

Quasi-Continuous Interpolation Scheme for Pathways between Distant Configurations

David J. Wales* and Joanne M. Carr*

University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, United Kingdom

ABSTRACT: A quasi-continuous interpolation (QCI) scheme is introduced for characterizing physically realistic initial pathways from which to initiate transition state searches and construct kinetic transition networks. Applications are presented for peptides, proteins, and a morphological transformation in an atomic cluster. The first step in each case involves end point alignment, and we describe the use of a shortest augmenting path algorithm for optimizing permutational isomers. The QCI procedure then employs an interpolating potential, which preserves the covalent bonding framework for the biomolecules and includes repulsive terms between unconstrained atoms. This potential is used to identify an interpolating path by minimizing contributions from a connected set of images, including terms corresponding to minima in the interatomic distances between them. This procedure detects unphysical geometries in the line segments between images. The most difficult cases, where linear interpolation would involve chain crossings, are treated by growing the structure an atom at a time using the interpolating potential. To test the QCI procedure, we carry through a series of benchmark calculations where the initial interpolation is coupled to explicit transition state searches to produce complete pathways between specified local minima.

1. INTRODUCTION

The first step in analyzing the kinetics for interconversion of different structures is usually the construction of an initial path between the regions of interest. For large-scale changes in morphology of a system treated in atomistic detail, this construction may itself become a challenging task. Obvious problems arise for methods where the initial interpolation is a straight line path between the product and reactant states, as atom clashes may occur. For systems with chains of covalently bonded atoms described by force fields such as CHARMM,^{1–4} AMBER,^{5–7} or coarse-grained bead models,^{8,9} more subtle problems can occur. Interpolation schemes that involve discrete images of the system bridging the product and reactant configurations can produce chain crossings, where the structure jumps between discrete images of the system in an unphysical manner. In particular, we have found that coarse-grained/united atom models can exhibit transition states and pathways corresponding to such chain crossings. These unphysical pathways generally lie significantly higher in energy than the physically relevant paths, where the chain-crossing is avoided. Nevertheless, they can lie below stationary points of the potential energy surface associated with the higher energy denatured or liquid-like phase of a finite system. This energy overlap makes the unphysical paths difficult to detect in automated procedures for constructing kinetic transition networks,^{10–14} such as discrete path sampling.^{11,15,16}

The real difficulty with interpolation via discrete images is that unphysical configurations can lie between them. Increasing the number of images, or the force constants of springs that connect them,^{17–22} will not change the chain-crossing topology once it has been established. What is really needed is some way to account for and avoid unphysical geometries that lie between the images. Here, we seek a solution that does not involve a fully continuous path^{23–25} but augments a discrete set of images with terms that represent the worst-case energetic contributions from the intervening straight line segments

between them. We therefore refer to this approach as quasi-continuous interpolation (QCI). This QCI framework enables the interpolation problem to be framed entirely in terms of geometry optimization, where a wide range of powerful techniques can be applied.^{26,27}

To benchmark the QCI procedure, we combine it with calculations of transition states and pathways to connect specified local minima (end points) in a variety of systems. The role of QCI within the context of an overall scheme to characterize geometrical pathways connecting local minima and transition states is discussed in section 2. Examples are provided for a coarse-grained model of protein L (section 3.3), the 12-residue tryptophan zipper peptide, trpzip2^{28,29} (section 3.1), and a change of morphology for an atomic cluster (section 3.2). Finally, more extensive benchmarking is conducted for 2100 pathways in the amyloidogenic GNNQQNY peptide^{30,31} ($\text{NH}_3^+\text{-Gly-Asn-Asn-Gln-Gln-Asn-Tyr-COO}^-$) and the tryptophan zipper peptide, trpzip1^{28,32} ($\text{NH}_3^+\text{-Ser-Thr-Trp-Glu-Asn-Gly-Lys-Trp-Thr-Trp-Lys-CH}_3$), for both the AMBER and CHARMM potentials. The general QCI approach is competitive with the schemes that exploit internal coordinates and can provide physically meaningful interpolations for the more complex rearrangements where chain-crossings otherwise result.

2. METHODS: FROM INITIAL INTERPOLATIONS TO CONNECTED PATHWAYS

2.1. Overview. The main innovation of the present work is the QCI procedure, which is discussed in section 2.3. However, to present this approach in context we consider complete calculations of pathways connecting specified local minima. For

Special Issue: Berny Schlegel Festschrift

Received: June 11, 2012

Published: August 28, 2012



the present purposes, we adopt a purely geometrical view, where pathways correspond to pairs of local potential energy minima, linked by steepest-descent paths from transition states, which are defined as stationary points of index one on the potential energy surface.³³ Complete pathways may consist of multiple minimum-transition state-minimum steps.

The role of QCI within this overall procedure is shown schematically in the flowchart (Figure 1). Here, the QCI step is

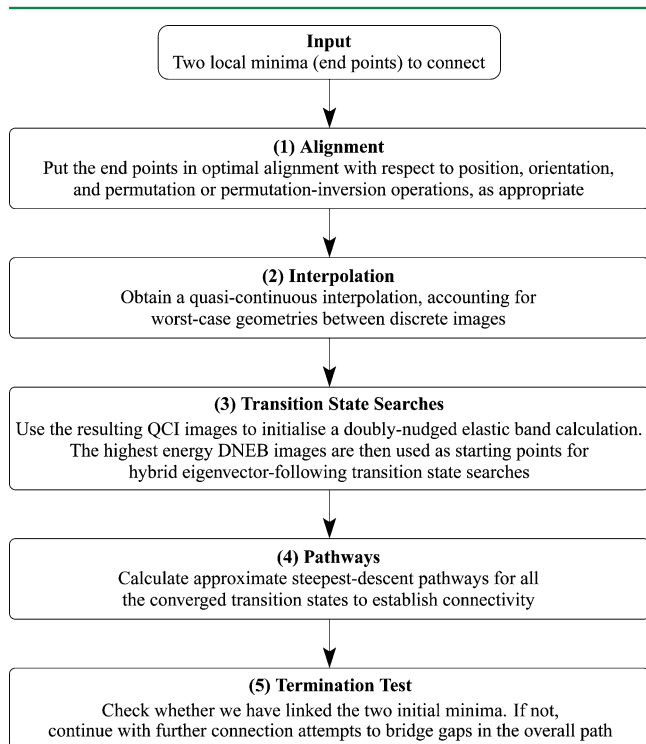


Figure 1. Schematic view of the overall double-ended connection strategy used in the present work. The QCI procedure represents an additional step between end point alignment and searches for transition states. A database of previously obtained minima can be used in steps 1 and 2, if available. In step 1, all minima are aligned to the same reference minimum if a database is available. Knowledge of a wide variety of local minima can be particularly helpful in defining the constraint potential for QCI in step 2, since nonbonded atoms are increasingly unlikely to appear at the same separation as more local minima are considered.

an addition to our usual connection-making framework. The objective is to avoid unphysical chain crossings or atom clashes for complex pathways, which sometimes arise in step 3. We also wish to achieve comparable or improved performance for simpler cases, where our existing framework already succeeds in creating kinetically relevant pathways.

1. Alignment. The first step in the connection procedure is always end point alignment. For an N -atom molecule containing no permutable atoms, this is a simple process, where the minimum Euclidean distance between end points in $3N$ dimensions can be obtained deterministically (section 2.2). Pathways between permutational isomers or enantiomers may be the focus of attention, for example, in predicting or interpreting tunneling splittings.³⁴ However, in all of the examples considered below we seek the simplest pathways possible between local minima. Optimal permutational alignment of identical atoms is essential in these calculations, and efficient methods based on overall and local distance metrics

are described in section 2.2. Without the permutational alignment, the resulting pathway may have additional steps that permute atoms of the same element in the two end points.

2. Interpolation. The QCI step provides a set of images and line segments that connect them, designed to optimize an interpolated pathway. Here, we note that all double-ended transition state search procedures^{17–22,35–37} already include an implicit interpolation. The novel aspect of QCI is that it accounts for the worst-case geometries that may arise between the discrete images. It is this feature that makes the method ‘quasi-continuous’. For pathways involving complex atomic motion, unphysical geometries can arise between images, for example, involving chain crossing or atom clashes. An initial quasi-continuous interpolation can avoid such problems, and the performance is comparable to existing methods for simpler interpolations, with improvements for pathways of increasing complexity. The examples discussed in section 3 include a case for protein L where the previous methodology fails, along with statistics for systems where comparisons of overall performance are possible.

Worst-case geometries on the line segments are found by identifying local minima in the interatomic distances for all pairs of atoms, between neighboring images. To improve efficiency, and avoid extracting specific energetic terms from external potentials such as CHARMM and AMBER, an auxiliary interpolation potential is constructed. This potential can be based on just the two end points in question or can employ data for any number of local minima previously obtained for the same system. Constraints are applied between atoms that are always separated by a fixed distance, to within a predetermined tolerance, and repulsive interactions are defined between all other atoms, with a relatively short cutoff. For these pairwise potentials, the smallest and largest distances between each pair of atoms on a synchronous straight line path between neighboring images can be calculated analytically. The total energy can therefore be augmented by terms corresponding to an internal maximum in the auxiliary interpolating potential, as described in section 2.3. This auxiliary interpolation potential, which incorporates some features of an elastic network model,³⁸ also enables the system to be grown a single atom at a time for the most difficult chain-crossing examples. A sensible guess for the interpolated position for the next atom can then be obtained in various ways from neighboring atoms for which constraints exist (see the Appendix).

3. Transition State Searches. The QCI images are employed to initialize a doubly-nudged³⁹ elastic band^{17–22} calculation, which is used to provide transition state candidates for hybrid eigenvector-following (section 2.4).^{19,40,41}

4. Pathways. The connectivity defined by all of the converged transition states obtained in step 3 is obtained. Approximate steepest-descent pathways are calculated following small perturbations of the geometry parallel and antiparallel to the eigenvector corresponding to the unique negative Hessian eigenvalue at the transition state.

5. Termination. Here, we check whether a complete path exists between the two end points in terms of connected —minimum—transition state—minimum—transition state—minimum— sequences. If not, we return to step 1, choosing new pairs of minima for connection attempts that will bridge gaps in the current database of stationary points. For pathways involving numerous transition states, this selection process can play an important role, and we employ a missing

connection algorithm⁴² based on Dijkstra's shortest path procedure.⁴³

Summary. The steps in parts 3–5 have all been discussed in detail in previous work. The permutational alignment procedure has been used before but has not been described in detail. We therefore explain the latter approach in section 2.2. The QCI procedure itself is discussed in section 2.3, and the associated formulas are presented in the Appendix.

