



HAL
open science

MinActionPath2: path generation between different conformations of large macromolecular assemblies by action minimization

Patrice Koehl, Rafael Navaza, Mustafa Tekpinar, Marc Delarue

► To cite this version:

Patrice Koehl, Rafael Navaza, Mustafa Tekpinar, Marc Delarue. MinActionPath2: path generation between different conformations of large macromolecular assemblies by action minimization. *Nucleic Acids Research*, 2024, pp.gkae421. 10.1093/nar/gkae421 . pasteur-04599508

HAL Id: pasteur-04599508

<https://pasteur.hal.science/pasteur-04599508>

Submitted on 3 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

MinActionPath2: path generation between different conformations of large macromolecular assemblies by action minimization

Patrice Koehl¹, Rafael Navaza², Mustafa Tekpinar³ and Marc Delarue^{1,3,*}

¹Department of Computer Science and Genome Centre, University of California, Davis, CA 95616, USA

²Plateforme de Cristallographie, C2RT, Institut Pasteur, Université Paris Cité, UMR 3528 du CNRS, 75015 Paris, France

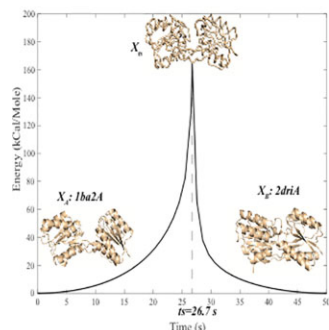
³Unité Architecture et Dynamique des Macromolécules Biologiques, Institut Pasteur, Université Paris Cité, UMR 3528 du CNRS, 75015 Paris, France

*To whom correspondence should be addressed. Tel: +33 1 45 68 86 05; Fax: +33 1 45 68 86 96; Email: delarue@pasteur.fr
Present address: Mustafa Tekpinar, Van Yüzüncü Yil University, Department of Physics, Faculty of Science, Kampus, Van 65080, Turkey.

Abstract

Recent progress in solving macromolecular structures and assemblies by cryogenic electron microscopy techniques enables sampling of their conformations in different states that are relevant to their biological function. Knowing the transition path between these conformations would provide new avenues for drug discovery. While the experimental study of transition paths is intrinsically difficult, in-silico methods can be used to generate an initial guess for those paths. The Elastic Network Model (ENM), along with a coarse-grained representation (CG) of the structures are among the most popular models to explore such possible paths. Here we propose an update to our software platform MinActionPath that generates non-linear transition paths based on ENM and CG models, using action minimization to solve the equations of motion. The new website enables the study of large structures such as ribosomes or entire virus envelopes. It provides direct visualization of the trajectories along with quantitative analyses of their behaviors at http://dynstr.pasteur.fr/servers/minactionpath/minactionpath2_submission.

Graphical abstract



MinActionPath2

Introduction

To accomplish their biological function, most of the enzymes and macromolecular assemblies involved in the fundamental processes of a cell cycle undergo a series of conformational changes during their functional cycle (1). For instance, most DNA polymerases switch from an open form to a closed form upon binding the correct incoming nucleotide. This serves as

a check point for the enzyme, as this transition occurs only rarely for the incorrect nucleotide and catalysis can occur only in the closed form (2). More complex nanomachines such as the ribosome go through a series of large-scale rearrangements that involve the sliding of the mRNA upon completion of the peptide synthesis step (3). Using cryogenic electron microscopy, it is now possible to generate experimentally the

Received: March 12, 2024. Revised: April 25, 2024. Editorial Decision: May 7, 2024. Accepted: May 7, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

different structures adopted by a given nanomachine in the presence of different substrates. The question remains as to how these conformational changes are achieved and which transition path is followed.

As transitions between allosteric conformations occur in the range of 1 μ s–1 ms, capturing experimentally transient structures at the atomic level along their trajectories is a difficult task. The knowledge of such structures, however, would considerably enlarge the scope of structure-inspired drug design. For example, drugs could be designed to prevent the transition in many more ways than permitted by the knowledge of just the endpoints of the transition. In addition, knowledge of the structure of the transition state itself would reveal how exactly catalysis is achieved by a given enzyme.

While computational methods can be used to simulate such transitions, it is important to remember that those transitions are intrinsically rare events. As such, sampling exhaustively the conformational space of the molecule of interest is essential, a sampling that is difficult to obtain with traditional molecular dynamics (MD) simulations. Many new methods are being developed to circumvent this problem (4). In practice, a full atom description of large macromolecular nanomachines is currently difficult to achieve and one must resort to coarse-grained models (5–7). Note that several methods have been developed to go from a coarse-grained representation of biological macromolecules back to their full atomic models (8–10).

In this paper, we derive transition paths between two conformations of a macromolecular system that correspond to a minimum of an action functional, governed by the overdamped Langevin equation. At finite temperature, it is known that such transition paths are found as the trajectories that minimize the Onsager–Machlup action functional (11). Finding such minimum action path (MAP) is equivalent to solving a two-point boundary differential equation. The complexity of this problem is directly related to the complexity of the potential energy considered. Many definitions of this potential have been considered, such as mixing two harmonic potentials centered on each structure (7,12). As our goal is to generate transition paths for large molecular systems (in the hundreds of thousands of atoms), we will consider coarse-grained potentials such as those associated with the elastic network model (ENM) (13).

