A Threshold-Minimization Scheme for Exploring the Energy Landscape of Biomolecules

Neelamraju, Sridhar; Johnston, Roy L; Schön, J Christian

DOI:
10.1021/acs.jctc.6b00118

Document Version
Peer reviewed version

Citation for published version (Harvard):
Neelamraju, S, Johnston, RL & Schön, JC 2016, 'A Threshold-Minimization Scheme for Exploring the Energy Landscape of Biomolecules: Application to a Cyclic Peptide and a Disaccharide', *Journal of Chemical Theory and Computation*, vol. 12, no. 5, pp. 2471-9. https://doi.org/10.1021/acs.jctc.6b00118

A threshold-minimization scheme for exploring the energy landscape of biomolecules: Application to a cyclic peptide and a disaccharide

Sridhar Neelamraju,*,†,¶ Roy L. Johnston,† and J. Christian Schön‡

School of Chemistry, University of Birmingham, Edgbaston, B15 2TT, Birmingham, and Max Planck Institute for Solid State Research, Heisenbergstrasse 1, D-70569, Stuttgart

E-mail: s.neelamraju@soton.ac.uk

Abstract

We present a scheme, called the threshold-minimization method, for globally exploring the energy landscape of small systems of biomolecular interest where typical exploration moves always require a certain degree of subsequent structural relaxation in order to be efficient, e.g. systems containing small or large circular carbon chains such as cyclic peptides or carbohydrates. We show that using this threshold-minimization method, we can reproduce not only the global minimum and relevant local minima but also overcome energetic barriers associated with different types of isomerism for the example of a cyclic peptide (cyclo-$(Gly)_4$). We then apply the new method to a disaccharide (α-D-Glucopyranose-1-2-β-D-Fructofuranose) and report energetically preferred configurations, barriers to boat-chair isomerisation in the glucopyranosyl-ring and discuss the energy landscape.

*To whom correspondence should be addressed
†School of Chemistry, University of Birmingham, Edgbaston, B15 2TT, Birmingham
‡Max Planck Institute for Solid State Research, Heisenbergstrasse 1, D-70569, Stuttgart
¶New Address: University of Southampton, Highfield Road, SO17 1BJ, Southampton

Introduction

Exploring the potential energy landscapes of biomolecules is of fundamental interest, and a number of global optimization and exploration methods have been introduced to facilitate the identification of the local minima of such systems.[1–10] The success of these methods often relies on employing clever ways for the generation of new conformations during the landscape exploration, combined with a local minimization of these structures. However, most global optimization algorithms do not permit the identification of energy and entropy barriers separating the various individual minima or basins, which are important for the kinetic stability of the conformations. To obtain information about such barriers, one needs to employ another class of algorithms such as the Nudged Elastic Band method[11] or other variants,[12–19] which produce some of the saddle points between the minima together with the corresponding energy barriers, or the threshold algorithm[20,21] that generates both energetic and entropic barriers plus the various minima that can be reached from a given minimum below an energy threshold. By repeating the threshold procedure for a large number of energy lids and using every minimum found as a starting minimum, we can also perform a global exploration of the landscape,[21] yielding both local minima, energy barriers and entropic barriers from the probability flows on the landscape.[20,22]

In the past, the threshold algorithm has been successfully employed to study the energy landscape of ionic clusters,[23,24] metallic clusters[25,26] and ionic solids.[22,27–31] However, the types of move employed during these threshold explorations have focussed on the displacement of individual atoms (plus changes in the variable simulation cell), in order to obtain useful barrier estimates. While these moves can also be employed for biologically relevant molecules, of course, they would not be computationally efficient, and it should pay to employ or develop more complex moves that are more suitable for threshold-type explorations of such molecules.

In this study, we describe a variant of the threshold algorithm that employs complex moves which allow an efficient landscape exploration, while at the same time obtaining reasonable barrier estimates. These so-called jump moves are a combination of multi-atom displacements inside the molecule plus short fast relaxations that remove only the largest unfavourable energy contributions after the jump, but are not full local minimizations. The threshold-minimization method is applicable to a large variety of biomolecular systems that can be described using force fields that allow gradient based minimizations,[1], such as e.g. the AMBER force-field.[32,33]

As example systems, we choose biologically relevant molecules, a cyclic peptide and a carbohydrate (disaccharide), which can contain rings of five, six, seven, eight or more atoms. The cyclic nature of these molecules reduces conformational flexibility and the complexity of the underlying landscape when compared to the corresponding acyclic molecules. Conformational searches based on a three-dimensional sweep of the torsion angles has been the preferred strategy for both cyclic peptides and carbohydrates.[34–36] We provide a short description of the methodology, followed by its application to the study of cyclo-$(Gly)_4$ and α-D-Glucopyranose-1-2-β-D-Fructofuranose and compare our results with previous studies in the literature.

2. Method

In order to be able to employ the threshold algorithm, it is useful to first identify some low-lying minima[2] that can serve as a first set of starting points for the threshold runs. For this purpose, we have employed a stochastic simulated annealing scheme with jump moves (see below) combined with periodically quenching the walker into a local minimum along the

[1]Note that the threshold algorithm as such does not require the availability of gradients - it is a stochastic exploration tool using only the energy as criterion.

[2]Of course, we could directly start with a local minimization from a randomly chosen configuration to identify a local minimum as the first starting point,[37] but it is more efficient to find some low-lying minima using a standard global optimization procedure like simulated annealing or basin-hopping.[1]

simulated annealing trajectory. In the main stage, we now employ the threshold algorithm to find the remaining local minima plus the effective barriers between them. Both the threshold algorithm and the jump move simulated annealing algorithm have been implemented in the G42+ code.[38] For the computation of the energy and the gradient minimizations in the case of organic (bio)molecules, we have added an AMBERTOOLS[33] interface to the G42+ code.[3]

2.1 Monte Carlo Simulated Annealing

We perform a short global optimization on the energy landscape using stochastic simulated annealing (SA). As usual, each SA run starts from a pre-defined conformation. Next, starting from a particular energy-minimized structure of the molecule of interest, i with energy E_i, the configuration i is modified by applying a deformation move, leading to a trial configuration j, which is accepted or rejected based on the metropolis criterion.[39] At each step of the annealing procedure, we perform either a small deformation move or a jump move (see moveclass section) that includes a short gradient descent using MINAB,[33] an implicit solvent minimization program written in NAB.[40] Note that these short gradient descents usually do not reach a local minimum, they just remove the energetically most unfavourable aspects of a configuration after the jump move (e.g. large overlaps of neighbouring atoms, etc.).

In order to identify a large set of local minima, we periodically stop along the trajectory of the SA run and perform multiple long random quenches into the nearby local minima. Note that during these quenches, no jump moves are performed. This yields a number of possible minima that serve as input for the threshold algorithm. The global search for minima during the SA-stage does not need to be exhaustive since the subsequent study via the threshold-minimization algorithm identifies most of the relevant low-energy local minima. The simulation parameters are listed in Table 1.

[3]Alternatively, one can employ various empirical potentials directly implemented in G42+ or ab initio (or DFT) energies via interfaces to quantum chemistry codes.

2.2 Threshold-Minimization algorithm

Starting points for the threshold exploration of the landscape are the minima obtained from the simulated annealing stage. For each local minimum, a series of thresholds slightly above the global minimum are chosen. Then, a random walk is performed in configuration space, restricting the total energy of the system after each move to be below the specified lid value. Else, the move is rejected. In the case of the threshold-minimization algorithm considered here, each move during the threshold part of the run is a jump move followed by an abbreviated gradient "minimization", i.e. we perform only a small number of gradient-based downhill steps to remove the most unfavorable aspects of the new configuration but do not reach a local minimum. Periodically, multiple long stochastic quenches are performed from holding points (usually with relatively high energies close to the threshold) along the walker's trajectory in configuration space, in order to check whether a walker has succeeded in reaching a neighbouring minimum while staying below the specified lid. It should be noted that the threshold-criterion is applied to the energy after the gradient-descent stage of the jump move. In this way, we ensure a thorough and efficient sampling of the low-energy conformations of the potential energy surface around the starting minimum, and reliable minima are obtained after the quench despite the simple nature of MINAB,[40] the inbuilt gradient-descent method in AMBERTOOLS.

