crystallographic data. Thus, for a given Si:Al ratio, our algorithm first randomly substitutes the required number of Si atoms with Al. Additional constraints have been applied, the most useful being that of Lowenstein's rule [12], where no Al–O–Al units are permitted. A cation is then inserted to compensate each Al, by placing it in the same plane as the Al and two randomly selected oxygens to which the Al is bonded. The cation–oxygen distance chosen is the average value found in crystallographic studies of high-silica zeolites [13]. Additional constraints are included to prevent unreasonable cation–cation distances. Each configuration is then output as an unique input file for our chosen energy minimisation code (in this case GULP) [14,15]. The energetics of the system are modelled using a well validated set of interatomic potentials [16,17], within a Born model of the lattice. The electrostatic contribution is treated using the Ewald method [18]. Such a potential model, which includes a description of polarisation of the framework oxygen, has proved to be reliable in describing the structure of siliceous and aluminosilicates zeolites, over a large compositional range.

2.2. Optimisation strategy

Whilst the MC method described provides reasonable starting geometries, we must recall that as the Si:Al ratio is reduced, the perturbation to the structure (local and long-range) becomes considerable. Furthermore, energy minimisation algorithms can only locate local minima and are often prone to numeric instability if the initial geometry is too far removed from a low energy configuration. Thus, simply attempting a full optimisation from such a starting structure often fails. Previous studies by Gray et al. [5] and Gray [6] have shown that for zeolites with low Si/Al ratios, where perturbation from the siliceous structure is greatest, conventional "constant pressure" optimisation (whereby all atom positions and cell parameters are optimised together) fail in almost 30% of cases. However, if a more measured approach is taken, that is a stepwise optimisation, taking into account now our analysis from this larger sample of 50,000 configurations, many of these configurations can also be converged. Note, that each optimisation (if successful) takes of the order of 120 min on a fast Pentium 4 processor and hence a serial (single queue) approach to such a study would require in excess of 12 years of CPU.

2.3. Database-driven job submission and steering

From the discussion above, it is clear that job management and submission is key to obtaining a configurational sample of value. Furthermore, it should be apparent that a simple "queue and wait" approach will severely compromise any queuing system due to the sheer number of jobs such a system would have to manage. Thus, we have developed software that will generate configurations according to user input and store the location of each configuration in a database. Each of these configurations can then be submitted to a distributed computing pool. Each job is optimised according to a stepwise strategy (discussed above), which is also stored in the database. On completion of each step of the optimisation strategy we can therefore determine if further submission is required, whether a particular configuration can be discarded or whether the optimisation strategy devised requires modification. This approach also allows us to "load" optimally a distributed computing pool. Any queuing system will become inefficient as the number of jobs increases. Thus, whilst it may seem attractive to simply generate a large number of configurations and submit them all, this strategy will reduce throughput on such a pool, particularly if many users submit large numbers of jobs. Similarly, we may wish to accelerate the progress of certain configurations, based on the results of our analysis, at the cost of other configurations, which we believe to be less important in the sample. Thus, we have devised a "drip-feed" strategy, whereby only a limited number of jobs are sent to the pool at any one time, but at the same time, we maintain the number of jobs, keeping the pool full, so as to maximise throughput. The workflow strategy is summarised in Fig. 1.

2.4. Analysis tools

Configurations are analysed immediately on return from the computing resource, to determine if they have completed or require further submission. Incomplete jobs can be assessed as to whether they will result in unfeasible geometries or are unlikely to converge. If convergence is likely, these jobs are returned to the job database, with the input file modified for the next applicable stage of the optimisation strategy in place. Moreover, a number of key data, e.g. changes in energy and gradient norm, are stored in the database at each stage, allowing convergence and other indicators to be constantly monitored and modified. Thus, if a stage in the optimisation strategy is not found to lead to any significant improvement in convergence, e.g. the stage is terminated after too few cycles or may have completed early, the strategy database can be modified to improve the process. The analysis strategy is summarised diagrammatically in Fig. 2.