In a preliminary study (14), we proposed MinActionPath, a method for computing the MAP between two conformations of a proteins on a simplified, two-wells free energy surface derived from the ENMs of the two conformations. The energy in each well is computed as a second order Taylor expansion of the corresponding ENM. Using this energy functional, the equations of motion corresponding to the MAP are solved analytically in each well, and continuity conditions at the crossing between the two wells define the transition point. There were, however, limitations to our implementation of this method. It could not handle structures with loosely connected domains, nor could it always maintain correct stereochemistry over the trajectory using a traditional Tirion elastic network (13). The new implementation, MinActionPath2, also referred to as MAP2, provides a solution to those two problems. First, it allows to define the neighbours in the pairwise sums of the energy function using Delaunay triangulation (15), and second, it provides the option of using a Go-like model to define the potential in proteins (16). It has been shown that the use of such potential reduces the number of covalent bonds with

unreasonable stereochemistry during the transition (17). The code has been entirely rewritten to enable the study of very large systems and the web interface now offers a user-friendly visualization of the resulting trajectory in the output.

MinactionPath2 is not the only tool currently available to generate transition paths between two conformations of a protein. The software ProDy is a very useful open-source package that is simple to use (18). It is available for download at <http://prody.csb.pitt.edu/comd/>. It includes coMD, a tool for generating trajectories using normal mode analysis with a Monte Carlo sampling algorithm that selects the modes that guide the initial structure towards the target structure (19). We note that coMD requires that the input files for the end points of the trajectory must contain the same number of atoms and that those atoms are in the same order. iMODS is a web server that allows for the study of protein dynamics using internal coordinates, i.e. dihedral angles. (20). It includes a morphing tool based on normal modes to generate trajectories, akin to coMD, but with the normal modes parameterized with internal coordinates; we note that it is limited by the size in large macromolecules. Recently, the web server eBDIMS, available at <https://ebd dims.biophysics.se/>, proposes to generate transition paths using Brownian dynamics (21). As MinActionPath2, it is based on elastic networks. It is currently limited to medium-size molecular systems and does not handle inputs with different numbers of atoms, two limitations that are relaxed in MinActionPath2.

In the Materials and methods section, we give details on the formalism and the algorithm itself, followed by a description of the web server and the needed input files and parameters. In the Result section, we present 3 examples: a protein-DNA complex, a ribosome and an entire virus shell. A short Discussion outlines the perspectives.

Materials and methods

A minimum action path between two conformations

The MinActionPath algorithm was described in details in (14) and (22). Here, we provide a brief overview for the sake of completeness.

Let us consider a molecule with N atoms, described by a position vector X . Assuming that we are only interested in effects over long time scales, the dynamic behavior of this molecule can be described by the overdamped Langevin equation.

$$\gamma \frac{dX}{dt} = -\nabla(U(X)) + B \quad (1)$$

where γ is a friction coefficient, $U(X)$ is the potential energy of the system, $\nabla(U(X))$ is the gradient of U , and B is an uncorrelated random force with zero mean. By rescaling the unit of time, it is possible to set $\gamma=1$. The probability p that this equation would have generated a path P over a time interval $[0, F]$ is given by (11)

$$p = \alpha \exp(-\beta S(P)) \quad (2)$$

where α is a normalizing constant, $\beta = 1/(k_b T)$, T is the temperature, k_b the Boltzmann constant, and $S(P)$ the action along the path P defined by:

$$S = \frac{1}{2} \int_0^F \left(\frac{dX}{dt} + \nabla(U(X)) \right)^2 dt \quad (3)$$

Given two possible configurations X_A and X_B for the molecule, the minimum action path that connects these two configurations is the one that maximizes the probability given by Equation (2). This path satisfies the Euler-Lagrange equation:

$$\frac{d^2 X}{dt^2} = \nabla \nabla (U(X)) \nabla (U(X)) \quad (4)$$

where $\nabla \nabla (U(X))$ is the Hessian of the potential energy computed at X .

In the neighborhood of X_A , the potential is given by a second-order expansion of the potential at X_A :

$$U_A(X) = (X - X_A)^T H_A(X_A) (X - X_A) \quad (5)$$

where $H_A(X_A)$ is the Hessian of the potential energy at position X_A . A similar equation defines the potential in the neighborhood of X_B . The total energy U is defined as:

$$U = \min(U_A + \Delta E, U_B) \quad (6)$$

where ΔE accounts for a difference in free energy between the two configurations X_A and X_B .

The format of the total energy (equation 6) provides a simple strategy for generating a trajectory under Equation (4): (i) find the transition time for which a trajectory starting at X_A under equation (4) based on U_A alone and a trajectory starting at X_B under equation (4) based on U_B alone meet. Note that there are analytical solutions to equation (4) at the left, and at the right of the transition (see 14,22). The point of intersection of the two trajectories needs to be continuous in position, velocity, and energy. We find this point using a conjugate gradient minimizer. It can usually be found in 10–15 iterations. (ii) Once the transition state is known, generate the trajectories on the left (from X_A) and on the right (from X_B) of the transition state.

Algorithm

The full algorithm to generate the minimum action path between two conformations based on elastic network models has been described elsewhere (14). Here, we describe a modification that has proved essential to enable generating paths for very large molecular systems (22).