On raising the lid beyond a certain threshold, different minima are obtained after the minimization step. The threshold-minimization algorithm ensures that the low-lying regions of the energy landscape are well sampled, and reliable local minima are obtained after multiple quenches (see Figure 2). Specifically, we employ only short local gradient descent runs instead of full minimizations, because we want to identify all major minimum basins in a fast and efficient way, without losing our ability to gain estimates on the relevant barriers and probability flows. However, there is a price to be paid: the landscape we obtain is only approximate, of course; for precise information on energetic barriers, one must perform only single point energy calculations during the random walk involving only small deforma-

tion moves. Nevertheless, the threshold-minimization scheme yields a useful quantitative overview of the large-scale features of the barrier structure of the landscape, together with the dominant minimum basins. The simulation parameters for the two molecules studied here are listed in Table 1.

2.3 Moveclasses

Both the global optimization and the threshold-minimization runs were performed without employing any symmetry constraints. Two classes of moves were employed during the explorations: Small local deformations involving only one or a few neighbouring atoms, and large jump moves affecting many atoms of the molecule. The jump move employed was a deformation of the flexible molecule by selecting two non-bonded atoms A and B within the ring, and then rotating one of the (possibly several) connected sets of atoms that lie between atoms A and B about the axis defined by these two atoms. The jump moves are always followed by a few downhill steps along the gradient, in order to remove the most unfavourable aspects of the new trial configuration. Here, the set of atoms A and B participating in a jump move can be pre-defined in order to make the moveclass even more efficient for more complex molecules. For e.g. one could include only the atoms belonging to a peptide bond in the case of peptide polymers or a protein, or only the glycosidic bond in the case of carbohydrates, etc. In the examples considered here, we let all atoms in the molecule take part in the deformation move.

2.4 Potentials

The AMBER ff03 force field[32,41] was used to evaluate energies at every step for the cyclic peptides. For the disaccharide, the GLYCAM06[42] force field compatible with AMBERTOOLS14[33,40] was employed. Solvent effects were modelled using the GB/SA implicit solvation method.[43] Water was the only solvent considered for both molecules during the threshold runs, though it is possible to re-minimise the structures for different values of the dielectric

constant. The force-field topology was generated using the tLEAP program[44] available with the AMBERTOOLS suite. Here, gradient descent minimizations are performed after each energy calculation within the threshold run with the MINAB[33] or FFGBSA[33] minimization routines implemented within AMBERTOOLS.[33] By default, the cut-off for non-bonded interactions (vdW and Coulomb) is set to 100 Å and that for the Generalized-Born approximation is fixed at 15 Å. For more information on default parameters, the reader is referred to Ref.[33]

Table 1: Simulation parameters for a simulated annealing and threshold-minimization runs for the cyclic peptide and the disaccharide molecules.

Simulation parameters	Value
Number of Monte Carlo steps per SA run	10^6
Number of independent SA runs	5
Initial temperature T_{init}	12, 000 K
Final temperature T_{final}	0.006 K
Holding points per simulated annealing run	100
Holding points per threshold run	50
Quenches per holding point	5
Steps per quench	10,000 steps
Number of minima generated per SA run	500
Number of minima generated per TM run	250
Minimum distance between atoms allowed during moves	0.55Å

3. Examples

3.1 Energy landscape of cyclo-[Gly]$_4$

There exist many cyclic peptides with diverse biological activities, including immuno-suppressive, anti-fungal and anti-bacterial properties.[45] To a large extent, the energy landscapes of cyclic peptides account for their conformational dynamics and provide some insight into their biological activity.[46,47] Thus, one way to better understand structural changes that occur in cyclic peptides as a function of amino-acid substitutions is to predict stable isomers of various combinations of amino acids in various solvents and establish structure-property

relationships.[46,48] Furthermore, cyclic peptides are interesting test systems for landscape exploration methods, because their isomerism is well understood through experiment[49–51] and theory.[2,46,52]

The cyclic tetraglycine system, cyclo-$[Gly]_4$, was chosen for two main reasons. First, it is the simplest cyclic tetrapeptide, with achiral and symmetric carbon centres that reduces the complexity of the landscape significantly. Secondly, it is a system whose energy landscape has been previously studied in detail.[46,53] Thus, it is a good system to test the applicability of the threshold minimization method to (bio)molecules in general, and to cyclic systems in particular.

In the past, Lousieau et al.[52] studied the conformers of naturally occurring cyclic tetrapeptides such as the natural herbicide Tentoxin, the anti-biotic Chlamydyocin, the anti-tumour agent Trapoxin, and the anti-malarial Apicidin, among others. Here, possible conformers were systematically built using SYBYL.[54] The barriers to isomerisation of Tentoxin were studied by Pinet et al.[50] using two-dimensional proton NMR and restrained molecular dynamics simulations. A detailed analysis of the energy landscape including local minima and energetic barriers for cis-trans isomerisation using discrete path sampling[55–58] for various cylcotetrapeptides of alanine, glycine and proline combined in different ratios and stereochemistries was performed by Oakley et al.[46] Monte Carlo multiple minimum $(MCMM)^2$ searches were performed by Che et al.[53] for heterochiral dipeptides of chimeric amino acids with energies compared for AMBER, CHARMM[59] and OPLS-AA[60] force-fields. Finally, the multidimensional conformational potential energy hypersurface of cyclo-$(Gly)_3$ was studied at the ab initio level by Tosso et al.[61]

3.1.1 Simulated annealing and threshold minimization

Our investigation has followed the approach described above in the methods section. The parameters for the simulated annealing (SA) and threshold minimization runs for the cyclopeptide are listed in Table 1. The SA run starts from a known structure of cyclo-$(Gly)_4$. Since the moves for generating the next configuration are quite large and involve rotation about a pre-defined axis, a single low-energy structure is all that is required to start the global optimization.[4]

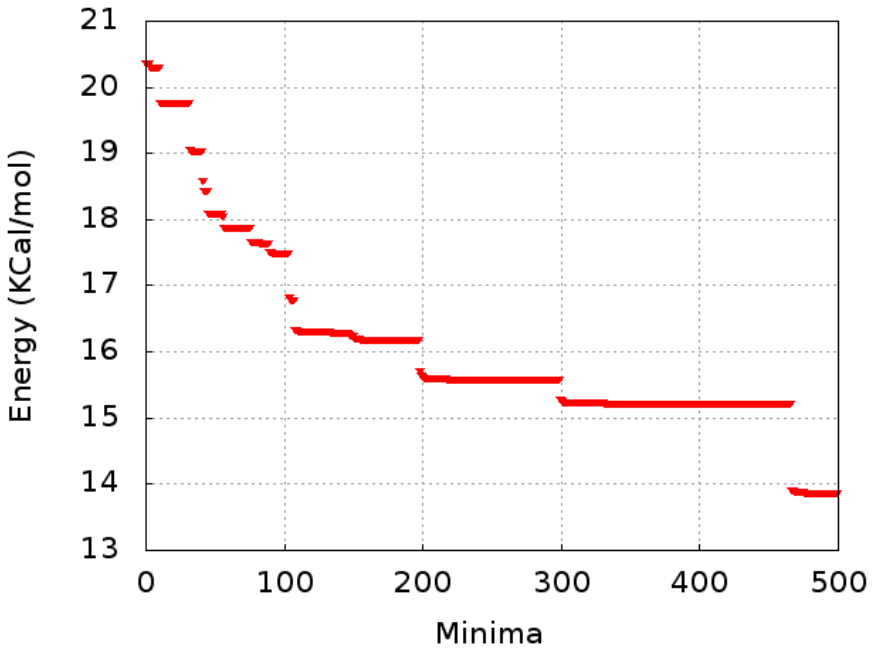

Figure 1: Energies of local minima found from one simulated annealing run on the cyclo-$(Gly)_4$ molecule. Note that some minima might have been observed several times..