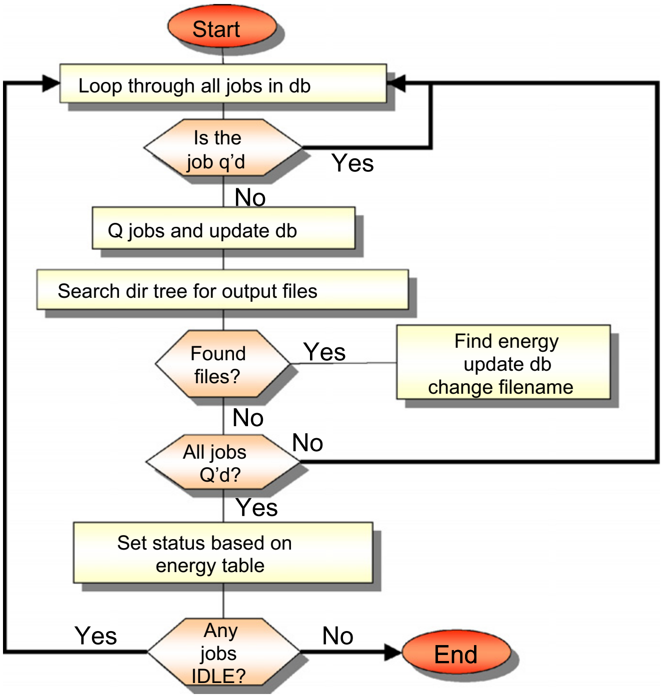
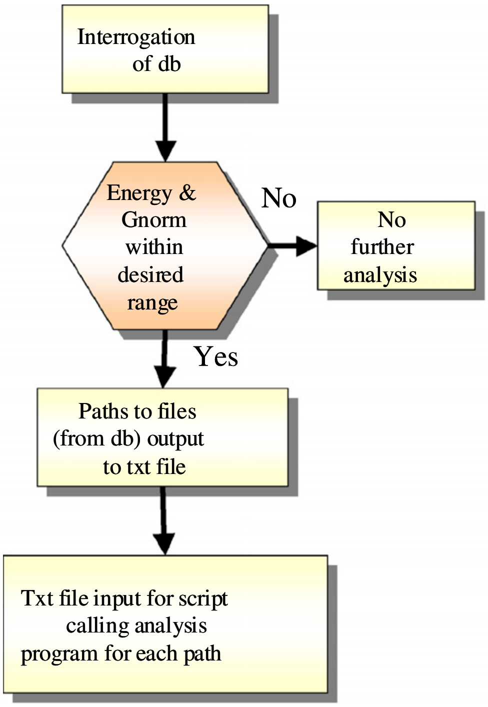


Fig. 1. Workflow for job submission control on a Condor pool. Jobs are stored in a database (db) and drip-fed into the Condor pool queue to maintain an optimal loading of the distributed computing resource.

Completed jobs, or indeed partially converged jobs, can also be more fully analysed. In addition to extracting energy and convergence data, we have written a flexible geometry analysis tool, which enables us to determine and store information on a wide range of parameters that are also stored in our database environment. We can now extract NMR data, T-site occupation by aluminium and cation-oxygen distances as a matter of course. We can also include comparisons with the cation site positions from an experimental unit cell and the coordination distances between specific atoms if required. This level of user-specification allows us to carry out a very thorough, but general analysis of a large number of configurations and on the basis of the results obtained compare some, none or all of the configurations with experimental cells. The process has been streamlined so that the user needs only one command line to carry out the complete analysis of all configurations and input to the database. From the work described here, using the results obtained with the aid of the analysis tools, we hope to improve our understanding of the key factors controlling the structure.

2.5. Condor pools

All the calculations have been performed on a Condor [19] pool at UCL [20]. The pool consists of approximately 1000 low specification



Standard Analysis	
Table	Description of Data
Energy	Id, path, initial &, energy, initial & final gradient norm, convergence status
Cell	Id, number of atoms, a,b, c, α, β, γ, cell volume, number each species, number of cations
Cat_O	Id, average cation-O distance, average cation-O distance, shortest & longest cation-O distance
T_O	Id, average T-O distance, average Al/Ai-O , shortest & longest Al/Si-O
Angle	Id, average T-O-T, min & max T-O-T
Tsites	Id, number of Tsites, occupation and percentage occupation of each site
NMR	Id, number of Al with xAl(4-x)Si neighbours (x=4,3,2,1,0), number of Si with xSi(4-x)Al neighbours (x=4,3,2,1,0)
User Specified Analysis	
Boltzmann	Boltzmann analysis, probability of population of each configuration.
Comparison	Id, label of species to be matched with expt. cell, average match distance, max & min match distance, number of matches match cutoff
Coordination	Id, average coordination, max & min coordination, number coordinated

Fig. 2. Illustration of analysis types available for "on-the-fly" processing of output files. Results can be stored directly in database tables for query construction.

desktop teaching PCs running Windows 2000, which act solely as a client for a Windows Terminal Server. Typically 90% of CPU cycles are available even when users are logged on. However, there are restrictions as to the type of application that suits the Condor architecture. The resource requirement for each individual job should be low and not require parallelisation. Ideal applications consist of a large number of similar size jobs that have slightly different input parameters. For this study of Mordenite we have performed in excess of 50,000 calculations for the specific ratio of Si:Al using approximately 12 CPU years of previously unused resource.