From Equations (4) and (5), the equation of motion close to the start conformation X_A is:

$$\frac{d^2 X}{dt^2} = H_A(X_A)^2 (X - X_A) \quad (7)$$

with the two boundary conditions $X(0) = X_A$ and $X(t_s) = X_{ts}$, where t_s is the transition time and X_{ts} the transition state. The matrix $H_A(X_A)$ is a real symmetric matrix. As such, it can be diagonalized:

$$H_A(X_A) = P_A D_A P_A^T \quad (8)$$

where P_A is an orthogonal matrix and D_A a diagonal matrix. The solution of Equation (8) is then:

$$X_A(t) = P_A f_A(D_A, t) P_A^T (X_{ts} - X_A) + X_A \quad (9)$$

where $f_A(x, t)$ is defined as:

$$f_A(x, t) = \frac{\sinh(xt)}{\sinh(xt_s)} \quad (10)$$

This approach is limited to macromolecules with less than a few thousand atoms as the diagonalization becomes impractical for larger structures in both computing time and

space constraints. There is, however, an alternate approach to diagonalization that is based on the concept of Krylov subspace. For the sake of simpler notations, let us define $A = H_A(X_A)$, $v = X_{ts} - X_A$, and $f = f_A(x, t)$. We build an orthogonal base $V_m = [v_1, \dots, v_m]$ of the Krylov space $K_m = \{v, Av, \dots, A^{m-1}v\}$ and a tridiagonal matrix T_m of dimension m such that

$$AV_m = V_m T_m + \beta_m v_{m+1} e_m^T \quad (11)$$

where e_i is the i th unit vector in \mathbb{R}^m , and β_m is an estimate of the error from this process. It then follows that,

$$f(A)v \approx \|v\| V_m f(T_m) e_1 \quad (12)$$

Note that m , the order of the Krylov space, is taken to be much smaller than the dimension of the matrix A , usually in the order of a few tens. In addition, computing V_m and T_m only requires computing matrix vector multiplications which can be performed efficiently using either the sparse structure of the Hessian A , or its tensor representation (23). Finally, the matrix T_m is a small tridiagonal matrix of size m that can easily be diagonalized.

Description of the web server

Initial and final states X_A and X_B

Input files

The two input files can be either in PDB format or mmCIF format. Hydrogen atoms and HETATM cards are ignored. There is no limit on the number of atoms or on the number of chains. Chain names are coded with only one character in PDB files, and this is obviously a limitation for macromolecules with more than 62 chains. Chain names using two characters are allowed in the PDBx file format, which is supported in our interface. We found that the best way to handle models with many chains (e.g. in ribosomes or viruses) is to use the mmCIF format with no limitation on the format of the chain name. While mmCIF files are preferred on input, both PDB and PDBx files are accepted and transformed into mmCIF files at this stage. If there are more than one model for each chain, the user can specify to keep only the first one.

Choosing the coarse graining

If requested by the user, the program will extract only CA atoms for proteins and C3' for DNA (or C4' for RNA); otherwise, the program will use all atoms present in the input files (again, except for hydrogen atoms and ligands identified with a HETATM card). This selection of one specific atom per residue or nucleotide is dictated by the NGL viewer (24). From a simple perspective of mass, however, a nucleotide is on average three times heavier than an amino acid. Users may want to account for this by representing each nucleotide with three atoms. Currently, this needs to be done as a preprocessing prior to using MinActionPath2. Among the three atoms, there should at least be one C3' atom per residue for DNA (C4' for RNA) and one CA atom for a protein, to have a correct trace representation of the chains by NGL.

Handling a different number of atoms in the two input files

In the previous version of this web site, there was a strict requirement that the two input files contain the same number of atoms. In the new version, we relax this constraint by preprocessing the input files through USAlign (25), with

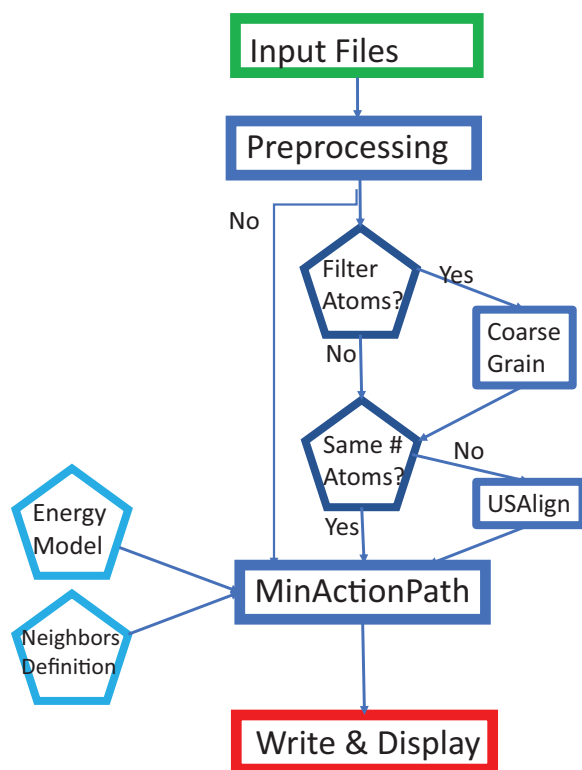


Figure 1. A flowchart of the MAP2 web site.

the option `-Tmscore 5`, which is essentially a sequence alignment, all chains against all chains (option `-ter 0`). Following this, the common atoms of the aligned residues with the same physico-chemical characters, defined by 5 classes ((P,A,G,S,T), (D,N,E,Q), (H,K,R), (F,Y,W), (V,L,I,C,M)), are extracted and forwarded to the algorithm generating the trajectory itself.