2.2. Permutational Alignment. The alignment of end points is critically important in obtaining efficient interpolation and pathways via geometry optimization. The alignment of N atoms that define rigid body geometries to minimize the Euclidean distance in $3N$ dimensions is easily achieved using a quaternion procedure.⁴⁴ However, when permutations of atoms of the same element are also included there is no longer a deterministic solution to this problem. For a fixed center of coordinates and orientation, the distance can be minimized using a shortest augmenting path algorithm,⁴⁵ and this is the procedure that has been used in the OPTIM program⁴⁶ for previous studies of biomolecules^{47–49} as well as clusters and condensed phases.⁵⁰

Suitably symmetrized modifications of the CHARMM and AMBER potentials were used throughout to ensure that permutational isomers of every stationary point have the same energy.^{39,51} The formulation of these force fields means that only limited subsets of atoms are actually permutable, giving a much smaller subset of the full nuclear permutation group. An auxiliary file specifying the allowed permutations is prepared via automated scripts.⁵¹ The entries in this file contain the number of permutable atoms in the primary group and the number of other sets of permutable atoms associated with this primary set, which can be zero. The indices of the atoms in the primary set are then provided, followed by the indices of the pairs of atoms in the secondary sets, which must be permuted in correspondence with the primary atoms. Examples are given for alanine, phenylalanine, and valine in Figure 2. The

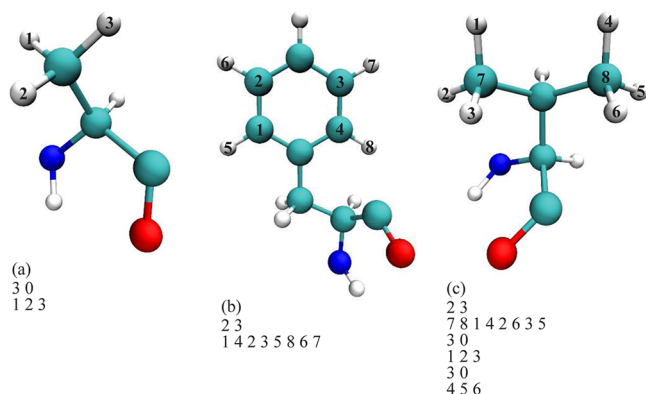


Figure 2. Allowed permutations and corresponding entries in the auxiliary file, illustrated for three amino acid side chains. (a) Alanine: the three hydrogens are permutable and comprise a single set. No other atoms are required to change places at the same time. (b) Phenylalanine: the aromatic ring can rotate, exchanging exactly four pairs of atoms. If atoms 1 and 4 comprise the primary set, then three other pairs 2–3, 5–8, and 6–7 define the secondary sets of atoms that must also be swapped. The definition of particular sets of permutable atoms as primary is arbitrary. (c) Valine: the two methyl groups can be exchanged via four simultaneous pair swaps (here, atoms 7 and 8 comprise the primary set). Additionally the three hydrogens within each methyl group can be permuted, as in alanine.

Euclidean distance metric is minimized for each primary set in turn, followed by overall translational and rotational alignment,⁴⁴ and this procedure is repeated until no further permutations result. The combination of permutational and translational/orientational alignment can produce a different result depending on the initial orientation, and it may therefore be necessary to start from a number of random initial orientations to find the global minimum for the Euclidean distance. For optimal alignment of atomic and molecular clusters, we also consider the enantiomer for one of the end points; this cycle is generally omitted for biomolecules. Further complications occur if the system is subject to an electric or magnetic field,^{52,53} a pulling potential,⁵⁴ or a central harmonic potential.^{55–58} In these situations, the overall translational and orientational alignment is restricted to operations that leave the Hamiltonian invariant.

In constructing kinetic transition networks via geometry optimization, it is also necessary to recognize permutation-inversion isomers whenever they are encountered. To achieve this goal we employ standard orientations before the shortest augmenting path procedure and diagnose equivalent isomers via 'zero' overall distance within a given tolerance following allowed permutations. The atom that is furthest from the center of coordinates is placed on the z axis and the atom that is then most distant from the z axis is moved into the xz plane by overall rotation about this axis. Due regard must be paid to the presence of external potentials. For example, only rotation about the axis of an applied electric or magnetic field is allowed in preparing the corresponding standard orientations. There is also a complication due to symmetry, which can be approximate symmetry in the local alignment procedure described below. In general, there could be more than one atom that is approximately the same distance from the center of coordinates or from the z axis. Permutational alignment must therefore be attempted for standard orientations based on all of the atoms in each of the corresponding orbits for the two end points, and a cutoff is required to decide whether the distances are considered the same or not. A tolerance of 0.2 Å generally seemed to be satisfactory.

Unfortunately, minimizing the total Euclidean distance in $3N$ dimensional space can actually result in incorrect local permutational alignment. For example, the terminal NH_3^+ groups of lysine residues that undergo a large spatial displacement in space can sometimes give a slightly shorter overall distance if they are permuted cyclically relative to the rest of the residue. Worse still, the optimal overall distance could correspond to swapping a pair of hydrogen atoms in the NH_3^+ group, which entails a very large barrier on the corresponding pathway. To overcome these problems, a local permutational alignment procedure has been implemented.⁵⁹ The standard orientation and shortest augmenting path procedures (including loops over atoms in the same orbit) are performed locally for each group of permutable atoms and n nearest neighbors. The neighbors are chosen from the atoms outside the permutable group (i.e., excluding all primary atoms in the group, as defined above) inside a cutoff distance from the center of coordinates of the permutable set for the two end points. A tolerance is also specified in terms of the minimal Euclidean distance for the permutable and neighbor atoms for an acceptable alignment. Atoms were added to a trial neighbor list one at a time in increasing order of the mean distance from the center of coordinates of the permutable set and accepted if the alignment threshold was achieved. Permutable neighbors

were added together with the rest of the atoms in their primary permutable set. This procedure was continued until a specified number of neighbors was reached, or no more candidate atoms remained within the cutoff distance. For the peptide benchmarks described in section 3.4, a maximum of 10 neighbor atoms, a threshold of 1.0 Å for the overall alignment, and a cutoff distance of 5 Å were found to give good results for the 2100 pathways considered.

Although the permutational alignment procedures entail a considerable amount of bookkeeping, they are generally fast compared to operations that involve evaluating the potential, especially when only small subsets of atoms are permutable. This effort can greatly reduce the computational expense of finding connected pathways, especially for large systems with many equivalent atoms, such as atomic clusters and condensed phases. Further complications arise for systems with periodic boundary conditions, where the cell symmetries and the center of coordinates must be accounted for. Our alignment procedures for condensed phases described by periodic supercells will be described in detail elsewhere.

2.3. The QCI Procedure. To implement the QCI procedure, we first constructed an auxiliary interpolation potential, designed to maintain a physically reasonable geometry for a continuous path between the end points. The auxiliary potential is defined for arbitrary subsets of atoms, enabling us to grow the interpolation an atom at a time. It is also computationally inexpensive compared to force fields such as CHARMM and AMBER, which are employed in some of the examples in section 3. In common with other double-ended schemes,^{17–22,35–37} the potential is defined and minimized over a series of images corresponding to intermediate geometries between the two end points. However, the QCI step does not aim to locate transition states directly. Instead, the objective is to avoid unphysical geometries that could arise between the discrete set of images. This goal is achieved by adding terms to the interpolation potential that correspond to minima in any interparticle distances that occur on straight line segments between consecutive images. Local minima of this type provide a worst-case condition for the pairwise-additive interpolation potential and enable us to diagnose chain crossings or atom clashes that could intervene between physically sensible geometries at the discrete image positions. An analogous problem arises when computing the potential between rod-like particles using the distance of closest approach;⁶⁰ details of the potential and its derivatives are provided in the Appendix.

The most important terms in the interpolation potential used in the present work are springs to constrain selected interatomic distances and short-range repulsions to avoid chain crossings and atom clashes. Conserved interatomic distances are diagnosed from the two end points, or from a database of local minima, if available. All unconstrained pairs of atoms are subject to short-range repulsion, handled via a neighbor list, as described in the Appendix. The distance constraints used in QCI are tight penalty functions, in contrast to the interpolations employed in the synchronous transit method.³⁵ The additional terms corresponding to local minima in the interparticle distances between images also distinguish the QCI approach from synchronous transit and nudged³⁹ elastic band^{17–22} (NEB) methods. These local minima will generally occur at different positions along the line segment between images for different pairs of atoms. All such terms are included. Furthermore, nudging is not required in QCI, and most of the examples described in section 3 work equally well

without additional springs between the images during the initial interpolation. Instead, springs are used to fix selected interatomic distances in the QCI interpolation potential (see Appendix).

The constrained atoms were chosen to define a percolating network to maintain the underlying structure of the molecule, which corresponds to covalently bonded atoms for the biomolecules considered. To test for percolation, the minimum number of steps via constraints from each atom to a reference atom was calculated using a depth first search.⁶¹ When only two end points are available, some caution is required to prevent interatomic distances being fixed when they actually need to change significantly and return to their starting values. Such problems generally disappear when a database of structurally diverse local minima is available for comparison, which can always be arranged in practice.

Having aligned the end points and defined a percolating network and the interpolation potential, the optimization of an interpolating set of images and line segments can begin. Placing initial guesses along a straight line path may succeed for simple pathways but can result in chain crossings and atom clashes, as illustrated in section 3.3. Since the interpolation potential can be defined for any subset of atoms in the molecule, we implemented a procedure to grow an interpolated structure progressively. Details of how atoms were added or removed during this procedure are provided in the Appendix. Knowledge of the local environment in the two end points was exploited to suggest initial positions for new atoms in the QCI images, and these guesses were informed by evaluating the interpolation potential. Atoms were removed via a backtracking procedure if minimization of the interpolation potential did not converge sufficiently, and some degree of randomness was introduced to avoid repeating failed attempts.

Several other measures were employed to improve the efficiency of the QCI step. Images were added and removed from the interpolation dynamically, according to distance criteria, and frozen if the root-mean-square gradient terms in the interpolation potential fell below a given threshold. Provision was also made to freeze the positions of certain atoms completely. For localized rearrangements in large systems, the majority of atoms may hardly move along the corresponding pathway once the end points have been suitably aligned. For the examples considered in the present work there is little advantage in identifying these ‘spectator’ atoms. A much larger gain in efficiency can result for molecules containing several thousand atoms, and we will report on pathway calculations for such systems elsewhere.

Overall, the QCI scheme is probably most closely related to the geometric targeting (GT) approach,^{62–64} where the aim is to generate plausible all-atom pathways for large conformational changes in proteins. For example, the GT approach has recently been employed to obtain pathways for nitrogen regulatory protein C,⁶⁴ which has also been treated using discrete path sampling.⁴⁷ The GT framework is based on geometrical constraints, as for the present QCI method, and employs more detailed information about protein stereochemical propensities. The aim of QCI is somewhat different, since we are using the interpolation to initialize transition state searches and characterize a connected path in terms of local minima and transition states of the potential energy surface. This calculation represents a single step in the construction of a kinetic transition network^{10–14} and the analysis of kinetics using discrete path sampling (see section 3).^{11,15,16} Most of the

computer time required for such calculations is used in refining transition states, and good initial guesses can speed up this process significantly. We therefore included a minimization phase in the QCI step where a fraction of the true potential corresponding to the force field in question was added to the interpolation potential.

QCI is also general enough for application to molecules without a fixed set of covalent bonds, as illustrated for an atomic cluster in section 3.2. Exploiting a computationally inexpensive potential for the interpolation provides an initial minimization problem, in contrast to targeted molecular dynamics approaches,^{65–67} and string methods.^{36,37} It would be interesting to compare methodologies that employ fully continuous pathways^{23–25,68} with the present approach, but this exercise lies beyond the scope of the present investigation.

2.4. Calculating Complete Pathways. In step 4 of the procedure outlined in Figure 1, where we deal directly with the true potential, V_{true} , a doubly nudged³⁹ elastic band^{17–22} (DNEB) optimization was conducted using starting images regularly spaced along the path (i.e., line segments and intervening images) obtained by the QCI procedure. Local maxima in the DNEB profile were then converged tightly to transition states using hybrid eigenvector-following (EF).^{19,40,41} The form of the uphill EF step was optimized in a series of investigations in earlier work.^{69–71} For each transition state, the connectivity was established in step 5 by calculating approximate steepest-descent paths using energy minimization following small displacements parallel and antiparallel to the eigenvector corresponding to the unique negative Hessian eigenvalue.