3. Results

In the following section we compare calculated properties with observed trends that allow us to ascertain the factors that differentiate between low and high energy structures. We then compare the properties of favourable and unfavourable structures with experiment and finally calculate the Boltzmann factors to arrive at a small number of structural motifs that we envisage being the ground state ensemble. With this information, we can then begin to establish criteria that allow us to distinguish between favourable and unfavourable configurations without the need for full geometry optimisation. Similarly, we attempt to establish whether particular configurations are therefore unlikely to be present in a real material.

As mentioned earlier, we have considered the zeolite Mordenite. A $2 \times 1 \times 1$ unit cell has been used to give sufficient T sites to allow us to simulate a Si:Al ratio of 11, chosen to allow direct comparison with experimental data [21,22] and also the initial study of Gray et al. [5] and Gray [6], allowing us to determine whether larger sampling is of importance. As mentioned earlier we will use the results obtained in this study to improve our understanding of the factors relating to structure. Therefore we have currently used only one constraint, i.e. that Lowenstein's rule should be obeyed during the MC configuration generation stage. Equally, as we wish ultimately to improve the efficiency of such calculations, both through improved configuration generation and by the use of strategies and drip feeding to improve overall calculation time, here we present an initial sample of results that provide the baseline with which we will be able to compare and contrast with in future work to assess the improvements in methodology.

We have therefore performed full geometry optimisation calculations on 50,000 configurations. With the data for these configurations stored in the database we are able to make detailed analysis of the structural features and compare many properties. As mentioned earlier we can easily screen unfavourable configurations from our final analysis. Approximately 25% of the configurations did not converge during geometry optimisation. Failure to converge is usually a consequence of the initial geometry being too strained for a simple optimisation procedure.

Considering the lattice energy as the primary indicator of structural stability, we observe that distinct steps (Fig. 3, top) of approximately 3 eV are formed along the profile. Furthermore we observe a correlation between cell volume and lattice energy: the structure becomes unstable as the unit cell volume increases, which we shall discuss below in terms of the location of cations. *However, such a clear trend is not observed when a more limited sample is constructed* (see Fig. 3, bottom). We have also considered a number of other geometric properties – e.g. the cation–framework oxygen bond length – and find similar steps in these properties, which correlate directly with the steps in lattice energy. Thus it is clear that we have distinct structure–stability correlations that are only observed when such large samples are considered.

Analysing a number of structures from each of the plateau between the steps it is clear that if the cations are placed in large channels there is little change in the cell volume, an example of which is shown on the upper left hand side of Fig. 4. However, if the MC generated structure has cations in the smaller cavities within the framework then they are trapped and therefore the lattice must expand to accommodate them, straining the Si–O and Al–O bonds and leading to a decrease in stability. The upper right hand side of Fig. 4 clearly shows the cation induced deformation