We maintain the possibility for the user to bypass this step, i.e. to generate a trajectory based on the input files without any processing by USAlign. This bypass assumes that the two files have the same number of atoms.

The general workflow and organization of the web server is summarized in Figure 1.

The elastic network model

Energy

We follow the concept of isotropic elastic network (5,13).

In this model, the energy of a macromolecule is set to be the harmonic energy associated with springs attached to a predefined set of pairs of atoms. Let us consider for example the start conformation X_A . The energy at a conformation X near X_A is set to:

$$V(X) = \frac{1}{2} k_1 \sum_{(i,j)} (r_{ij} - r_{ij}^0)^2 \quad (13)$$

In this equation, k_1 is the isotropic force constant of all the springs formed by the pairs of atoms i and j , r_{ij} and r_{ij}^0 are the distances between i and j in the conformation X and in the reference conformation X_A , respectively, and the summation extends to all pairs (i,j) included in the network. A similar equation describes the energy of a conformation X near X_B , the target conformation, with k_1 replaced by a dif-

ferent constant k_2 . We will refer to this energy as the Tirion energy.

Note that the energy defined by Equation (7) does not account for stereochemistry and as such dynamics that are based on such an energy may undergo too large deviations from acceptable stereochemistry, especially if the two end conformations differ by a large rmsd value. To circumvent this problem, more terms are needed in the energy. The Go Model has been parametrized with this purpose in mind, including terms that are based on bond lengths, torsion angles, dihedral angles, and a van-der-Waals-like term (16,17). Note that the Go model only considers CA atoms and currently works only for proteins. It is available as an option for MinActionPath2.

The geometry of the elastic network

Several criteria have been used to define the set of atom pairs that are included in Equation (6). In standard elastic network models, a cutoff distance R_c is defined such that all pairs of atoms separated by less than this cutoff are included in the network. There are, however, no guidelines as to which values for R_c are best. Typical values for R_c are in the range 13–15 Å when the network is based on CA only, and in the range 7–9 Å for all atoms (26). An alternate method is to build a geometric structure on the sets of positions of the atoms; the Delaunay complex is well suited for this purpose (15,27). The advantage of the last criterion is that it is a parameter-free method that handles well possible ‘dangling parts’ of the model, which appear for instance in AlphaFold models for loosely connected domains of large proteins or macromolecular complexes. These two criteria, i.e. a cutoff-based or Delaunay-based elastic network (the default choice), are available in MinActionPath2.

Choosing k_1 and k_2 for the Tirion energy

In principle k_1 can have a different value from k_2 : this will influence the position of the transition point, as described in the Results section. The usual way to choose an isotropic k value for the Tirion energy is to impose the relationship,

$$k = \frac{A}{\langle B \rangle} \quad (14)$$

in kcal/mol/Å² where $\langle B \rangle$ is the average B -factor of the structure and A is a proportionality factor that contains the temperature-dependence of the elastic constant. The smaller the average B -factor, the higher the elastic constant k , expected to be in the range 0.16 ± 0.09 kcal/mol/Å².

Note that MinActionPath2 calculates $\langle B \rangle$ for both the start and target structures, if those values are available in the input files, and these values are reported in the log file, so that they may be used to define the ratio of k_1 and k_2 in a subsequent run.

Output

First the matching of chains found by USAlign is provided, including the sequence alignment of each chain; then the log file of MinActionPath2 is given, which contains details on the convergence of the algorithm as well as a quick analysis of the trajectory in terms of the energy of each frame, its rmsd versus each of the starting and ending states, as well as the Q_1 and Q_2 quantities (percentage of specific contacts present in the initial or final state(s), respectively). This allows Q_1 ver-