For the DNEB calculations, we employed a dynamic adjustment of the force constant for the springs between the DNEB images. This adjustment improved all the benchmark timings reported in section 3.4. The force constant was increased or decreased by 5% every five DNEB steps depending on the average spacing of the images. If the average deviation of the spacing from the mean exceeded 5%, the force constant was increased, and otherwise it was decreased. Variable spring constants between different images have previously been used to increase the density of images near a transition state.²⁰ However, the uniform rescaling employed in the present work was intended to obtain an equally spaced image distribution more rapidly. For pathways involving both high energy and low energy transition states, our observation is that the low-lying structure for the DNEB images can be unresolved if the spacings are not uniform enough. The rescaling helped to converge some pathways of this sort much more efficiently.

In step 5 of the overall connection procedure (Figure 1), we test to see whether a complete connected path between the specified end points has been obtained. Depending on the separation of the end points, further connection attempts may be needed. Pairs of unconnected minima were chosen for subsequent pathway searches using our missing connection approach,⁴² which is based on Dijkstra's algorithm.⁴³ To define a metric for the missing connection algorithm, we normally employ a function of the Euclidean distance in 3N dimensional space, minimized with respect to overall translation and rotation as well as permutation of equivalent atoms. However, having constructed the interpolation potential, V_{int} , we can employ it to guide the selection of pairs of minima that are most likely to produce a connected path. The metric used in the missing connection algorithm was chosen as the minimal distance plus $V_{\text{int}}/10^3$, with the result exponentiated or raised to

a power such as 10 or 20. The latter rescaling tends to disfavor single connection attempts between minima separated by large metric values relative to multiple connection attempts between 'closer' minima. The value of V_{int} was calculated using the two minima in question as the only images. This choice of metric proved to be particularly important for pathways where a straight line interpolation would result in chain crossing. In such cases, the minimized distance for one chain passing through another chain could be quite small, whereas V_{int} should be large. Hence it is possible to work around crossings via longer pathways that correspond to lower energy and physically reasonable transition states. Dividing V_{int} by 10^3 means that the metric is dominated by V_{int} for unphysical straight line interpolations but correlates mainly with the Euclidean distance for more realistic interpolations. True transition states corresponding to unphysical chain crossings seem to exist for all of the biomolecular force fields we have investigated and cannot generally be eliminated on the basis of the true potential.

3. RESULTS

Several examples that employ the QCI procedure to obtain initial pathways are described in the subsections below. These initial paths would generally require further refinement using discrete path sampling^{11,15,16} to build up a kinetic transition network.^{10–14} For distant end points, we usually find that the initial path can be improved upon significantly by additional pathway searches. Nevertheless, the efficiency with which kinetically relevant paths are located depends strongly on the initial path.

3.1. Trpzip2. As a straightforward test case, we first considered the 12-residue tryptophan zipper peptide, trpzip2^{28,29} (NH_3^+ -Ser-Thr-Trp-Glu-Asn-Gly-Lys-Trp-Thr-Trp-Lys- CH_3). The force field employed was a symmetrized^{39,51} version of CHARMM19² together with the implicit solvation potential EEF1.⁷² Eleven different pairs of local minima were selected from a large database, which will be described elsewhere in terms of a detailed kinetic analysis. These pairs were selected as cases where relatively high barriers remained between low-lying minima for end point alignment schemes involving internal coordinates.⁷³ The QCI calculations therefore served to check whether pathways with lower barriers exist.

The QCI procedure either reproduced the pathway with the lowest barrier found using interpolation in internal coordinates⁷³ or, in a few cases, improved upon it. The number of transition states in the resulting paths varied from one to 40, depending on the end points in question. Two of the shorter paths are illustrated in Figures 3 and 4; the VMD program was used to visualize structures throughout the present work.⁷⁴ The first path is obtained via a single interpolation and corresponds to a conformational change in the lys side chain, which is involved in a salt bridge to glu (Figure 3). The second path consists of a more complicated reorganization of the trp side-chains, with eight transition states linking nine local minima (Figure 4). A complete path was obtained in three cycles of the missing connection procedure,⁴² which required 700 s of CPU time on a laptop computer (using one core of a 2.50 GHz Intel T9300 processor), including tight convergence to an RMS gradient of 10^{-6} kcal/mol and normal-mode analysis for all stationary points. The overall path involves rearrangements on rather different length and energy scales, which can cause problems with automated geometry optimization procedures.

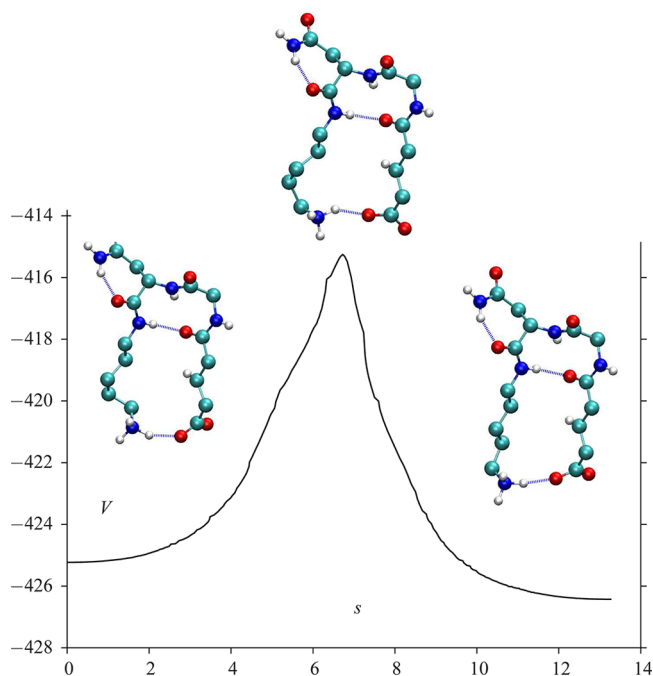


Figure 3. Potential energy, V (kcal/mol), plotted as a function of the integrated path length, s (Å), for a single transition state pathway of the trpzip2 peptide. The two local minima and the transition state are superimposed above the path. This pathway corresponds to an internal rotation of the lys side chain constrained by the hydrogen-bonding contact in the glu/lys salt bridge. The snapshots include the C_{α} carbon atoms and side chains of glu and lys, along with the intervening gly and asn residues.

For example, the first step corresponds to rotation of the hydroxyl group in ser, while the second step involves a much higher barrier for restacking trp side chains.

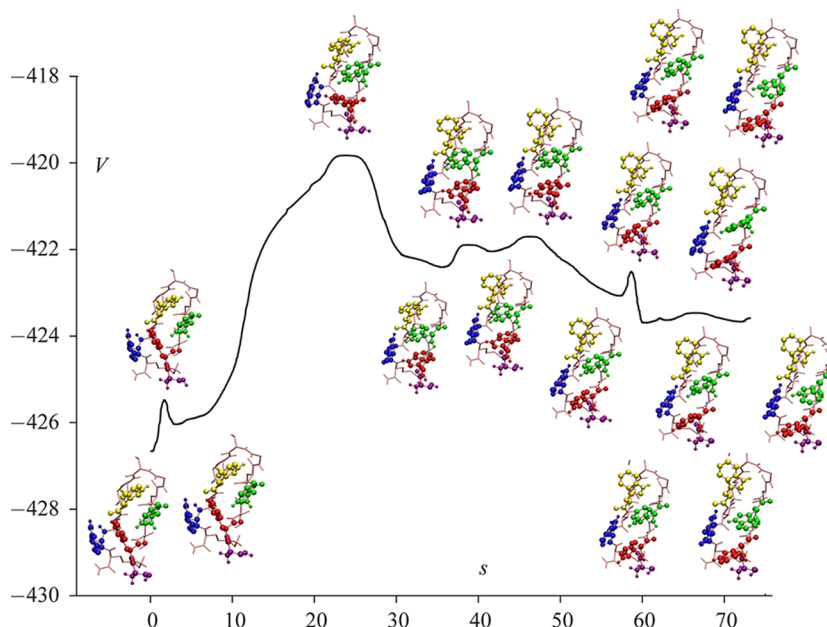


Figure 4. Potential energy, V (kcal/mol), plotted as a function of the integrated path length, s (Å), for a path involving trp side chain reorganizations in the trpzip2 peptide. The relevant atoms of the nine local minima and the eight transition states are illustrated below and above the path, respectively, at approximately the corresponding path length. The trp residues are all colored differently to distinguish them, and the ser residue is also highlighted, since the first step involves a conformational change in the corresponding OH group.

3.2. An Atomic Cluster. The QCI scheme was also implemented for an atomic cluster, to test whether useful interpolations could be obtained for systems without a fixed connectivity defined by covalent bonds. In this case, the interpolation potential was defined using the usual Lennard-Jones form⁷⁵ plus additional repulsive terms, which were only included when minima in the interatomic distances occurred between the images; details are provided in the Appendix. The pairwise form of the true potential in this case meant that a simpler QCI step could be employed. In particular, it was not necessary to grow the interpolated images by adding atoms progressively. The interpolation potential also contains the true potential throughout.

The pathway illustrated in Figure 5 corresponds to a change in morphology from the face-centered-cubic global minimum of the 38-atom cluster LJ_{38} bound by the Lennard-Jones potential.⁷⁵ This system exhibits a double-funnel potential energy landscape^{76–81} and has served as a benchmark for global optimization,^{82,83} thermodynamics,^{77–80} and rare event dynamics.^{15,16,84,85} The pathway calculation based on QCI required 6 s of CPU time (on one core of a 2.50 GHz Intel T9300 processor) and four cycles of the missing connection procedure.⁴² For comparison, when a DNEB interpolation was used via an initial straight line guess the CPU time was practically the same, although a slightly different pathway was produced. The precise sequence of transition states and minima is generally quite sensitive to parameters of the interpolation procedure in all of these calculations, and small changes can significantly affect the number of cycles required to obtain a complete connected path.

3.3. Protein L. Protein L exhibits a ubiquitin fold with a central α helix packed against a mixed four-strand β sheet.⁸⁶ This system was modeled using the sequence-dependent BLN potential of Brown and co-workers,^{8,9} where each of the 56 amino acids is represented by a hydrophobic (B), hydrophilic

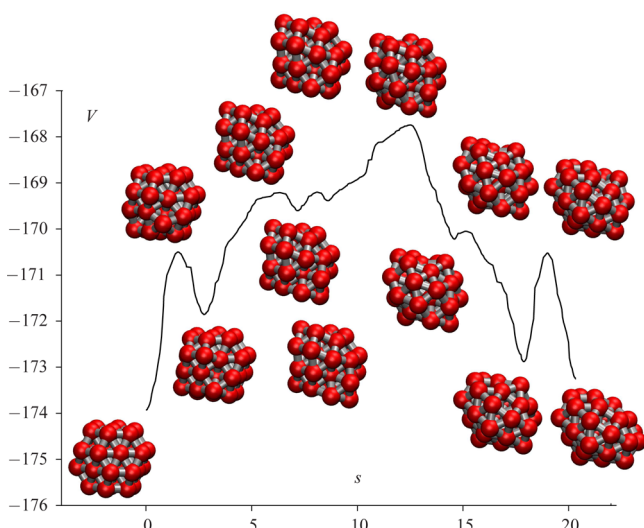


Figure 5. Potential energy, V (e), plotted as a function of the integrated path length, s (σ), for a path connecting the global minimum and second-lowest minimum of the LJ_{38} cluster. This path corresponds to an overall change in morphology from face-centered-cubic to icosahedral packing and involves six transition states and seven minima. These stationary points are illustrated above and below the path, respectively, at approximately the corresponding path length.