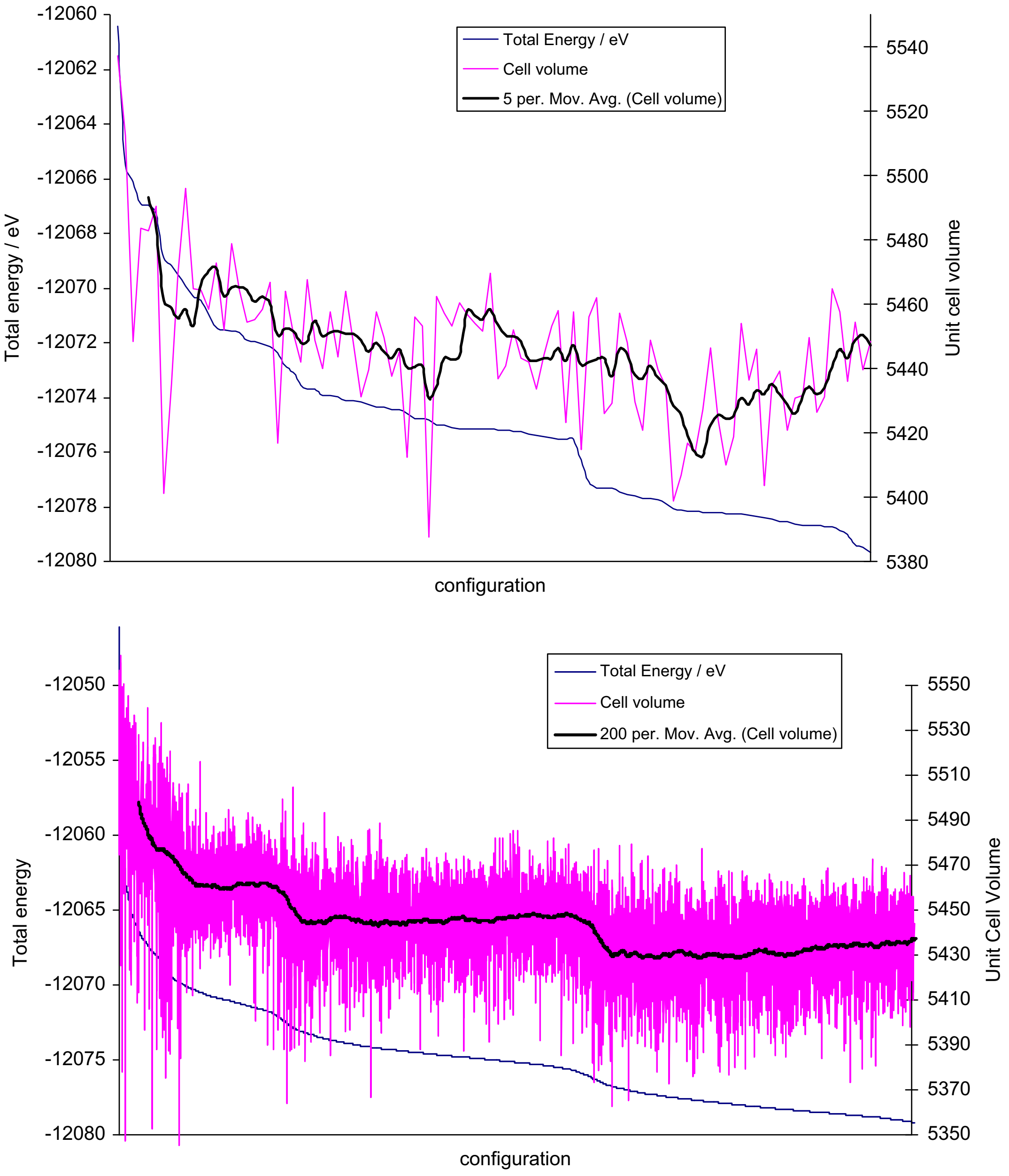


Fig. 3. (top) Cell volumes and total energies taken from the analysis of data from 100 configurations. (bottom) Results from 10,000 configurations with again cell volumes and total energies displayed. In both cases the configurations are ordered with reference to the total energy, the most stable structures being on the right hand side of both figures. Appropriate moving averages are applied to aide the eye.

of the central channel, which is now oval. Thus, we may now modify our structure generation algorithm to exclude these small cavities as viable sites.

From these initial results, we are able to provide crude selection criteria prior to further analysis. For example, we could exclude all those configurations with a lattice energy greater than the value at which we observe a significant step. Similar criteria can also be made for exclusion based on cell volume and average Al–Na distances (an indicator of cation coordination). Of course, it would be more useful to use such criteria to select configurations for further analysis. Again, such selection is facilitated by our use of a database for storing the results. Table 1 lists criteria by which we can construct a detailed analysis, a statistical sample of not only those configurations that we consider to be a good representation of the structure but also those that are less representative. Construction of such extreme samples will allow us to identify any specific geometric features that generate “favoured” or “unfavoured” configurations. Once identified these additional criteria can be introduced as constraints in our initial configuration generation algorithms.

We have then compared the calculated cation positions with experimentally determined [21,22] cation positions for a material at a similar ratio of Si–Al, although crucially the experimental cell is hydrated. 26.9% of the “favoured” configuration cations found matches with a match radius of 2 Å whilst only 15.1% of the “unfavoured” configurations did so. Close inspections of the experimental unit cell with those of a typical “unfavoured” configuration and a typical “favoured” configuration suggest that the match is even better than suggested as sites in the favoured cell where there is no match found are populated with water in the experimental cell and empty in the “unfavoured” configuration, indicating that an even more marked difference between the “favoured” and “unfavoured” configuration would be observed if the issue of hydration was overcome. Hydration effect can be modelled using a continuum