The minima obtained from one SA run for cyclo-$(Gly)_4$ are shown in Figure 1. 24 different minima were obtained after the first SA run. Each minimum obtained through SA serves as a starting point for threshold-minimization (TM). New minima, below the specified threshold emerge. Since each energy calculation includes a short gradient descent minimization with MINAB, the holding points can sometimes nearly co-incide with other local minima as shown

[4]The set of atom pairs, triplets and quartets in the molecules needed for constructing the various moves and jump moves are derived from this starting configuration.

in Figure 2. However, since the MINAB gradient descent routine is strictly used only for structure refinement purposes, it is not very good at finding low-lying minima on a consistent basis.[33] Local minima identified after the quench are, therefore, typically lower in energy.

Threshold-minimization runs were performed, starting from the global minimum at different threshold values. These vary from 0.024 to 0.032 eV/atom in steps of 0.001 eV/atom and from 0.04 to 0.07 eV/atom in steps of 0.01 eV/atom. Figure 2 shows trajectories of three thresholds picked from the latter set (0.05, 0.06 and 0.07 eV/atom).

3.1.2 Minima and energies

Quite generally, a tetrapeptide can be described by up to four different sequences because the starting point for the sequence in a cyclic peptide is arbitrary. Furthermore, as cyclic peptides do not present N- and C- termini, the standard numbering of residues which is valid for linear peptides, does not hold. We use the naming scheme for cyclic peptides suggested by Loiseau et al.[52] Amino acids are numbered in the direction N- to C-edges, starting from the residue whose symbol is the letter that appears first in the alphabet. Conformational isomers of cyclopeptides can essentially be classified into two categories (cis-trans isomerism and orientational isomerism). Following Loiseau et al., the sequence is identified from the α-Carbon belonging to the first residue and follows the N- to C-terminal rule clockwise. Letters **C** and **T** are used to identify cis and trans conformations, respectively. The second type of conformational isomers are orientational isomers, i.e. they are categorized on the basis of the orientation of the peptide carbonyl groups, either upward or downward with regard to the mean plane of the ring as defined by the mean plane of Cα positions (depicted as the yellow plane in Figure 3), with the residues arranged clockwise. The letters **U** and **D** denote carbonyls which are pointing upwards and downwards with respect to this mean plane, respectively.

Following this scheme, the global minimum shown in Figure 3 obtained through threshold-

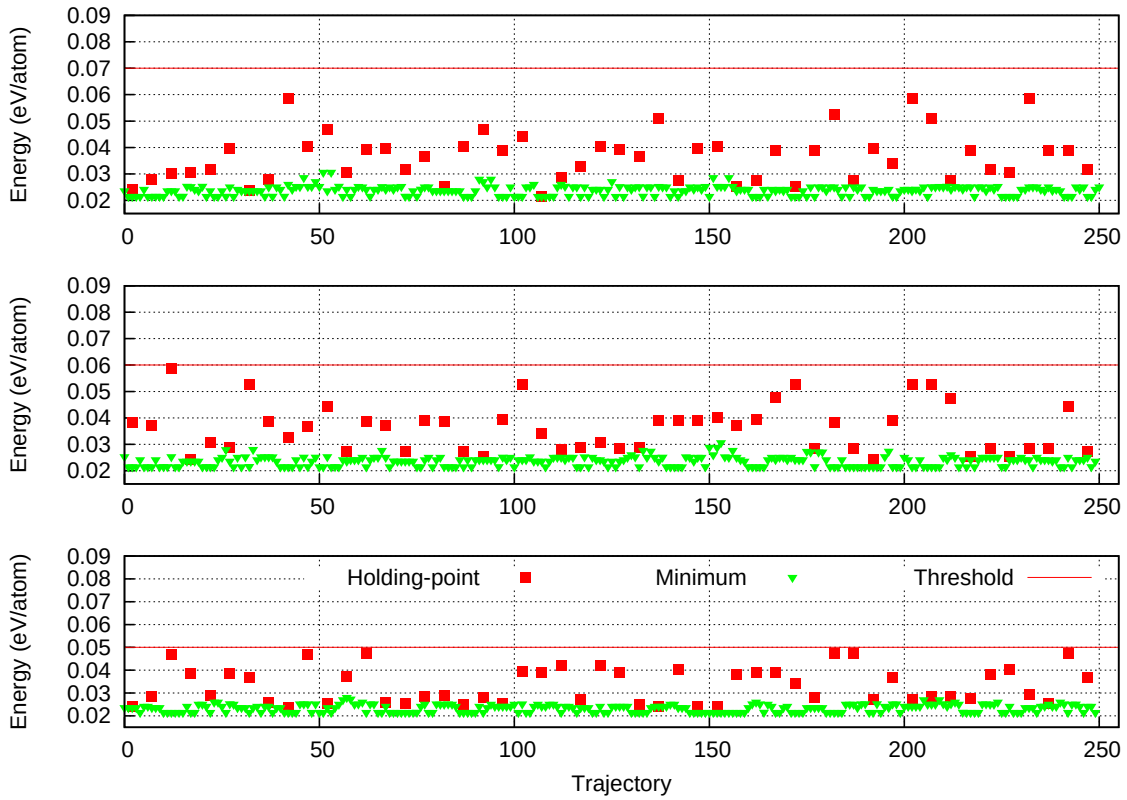

Figure 2: Threshold-Minimization runs starting from the putative global minimum structure located from simulated annealing. The red line indicates the threshold energy (0.05, 0.06 and 0.07 eV/atom), the red squares indicate the holding points and the green triangles indicate the minimum energy structures found after quenching from the holding points. Five quenches per holding point are performed.

minimization has the notation TTTT-UUUU. This was also identified as the global minimum in Ref.[46,52] The four lowest minima for the all trans cyclo-$(Gly)_4$ system identified through our scheme are depicted in Figure 3. The black atom in the figure is the α-Carbon of the reference peptide from where the numbering begins. The lowest energy isomers with one cis (TTTC-DDUD) and two cis peptide bonds (CTCT-UDUU) are also depicted.

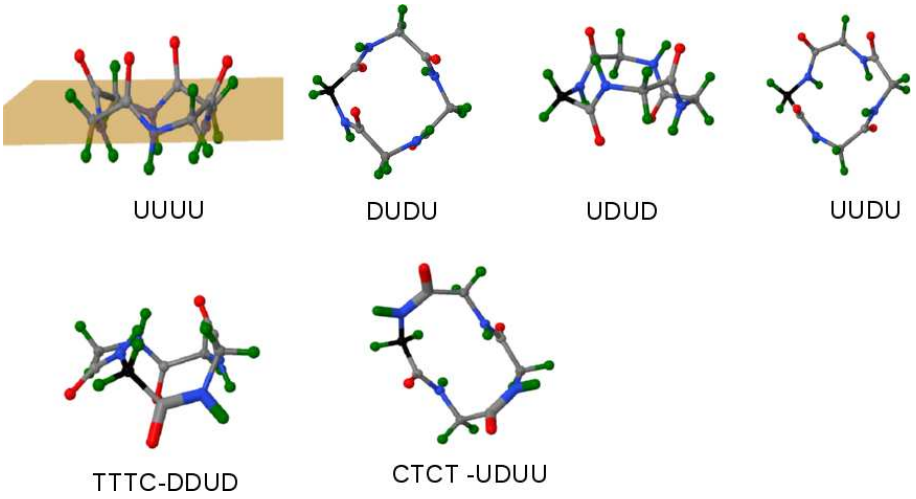

Figure 3: The four lowest minima identified for cyclo-$(Gly)_4$ in water using the SA-TM approach (top); the TTTT-component of the sequence name is dropped. Lowest energy isomers with one cis peptide bond (TTTC-DDUD) and two cis peptide bonds (CTCT-UDUU) (bottom). The black atom is the α-carbon from where the numbering begins.The slab in UUUU represents the average plane formed by the α-carbons.