(L), or neutral (N) bead. A detailed analysis of the energy landscape for this protein as a function of static pulling force is presented elsewhere.⁵⁴ The QCI procedure proved to be very effective in providing pathways between distant structures that avoid chain crossings and has been tested in millions of connection attempts.⁵⁴ Two examples will be illustrated here. In the first case, standard DNEB calculations without QCI result in an unphysical chain crossing, which is avoided when the QCI step is included. The second example shows how a complete folding pathway obtained using QCI compares with the most favorable kinetic pathway following more extensive pathway optimization.

The first connection involves a low-lying minimum (the sixth lowest for zero force) and a misfolded structure that lies significantly higher in energy. In the latter minimum, the C-terminus is threaded through beneath part of the β_4 strand, and it is this partially knotted structure that causes problems when QCI is not used. Complete paths are obtained both with and without QCI, but without QCI a chain crossing is observed (Figure 6a). The two runs took 7619 and 7310 s of CPU time on an Intel Xeon E5404 processor (running at 2.0 GHz), and the paths involve 56 and 38 transition states, respectively, for Figures 6a and b. Eighty-four cycles of the connection procedure were needed when QCI was not used, and nine when the QCI step was included. Similar pathways are obtained for a range of QCI parameters, and it is not necessary to include springs between the QCI images to produce a physically meaningful pathway.

For the folding pathway illustrated in Figure 7, the two minima correspond to the global minimum for zero force and the global minimum that first appears when a static force is applied to residues 1 and 56, relaxed for zero force. The pathway in Figure 7 involves 41 transition states and required 12 cycles of the QCI procedure, which took 3045 s of CPU time on an Intel Xeon E5404 processor (running at 2.0 GHz). The folding path involves initial formation of the N-terminal hairpin, followed by association of the hairpin with the helix.

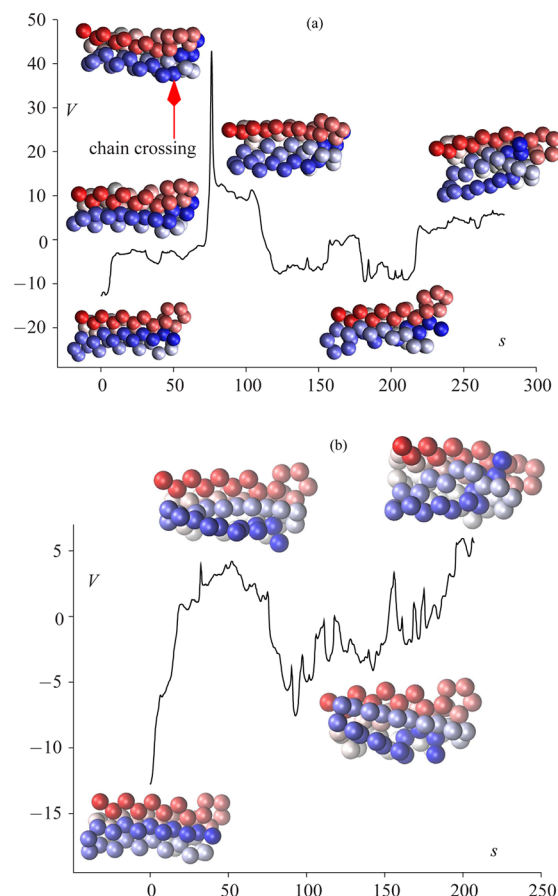


Figure 6. Potential energy, V (e), plotted as a function of the integrated path length, s (σ), for paths connecting a low-lying minimum and a partially knotted minimum of protein L represented by a coarse-grained potential. Selected structures are illustrated, colored from red to blue (N-terminus, β_1 , to C-terminus, β_4) according to the position in the chain. These structures are positioned at approximately the corresponding path length. (a) Pathway with chain crossing obtained without QCI. (b) Pathway obtained with QCI.

Strand β_1 then dissociates, and strand β_3 associates with the helix and strand β_2 . Strands β_1 and β_4 then form a connection and associate with the helix in a concerted fashion. For reference, the path that makes the largest contribution to the overall rate constant for transitions between these two minima after refinement of the kinetic transition network involves 22 transition states and a similar overall barrier.⁵⁴ In the folding direction, the N-terminal hairpin forms first, followed by association with the helix and strand β_3 . Finally, strand β_4 joins the β sheet. Hence, the fastest path corresponds to the same mechanism as the longer initial path, but with some unnecessary motion cut out.

3.4. Benchmarks for the Trpzip and GNNQQNY Peptides. In previous work,⁷³ we tested seven different combinations of internal and Cartesian coordinates for interpolation and alignment of the amyloidogenic GNNQQNY peptide^{30,31} (NH_3^+ -Gly-Asn-Asn-Gln-Gln-Asn-Tyr-COO $^-$) and the tryptophan zipper peptide, trpzip1^{28,32} (NH_3^+ -Ser-Thr-Trp-Glu-Asn-Gly-Lys-Trp-Thr-Trp-Lys-CH $_3$). Details of these calculations can be found in the original reference;⁷³ the abbreviations are summarized in Table 1.

For each peptide, we chose the same pairs of minima from existing databases for connection attempts as in previous work. Pairs known to be connected by pathways involving at least

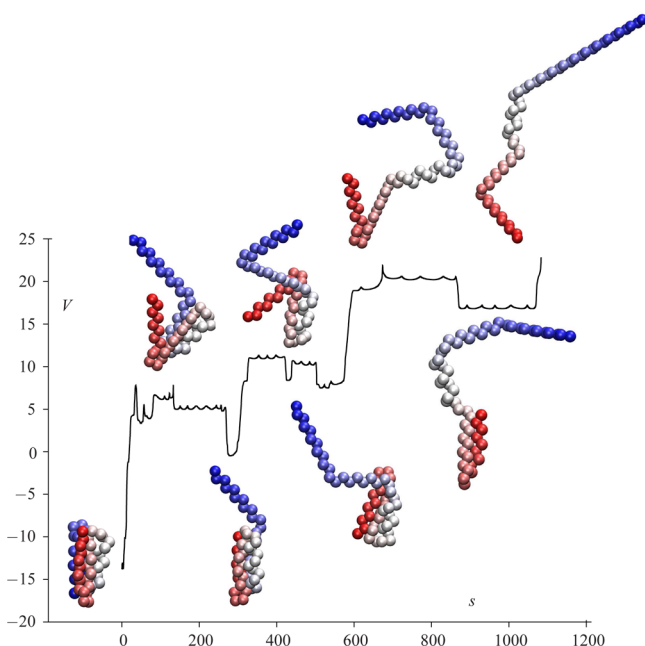


Figure 7. Potential energy, V (ϵ), plotted as a function of the integrated path length, s (σ), for a path connecting the global minimum and an extended minimum for protein L represented by a coarse-grained potential. Selected local minima are illustrated, colored from red to blue (N-terminus, β_1 , to C-terminus, β_4) according to the position in the chain. These structures are positioned at approximately the corresponding path length.

Table 1. Summary of the Different Alignment and Interpolation Schemes Used in Previous Work⁷³

method	alignment	interpolation
BCSC	Cartesian	Cartesian for backbone and side chains
BCSI	Cartesian	Cartesian for backbone, CHARMM internal for side chains
BISI	Cartesian	CHARMM internal for backbone and side chains
NONI	Cartesian	natural internal
NI	natural internal	natural internal
NIS	natural internal	natural internal, using nonequispaced images
NIC	natural internal	natural internal or Cartesian, according to the lowest energy

one, two, and three transition states were selected to give test sets denoted TS1, TS2, and TS3. The test sets for CHARMM involved connections between 150 pairs of local minima, except for GNNQQNY TS1 and TS2, where 300 examples were chosen, giving 1200 in total. For the AMBER benchmarks, 150 pairs were considered in each case, i.e. 900 paths in total. The CHARMM calculations employed the united-atom force field CHARMM19² with the implicit solvation potential EEF1,⁸⁷ while the AMBER calculations used the ff03 parameters⁸⁸ and the generalized Born solvation model GB^{OBC}.⁸⁹ Both potentials were symmetrized so that permutational isomers would have the same energy.^{39,51} The mean CPU times required to achieve a reconnection for each procedure and test set are summarized in Table 2 and displayed graphically in Figure 8. A consistent set of parameters was employed for the OPTIM program throughout these tests, the only differences being the keywords associated with the interpolation procedure. The neighbor list for nonbonded interactions in CHARMM was updated every

100 geometry optimization steps, after noting an apparent discontinuity that occurred in one test case when the check was every 1000 steps.

The timings for these 2100 connection attempts, referred to hereafter as the MSB test set, are summarized in Table 2 and plotted in Figure 8. These timings refer to the local permutational alignment procedure described in section 2.2, which improved the overall performance in virtually every case. In particular, the QCI approach is generally as good as the methods that use internal coordinates, and even DNEB searches starting from straight line interpolations can be competitive when the permutational alignment is optimal. The overhead for the QCI procedure is visible in the simple connections that run very quickly with all the interpolation procedures. For the most difficult connections, namely TS3 with AMBER and trpzip1, the QCI interpolation is the fastest among the methods that give zero failures. In general we would expect the biggest improvements to arise for the most complicated pathways. The results for different potentials can be quite system specific, since the alternative parametrizations could give very different initial paths when the end points are distant in configuration space. A detailed discussion of the characteristics observed for different potentials can be found elsewhere.⁷³ The principal advantage of QCI is that it does not require prior identification of internal coordinates, which may be problematic for systems containing separate molecules, for example. To facilitate future comparisons, all the necessary input files, along with the OPTIM output in each case, will be made available for download from the OPTIM Web site.⁴⁶

4. CONCLUSIONS

The quasi-continuous interpolation (QCI) procedure can identify physically realistic initial paths between fixed end point configurations in a variety of systems. QCI is a chain-of-states method, involving geometry optimization of all of the coupled images of the system to locate a pathway that satisfies various constraint conditions. During the minimization procedure for the chain of states, additional terms are added to the potential energy function, which account for constraint violations between the images. For synchronous motion between images, local minima in the distance between any given pair of atoms can be identified analytically. The auxiliary interpolation potential is evaluated for these minimum distances, and corresponding penalty terms are added to the total energy and gradient. The constraint violations may occur at different points along the line segment between images for different pairs of atoms, and all such terms are included. Hence we account for atom clashes or chain crossings that occur between images.