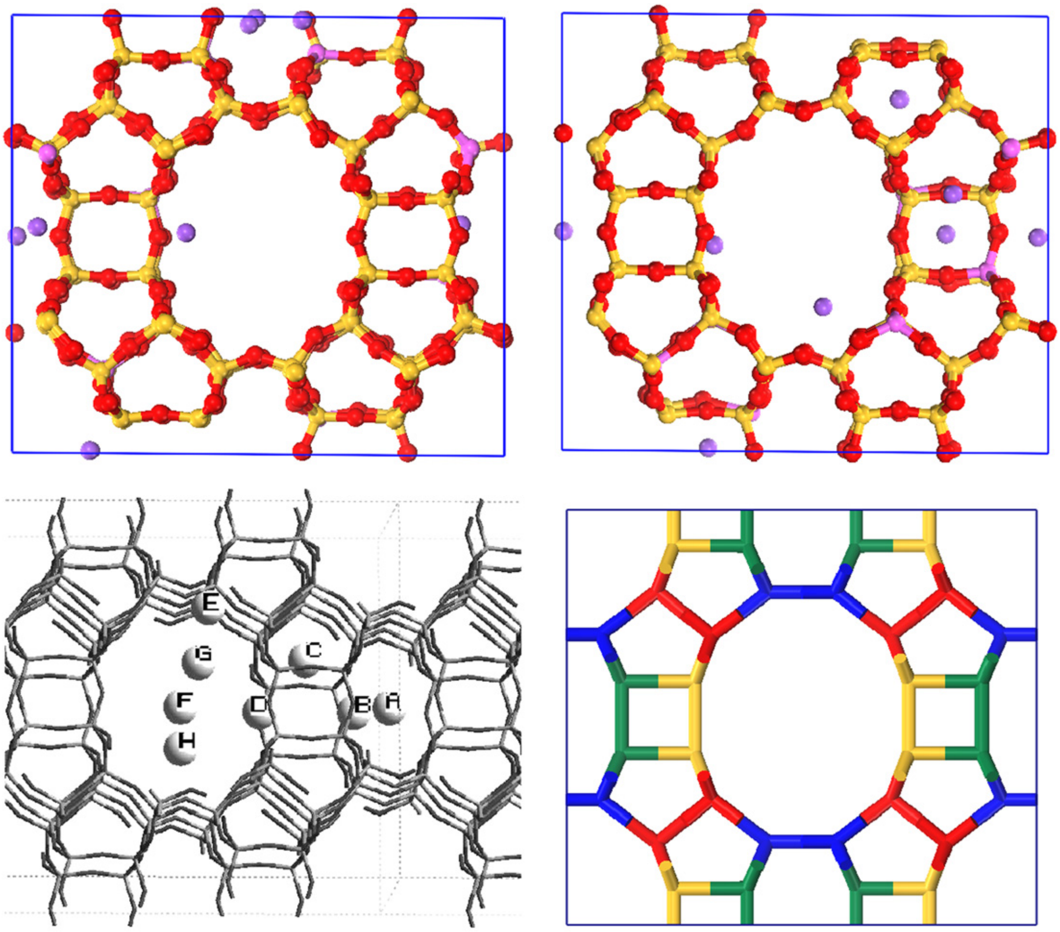


Fig. 4. Top Left—The unit cell of the lowest energy structure found in our study. Top Right—An "unfavoured" structure chosen at random from the second step of Fig. 3. Bottom Left—Position of cations in experimental unit cell. Bottom Right—A map of the four different T sites of Mordenite [21]. Blue = Si1, red = Si2, green = Si3, yellow = Si4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Comparison of properties for favoured and unfavoured configurations.

Property	Favoured	Unfavoured
Lattice energy (eV)	$\leq -12{,}070$	$\geq -12{,}070$
Al–Na average distance (Å)	>3.4	<3.4
Cell volume (Å^3)	<5420	>5475
Average cation–O distance (Å)	>2.65	<2.65

model of hydration as illustrated by the case of Goosecreekite [23], where such a model correctly predicted order aluminium distribution or explicitly, although the latter requires considerable care and additional experimental input.

Fig. 5 shows that the "unfavoured" configurations tend to have a much lower number of cation matches with experiment than the "favoured" configurations. We can also see a strong correlation between the energy and the number of cation matches in the "favoured" configuration plot. The number of matches reduces to close to that for the unfavoured matches at higher energies (in the "favoured" plot), which is at approximately the position of the first step in Fig. 3, bottom, indicating that any higher energy configurations may not be of any interest for further refinement.

We can further reduce the number of "favoured" configurations by performing a Boltzmann population analysis. The energy of each of the configurations is calculated relative to the lowest energy configuration. These relative energies were then used to determine a probability of the configuration being populated at a temperature of 473.15 K (chosen as typical of synthesis conditions) by way of a Boltzmann distribution:

$$P=[\exp(-E_{\mathrm{rel}}(n)/kT)]/\sum\exp(-E_{\mathrm{rel}}(n)/kT$$

$$E_{\mathrm{rel}}(n)=E_n-E_G$$

where E_G is the lowest energy configuration.