Table 2: Orientational isomers for the all-trans (TTTT) peptide. (a=15.49 kcal/mol, b=16.14 kcal/mol and c=16.78 kcal/mol)

Isomer	Energy (kcal/mol)	Threshold
UUUU	13.869	a
DUDU	15.238	a
UDUD	15.239	b
UUDU	15.591	b
DUUU	15.591	b
DDDU	15.592	b
DUUU	15.592	b
UUUD	15.593	b
DUDD	15.594	c
UDDD	15.596	c
DDUD	15.604	c

The orientational isomers of the all-trans (TTTT) peptide in increasing order of energy

are listed in Table 2. A set of 250 minima, each obtained after four separate threshold-minimization runs at threshold of (a) 15.49 kcal/mol, (b) 16.14 kcal/mol, (c) 16.78 kcal/mol and (d) 20.01 kcal/mol, are depicted in Figure 4. The first threshold is set where we see the emergence of one new conformational isomer. The TTTT-UUUU and TTTT-DUDU isomers are the global minimum in water, and the global minimum in vacuum, respectively. At the second threshold value of 16.14 kcal/mol most of the orientational isomers appear with all peptide bonds in the trans form. The third threshold value of 16.78 kcal/mol (corresponding to 2.91 kcal/mole above the global minimum) is where we see the occurrence of the first cis peptide bond (TTTC-DDUD). For comparison, the barrier for the lowest cis-trans transition state in cyclo-$(Gly)_4$ was calculated as 16.4 kcal/mol in a previous study[46] where discrete path sampling calculations were employed. Finally, the energetic barrier for the two cis-peptide bonds (TCTC-DUUD) to emerge is breached at the threshold value of 20.01 kcal/mol (6.141 kcal/mol above the global minimum), i.e. another 3.31 kcal/mol above the barrier for the first cis-trans isomerisation.

This analysis clearly indicates that even though we partly minimize the energy at every step (through MINAB), we are able to efficiently cross energetic barriers and reproduce the minima and energy barriers found in the literature,[46] for the energy landscape of cyclo-$[Gly]_4$. Thus, we can be confident of the proposed methodology as a tool for the efficient exploration of the energy landscapes of (bio)molecules, and we can analyse the additional data provided by the threshold-minimization runs such as the probability flows and phase space volumes. This allows us to derive approximate estimates of basin sizes corresponding to each minimum and to understand how these basins interact with each other as a function of the threshold. Of course, for such estimates to become more quantitative, thorough sampling of the threshold runs from each orientational and cis-trans isomer is required.

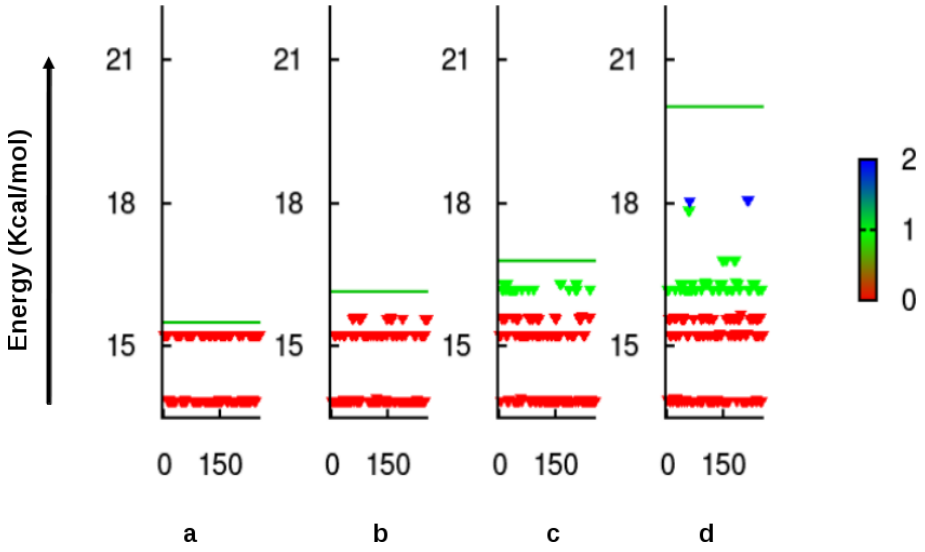

Figure 4: A set of 250 minima obtained after four separate threshold-minimization runs at threshold of (a)15.49 kcal/mol, (b) 16.14 kcal/mol, (c) 16.78 kcal/mol and (d) 20.01 kcal/mol. Structure with all trans (red), one cis (green) and two cis (blue) peptide bonds are shown.

3.1.3 Probability flows and analysis of holding points

Along with estimates of the energetic barriers, the threshold algorithm also provides important information about the stability of the minima and basins corresponding to potential energy minima by determining the likelihood of a given minimum configuration to transform to a new minimum configuration after a certain energy barrier is crossed. Of course, exhaustive sampling is required to draw firm conclusions.

Here, we summarize the probability flows between the minima containing all-trans peptides, one-cis peptide and two-cis peptides[5] in Figure 5. Each run starts from the lowest energy conformer for each type of cis-trans arrangement in the equivalent acyclic peptide (i.e. 1 with TTTT, 4 with TTTC, 2 with TCTC and 4 with TCCT). 250 minima per starting structure per threshold per minimum are generated. Therefore, each dot on the chart represents the results from one threshold-minimization run per minimum.

From Figure 5, we see that as the threshold is raised, new minima begin to be found. As

[5]Four possible acyclic arrangements of CTTT, TCTT, TTCT and TTTC are considered identical since the ring is cyclic. With two-cis peptide bonds six possible acyclic arrangements of CCTT, TCCT, TTCC, CTTC, CTCT and TCTC are reduced to two possibilities of: TCTC and TTCC. Thus a total of four possible cis-trans arrangements are possible at the threshold values considered.

noted above, the first cis peptide bond occurs at 16.78 kcal/mol (2.911 kcal/mol above the global minimum). As the threshold value is increased, the percentage of minima with cis peptide bonds increases from $\approx 10\%$ at threshold (a) to $\approx 30\%$

Furthermore, an analysis of the holding points reveals that there is a considerably greater number of holding points with cis peptide bonds that change to the trans form on quenching. Therefore the transition region that connects various types of all-trans conformers should contain isomers with one or more cis bonds. An example is provided in Figure 6 where we plot the trajectory of a threshold run at a threshold value of 19.34 kcal/mol. The holding points and minima are coloured according to the number of cis peptide bonds found. This is the highest threshold studied, where the minima exhibit at most one cis peptide. But even though no minima with two cis peptide bonds are found, there exist multiple holding point configurations with more than one cis peptide bond. These, however, relax to minima with only one cis peptide or no cis peptides upon quenching.

With the application of the threshold-minimization algorithm to the well-studied cyclo-Gly_4 system, we are able to reproduce known minima and energetic barriers. Additionally, we show that, we can extract probability flows between minima that give us an approximate idea of the size of the energetic basins the minima correspond to, and the height of the barriers separating these minima.

3.2 Energy landscape of α-D-Glucopyranose-1-2-β-D-Fructofuranose

We now apply the same methodology to study another class of biomolecules that contain cyclic rings: the carbohydrates, specifically the sucrose (α-D-Glucopyranose-1-2-β-D-Fructofuranose) molecule. Details of the potentials and parameters used are presented in the Methods section.