The pairwise interpolation potential can be evaluated for any subset of atoms, which enables paths to be constructed by adding one or more atoms at a time. Once a path that satisfies all the bond distance and atom–atom repulsion constraints has been obtained, further refinement is allowed using a combination of the interpolation function and the true potential. Regularly spaced configurations along the path between images are then used as the starting guess for a doubly-nudged³⁹ elastic band^{17–22} calculation, and local maxima in the DNEB profile are tightly converged to transition states using hybrid eigenvector-following.^{19,40,41} This approach has been tested and benchmarked for an atomic cluster, selected pathways of trpzip2, a coarse-grained model of protein L, and atomistic representations of two peptides using both

Table 2. Timings for the MSB Test Set Introduced in Previous Work^{73a}

interpolation	GNNQQNY			trpzip1		
	TS1	TS2	TS3	TS1	TS2	TS3
CHARMM						
QCI	6.3	14.0	20.4	29.6	57.8	91.1
DNEB	3.0	8.6	14.7	30.0	62.9	115.0
BCSC	3.1	8.8	14.7	30.7	62.0	112.4
BCSI	3.1	9.4	13.1	30.6	50.8	90.8
BISI	3.7	12.0	19.7	47.9	96.9(1)	116.8
NONI	3.4	9.1	13.6	30.2	78.0	92.9(1)
NI	3.7	9.1	13.5	30.7	78.1	90.9(1)
NIS	3.7	8.8	14.4	30.9	74.9(1)	94.9(1)
NIC	3.7	8.5	13.6	29.9	55.9	106.1(1)
AMBER						
QCI	37.5	79.7	92.2	214.2	412.4	562.8
DNEB	31.5	77.2	132.0(1)	251.5	516.2	576.7(3)
BCSC	31.6	76.1	131.4(1)	243.8	511.7	590.2(3)
BCSI	31.5	77.4	131.2(1)	245.2	518.2	590.5(3)
BISI	31.5	76.2	131.5(1)	244.8	524.2	584.6(3)
NONI	31.9	82.7	79.6(1)	177.4(2)	439.2	703.7
NI	32.1	82.4	79.9(1)	175.4(2)	445.7	690.7
NIS	31.8	72.2(1)	80.0	197.4	420.5	730.5
NIC	30.4	78.4	75.7	205.2(1)	360.9	614.2

^aQCI and DNEB refer to interpolation using the QCI approach described in the text and to images placed along an initial straight line path for the doubly-nudged³⁹ elastic band^{17–22} procedure, respectively. The other abbreviations are defined in Table 1. The average CPU time (s) for an Intel Xeon E5404 processor (running at 2.0 GHz) is given in each case; a number in parentheses indicates that there were one or more failures for this combination when a maximum of 50 cycles of the missing connection algorithm⁴² was allowed. The number of tests in each category is 150, except for sets TS1 and TS2 for GNNQQNY with CHARMM, where 300 paths each were considered. All the input files required to run these OPTIM benchmarks will be made available from the OPTIM Web site,⁴⁶ together with the corresponding output.

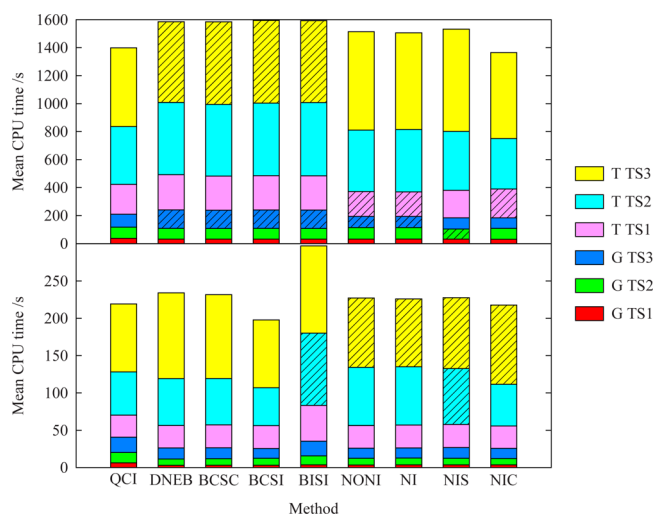


Figure 8. Stacked bar charts displaying the results from Table 2. The average CPU time (s) using an Intel Xeon E5404 processor (running at 2.0 GHz) is plotted for the pathways in the MSB test set⁷³ and each interpolation method. The results with AMBER are shown in the upper panel, and those with CHARMM are in the lower panel. In the legend, G and T refer to GNNQQNY and trpzip1, respectively. Stripes within a box indicate that there were one or more failures for this combination within 50 cycles of the missing connection algorithm⁴² and therefore that the box height shown is a lower bound.

AMBER^{5–7} and CHARMM^{1–4} force fields. For the most difficult peptide connections the QCI approach was somewhat faster than schemes that use internal coordinates. However, the main benefit is realized in the construction of kinetic transition

networks for folding and unfolding of protein L and protein G, where chain crossings were frequently obtained without QCI.⁵⁴

Efficient construction of pathways using geometry optimization for atomistic representations of peptides and proteins depends upon proper treatment of permutational isomers. In previous work, we have shown how to automate the symmetrization of CHARMM and AMBER force fields so that consistent energies are obtained for equivalent configurations.^{39,51} It is also necessary to align the end points in double-ended calculations, to avoid unnecessary and potentially unphysical rearrangements. The shortest augmenting path procedure employed in previous work^{47–49,73} is improved in the present contribution using local distance metrics by defining suitable neighborhoods. The local permutational alignment solves the initial interpolation problem for cases where the overall minimum distance for all atoms produces additional permutational rearrangements on the pathway. For difficult cases involving distant end points local permutational alignment can produce successful connections where previous procedures have failed or required excessive computer time.

■ APPENDIX: DETAILS OF THE QCI PROCEDURE

The Interpolation Potential

Here, we define the auxiliary interpolation potential employed for all of the biomolecular examples in section 3, which consists of constraint and repulsive terms for pairs of atoms, α and β . The complete potential for a set of interpolating images, labelled by superscripts i and j , with configurations \mathbf{r}^i and \mathbf{r}^j , includes all of the pairwise terms evaluated at each image. The potential also includes terms corresponding to local minima in the distance that appear for pairs of atoms on straight line paths

between the images, as explained below. The latter feature provides the quasi-continuous part of the interpolation.

We denote the three-dimensional position vectors for atoms α and β in configuration i as \mathbf{r}_α^i and \mathbf{r}_β^i etc. The line segments joining the atomic images in two configurations i and j can be written as

$$\begin{aligned}\mathbf{r}_\alpha &= \mathbf{r}_\alpha^i \sin^2 \theta + \mathbf{r}_\alpha^j \cos^2 \theta \\ \mathbf{r}_\beta &= \mathbf{r}_\beta^i \sin^2 \theta + \mathbf{r}_\beta^j \cos^2 \theta\end{aligned}\quad (1)$$

so the distance between atoms α and β as a function of the interpolation angle $0 \leq \theta \leq \pi/2$ is

$$\begin{aligned}[d_{\alpha\beta}^{ij}(\theta)]^2 &= (\mathbf{r}_\alpha - \mathbf{r}_\beta)^2 \\ &= \mathbf{r}_\alpha^2 + \mathbf{r}_\beta^2 - 2\mathbf{r}_\alpha \cdot \mathbf{r}_\beta \\ &= |\mathbf{r}_\alpha^i|^2 \sin^4 \theta + |\mathbf{r}_\alpha^j|^2 \cos^4 \theta + 2\mathbf{r}_\alpha^i \cdot \mathbf{r}_\alpha^j \sin^2 \theta \cos^2 \theta + |\mathbf{r}_\beta^i|^2 \sin^4 \theta + |\mathbf{r}_\beta^j|^2 \cos^4 \theta \\ &\quad + 2\mathbf{r}_\beta^i \cdot \mathbf{r}_\beta^j \sin^2 \theta \cos^2 \theta \\ &\quad - 2(\mathbf{r}_\alpha^i \cdot \mathbf{r}_\beta^i \sin^4 \theta + \mathbf{r}_\alpha^j \cdot \mathbf{r}_\beta^j \cos^4 \theta + \sin^2 \theta \cos^2 \theta [\mathbf{r}_\alpha^i \cdot \mathbf{r}_\beta^j + \mathbf{r}_\alpha^j \cdot \mathbf{r}_\beta^i])\end{aligned}\quad (2)$$

Solutions of $\partial[d_{\alpha\beta}^{ij}(\theta)]^2/\partial\theta = 0$ always exist for $\sin \theta = 0$ and $\cos \theta = 0$, i.e., $\theta = 0$ and $\pi/2$ in the range considered. The third root of the cubic equation is θ^* , where

$$\begin{aligned}\cos^2 \theta^* &= \frac{(\mathbf{r}_\alpha^i - \mathbf{r}_\beta^i) \cdot (\mathbf{r}_\alpha^i - \mathbf{r}_\beta^j - \mathbf{r}_\alpha^j + \mathbf{r}_\beta^j)}{|\mathbf{r}_\alpha^i - \mathbf{r}_\beta^j - \mathbf{r}_\alpha^j + \mathbf{r}_\beta^j|^2} \text{ or} \\ \sin^2 \theta^* &= \frac{(\mathbf{r}_\alpha^j - \mathbf{r}_\beta^j) \cdot (-\mathbf{r}_\alpha^i + \mathbf{r}_\beta^i + \mathbf{r}_\alpha^i - \mathbf{r}_\beta^i)}{|\mathbf{r}_\alpha^i - \mathbf{r}_\beta^j - \mathbf{r}_\alpha^j + \mathbf{r}_\beta^j|^2}\end{aligned}\quad (3)$$

with $\cos^2 \theta^* + \sin^2 \theta^* = 1$.

The solutions for $\theta = 0$ and $\theta = \pi/2$ correspond to the limits of the range with

$$\begin{aligned}[d_{\alpha\beta}^{ij}(0)]^2 &= |\mathbf{r}_\alpha^j|^2 + |\mathbf{r}_\beta^j|^2 - 2\mathbf{r}_\alpha^j \cdot \mathbf{r}_\beta^j \equiv [d_{\alpha\beta}^j]^2 \text{ and} \\ [d_{\alpha\beta}^{ij}(\pi/2)]^2 &= |\mathbf{r}_\alpha^i|^2 + |\mathbf{r}_\beta^i|^2 - 2\mathbf{r}_\alpha^i \cdot \mathbf{r}_\beta^i \equiv [d_{\alpha\beta}^i]^2\end{aligned}\quad (4)$$

respectively, where $d_{\alpha\beta}^i$ is the distance between atoms α and β in configuration i . If the solutions for $\cos^2 \theta^*$ and $\sin^2 \theta^*$ lie between 0 and 1, then $[d_{\alpha\beta}^{ij}(\theta)]^2$ exhibits a minimum in the range $0 < \theta < \pi/2$, with local maxima at the end points (Figure 9). Otherwise, one end point is a local maximum and the other is a local minimum. For the internal minimum, we find

$$\begin{aligned}\frac{\partial[d_{\alpha\beta}^{ij}(\theta^*)]^2}{\partial x_\alpha^i} &= \frac{2[(\mathbf{r}_\alpha^i - \mathbf{r}_\beta^i) \cdot (\mathbf{r}_\alpha^j - \mathbf{r}_\beta^j)]^2 - |\mathbf{r}_\alpha^i - \mathbf{r}_\beta^j|^2 |\mathbf{r}_\alpha^j - \mathbf{r}_\beta^j|^2 [x_\alpha^i - x_\beta^j - x_\alpha^j + x_\beta^i]}{|\mathbf{r}_\alpha^i - \mathbf{r}_\beta^j - \mathbf{r}_\alpha^j + \mathbf{r}_\beta^j|^4} \\ &\quad + \frac{2[(x_\alpha^i - x_\beta^j) |\mathbf{r}_\alpha^j - \mathbf{r}_\beta^j|^2 + (\mathbf{r}_\alpha^i - \mathbf{r}_\beta^i) \cdot (\mathbf{r}_\alpha^j - \mathbf{r}_\beta^j) (x_\beta^j - x_\alpha^i)]}{|\mathbf{r}_\alpha^i - \mathbf{r}_\beta^j - \mathbf{r}_\alpha^j + \mathbf{r}_\beta^j|^2}\end{aligned}\quad (7)$$

with analogous expressions for y_α^i and z_α^i .