Those configurations that had a probability greater than 5% were then further analysed as they were considered to be structures most likely to correspond to those synthesised experimentally. The result obtained for the cation match for these structures was much greater than the average of all the "favourable" structures, 41.7% and 22.9% cation matches per unit cell, respectively. We also saw an improvement in the agreement of the NMR data to that of experiment, both with respect to the complete set of structures that have converged with an energy lower than −12,073 eV and the calculations carried out by Gray et al. [5] and Gray [6] on a smaller set of data. Agreement with T-site occupation is reasonable, see Table 2, but we would not anticipate reproducing the experimental data more accurately as the models considered are conceptually different, as now discussed.

As we are now sampling quite extensively, we might expect disparity between calculation and experiment to reduce, and indeed the differences in T site occupation between experiment and calculation are now reduced compared to Gray et al. [5] and Gray [6]. However, first we must ensure that we are comparing similar quantities. Our calculations show that T site occupations

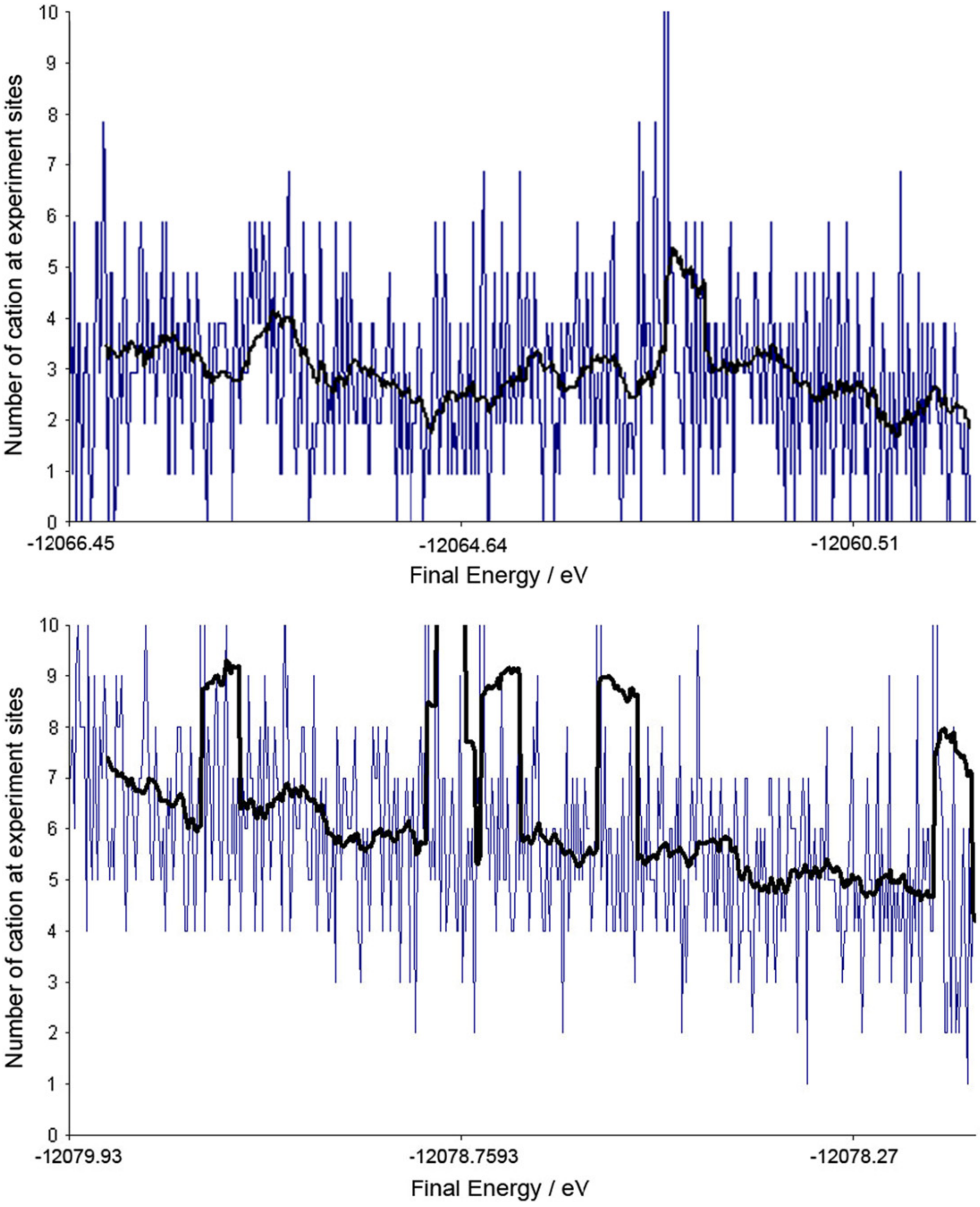


Fig. 5. Comparison of calculated sodium positions with those from experiment for (top) unfavoured and (bottom) favoured configurations. Note that most stable configuration in each is at the left side of the *x*-axis. It is clear that the number of cations located at experimental sites is higher in the more stable configurations.