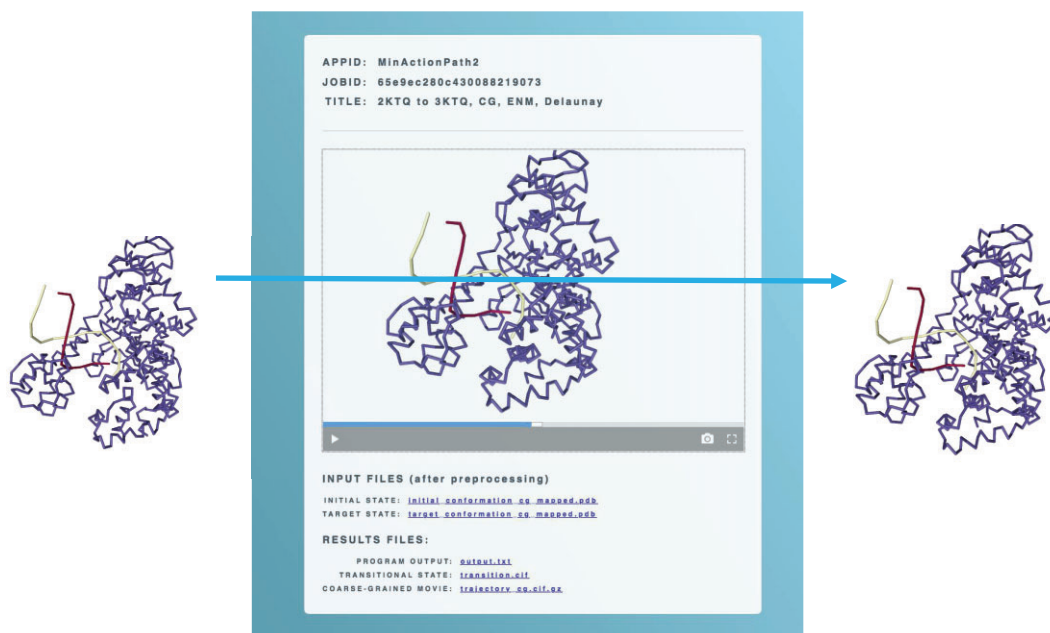


Figure 2. Illustration of the Example 1 with the Klenow fragment of DNA Polymerase I from *E. coli* in complex with DNA.

sus Q_2 plots to be built; those plots have been found useful when comparing geometrically all sorts of trajectories (14,28). A link to download the mmCIF file of the transition state as well as of the trajectory itself is provided. The entry files to MinActionPath2 (mmCIF) are also provided.

Finally, a window based on the NGL viewer (24) is automatically open on output, where the trajectory can be played, with the possibility to rotate, translate and zoom in the molecule(s).

Results

MinActionPath2 generates a transition path with minimum action between two conformations of a molecular systems. Its interface has been designed to limit the number of parameters required from the user, to improve the ease of use, without loss of quality. The main inputs include the files containing the start and end conformations for the trajectory, the definition of the elastic networks that represent the structure, as well as the length of time for the trajectory. As described above, the elastic networks can be derived from a cut-off criterium, or by defining a Delaunay complex over the molecule. The energies associated with those networks are based either on a Tirion potential, in which case the user is expected to provide values for the elastic constants associated with the start and target conformations, or a Go-like potential. In the following, we describe a set of worked examples available on the MinActionPath2 web page that are associated with proteins and protein–DNA complexes, as well as brief analyses of the importance of the choices the user needs to make when running our program.

Worked examples

Protein–DNA complex

The first example is the Klenow fragment of DNA polymerase pol I from *Escherichia coli* in a binary or ternary complex with DNA (Figure 2). The input files for the start and target conformations

have the PDB codes 2KTQ and 3KTQ, respectively. The two structures have 551 atoms in common when keeping only one atom per residue. The superposition by USAlign allows the extraction of the common residues in the protein (chain A) and in the DNA (Chains B and C, or B and D, which had to be re-matched). The DNA has the same sequence. We use a Delaunay elastic network and the Tirion potential with the default values for elastic constants $k_1 = 0.1$ and $k_2 = 0.1$. Computing the trajectory between conformations 2KTQ and 3KTQ takes only 17 s. This trajectory, and the associated information such as rmsd between snapshots and the initial and final conformations, are available on the MAP2 web server under ‘Examples’. In comparison, computing the trajectory based on all atoms (4401 common atoms) takes 108 s.

Ribosome

The second example concerns two conformations of the same ribosome of *E. coli* (8G34.cif and 8G31.cif). Those two conformations contain 52 chains and a total of 10474 atoms in common, after filtering and retaining one atom per residue. The mapping of chains is successfully found by USAlign. We use a Delaunay elastic network and the Tirion potential with the default values for elastic constants $k_1 = 0.1$ and $k_2 = 0.1$. The algorithm finds the transition time in 14 conjugate gradient cycles. The total time to find the trajectory is 263 s. The trajectory is displayed as a movie in the result web page, with each chain colored with a different color. In comparison, computing the trajectory based on all atoms (144 244 common atoms) takes 9125 s. We were able to compute trajectories between eukaryotic ribosomes whose sizes were at least up to 200 000 atoms.

Zika virus

We considered two conformations of the entire envelope of the Zika virus, one at pH 5 (PDB code: 5H32) the other at pH 8 (PDB code: 5H37). Both structures represent the virus envelope in the presence of antibody C10 (29). We only retained the coordinates of the CA of the virus envelopes. The

