POLARIS: Path of Least Action Analysis on Energy Landscapes

Evan Seitz and Joachim Frank*

Cite This: https://dx.doi.org/10.1021/acs.jcim.9b01108

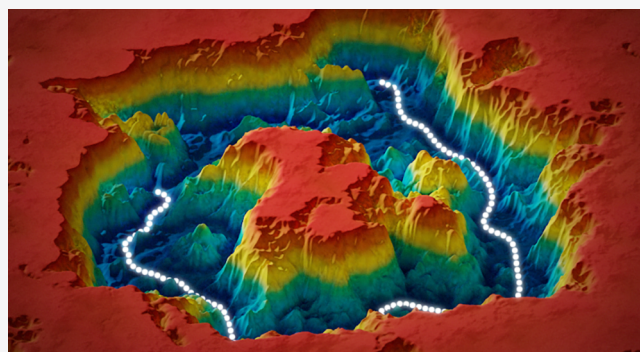Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Free-energy landscapes are a powerful tool for analyzing dynamical processes - capable of providing a complete mapping of a system's configurations in state space while articulating its energetics topologically in the form of sprawling hills and valleys. Within this mapping, the path of least action can be derived - representing the most probable sequence of transitions taken between any two states in the landscape. In this article, POLARIS (Path of Least Action Recursive Survey) is presented as a dynamic, global approach that efficiently automates the discovery of the least action path on previously determined 2D energy landscapes. Important built-in features of this program include plotting of landscape trajectories and transition state theory diagrams, generation of text files with least action coordinates and respective energies, and bifurcation analysis tools that provide downstream versatility for comparing most probable paths and reaction rates.

INTRODUCTION

The construction and exploration of energy landscapes is essential for understanding the complex dynamics of molecular systems.[1] Since the initial underpinnings of potential energy surfaces over a century ago,[2] many attempts have been made to articulate its required properties. Fundamentally, the free-energy landscape is understood to be an intrinsic property of a given molecular system, independent of the experimental method used to obtain it.[3] As these systems dynamically operate between distinct conformational states over many time and length scales, any mapping to the corresponding energy landscape must also be multidimensional and hierarchical in nature.[4] Within this landscape, the sprawling layout of energy hills and valleys characterizes the system's navigational probabilities - with deep wells representing distinct conformational states and peaks or ridges acting to constrain the transitions between them. Further, since specific sequences of conformations give rise to biomolecular function, functional dynamics should be accounted for within this descriptive topology.[5]

For decades, a variety of methods have existed to analyze energy landscapes,[6] including global optimization algorithms,[7] transition state search schemes,[8,9] and coordinate transformations,[10] to name a few. During that time, many landscapes were successfully modeled for protein folding and related dynamics,[11−13] while applications to larger, more complex macromolecular assemblies remained in their infancy.[4] In recent years, however, technological advances in single-particle cryo-EM[14−17] have opened the door for constructing the free-energy landscape across a number of these previously incomprehensible systems.[18−21]

Cryo-EM allows macromolecules to be experimentally visualized en masse via electron microscopy after being rapidly frozen in vitreous ice (quenching) at a rate assumed to be faster than the overall reconfiguration time of the system. As a result, the ensemble of frozen molecules closely approximates the Boltzmann distribution of states of the system immediately prior to freezing.[3] Virtually all projection directions of the particle are obtained for each conformation in its state space. These data are then analyzed by ManifoldEM - a new computational technique employing manifold embedding[19−21] to construct a lower-dimensional state space of the system from its experimental sightings. The number of particles found in each state is then compared to the occupancy of the most populated conformation, with this ratio converted into standard free-energy differences via the Boltzmann factor to create the free-energy landscape.[18] In this way, the landscape's minimum-energy wells represent the system's most highly preferred conformations as witnessed during the experiment. Through this method, the potential now exists for creating the conformational free-energy landscape for any molecular machine. With this powerful tool, new methods must be devised to extract valuable information from these exper-

imentally determined landscapes for further biological elucidation.

Existing on a scale where thermal and deterministic forces act on equal footing, the macromolecules are constantly buffeted by the random motions of nearby solvent molecules as they compete to internally perform the sequence of motions required to operate. As a result, transitions between distinct states occur via a series of thermally driven steps. This diffusive process can occur through a multitude of structural pathways in the energy landscape, ultimately creating an ensemble of transitory routes between any two configurations.[5] Each of these available pathways represents a unique sequence of conformational events - with the probability for a transition to take one such pathway dependent on both the surrounding highest-energy peaks[4] and the total integrated energy along that path's diverse range of intermediate states.