of the top six structures vary substantially: each of the configurations has a calculated population at 200 °C of between 5% and 14%. If these six structures are all populated in experiment, we see an averaging in the experimental T-site occupation. Thus lack of agreement with experiment for these parameters is not necessarily indicative of poor methodology, rather the experimental data cannot be replicated computationally as the experiment and calculations consider different scenarios. In the crystal structure, the configuration is determined by the average of all configurations populated at the conditions imposed, whilst the calculations consider each configuration exclusively. The cation positions are therefore "smeared" in experiment, while in calculations we do not allow the possibility of unit cell to unit cell content change as would be anticipated if many configurations could be populated. However, the main differences that remain are likely to be a limitation of the model considered here. We have constructed and calculated the properties of *anhydrous* unit cells: crystals of mordenite form in the presence of and contain significant quantities of water. Including water would make the process significantly more computationally intensive. However, some of us have shown how approximate models of water – modelled only as a dielectric screen – can allow aluminium distributions in zeolites to be accurately obtained [23]. Such an approach may be valuable here.

By performing the calculations on 50,000 configurations we are now in the position to discount sampling as being the root cause of the discrepancy between experiment and the work of Gray et al. [5] and Gray [6]. It can be seen that we improve the agreement of the local structure between calculation and experiment with both the ensembles of configurations, labelled "−12,073 eV set" and "Boltzmann" in Table 2, but the improvement is small and the same trends persist with 4Si0Al being slightly overpopulated in calculations compared to experiment at

Table 2
Comparison of NMR and T-site occupations for computed sets and experiment. The NMR notation refers to the T-site neighbours surrounding a given Si. Thus 3Si1Al refers to a Si site which has as nearest neighbour T-sites, 3 silicon an 1 aluminium.

Quantity	Experiment (%) [22,23]	Gray et al. [5] and Gray [6] (%)	Lowest energy set[a] (%)	Boltzmann (%)
4Si0Al	63.3	70.6	68.2	67.8
3Si1Al	34.2	25.6	28.4	28
2Si2Al	2.5	3.7	4	3.8
1Si3Al	0	0.1	0.1	0.4
T-site 1	13	12	13	14
T-site 2	12	11	11	10
T-site 3	29	26	24	34
T-site 4	21	26	16	13

[a] The lowest energy set are those configuration with a lattice energy lower than −12,073 eV discussed in the text.

the expense of the 3Si1Al configuration. When we turn to the long range structure as shown by XRD experimentally and defined here as the T site occupation we can see that the best agreement between calculations and experiment is that of the 12,073 eV set with both our new sites correctly favouring T3 over T4, which was not observed in the study of Gray et al. [5] and Gray [6]. We now correctly order the T site occupation with respect to experiment although T4 is out by 8%.

The other main features noted to be common in all of the six lowest energy configurations are the positions of the cations, which exclusively populate the larger cavities of the mordenite framework. Structures with lattice energies only slightly higher than these configurations showed some cations located in the (experimentally less populated) 8-membered ring channels while, in contrast, the high energy configurations showed an even greater tendency for the cations to be found at these sites.

4. Discussion and conclusions

It is of interest to compare the structural properties that make up the 6 lowest energy structures and compare to those of the unfavoured set, shown in Table 3.

While we certainly see trends in properties the analysis of certain configurations picked at random from amongst the most stable configurations that would suggest that there are certain ranges of properties that can be imposed on the starting configuration, we must be careful not to avoid discounting possible configurations of interest. The complexity of the system means that local distortions can be accommodated and still lead to population of configurations with individual descriptors outside the ranges shown in our trends.

Before we can really influence the configuration generation with knowledge learnt in these studies we also need to step back to the starting structure. Again, it will be counterproductive if we eliminate a configuration from our ensemble because of properties in the starting structure that then change dramatically in the course of the simulation. It is, however, very clear that all stable structures have cations in the large channels and *not* in the 8-membered rings. We are therefore currently adding further functionality to allow the exclusion of configurations where there are a large number of cation–oxygen contacts, characteristic of a cation in a confined position. Fixing cation position to be in large rings is a clear parameter and there is no question of cutoffs or ranges of values with the arbitrariness that this brings. The data we have shows that if the configuration is stable then the cations are only found in the larger rings, in agreement with experiment.