Various possibilities have been tested for the construction of the interpolation function. For example, harmonic springs were included by adding contributions proportional to $|\mathbf{r}_\alpha^i - \mathbf{r}_\alpha^{i+1}|^2$ for adjacent images, as for previous chain-of-states methods.^{17–22,39} Aside from the all-atom representation of trpzp1 with AMBER,

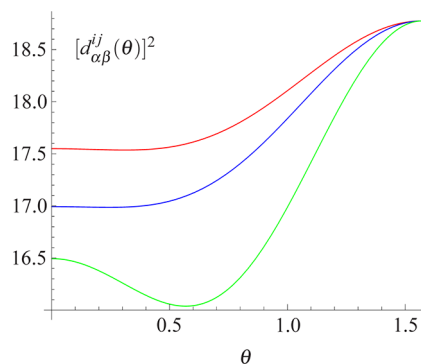


Figure 9. Squared distance (arbitrary units) between two atoms \mathbf{r}_α and \mathbf{r}_β as a function of θ (radians). As the z coordinate of \mathbf{r}_β^j , atom β in configuration j , varies, an internal minimum develops.

$$\begin{aligned}[d_{\alpha\beta}^{ij}(\theta^*)]^2 &= \frac{|\mathbf{r}_\alpha^i - \mathbf{r}_\beta^j|^2 |\mathbf{r}_\alpha^j - \mathbf{r}_\beta^i|^2 - [(\mathbf{r}_\alpha^i - \mathbf{r}_\beta^j) \cdot (\mathbf{r}_\alpha^j - \mathbf{r}_\beta^i)]^2}{|\mathbf{r}_\alpha^i - \mathbf{r}_\beta^j - \mathbf{r}_\alpha^j + \mathbf{r}_\beta^i|^2}\end{aligned}\quad (5)$$

In the quasi-continuous interpolation (QCI) procedure, the basic idea is to add contributions to the energy corresponding to every internal minimum in the distances $[d_{\alpha\beta}^{ii+1}(\theta^*)]^2$ for images i and $i + 1$. The continuous path between end points consists of the M images and the straight line segments between them. The associated energy functional includes contributions from all of the images, along with terms corresponding to minima, occurring between images, in the distances between pairs of atoms. These local minima will generally occur for different configurations along each line segment. If the interpolation potential, V_{int} , is a sum over pairwise additive functions of the interatomic distances, V_1 , V_2 , etc., and we have an ordered chain of M configurations $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_M$, then we minimize the sum

$$V_{\text{int}} = \sum_{i=2}^{M-1} \sum_p \sum_{\alpha < \beta} V_p[d_{\alpha\beta}^i] + \sum_{i=1}^{M-1} \sum_p \sum_{\alpha < \beta}^* V_p[d_{\alpha\beta}^{ii+1}(\theta^*)]\quad (6)$$

where configurations 1 and M are the fixed end points and the second term is over all of the distances that exhibit internal minima between consecutive images. The sum over p accounts for different types of contribution to V_{int} (see below). To minimize V_{int} , we require the derivatives with respect to all the Cartesian coordinates of atoms in the different images. Let x_α^i be the x coordinate of atom α in image i . Then,

these spring terms were not required. However, the examples provided for benchmarking in section 3.4 employ the same set of parameters for each QCI calculation, and springs were added between QCI images with a force constant of unity in each case. The parameter sets are therefore closest to optimal for the

AMBER trpzip1 benchmarks but provide acceptable performance for the simpler examples as well.

In addition to the spring terms between images, the interpolation potential, V_{int} , consisted of two further contributions, the first to maintain the conserved structure, and the second to keep unconstrained atoms apart. The first, constraint, term uses harmonic springs between individual atoms to fix interatomic distances that are preserved to within a given tolerance in the two end points. Related constructions have previously been employed in elastic network-type³⁸ treatments of large-amplitude motion^{90–107} and double-well Gō-type models.^{108–112} Constraints were applied for atoms α and β only if the distances in the end points, $d_{\alpha\beta}^1$ and $d_{\alpha\beta}^M$, were both less than 5 Å (5 σ for the atomic cluster), and the change in distance $|d_{\alpha\beta}^1 - d_{\alpha\beta}^M| < \Delta$. For the peptides and proteins considered in the present work, the constraints need to define a percolating network, and the input parameter Δ was increased by 10% until this condition was achieved. Percolation was diagnosed by calculating the number of steps from one selected atom to all the others using a depth first search.⁶¹ Here, each constraint was taken as an edge connecting nodes of a graph defined by all the atoms. The reference separation for constrained atoms was taken as the average value $\bar{d}_{\alpha\beta} = (d_{\alpha\beta}^1 + d_{\alpha\beta}^M)/2$. For each constraint the corresponding term in the interpolation potential was set to

$$V_{\text{con}}(d_{\alpha\beta}^i) = \begin{cases} \frac{\epsilon_{\text{con}}[(d_{\alpha\beta}^i - \bar{d}_{\alpha\beta})^2 - (C_{\alpha\beta}^{\text{con}})^2]^2}{2(C_{\alpha\beta}^{\text{con}})^2} & |d_{\alpha\beta}^i - \bar{d}_{\alpha\beta}| > C_{\alpha\beta}^{\text{con}} \\ 0 & |d_{\alpha\beta}^i - \bar{d}_{\alpha\beta}| \leq C_{\alpha\beta}^{\text{con}} \end{cases} \quad (8)$$

where parameters ϵ_{con} and $C_{\alpha\beta}^{\text{con}}$ determine the strength of the constraint and the width of the interval around $\bar{d}_{\alpha\beta}$ for which V_{con} vanishes. $C_{\alpha\beta}^{\text{con}}$ is chosen large enough so that the constraint term vanishes for the end points and is increased if necessary when new minima are found so that V_{con} also vanishes for these structures. Initially, we simply set $C_{\alpha\beta}^{\text{con}} = |d_{\alpha\beta}^1 - d_{\alpha\beta}^M|/2$. The chosen form provides an even function of $d_{\alpha\beta}^i - \bar{d}_{\alpha\beta}$ for which $V_{\text{con}}(d_{\alpha\beta}^i)$ and its first derivative vanish at $d_{\alpha\beta}^i = \bar{d}_{\alpha\beta} \pm C_{\alpha\beta}^{\text{con}}$. These properties were found to improve the efficiency in numerical minimizations of V_{int} ; distances that are shorter or longer than $\bar{d}_{\alpha\beta}$ are penalized in a symmetrical fashion.

For biomolecules, constraints were restricted to atoms sufficiently close in sequence, as defined by a parameter n_1 . Here, the sequence corresponds to the atom ordering in the CHARMM or AMBER input files, which is based upon the molecular topology. A value of $n_1 = 15$ was used in calculations for peptides and proteins when the constraints were defined on the basis of the separations in the two end points alone. This restriction avoids constraining atoms distant in sequence that happen to have very similar separations in the reference minima. When constraints are defined using a larger database of local minima, this condition is very unlikely to occur, and n_1 was set to the total number of atoms.

The repulsive contribution to V_{int} was applied to prevent unphysical geometries appearing along the line segments that define the interpolating path. Repulsions were only included for atoms far enough apart in sequence, defined by the condition $|\alpha - \beta| > n_2$ for a second sequence separation parameter n_2 . In fact, setting $n_2 = 0$ proved to be satisfactory for all the tests described in section 3. Repulsions were not allowed for atoms involved in any of the distance constraint terms

described above. A cutoff distance was defined in terms of the minimum end point separation minus a small displacement, $0 < \delta \ll 1$, and an input parameter, C , with $C_{\alpha\beta}^{\text{rep}} = \min(d_{\alpha\beta}^1 - \delta, d_{\alpha\beta}^M - \delta, C)$. For this choice, the atoms in question do not repel one another in the fixed images, 1 and M . For each pair of atoms satisfying these conditions, the repulsion was defined as

$$V_{\text{rep}}[d_{\alpha\beta}^i] = \begin{cases} \frac{\epsilon_{\text{rep}}((C_{\alpha\beta}^{\text{rep}})^3 - 3C_{\alpha\beta}^{\text{rep}}(d_{\alpha\beta}^i)^2 + 2(d_{\alpha\beta}^i)^3)}{(C_{\alpha\beta}^{\text{rep}})^3(d_{\alpha\beta}^i)^2} & d_{\alpha\beta}^i \leq C_{\alpha\beta}^{\text{rep}} \\ 0 & d_{\alpha\beta}^i > C_{\alpha\beta}^{\text{rep}} \end{cases} \quad (9)$$

This form again ensures that V_{rep} and its gradient vanish at the cutoff $C_{\alpha\beta}^{\text{rep}}$, which can depend upon the pair of atoms in question. Analytical derivatives of both V_{con} and V_{rep} are easily formulated using the chain rule and eq 7. A neighbor list was constructed for V_{rep} to avoid referencing pairs that lie outside the cutoff in every iteration. The list included all pairs of atoms approaching within a distance of $\zeta C_{\alpha\beta}^{\text{rep}}$ for any image or intervening minimum. $\zeta = 2$ proved to be generally effective, with the neighbor list updated every 20 minimization steps and whenever the number of active atoms changed.

When only two end points were specified, constraints were only included for atoms close enough in sequence by the condition $|\alpha - \beta| < n_1$ for the sequence separation parameter n_1 , introduced above. However, a more general procedure makes use of previously determined stationary points or configurations to define the constraint and repulsive terms. This approach was employed for the benchmarking described in section 3.4 for the peptides GNNQQNY and trpzip1. Constraints were applied using the percolation test described above for all of the minima in the test set, and the corresponding pairs and distances were saved to define the potential in all subsequent runs. The pairs of atoms corresponding to repulsive interactions were also recorded, along with the reference constraint distances $\bar{d}_{\alpha\beta}$ and cutoffs $C_{\alpha\beta}^{\text{con}}$ and $C_{\alpha\beta}^{\text{rep}}$. All cutoffs were defined large enough to give $V_{\text{con}} = 0$ and $V_{\text{rep}} = 0$ for all the local minima encountered and increased if necessary during the run. The reference distance for constrained atoms α and β was set to $\bar{d}_{\alpha\beta} = (\max_q d_{\alpha\beta}^q - \min_q d_{\alpha\beta}^q)/2$ with cutoff $C_{\alpha\beta}^{\text{con}} = (\max_q d_{\alpha\beta}^q - \min_q d_{\alpha\beta}^q)/2$, where \max_q and \min_q refer to the maximum and minimum separation over the set of reference structures, q . The cutoff for repulsive interactions between any pair of atoms was set to the smallest separation between the pair encountered in any of the reference minima, or in any new minima subsequently encountered.

Since the interpolation potential is defined with respect to specific atom–atom contacts, it is essential to employ consistent permutational isomers throughout the procedure. When multiple minima were used to define the potential, all structures were aligned with respect to the first reference minimum using the procedures described in section 2.2. When only two end points were available, the first one was taken as the reference.

The formulation of the interpolation potential to include contributions corresponding to local minima in the interatomic distance prevents the line segments between images from cutting through unfavorable regions of the true potential. Hence, the problems of corner-cutting and sliding-down¹¹³ that occur in discrete elastic band-type representations^{17–22,39} do not appear directly for the QCI potential. To prevent images becoming too close together or too far apart, new images were

added where large gaps appeared, and images were combined according to distance thresholds. This procedure is described below.

Growing an Interpolated Structure

The next step, also programmed in the OPTIM package,⁴⁶ is to define the percolating network of distance constraints in V_{con} . An initial guess could then be generated from a straight line interpolation between the two end points using all of the atoms. However, for the most difficult cases involving chain crossings, it proved more effective to add one atom at a time to a growing chain. The pairwise design of V_{int} makes it straightforward to deal with a subset of atoms, where the constraints and repulsions are progressively turned on as atoms are added.