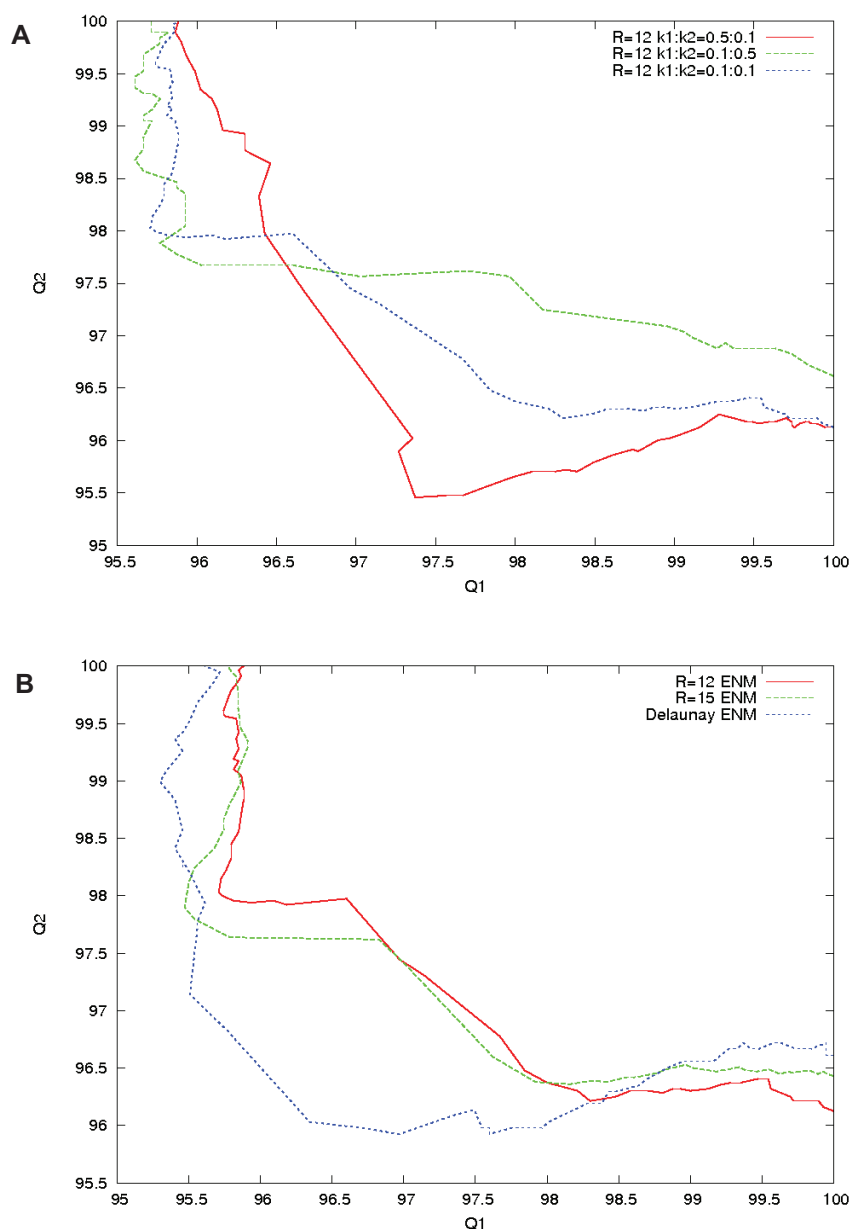


Figure 3. Comparison of the trajectories generated by MinActionPath2 for the pair of structures 1ANF and 1OMP of maltodextrin binding protein. **(A)** Q_1 versus Q_2 plots (where Q_1 and Q_2 are the percentage of native contacts within the structures of the trajectories with respect to the start and target structures) showing the effect of different elastic constants in the ENM of the initial and final states. **(B)** Same Q_1 versus Q_2 plots to compare the effect of using a Delaunay-based elastic network compared with a cutoff-based elastic network.

corresponding conformations include 180 chains (E proteins) for a total of 71340 atoms after retaining only one atom per residue. We used the Delaunay triangulation to define the geometry of the elastic network. The Tirion potential was set with default values $k_1 = 0.1$ and $k_2 = 0.1$. The algorithm found a trajectory in 1984 seconds. The trajectory and associated information are available on the MAP2 web server under 'Examples'. Note that in the NGL window, each chain is colored with a different color.

Influence of k_1 versus k_2 for an ENM model

A test was made to see the influence of choosing different elastic constants k_1 and k_2 for the initial and final states of the molecule of interest. We considered the pair of structures 1ANF and 1OMP of maltodextrin binding protein. Those conformations have 370 residues in common. The rmsd be-

tween the start and target conformations is 3.77 Å. As can be seen in Figure 3A, the two Q_1 versus Q_2 plots based on $k_1 = 5 \cdot k_2$ and $k_2 = 5 \cdot k_1$ vary above and below the plot $k_1 = k_2$, respectively. For those two settings of k_1 and k_2 , the Energies at the transition point are much higher (about 60 kcal/mol) compared to the equivalent energy when k_1 is set equal to k_2 (30 kcal/mol). The transition times when $k_1 = 5 \cdot k_2$ and $k_2 = 5 \cdot k_1$ are shifted to the left, or to the right, respectively, compared to $k_1 = k_2$, as expected.

Influence of Delaunay- versus cutoff-defined neighbors

In Figure 3B, for the same system of maltodextrin binding protein, we show results obtained with elastic networks defined with a cutoff R_c set to either 12 or 15 Å, with those obtained with an elastic network obtained from a Delaunay triangu-

lation. The first two cases show similar behaviors, illustrating the fact that similar results are obtained with the cutoff-defined method of defining neighbors provided that $R_c > 12$, while the latter (i.e. the Delaunay EN) departs more from the diagonal, with a more non-linear behavior. Note that the energies at the transition point cannot be strictly compared as the number of edges in the networks are different.

Discussion

Action minimization is recognized as an efficient method to generate trajectories described by a Langevin equation (with noise) under the constraint of a known starting and ending points (4), with currently few practical implementations (30). We have described a web site implementation of a faster and more user-friendly version of our initial implementation of the method under a simplified coarse-grained potential (14). The new version, MinActionPath2 can handle large macromolecular assemblies, including entire shells of viruses (>200 000 atoms). It allows for a direct visualization of the trajectory with graphical tools on the output, as well as provides the option of downloading the full trajectory.