Neutron diffraction[62] and X-Ray[63] analyses of sucrose have shown that the glucopyranose

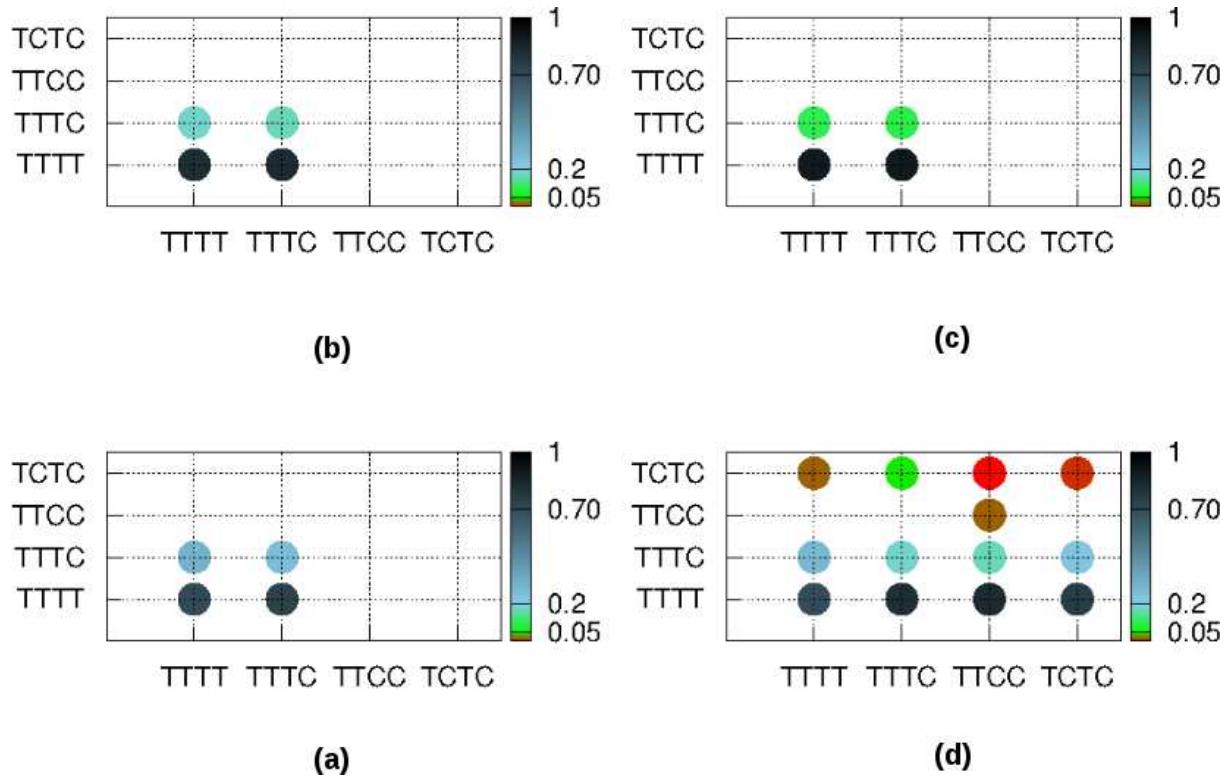

Figure 5: (a) Probability flow chart at four different threshold values of (a) 16.788 kcal/mol (0.026 eV/atom), (b)17.434 kcal/mol (0.027 eV/atom) (c) 19.37 kcal/mol (0.030 eV/atom)and (d) 20.01 kcal/mol (0.031 eV/atom). X axis: Starting configuration for the threshold runs. The lowest energy isomer for each of the four distinct arrangements of the peptide bond are chosen. Y axis: Final configuration obtained after threshold-minimization. Red indicates less than 5 %, green between 5 % and 20 % blue between 20 % and 70 % and black between 70% and 100 % probability flow from the starting configuration to the final configuration.

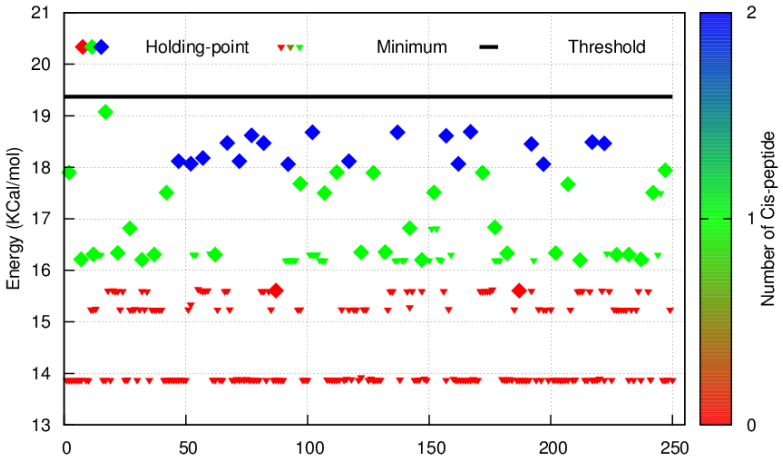

Figure 6: Single threshold-minimization trajectory at a threshold of 19.34 kcal/mol (5.47 kcal/mol above the global minimum). Holding points are indicated by diamonds, minima obtained after quench are indicated by triangles. The colour shows the number of cis peptide bonds in the structure. Red = all trans, green = one cis and blue = two cis peptide bonds. While a majority of the minima obtained (triangles) are in the all-trans (TTTT) conformation (red), the holding points (diamonds) can contain both one cis peptide bond (green diamonds) and two cis peptide bonds (blue diamonds).

ring adopts a $^4C_1$ conformation, and the fructofuranose residue adopts a $^4T_3$ twist form in water. The overall structure is determined by two inter-residue intra-molecular hydrogen bonds. The crystal structure of sucrose and its molecular conformers have been extensively studied experimentally using Raman,[64] solution X-Ray diffraction,[65] NMR[66–69] including Nuclear Overhauser Effect[68] (NOE) measurements. Finally, Marszalek et al.[70] performed single molecule force measurements on various single polysaccharide molecules under an atomic force microscope and found that the glucopyranosyl ring can switch from chair to boat-like or inverted chair conformations under an applied force of about 200 pN.

Herve et al.[68] suggested the existence of six isomers through a combination of NMR and molecular modelling. They also confirmed that the glucopyranosyl ring exists in a chair conformation in a dilute water solution. Immel et al.[71] studied three different isomers of sucrose molecules through molecular dynamics and umbrella sampling methods in aqueous solution and in vacuum and found three major conformers (called S1, S2 and S3). These suggested a very limited conformational flexibility of the torsion angle about the C-1^g-O-1^g bond in vacuum using the GROMOS force-field.[72] However, to our knowledge, a detailed energy landscape study that systematically discusses energetic barriers between conformers of sucrose has not yet been performed.

3.2.1 Minima and energies

Concerning the analysis of the isomers found, we have used the IUPAC nomenclature for five and six-membered ring forms of mono-saccharides and their derivatives was used. Six-membered rings are classified into chairs (C), Boats (B), Skew-boats (S) and Half-Chairs (H). Five-membered rings are classified into Envelopes (E) and Twists (T).[73]

First, we perform simulated annealing runs using the parameters listed in the Methods section. Energies of local minima obtained from these are depicted in Fig. 7. In addition to the already known structures, several new low-energy minima were identified, in particular

low-energy chair ($^4C_1$, $^1C_4$) and boat ($^{0,3}B$) conformations of the glucopyranosyl ring. Figure 8 depicts examples of a few low-energy isomers.

Multiple chair conformers can exist in water within an energy range of 2-3 kcal/mol. A few boat-up conformers are also lower in energy than the corresponding chair conformers. Ramachandran plots[74] (c.f. Fig. 10) for comparison with previous results of Immel et al. are also constructed. Most computational studies on the sucrose molecule[65,71] have considered such plots in considerable detail.