To start the initial interpolation procedure, the two atoms involved in a distance constraint that move least between the end points in the fixed images \mathbf{r}^1 and \mathbf{r}^M were identified. Having found the constrained atoms that minimize $|\mathbf{r}_\alpha^1 - \mathbf{r}_\alpha^M| + |\mathbf{r}_\beta^1 - \mathbf{r}_\beta^M|$, $M - 2$ images were created for atoms α and β at regular intervals between the end points by linear interpolation. Atoms were then added sequentially when local convergence criteria for V_{int} were satisfied for the current set of active atoms, N_{active} . It is usually possible to achieve paths with $V_{\text{int}} = 0$, where all the distance constraints are satisfied within the tolerance $C_{\alpha\beta}^{\text{con}}$, and all the repulsive interactions lie outside $C_{\alpha\beta}^{\text{rep}}$. A new atom was therefore added when $V_{\text{int}} < V_{\text{int}}^{\text{max}}$ and the root mean-square (RMS) gradient for the $3(M - 2) \times N_{\text{active}}$ degrees of freedom fell below a tolerance parameter $G_{\text{int}}^{\text{max}}$. Some element of randomness was introduced into this selection, and the initial interpolation for the new atom, to avoid repeating failed interpolation attempts in the backtracking procedure described below. For each inactive atom, γ , the number of constraints to current active atoms, ε , and the sum of reciprocal reference distances, $s_\gamma = \sum_{\varepsilon \in \text{active}} 1/\bar{d}_{\varepsilon\gamma}$, were calculated. The probability of adding atom γ to the active list was then chosen proportional to s_γ/t_γ^4 , where t_γ was the number of times this atom had been tried before plus one. This scheme was chosen empirically after comparing various other possibilities.

To provide an interpolation for the next active atom, several schemes were compared. For $N_{\text{active}} \geq 3$, various possibilities were considered using active atoms to define a local orthogonal coordinate system. A sorted list of all of the active atoms was constructed on the basis of the average distance to the new atom for the two fixed end points. For three selected active atoms, the components of the displacement vector for the new atom were calculated for the first end point. The interpolated position of the new atom was then obtained using the same components in the orthogonal coordinate system constructed from the same three active atoms in each image configuration 2, 3, ..., $M - 1$. For $N_{\text{active}} > 3$, active atoms were selected according to probabilities based on $1/o_\varepsilon^2$, where o_ε was the order in which atom ε appeared in the sorted list of average end point distances. Various other selection schemes were also tested. For example, active atoms were also ordered according to how well their distance to the new atom was preserved in the end points, using $|d_{\varepsilon\gamma}^1 - d_{\varepsilon\gamma}^M|$. In addition, the V_{int} potential was also calculated using a coordinate system constructed from the three closest active atoms involved in constrained distances to the new atom. A fourth scheme simply interpolated the position based on the position of the active atom with the shortest $\bar{d}_{\varepsilon\gamma}$ constraint distance in each image. The initial position of the new atom in each image was then taken as a displacement from this active atom, weighting the displacement vectors in images 1

and M according to the image position in the chain of configurations.

The initial interpolation for each new atom was chosen according to the lowest initial value of V_{int} for the four schemes described above. V_{int} was then minimized using the limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm of Liu and Nocedal,^{114,115} reinitializing each time N_{active} increased. For examples involving possible chain crossings, some cases were found where the convergence conditions on V_{int} and the RMS gradient were not achieved within the permitted number of refinement steps, $L_{\text{max}}^{(1)}$. Such cases were treated by backtracking, removing the last N_{back} atoms added to the active set, and adding one new one using the above scheme. N_{back} was increased by one for every convergence failure but was not allowed to exceed $\min(20, N_{\text{active}} - 2)$.

If N_{active} failed to reach N before a total of $L_{\text{max}}^{(2)}$ LBFGS iterations, then the interpolation was abandoned. Otherwise, the interpolation proceeded in a second phase for a further $L_{\text{max}}^{(3)}$ steps using a new objective function $(1 - x)V_{\text{int}} + xV_{\text{true}}$. The parameter x is the fraction of the true potential to use, which can be one. However, the overhead for using the relatively simple interpolating function in this stage was generally small, and retaining the fraction $1 - x$ of V_{int} was helpful in preventing the interpolation from collapsing at the start of the refinement, when some of the derivatives of V_{true} could still be large.

Separate convergence parameters were also specified for the minimization of $(1 - x)V_{\text{int}} + xV_{\text{true}}$ in all $3N(M - 2)$ degrees of freedom, with a maximum number of additional steps, $L_{\text{max}}^{(3)}$. The objective here is not to converge the interpolation accurately but to supply starting configurations for further refinement.^{39,116} In fact, the value of the objective function must be allowed to rise during the minimization procedure, since internal minima can appear (and disappear) in the interatomic distances.

Adding, Deleting, and Freezing QCI Images

Some further refinements of the interpolation procedure were used to improve efficiency, namely dynamic removal and addition of images and provision for frozen images. Starting from a specified number of images, M , images were removed if an image spacing became too close (less than a parameter D_{min}), and a bisecting image was added for image spacings that exceeded a specified distance, D_{max} , up to a specified maximum number of images, M_{max} . This dynamic adjustment was also permitted during the second-phase minimization of the function $(1 - x)V_{\text{int}} + xV_{\text{true}}$. The corresponding degrees of freedom were simply removed from the LBFGS memory, and new degrees of freedom were initialized using average values from the neighboring images. Images were frozen if the RMS gradient fell below a tolerance of $G_{\text{int}}^{\text{freeze}}$. The gradient terms were simply set to zero for frozen images with frozen neighbors but calculated normally for frozen images with at least one unfrozen neighbor to check whether any images should be unfrozen.

Freezing Atoms

For large systems, where the majority of atoms do not move much throughout the pathway, the interpolation can be speeded up significantly by treating these ‘spectator’ atoms separately. A simple linear interpolation was used for such atoms, and their interpolated positions were frozen during the initial phase (i.e. using only V_{int}), if the atomic displacement between aligned end points fell below a specified cutoff. To define constraints in this framework, the depth first search

employed to diagnose a percolating network was modified to admit all of the frozen atoms as root nodes. Constraint and repulsive terms were included between active and frozen atoms but not between pairs of frozen atoms, thus reducing the computational cost of evaluating the interpolation potential. However, it was also necessary to keep a minimum number of active atoms, since some parts of the molecule can end up back almost where they started, but must be allowed to move significantly to accommodate the motion of interest. For the benchmarking in section 3.4, linear interpolation was used if the aligned end point–end point distance for a given atom fell below 0.5 Å, and a minimum of 15 active atoms were retained.

In trial calculations, schemes involving a maximum number of distance constraints were considered for each active atom. In some cases, the interpolation phase can succeed using all of the constraints identified. However, the most efficient setting identified for the benchmarks in section 3.4 used a maximum of six constraints per atom. Three constraints proved to be insufficient to prevent spurious rotation of trp groups in the benchmarking for trpzip1 (section 3.4). Interpolations were also considered in which contributions from internal minima in $d_{\alpha\beta}^{ij}$ were only added for V_{rep} . However, the results generally appear better when the corresponding terms are also included for V_{con} .

Interpolation Potential for an Atomic Cluster

For clusters bound by the Lennard-Jones potential,⁷⁵ V_{int} was defined as

$$V_{\text{int}} = \sum_{i=2}^{M-1} \sum_{\alpha < \beta} V_{\text{LJ}}[d_{\alpha\beta}^i] + \sum_{i=1}^{M-1} \sum_{\alpha < \beta}^* V_{\text{rep}}[d_{\alpha\beta}^{ii+1}(\theta^*)] \quad (10)$$

where V_{LJ} is the usual Lennard-Jones form and

$$V_{\text{rep}}[d_{\alpha\beta}^{ii+1}(\theta^*)] = \frac{\epsilon_{\text{rep}}(d_{\alpha\beta}^i - d_{\alpha\beta}^{ii+1}(\theta^*) - \delta)^2 (d_{\alpha\beta}^{i+1} - d_{\alpha\beta}^{ii+1}(\theta^*) - \delta)^2}{[d_{\alpha\beta}^{ii+1}(\theta^*)]^4} \quad (11)$$

for $d_{\alpha\beta}^i - d_{\alpha\beta}^{ii+1}(\theta^*) \geq \delta$ and $d_{\alpha\beta}^{i+1} - d_{\alpha\beta}^{ii+1}(\theta^*) \geq \delta$ and zero otherwise. Reduced units for the LJ potential are employed throughout this section. The form of V_{rep} was chosen to provide a smooth function and first derivatives. The parameter δ was included to prevent numerical instabilities that might occur in the limits where the internal minimum is very shallow and close to one of the images. The results reported above employed $M = 11$ initial images, a maximum of $L_{\text{max}}^{(2)} = 2000$ LBFGS steps, $G_{\text{int}}^{\text{max}} = 10^{-3}$, $\epsilon_{\text{rep}} = 1$, and $\delta = 10^{-6}$. However, sensible pathways were also obtained with $\delta = 0$. No constraints were applied, and the complete V_{int} including all atoms was used throughout the minimization. This procedure is therefore simpler than the two-phase scheme employed for the biomolecules considered in this report, in which firstly V_{int} and secondly $(1 - x)V_{\text{int}} + xV_{\text{true}}$ were taken as the functions to be minimized. Dynamic addition, deletion, and freezing of images was included, as described above, with $D_{\text{min}} = 0.1$, $D_{\text{max}} = 2.5$, and $G_{\text{int}}^{\text{freeze}} = 10^{-10}$. The QCI images were again used to seed DNEB calculations followed by single-ended hybrid eigenvector-following transition state refinement.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dw34@cam.ac.uk; jmc49@cam.ac.uk.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to Dr. Thomas Hofer and Dr. Christopher Whittleston for supplying the trpzip2 end points and to Dr. Tomas Oppelstrup for his interface to the shortest augmenting path procedure.

REFERENCES

- (1) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (2) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902–1921.
- (3) MacKerell, A. D., Jr.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (4) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86–104.
- (5) Case, D. et al. *AMBER 9*; University of California: Oakland, CA, 2006.
- (6) Pearlman, D.; Case, D.; Caldwell, J.; Ross, W.; Cheatham, T., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1–41.
- (7) Case, D.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (8) Brown, S.; Fawzi, N. J.; Head-Gordon, T. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 10712–10717.
- (9) Brown, S.; Head-Gordon, T. *Protein Sci.* **2004**, *13*, 958–970.
- (10) Noé, F.; Krachtus, D.; Smith, J. C.; Fischer, S. *J. Chem. Theory Comput.* **2006**, *2*, 840–857.
- (11) Wales, D. J. *Int. Rev. Phys. Chem.* **2006**, *25*, 237–282.
- (12) Noé, F.; Fischer, S. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154–162.
- (13) Prada-Gracia, D.; Gómez-Gardenes, J.; Echenique, P.; Fernando, F. *PLoS Comput. Biol.* **2009**, *5*, 1–9.
- (14) Wales, D. J. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3–10.
- (15) Wales, D. J. *Mol. Phys.* **2002**, *100*, 3285–3306.
- (16) Wales, D. J. *Mol. Phys.* **2004**, *102*, 891–908.
- (17) Elber, R.; Karplus, M. *Chem. Phys. Lett.* **1987**, *139*, 375–380.
- (18) Czerminski, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580–5601.
- (19) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **1999**, *111*, 7010–7022.
- (20) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9901–9904.
- (21) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9978–9985.
- (22) Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **2001**, *115*, 9657–9666.
- (23) Nguyen, C. N.; Stratt, R. M. *J. Chem. Phys.* **2010**, *133*, 124503.
- (24) Faccioli, P. *J. Phys. Chem. B* **2008**, *112*, 13756–13764.
- (25) Faccioli, P.; Lonardi, A.; Orland, H. *J. Chem. Phys.* **2010**, *133*, 045104.
- (26) Schlegel, H. B. Geometry Optimization: 1. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., Schreiner, P. R., Eds.; John Wiley and Sons: Chichester, West Sussex, U.K., 1998; Vol. 2, pp 1136–1142.
- (27) Farkas, O.; Schlegel, H. B. *J. Mol. Struct.* **2003**, *666–667*, 31–39.
- (28) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 5578–5583.
- (29) Snow, C. D.; Qiu, L.; Gai, F.; Hagen, S. J.; Pande, V. S. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 4077–4082.
- (30) Balbirnie, M.; Grothe, R.; Eisenberg, D. S. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 2375–2380.
- (31) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D. *Nature* **2005**, *435*, 773–778.
- (32) Snow, C. D.; Qiu, L.; Du, D.; Gai, F.; Hagen, S. J.; Pande, V. S. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 4077–4082.