Table 3
Comparison of structural properties for the 6 lowest energy structures and "unfavoured" structures.

Property	Average of 6 lowest energy structures	Unfavoured
Lattice energy (eV)	−12,079	≥ 12,068
Al–Na average distance (Å)	3.5	< 3.4
Cell volume (Å^3)	5461	> 5475
Average cation–O distance (Å)	2.69	< 2.65

The approach adopted in this paper has general implications: we have developed an understanding of the important structural constraints imposed on the structure using tools, both software and hardware, at our disposal and deploying them in a novel MC/EM hybrid scheme, allowing us to cover a vast phase space. As the deployment is over previously unused low specification compute resource, it provides a very low budget method of simulating in excess of 50,000 unit cells (each of which consists of 296 atoms) and will provide an effective procedure for modelling materials exhibiting partial site occupancy—a long standing challenge in solid state modelling.

Acknowledgments

We acknowledge the Nuffield Undergraduate Bursary scheme for funding RC. This work was also part funded by EPSRC grant code GR/R97207/01. We would like to thank Prof. J. Brodholt and the team for the support with the Condor pool at UCL.

References

[1] J.A. Purton, J.D. Blundy, M.B. Taylor, G. Barrera, N.L. Allan, Chem. Commun. 5 (1998) 627–628.

[2] R. Grau-Crespo, S. Hamad, C.R.A. Catlow, N.H. de Leeuw, J. Phys. C 19 (2007) 256201.

[3] A.R. Ruiz-Salvador, A. Gomez, D.W. Lewis, C.R.A. Catlow, L.M. Rodriguez-Albelo, L. Montero, G. Rodriguez-Fuentes, Phys. Chem. Chem. Phys. 2 (2000) 1803–1813.

[4] A.R. Ruiz-Salvador, D.W. Lewis, J. Rubayo-Soneira, G. Rodriguez-Fuentes, L.R. Sierra, C.R.A. Catlow, J. Phys. Chem. B 102 (1998) 8417–8425.

[5] A.E. Gray, A. O'Brien, D.W. Lewis, Stud. Surf. Sci. Catal. 135 (2001) 11–19.

[6] A.E. Gray, Ph.D. Thesis. University of London, 2003.

[7] W.G. Richards, Nat. Rev. Drug Discov. 1 (2002) 551–555.

[8] 〈www.mysql.com〉.

[9] J.B. Jones, Acta Crystallogr. B 24 (1968) 355–358.

[10] A. Alberti, G. Gottardi, Z. Kristallogr. 184 (1988) 49–61.

[11] A. Alberti, G. Gottardi, T. Lai, The Determination of (Si,Al) Distribution in Zeolites, in: D. Barthomeuf, E.G. Derouane, W. Holderich (Eds.), Guidelines for Mastering the Properties of Molecular Sieves, Plenum Press, New York, 1990, p. 145.

[12] W. Lowenstein, Am. Mineral. 39 (1954) 92.

[13] W.J. Mortier, Compilation of Extra Framework Sites in Zeolites, Butterworth & Co., Guildford, 1982.

[14] J.D. Gale, J. Chem. Soc. Faraday Trans. 93 (1997) 629–637.

[15] J.D. Gale, A.L. Rohl, Molec. Sim. 29 (2003) 291–341.

[16] R.A. Jackson, C.R.A. Catlow, Molec. Sim. 1 (1988) 207–224.

[17] M.J. Sanders, M. Leslie, C.R.A. Catlow, J. Chem. Soc. Chem. Commun. 19 (1984) 1271–1273.

[18] R.P. Ewald, Ann. Physik 64 (1921) 253–287.

[19] M. Litzkow, M. Livny, M. Mutka, Proceedings of the 8th International Conference on Distributed Computing Systems, 1988, pp. 104–111. 〈www.cs.wisc.edu/condor/〉.

[20] Chapman P. Wilson, T. Tannenbaum, M. Farrellee, M. Livny, J. Brodholt, W. Emmerich, Proceedings of the UK E-Science All Hands Meeting, 2004, pp. 870–878.

[21] W.M. Meier, Z. Kristallogr. 115 (1961) 439–450.

[22] K. Shiokawa, M. Ito, K. Itabashi, Zeolites 9 (1989) 170–176.

[23] A.R. Ruiz-Salvador, N. Almora-Barrios, A. Gomez, D.W. Lewis, PCCP 9 (2007) 521–532.