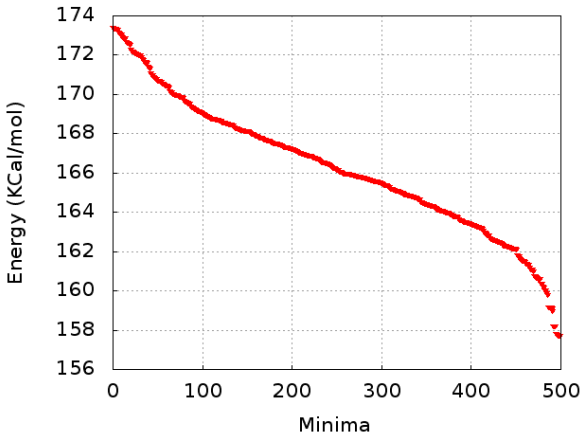

Figure 7: Energies of local minima found from one simulated annealing run on the sucrose molecule in water. The known global minimum was reproduced.

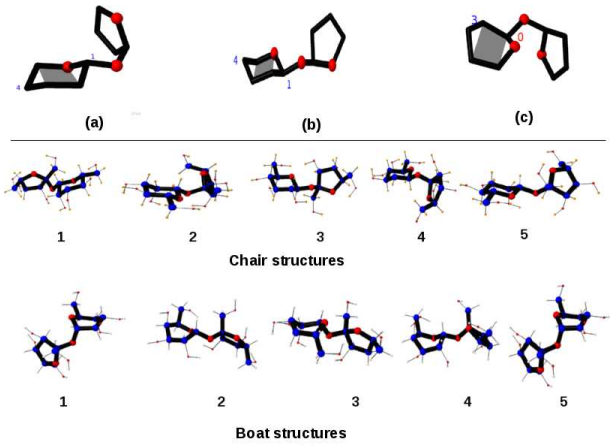

Figure 8: Low-energy conformers of the sucrose molecule found after MC/SA. (Top) The glucopyranosyl ring can take three forms: (a) $^1C_4$ (b) $^4C_1$ and (c)$^{0,3}B$. Only the ring skeletons are shown.(Bottom) Low-energy chair and boat conformations for the sucrose molecule.

After the simulated annealing (SA) stage, we performed threshold- minimizations taking

the lowest minima of the SA as starting points. A set of 250 minima were obtained after each threshold run. Three such runs are plotted in Figure 9 at threshold values of a) 159.8 kcal/mol, b) 161.88 kcal/mol and c) 166.03 kcal/mol (corresponding to 0.154, 0.156 and 0.160 eV/atom, respectively). The energy landscape of sucrose is clearly much more complex than that of cyclo-$(Gly)_4$, as can be seen from the much larger number of distinct minima obtained from each threshold run. In total, around 2500 minima were found for this system after threshold-minimization runs at seven different thresholds.

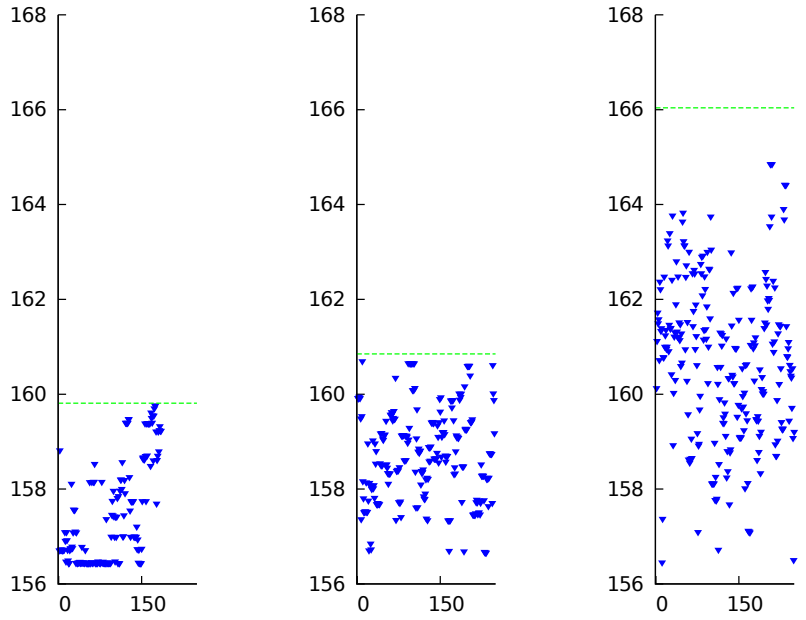

Figure 9: Threshold minimization runs performed with water as solvent at three different threshold values of a) 159.8 kcal/mol, b) 161.88 kcal/mol and c) 166.03 kcal/mol.

We can "automatically" determine whether a given glucopyranosyl ring adopts the chair or boat form by measuring the maximum distance d_{max} between the six atoms belonging to the ring ($d_{max} \approx$ 2.65 - 2.77 Å for the boat and $d_{max} \approx$ 2.85 Å for the chair form, respectively). This is useful since different atoms in the glucopyranosyl ring can reside above or below the reference plane of the chair configuration. The same holds true for the boat conformer. This maximum distance measure is plotted with respect to energy in Figure 11 for two different thresholds as Ramachandran plots.

From this figure, we can see that at a low threshold of 160.8 kcal/mol (4.2 kcal/mol above

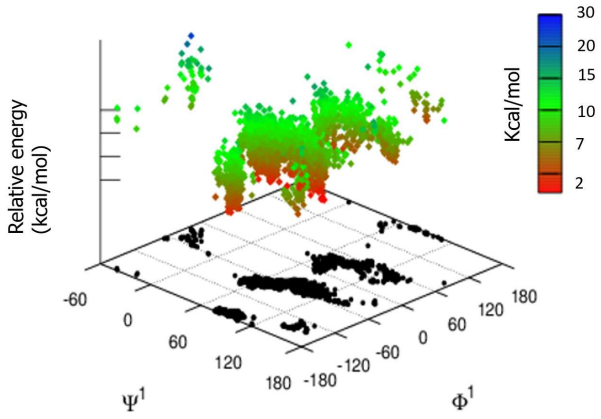

Figure 10: Energy landscape projected on the Ramachandran plot for 2500 minima obtained after threshold-minimization runs. Three clear regions that were designated S1, S2 and S3 by Immel et al.[71] are seen. X-axis: Φ^1 is the dihedral C2^f-O1-C1^g-O5^g, Y-axis: Ψ^1 is the dihedral O5^f-C2^f-O1-C1^g. ^{g} and ^{f} indicate atoms belonging to the glucopyranosyl and fructuranosyl ring, respectively. The z-axis gives the energy relative to global minimum.

the global minimum), all minima obtained are in the chair form. However, at the threshold above 166.01 kcal/mol (9.472 kcal/mol above global minima), many conformers in the boat conformation begin to appear quite clearly. On analysing the Ramachandran plot along the (O-5^g-C-1^g-O-1^g-C-2^f and C-1^g-O-1^g-C-2^f-O-5^f dihedrals about the glycosidic linkage, we see from figure 11.1a, that at the low threshold of 160.8 kcal/mol, only two specific regions ($(\phi, \psi) \approx (-60 \pm 5°, -40°)$, and $(\phi, \psi) \approx (60 \pm 5°, +40 \pm 2°)$, respectively) are populated at the lowest energies. At the higher threshold more high energy minima emerge. The ψ dihedral angle varies from $-180°$ to $+180°$ in steps of about $120°$, as can be seen in Figure 11.2a. This clustering of local minima about the two dihedrals was also observed by Immel et al.[71]