- (33) Murrell, J. N.; Laidler, K. J. *Trans. Faraday Soc.* **1968**, *64*, 371–377.
- (34) Walsh, T. R.; Wales, D. J. *J. Chem. Soc., Faraday Trans.* **1996**, *92*, 2505–2517.
- (35) Halgren, T. A.; Lipscomb, W. N. *Chem. Phys. Lett.* **1977**, *49*, 225–232.
- (36) E, W.; Ren, W.; Vanden-Eijnden, E. *Phys. Rev. B* **2002**, *66*, 052301.
- (37) Peters, B.; Bell, A. T.; Chakraborty, A. *J. Chem. Phys.* **2004**, *121*, 4453–4460.
- (38) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77*, 1905–1908.
- (39) Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082–2094.
- (40) Munro, L. J.; Wales, D. J. *Phys. Rev. B* **1999**, *59*, 3969–3980.
- (41) Kumeda, Y.; Munro, L. J.; Wales, D. J. *Chem. Phys. Lett.* **2001**, *341*, 185–194.
- (42) Carr, J. M.; Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2005**, *122*, 234903.
- (43) Dijkstra, E. W. *Numer. Math.* **1959**, *1*, 269–271.
- (44) Kearsley, S. K. *Acta Crystallogr., Sect. A* **1989**, *45*, 208–210.
- (45) Jonker, R.; Volgenant, A. *Computing* **1987**, *38*, 325–340.
- (46) Wales, D. J. *OPTIM: A program for optimizing geometries and calculating pathways*.
- (47) Khalili, M.; Wales, D. J. *J. Phys. Chem. B* **2008**, *112*, 2456–2465.
- (48) Carr, J. M.; Wales, D. J. *J. Phys. Chem. B* **2008**, *112*, 8760–8769.
- (49) Carr, J. M.; Wales, D. J. *J. Phys. Chem. Chem. Phys.* **2009**, *11*, 3341–3354.
- (50) This procedure corresponds to the keyword PERMDIST in the GMIN, OPTIM, and PATHSAMPLE programs.
- (51) Malolepsza, E.; Strodel, B.; Khalili, M.; Trygubenko, S.; Fejer, S.; Wales, D. J. *J. Comput. Chem.* **2010**, *31*, 1402–1409.
- (52) James, T.; Wales, D. J.; Rojas, J. H. *J. Chem. Phys.* **2007**, *126*, 054506.
- (53) Chakrabarti, D.; Wales, D. J. *Soft Matter* **2011**, *7*, 2325–2328.
- (54) Wales, D. J.; Head-Gordon, T. *J. Phys. Chem. B* **2012**, *116*, 8394–8411.
- (55) Rafac, R.; Schiffer, J. P.; Hangst, J. S.; Dubin, D. H. E.; Wales, D. J. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 483–486.
- (56) Wales, D. J.; Lee, A. M. *Phys. Rev. A* **1993**, *47*, 380–393.
- (57) Yurtsever, E.; Calvo, F.; Wales, D. J. *Phys. Rev. E* **2005**, *72*, 026110.
- (58) Calvo, F.; Yurtsever, E.; Wales, D. J. *J. Chem. Phys.* **2012**, *136*, 024303–024303.
- (59) This procedure corresponds to the keyword LPERMDIST in the GMIN, OPTIM, and PATHSAMPLE programs.
- (60) Vega, C.; Lago, S. *Comput. Chem.* **1994**, *18*, 55–59.
- (61) Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; Stein, C. *Introduction to Algorithms*; MIT Press: Cambridge, MA, 2003; pp 540–547.
- (62) Wells, S.; Menor, S.; Hespenheide, B.; Thorpe, M. F. *Phys. Biol.* **2005**, *2*, S127–S136.
- (63) Farrell, D. W.; Speranskiy, K.; Thorpe, M. F. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2908–2921.
- (64) Farrell, D. W.; Lei, M.; Thorpe, M. F. *Phys. Biol.* **2011**, *8*, 026017.
- (65) Schlitter, J.; Engels, M.; Krüger, P.; Jacoby, E.; Wollmer, A. *Mol. Simul.* **1993**, *10*, 291–308.
- (66) Schlitter, J.; Engels, M.; Krüger, P. *J. Mol. Graphics* **1994**, *12*, 84–89.
- (67) Ma, J.; Karplus, M. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 11905–11910.
- (68) Elber, R.; Shalloway, D. *J. Chem. Phys.* **2000**, *112*, 5539–5545.
- (69) Wales, D. J. *Mol. Phys.* **1991**, *74*, 1–25.
- (70) Wales, D. J. *J. Chem. Soc., Faraday Trans.* **1992**, *88*, 653–657.
- (71) Wales, D. J. *J. Chem. Soc., Faraday Trans.* **1993**, *89*, 1305–1313.
- (72) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928–1931.
- (73) Bauer, M. S.; Strodel, B.; Fejer, S. N.; Koslover, E. F.; Wales, D. J. *J. Chem. Phys.* **2010**, *132*, 054101.
- (74) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (75) Jones, J. E.; Ingham, A. E. *Proc. R. Soc. London, Ser. A* **1925**, *107*, 636–653.
- (76) Wales, D. J.; Miller, M. A.; Walsh, T. R. *Nature* **1998**, *394*, 758–760.
- (77) Doye, J. P. K.; Miller, M. A.; Wales, D. J. *J. Chem. Phys.* **1999**, *110*, 6896–6906.
- (78) Neirotti, J. P.; Calvo, F.; Freeman, D. L.; Doll, J. D. *J. Chem. Phys.* **2000**, *112*, 10340–10349.
- (79) Calvo, F.; Neirotti, J. P.; Freeman, D. L.; Doll, J. D. *J. Chem. Phys.* **2000**, *112*, 10350–10357.
- (80) Frantz, D. D. *J. Chem. Phys.* **2001**, *115*, 6136–6157.
- (81) Wales, D. J. *Energy Landscapes*; Cambridge University Press: Cambridge, U. K., 2003.
- (82) Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.
- (83) Wales, D. J.; Scheraga, H. A. *Science* **1999**, *285*, 1368–1372.
- (84) Miller, M. A.; Doye, J. P. K.; Wales, D. J. *J. Chem. Phys.* **1999**, *110*, 328–334.
- (85) Picciani, M.; Athenes, M.; Kurchan, J.; Tailleur, J. *J. Chem. Phys.* **2011**, *135*, 034108.
- (86) Wikstrom, M.; Drakenberg, T.; Forsen, S.; Sjöbring, U.; Bjork, L. *Biochemistry* **1994**, *33*, 14011–14017.
- (87) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 133–152.
- (88) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (89) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383–394.
- (90) Tama, F.; Sanejouand, Y. H. *Protein Eng.* **2001**, *14*, 1–6.
- (91) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demeril, M. C.; Keskin, O.; Bahar, I. *Biophys. J.* **2001**, *80*, 505–515.
- (92) Kim, M. K.; Chirikjian, G. S.; Jernigan, R. L. *J. Mol. Graphics Modell.* **2002**, *21*, 151–160.
- (93) Ming, D.; Kong, Y.; Lambert, M. A.; Huang, Z.; Ma, J. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 8620–8625.
- (94) Delarue, M.; Sanejouand, Y. H. *J. Mol. Biol.* **2002**, *320*, 1011–1024.
- (95) Ikeguchi, M.; Ueno, J.; Sato, M.; Kidera, A. *Phys. Rev. Lett.* **2002**, *94*, 078102.
- (96) Zheng, W.; Doniach, S. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13253–13258.
- (97) Reuter, N.; Hinsén, K.; Lacapère, J. J. *Biophys. J.* **2003**, *85*, 2186–2197.
- (98) Xu, C.; Tobi, D.; Bahar, I. *J. Mol. Biol.* **2003**, *333*, 153–168.
- (99) Tama, F.; Valle, M.; Franks, J.; Brooks, C. L. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9319–9323.
- (100) Maragakis, P.; Karplus, M. *J. Mol. Biol.* **2005**, *352*, 807–822.
- (101) Bahar, I.; Rader, A. J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586–592.
- (102) Whitford, P. C.; Miyashita, O.; Levy, Y.; Onuchic, J. N. *J. Mol. Biol.* **2007**, *366*, 1661–1671.
- (103) Zheng, W.; Brooks, B. R.; Hummer, G. *Proteins* **2007**, *69*, 43–57.
- (104) Korkut, A.; Hendrickson, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 15667–15672.
- (105) Lu, M.; Ma, J. A minimalist network model for studying biomolecular vibration. In *Energy Flow in Proteins*; Leitner, D., Straub, J., Eds.; CRC Press: Boca Raton, FL, 2009; pp 229–245.
- (106) Peng, C.; Zhang, L.; Head-Gordon, T. *Biophys. J.* **2010**, *98*, 2356–2364.
- (107) Batista, P.; Robert, C.; Marechal, J. D.; Hamida-Rebai, M.; Pascutti, P.; Bisch, P.; Perahia, D. *Phys. Chem. Chem. Phys.* **2010**, *12*, 2850–2859.
- (108) Best, R. B.; Chen, Y.-G.; Hummer, G. *Structure* **2005**, *13*, 1755–1763.

- (109) Okazaki, K.-i.; Koga, N.; Takada, S.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11844–11849.
- (110) Okazaki, K.-i.; Takada, S. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 11182–11187.
- (111) Lu, Q.; Wang, J. *J. Am. Chem. Soc.* **2008**, *130*, 4772–4783.
- (112) Lai, Z.-Z.; Lu, Q.; Wang, J. *J. Phys. Chem. B* **2011**, *115*, 4147–4159.
- (113) Jónsson, H.; Mills, G.; Jacobsen, K. W. Nudged elastic band method for finding minimum energy paths of transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, B. J., Ciccotti, G., Coker, D. F., Eds.; World Scientific: Singapore, 1998; Chapter 16, pp 385–404.
- (114) Nocedal, J. *Math. Comput.* **1980**, *35*, 773–782.
- (115) Liu, D.; Nocedal, J. *Math. Prog.* **1989**, *45*, 503–528.
- (116) Sheppard, D.; Terrell, R.; Henkelman, G. *J. Chem. Phys.* **2008**, *128*, 134106.