The computed trajectory is usually characterized by a very high energy of the transition state, which could appear at first sight as a severe limitation of the program. We note, however, that it is possible to relax this trajectory to improve the energy at the transition, without modifying the overall trajectory significantly (22). We intend to enable such relaxation with MinActionPath2 as future work.

There are many applications that could easily benefit from MinActionPath2. For example, the structures of snapshots along a trajectory between two conformations of a protein can be used to discover new ligand-binding pockets, in very much the same way that normal modes are used to reveal cryptic pockets (31). In addition, the MinActionPath2 trajectories can be used as an input for other methods that requires an initial guess of the trajectory, such as the string methods, for further sampling of possible paths using an all-atom representation and a more physical energy (32,33).

We point out that new methods based on AlphaFold2 are being developed to predict new structures from their sequence, not only in the apo form but also in a holo form (bound to a ligand), with encouraging results (34). Hence, the potential of application of MAP2 is considerably larger than the pairs of similar structures deposited in the PDB.

MinActionPath2 was designed with minimal user inputs in mind, namely the input files for the two structures considered, information on the geometry of the elastic networks (cutoff-based, or Delaunay-based), and information on the energies associated with those networks (Tirion-based with isotropic elastic constants, possibly different for the two input structures, or Go-like potentials that are parameter-free). There is room for improvement with all those choices we have made. First, note that we have chosen isotropic values for the elastic constants associated with the Tirion potential. It is possible to use instead anisotropic elastic constants (i.e. with the option of setting different constants for different pairs of atoms), and even refine those elastic constants based on the *B*-factors associated with the structures (when those *B*-factors are available). We refer the reader to (23) and (35) for a full discussion on how to choose the elastic constants. We note that in this last case the local elastic constants are set using the surface of interaction between atoms, calculated after Delaunay tessellation:

hence, we already have the elements to calculate and calibrate such local elastic constants, both for proteins and nucleic acids (and complexes thereof). Second, our current implementation of the Go-model is parameter free. While this reduces the number of parameters, it may also negatively impact the quality of the trajectory (for example it would not correctly capture difference in the internal dynamics of the two structures that is reflected in the *B*-factor). In future work, we will propose a Go-like potential that can be parameterized. On a side note, our implementation of the Go-like potential is specific to proteins. As a result, the correct stereochemistry may not be maintained during the trajectory, especially in complexes containing nucleic acids. We will parameterize an equivalent potential for nucleic acids and complexes between them and proteins (see for example, (36)).

Data availability

MinActionPath2 (MAP2) is free and open to all users at http://dynstr.pasteur.fr/servers/minactionpath/minactionpath2_submission/.

Acknowledgements

We thank Marat Yusupov and Yury Zgadzay at IGBMC, Strasbourg for sharing recent PDB files of *C. albicans* ribosomes to test the program. We thank Felix Rey (IP) for sharing interest in simulating transitions in whole viruses' structures. We thank Henri Orland (IPhT, CEA) for making the stay of PK in Paris possible during the course of this work and for interesting discussions.

Funding

P. Koehl was the recipient of a CEA fellowship while working at the Institut de Physique Theorique (IPhT), CEA, Gif/Yvette during the fall of 2023. Part of this work was done during this period. MinActionPath2 (MAP2) is free and open to all users. Web site address: http://dynstr.pasteur.fr/servers/minactionpath/minactionpath2_submission/. Funding for open access charge: Internal Funds from Institut Pasteur.

Conflict of interest statement

None declared.

References

1. Steven, A.C., Baumeister, W., Johnson, L.N. and Perham, R.N. (2016) In: *Molecular Biology of Assemblies and Machines*. Garland Science, NY.
2. Steitz, T.A. (1999) DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.*, **274**, 17395–17398.
3. Noller, H.F., Lancaster, L., Mohan, S. and Zhou, J. (2017) Ribosome structural dynamics in translocation: yet another functional role for ribosomal RNA. *Q. Rev. Biophys.*, **50**, e12.
4. Elber, R., Makarov, D. and Orland, H. (2020) In: *Molecular Kinetics in Condensed Phases: Theory, Simulation, and Analysis*. Wiley, Hoboken, NJ.
5. Zheng, W. and Wen, H. (2017) A survey of coarse-grained methods for modeling protein conformational transitions. *Curr. Opin. Struct. Biol.*, **42**, 24–30.