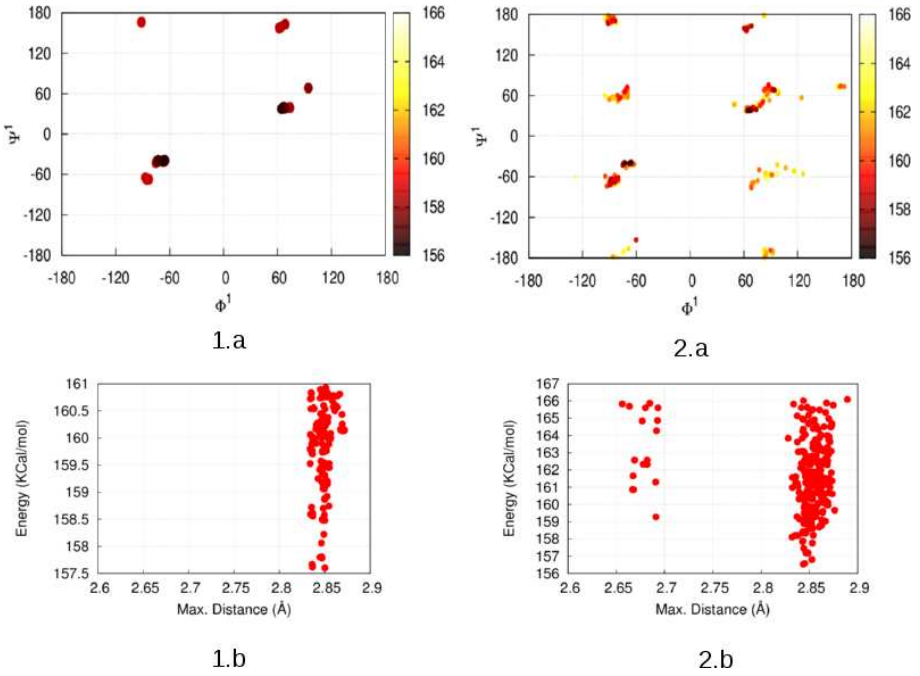

Figure 11: a) Ramachandran plots obtained for all minima found along two dihedral angles (ϕ^1 - O-5^g-C-1^g-O-1^g-C-2^f and ψ^1 - C-1^g-O-1^g-C-2^f-O-5^f). b) Maximum distance between six atoms belonging to the glucopyranosyl ring. The boat and chair forms can be distinguished at threshold values of 1) 160.8 kcal/mol (0.155 eV/atom) and 2) 166.01 kcal/mol (0.16 eV/atom).

Discussion and Conclusion

Using two examples of biomolecules, we have shown that we are able to sample the low-energy region of the energy landscape of such systems efficiently with the threshold-minimization algorithm. For the cyclic peptide cyclo-$(Gly)_4$, we are able to reproduce all the important minima and energetic barriers reported in the literature, and in addition estimate the probability flows among the major minimum basins on the energy landscape. Regarding the landscape of the sucrose molecule, we reproduce the earlier work and furthermore report additional minima such as two different chair conformations, $^1C_4$, $^4C_1$ and a boat conformation $^{0,3}B$ for the glucopyranosyl ring. The boat conformation in the low-energy region has not been reported in earlier studies that use potentials similar to ours. A thorough and systematic study of energetic barriers separating various conformers of the sucrose molecule has also not been performed to our knowledge. This will be a focus of investigations in the future.

The threshold-minimization method combines central features of the threshold algorithm - a careful exploration of the low-energy region of a complex energy landscape and estimation of probability flows on the landscape - with the advantages of the large moves on the landscape used in many global optimization techniques. We would like to note that we are letting all the atoms participate in the global optimization procedure, i.e. we did not preselect any atoms and atom groups used to define the jump moves. This is an additional option available in the current implementation of the algorithm, and thus there is still a large potential for better efficiency.

There is growing evidence that landscape exploration with judiciously chosen large moves followed by full local minimizations, is an efficient strategy for global optimization. The price paid is a loss of information about barriers, probability flows, and basin sizes. The threshold-minimization method aims to address this problem while still yielding the low-energy minima of the system. As a consequence, the jump moves employed need to be big enough to allow a fast sampling of a large section of the landscape, and yet must be small enough so that we

can extract approximate dynamics. In the current study, we achieve this by keeping the size of the new complex moves (mostly rotations about an axis connecting two non-neighbouring atoms, and rotations of pre-defined groups of atoms) just large enough to provide relatively efficient sampling, with acceptance rates of about 0.8 - 0.95 for the jump moves. Of course, this acceptance rate depends on the threshold used for a particular run. The optimal size or range of the move depends on the system in question.

Because of the short fast relaxations employed after every jump move, the barrier estimates and probability flows are not as accurate as they would be with the threshold algorithm, where single point energies are used. However, we are able to sample a larger section of the low-energy region of the landscape and we show that the barrier estimates and probability flows obtained for the systems studied here are reasonable starting points for a more involved investigation. Here, we note that while direct calculation of transition states with e.g. a doubly nudged elastic band approach[16] are very useful, they do not provide a quantitative estimate of probability flows and are not suited to estimate the size of basins surrounding local minima and of transition regions on the landscape.

Furthermore, various other parameters important for more involved energy landscape explorations using the threshold algorithm or discrete path sampling, e.g. an exhaustive set of local minima to start the runs from, upper estimates on the energetic barriers, the first (i.e. lowest) threshold value to consider, ideal resolution of the thresholds and an approximate idea of interesting features of the energy landscape become available from the threshold-minimization runs. Thus, threshold-minimization appears to be a suitable method for the efficient global exploration of the energy landscape of complex systems like biomolecules, since it allows us to combine global optimizations and barrier structure investigations.

Acknowledgements

SN would like to thank the German Academic Exchange Service (DAAD), Germany for providing a research grant for young academics under funding program 50015537. We also acknowledge the Engineering and Physical Sciences Research Council, UK (EPSRC) for funding under Programme Grant EP/I001352/1. The calculations described in this paper were performed using the University of Birmingham's BlueBEAR HPC facility (see http://www.bear.bham.ac.uk).

References

(1) Li, Z.; Scheraga, H. A. *Proc. Nat. Acad. Sci.* **1987**, *84*, 6611–6615.

(2) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.

(3) Putz, H.; Schön, J. C.; Jansen, M. *Ber. Bunsenges.* **1995**, *99*, 1148,–1153.

(4) Schön, J. C.; Jansen, M. In *Pauling's Legacy: Modern Modeling of the Chemical Bond*; Maksic, Z., Orville-Thomas, W., Eds.; Elsevier: Amsterdam, 1999; pp 103–127.

(5) Shayeghi, A.; Götz, D.; Davis, J.; Schaefer, R.; Johnston, R. L. *Phys. Chem. Chem. Phys.* **2015**, *17*, 2104–2112.

(6) Hu, J.; Ma, A.; Dinner, A. R. *J. Chem. Phys.* **2006**, *125*, 114101.

(7) Borrelli, K. W.; Vitalis, A.; Alcantara, R.; Guallar, V. *J. Chem. Theory Comput.* **2005**, *1*, 1304–1311.

(8) Wales, D. J.; Doye, J. P. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.

(9) Schön, J. C.; M. Jansen, *Z. Krist.* **2001**, *216*, 307–325.

(10) Schön, J. C.; Jansen, M. *Z. Krist.* **2001**, *216*, 361–383.

(11) Jónsson, H.; Mills, G.; Jacobsen, K. W. *Classical and quantum dynamics in condensed phase simulations* **1998**, *1*, 385–404.

(12) Voter, A. F. *Phys. Rev. Lett.* **1997**, *78*, 3908.

(13) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **1999**, *111*, 7010–7022.

(14) Munro, L. J.; Wales, D. J. *Phys. Rev. B* **1999**, *59*, 3969–3980.

(15) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

(16) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082–2094.

(17) Sheppard, D.; Terrell, R.; Henkelman, G. *J. Chem. Phys.* **2008**, *128*, 134106.

(18) Zagorac, D.; Schön, J. C.; Jansen, M. *J. Phys. Chem. C* **2012**, *116*, 16726–16739.

(19) Zheng, Y.; Xiao, P.; Henkelman, G. *J. Chem. Phys.* **2014**, *140*, 044115.

(20) Schön, J. C.; Putz, H.; Jansen, M. *J. Phys. Cond. Mat.* **1996**, *8*, 143–156.

(21) Schön, J. C. *Ber. Bunsenges.* **1996**, *100*, 1388–1391.

(22) Wevers, M. A. C.; Schön, J. C.; Jansen, M. *J. Phys.: Cond. Matt.* **1999**, *11*, 6487–6499.

(23) Doll, K.; Schön, J. C.; Jansen, M. *J. Chem. Phys.* **2010**, *133*, 024107.

(24) Neelamraju, S.; Schön, J. C.; Doll, K.; Jansen, M. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1223–1234.

(25) Heard, C. J.; Johnston, R. L.; Schön, J. C. *ChemPhysChem* **2015**, *16*, 1461–1469.

(26) Pacheco-Contreras, R.; Dessens-Félix, M.; Borbón-González, D. J.; Paz-Borbón, L. O.; Johnston, R. L.; Schön, J. C.; Posada-Amarillas, A. *J. Phys. Chem. A* **2012**, *116*, 5235–9.

(27) Schön, J. C.; Jansen, M. *Int. J. Mat. Res.* **2009**, *100*, 135–152.

(28) Doll, K.; Jansen, M. *Angew. Chem.* **2011**, *123*, 4723–4728.

(29) Schön, J. C.; Jansen, M. *Comp. Mater. Sci.* **1995**, *4*, 43–58.

(30) Putz, H.; Schön, J. C.; Jansen, M. *Comp. Mater. Sci.* **1998**, *11*, 309–322.

(31) Schön, J. C.; Jansen, M. In *Mat. Res. Soc. Symp. Proc. Vol. 848: Solid State Chemistry of Inorganic Materials V*; Li, J., Brese, N. E., Kanatzidis, M. G., Jansen, M., Eds.; MRS: Warrendale, 2005; p 333.

(32) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comp. Chem.* **2005**, *26*, 1668–1688.

(33) AmberTools 14 User Manual.

(34) Wu, M. G.; Deem, M. W. *J Chem. Phys.* **1999**, *111*.

(35) Dowd, M. K.; Kiely, D. E.; Zhang, J. *Carbohydr Res.* **2011**, *346*, 1140–8.

(36) French, A. D.; Kelterer, A.-M.; Johnson, G. P.; Dowd, M. K.; Cramer, C. J. *J. Mol. Graph. Mod.* **2000**, *18*, 95–107.

(37) Sibani, P.; van der Pas, R.; Schön, J. C. *Comp. Phys. Comm.* **1999**, *116*, 17–27.

(38) Schön, J. C. *Proc. Appl. Ceram.* **2015**, *9*, 157–168.

(39) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–1092.

(40) Thomas J. Macke,; David A. Case, In *Molecular Modeling of Nucleic Acids*; Leontis, N. B., SantaLucia, J., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 1997; Vol. 682; pp 379–393.

(41) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comp. Chem.* **2003**, *24*, 1999–2012.

(42) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comp. Chem.* **2008**, *29*, 622–655.

(43) Mongan, J.; Case, D. A.; McCammon, J. A. *J. Comp. Chem.* **2004**, *25*, 2038–2048.

(44) Schafmeister, C. E. A. F.; Ross, W. S.; Romanovski, V. LEAP. University of California, San Francisco(1995).

(45) Joo, S. H. *Biomol Ther (Seoul)* **2012**, *20*, 19–26.

(46) Oakley, M. T.; Johnston, R. L. *J. Chem. Theory Comput.* **2014**, *10*, 1810–1816.

(47) Oakley, M. T.; Oheix, E.; Peacock, A. F. A.; Johnston, R. L. *J. Phys. Chem. B* **2013**, *117*, 8122–8134.

(48) Neelamraju, S.; Oakley, M. T.; Johnston, R. L. *The Journal of chemical physics* **2015**, *143*, 165103.

(49) Kopple, K. D.; Wang, Y. S.; Cheng, A. G.; Bhandary, K. K. *J. Am. Chem. Soc.* **1988**, *110*, 4168–4176.

(50) Pinet, E.; Neumann, J. M.; Dahse, I.; Girault, G.; André, F. *Biopolymers* **1995**, *36*, 135–152.

(51) Kessler, H. *Angew. Chem. Intl. Ed.* **1982**, *21*, 512–523.

(52) Loiseau, N.; Gomis, J.-M.; Santolini, J.; Delaforge, M.; André, F. *Biopolymers* **2003**, *69*, 363–85.

(53) Che, Y.; Marshall, G. R. *J. Med. Chem.* **2006**, *49*, 111–124.

(54) Ash, S.; Cline, M. A.; Homer, R. W.; Hurst, T.; Smith, G. B. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 71–79.

(55) Wales, D. J. *Mol. Phys.* **2002**, *100*, 3285–3305.

(56) Wales, D. J. *Mol. Phys.* **2004**, *102*, 891–908.

(57) Wales, D. J. *Int. Rev. Phys. Chem.* **2006**, *25*, 237–282.

(58) Wales, D. J. PATHSAMPLE: A program for refining and analysing kinetic transition networks. `http://www-wales.ch.cam.ac.uk/PATHSAMPLE/`, (accessed March 31, 2016).

(59) Brooks, B. R. et al. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(60) Kony, D.; Damm, W.; Stoll, S.; Van Gunsteren, W. F. *J. Comp. Chem.* **2002**, *23*, 1416–1429.

(61) Tosso, R. D.; Zamora, M. A.; Suvire, F. D.; Enriz, R. D. *J. Phys. Chem. A* **2009**, *113*, 10818–10825.

(62) Brown, G. M.; Levy, H. A. *Acta Cryst. B* **1973**, *29*, 790–797.

(63) Hanson, J.; Sieker, L.; Jensen, L. *Acta Cryst. B* **1973**, *29*, 797–808.

(64) Mathlouthi, M.; Luu, D. V. *Carbohydr Res* **1980**, *81*, 203–212.

(65) Bock, K.; Lemieux, R. U. *Carbohydr Res* **1982**, *100*, 63–74.

(66) McCain, D. C.; Markley, J. L. *J. Am. Chem. Soc.* **1986**, *108*, 4259–4264.

(67) Duker, J. M.; Serianni, A. S. *Carbohydr Res* **1993**, *249*, 281–303.

(68) Herve du Penhoat, C.; Imberty, A.; Roques, N.; Michon, V.; Mentech, J.; Descotes, G.; Perez, S. *J. Am. Chem. Soc.* **1991**, *113*, 3720–3727.

(69) Poppe, L.; Van Halbeek, H. *J. Am. Chem. Soc.* **1992**, *114*, 1092–1094.

(70) Marszalek, P. E.; Oberhauser, A. F.; Pang, Y.-P.; Fernandez, J. M. *Nature* **1998**, *396*, 661–664.

(71) Immel, S.; Lichtenthaler, F. W. *Liebigs Ann.* **1995**, *1995*, 1925–1937.

(72) van Gunsteren, W. F.; Daura, X.; Mark, A. E. *Encyclopedia of Computational Chemistry* **1998**, *2*, 1211.

(73) McNaught, A. D. *Carbohydr Res.* **1997**, *297*, 1–92.

(74) Ramachandran, G.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95–99.

Table of contents graphic

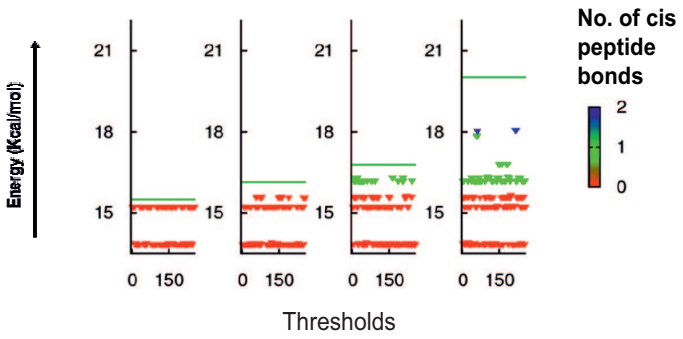

Figure 12: Threshold minimization runs on cyclo-$(Gly)_4$ at four different energy threshold values.