6. Zheng,W., Wen,H., Iacobucci,G.J. and Popescu,G.K. (2017) Probing the structural dynamics of the NMDA receptor activation by coarse-grained modeling. *Biophys. J.*, **112**, 2589–2601.
7. Delarue,M., Koehl,P. and Orland,H. (2017) Ab initio sampling of transition paths by conditioned Langevin dynamics. *J. Chem. Phys.*, **147**, 152703.
8. Jones,M.S., Shmilovich,K. and Ferguson,A.L. (2023) DiAMoNDBack: diffusion-denoising autoregressive model for non-deterministic backmapping of calpha protein traces. *J. Chem. Theory Comput.*, **19**, 7908–7923.
9. Shmilovich,K., Stieffenhofer,M., Charron,N.E. and Hoffmann,M. (2022) Temporally coherent backmapping of molecular trajectories from coarse-grained to atomistic resolution. *J. Phys. Chem. A*, **126**, 9124–9139.
10. Perry,Z.R., Pyle,A.M. and Zhang,C. (2023) Arena: rapid and accurate reconstruction of full atomic RNA structures from coarse-grained models. *J. Mol. Biol.*, **435**, 168210.
11. Onsager,L. and Machlup,S. (1953) Fluctuations and irreversible processes. *Phys. Rev.*, **91**, 1505–1512.
12. Das,A., Gur,M., Cheng,M.H., Jo,S., Bahar,I. and Roux,B. (2014) Exploring the conformational transitions of biomolecular systems using a simple two-state anisotropic network model. *PLoS Comput. Biol.*, **10**, e1003521.
13. Tirion,M.M. (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, **77**, 1905–1909.
14. Franklin,J., Koehl,P., Doniach,S. and Delarue,M. (2007) MinActionPath: maximum likelihood trajectory for large-scale structural transitions in a coarse-grained locally harmonic energy landscape. *Nucleic Acids Res.*, **35** W477–W482.
15. Xia,F., Tong,D., Yang,L., Wang,D., Hoi,S.C., Koehl,P. and Lu,L. (2014) Identifying essential pairwise interactions in elastic network model using the alpha shape theory. *J. Comput. Chem.*, **35**, 1111–1121.
16. Clementi,C., Nymeyer,H. and Onuchic,J. (2000) Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.*, **298**, 937–953.
17. Hsieh,Y.C., Delarue,M., Orland,H. and Koehl,P. (2023) Analyzing the geometry and dynamics of viral structures: a review of computational approaches based on alpha shape theory, normal mode analysis, and Poisson-Boltzmann theories. *Viruses*, **15**, 1366.
18. Zhang,S., Krieger,J.M., Zhang,Y., Kaya,C., Kaynak,B., Mikulska-Ruminska,K., Doruker,P., Li,H. and Bahar,I. (2021) ProDy 2.0: increased scale and scope after 10 years of protein dynamics modelling with Python. *Bioinformatics*, **20**, 3657–3659.
19. Gur,M., Madura,D.J. and Bahar,I. (2013) Global transitions of proteins explored by a multiscale hybrid methodology: application to adenylate kinase. *Biophys. J.*, **105**, 1643–1165.
20. Lopez-Blanco,J.R., Aliaga,J.I., Quintana-Orti,E.S. and Chacon,P. (2014) iMODS : internal coordinates normal mode analysis server. *Nucleic Acids Res.*, **42**, W271–W276.
21. Orellana,L., Yoluk,O., Carrillo,O., Orozco,M. and Lindahl,E. (2016) Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations. *Nat. Comm.*, **7**, 12575.
22. Koehl,P. (2016) Minimum action transition paths connecting minima on an energy surface. *J. Chem. Phys.*, **145**, 184111.
23. Koehl,P., Orland,H. and Delarue,M. (2021) Parameterizing elastic network models to capture the dynamics of proteins. *J. Comput. Chem.*, **42**, 1643–1661.
24. Rose,A.S. and Hildebrand,P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.*, **43** W576–W579.
25. Zhang,C., Shine,M., Pyle,A.M. and Zhang,Y. (2020) US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods*, **19**, 1109–1115.
26. Eyal,E., Yang,L.W. and Bahar,I. (2006) Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics*, **22**, 2619–2627.
27. Koehl,P. and Delarue,M. (2019) Coarse-grained dynamics of supramolecules: conformational changes in outer shells of Dengue viruses. *Prog. Biophys. Mol. Biol.*, **143**, 20–37.
28. Seyler,S.L., Kumar,A., Thorpe,M.F. and Beckstein,O. (2015) Path similarity analysis: a method for quantifying macromolecular pathways. *PLoS Comput. Biol.*, **11**, e1004568.
29. Zhang,S., Kostyuchenko,V.A., Ng,T.S., Lim,X.N., Ooi,J.S., Lambert,S., Tan,T.Y., Widman,D.G., Shi,J., Baric,R.S., *et al.* (2016) Neutralization mechanism of a highly potent antibody against Zika virus. *Nat. Commun.*, **7**, 13679.
30. Lee,J., Lee,I.H., Joung,I., Lee,J. and Brooks,B.R. (2017) Finding multiple reaction pathways via global optimization of action. *Nat. Commun.*, **8**, 15443.
31. Zheng,W. (2021) Predicting cryptic ligand binding sites based on normal modes guided conformational sampling. *Proteins*, **89**, 416–426.
32. Pan,A.C., Sezer,D. and Roux,B. (2008) Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B.*, **112**, 3432–3440.
33. Lev,B., Murail,S., Poitevin,F., Cromer,B.A., Baaden,M., Delarue,M. and Allen,T.W. (2017) String method solution of the gating pathways for a pentameric ligand-gated ion channel. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E4158–E4167.
34. Sala,D., Engelberger,F., Mchaourab,H.S. and Meiler,J. (2023) Modeling conformational states of proteins with AlphaFold. *Curr. Opin. Struct. Biol.*, **81**, 102645.
35. Mailhot,O., Frappier,V., Major,F. and Najmanovich,R. (2022) Sequence-sensitive elastic network captures dynamical features necessary for miR-125a maturation. *PLoS Comput. Biol.*, **18**, e1010777.
36. Chakraborty,D., Mondal,B. and Thirumalai,D. (2024) Brewing COFFEE: a sequence-specific coarse-grained energy function for simulations of DNA-protein complexes. *J. Chem. Theory. Comput.*, **20**, 1398–1413.