Historically, such notions stem from Maupertuis' *principle of least action*,[22] which defined the natural tendency of a system to choose the path among all available paths connecting two specified points that ultimately minimizes its *action*. Maupertuis worked closely with Euler, who continued to advance this concept as a "minimization of effort"[23] - an expression corresponding to our present-day understanding of potential energy. From this, the *path of least action* can be defined and computationally generated for the energy landscape - a path representing the most probable sequence of unique conformational changes taken by the system to navigate between two distinct configurations.

In this study we assume that the free-energy landscape is already given in the form of a two-dimensional array, determined from cryo-EM data using the method outlined above. Associated with each array point is a set of reaction coordinates, the value of the free energy, and a "local" 3D density map representing the local conformation of the molecule. POLARIS aims to find the path of least action between any two points within the energy landscape, using only information contained within it (i.e., the energy of each state, with all states spatially organized along the reaction coordinates).

Many general techniques already exist to algorithmically solve pathfinding problems - seeking to identify the best path between any two points as defined by some overhead metric (e.g., shortest or fastest route). In the majority of these, graph theory plays a prominent role, with hallmark algorithms such as Dijkstra's[24] and many variants[25−27] proven successful. However, these methods rely on the assumption that the kernel being used (representing the distance between any two points) is deterministically correct. By calculating this kernel across every combination of points, a weighted graph is produced, where a value (weight) is given to every edge connecting the set of vertices in the graph. A collection of decisions is subsequently made to find the collection of intermediate vertices such that the sum of their constituent edge weights is minimized.

However, in situations where the value of each edge weight is not conveniently predefined (as is the case in energy landscapes), uncertainty arises from the existence of a theoretically infinite number of kernel functions one could use to define such a weighting.[28] For example, Graph Spectral Image Processing[29] aims to construct a graph connecting pixels with weights via a bilaterally filtered Gaussian kernel incorporating pixel intensity and distance, while also requiring two additional parameters. Alternatively, edge weights can be defined based on such metrics as local pixel patches or overhead features.[28] Ultimately, the definition of these edge weights is application-dependent,[29] with the solution to the pathfinding problem changing as the underlying relationships between each set of points is altered. As such, using edge weights to find the absolute path of least energy (in the context of energy landscape problems) requires that the optimized output from one arbitrary weighting function be compared to an infinite number of competing weighting functions - presenting an endless search for each graph-based approach.

Further, the very choice of overhead metric within the contextual demands of each system is a source of uncertainty. In the case of the macromolecular assembly driven by thermal motion, such a metric should reflect the total integrated energy of each possible path, while also analyzing these paths within the landscape's hierarchy of distinguished scales (macroscopic, mesoscopic, and microscopic[4]). Regardless, algorithms do exist that use broad, graph-based approaches to navigate between states in the energy landscape. One such tool is the Dijksta-inspired MEPSA (Minimum Energy Pathway Analysis[30]) which will be analyzed and compared in detail with our method.

Another class of algorithms includes the chain-of-states method,[31,32] whereby a set of nodes along an initially straight pathway between two local minima is relaxed to find the minimum-energy path. These nodes are optimized with physical constraints such that the chain-of-states are connected by spring forces to ensure equal spacing along the reaction path.[33] Uncertainty in this class' performance is first introduced in the initial choice among several competing methods, which include such options as "nudged elastic band",[34] "doubly nudged elastic band",[35] "string",[36] and "simplified string".[37] Once chosen, certainty in each method is additionally limited by a subsequent choice in the applied force-based optimizer, including such as options as "steepest-descents", "quick-min", "fast inertial relaxation engine", "conjugate gradients", and 'limited memory Broyden-Fletcher-Goldfarb-Shannon" - with each optimizer presenting its own strengths and limitations.[33] Additionally, for highly complex energy landscapes having several competing global pathways separated by massive maximum-energy regions, without foresight, this class of methods will become globally restricted to the linear region between the local minima they happen to be initiated between.

POLARIS (Path of Least Action Recursive Survey)[38] provides an alternative approach to these minimum-energy pathfinding algorithms by avoiding the arbitrary assignment of edge weights, physics-based optimizers, or globally restrictive strings. Instead, POLARIS aims to prioritize and isolate the most energetically favorable coordinates in a given landscape - as these represent highly occupied transit regions on the macroscopic scale. This initial method reflects the paradigm used by hierarchical pathfinders,[39] whereby the high-level overview route is assigned first by analyzing the most favorable transit locations in between. Between these anchors in the landscape, all permutations of a set of higher-order lines are drawn with the net energies compared from each. The algorithm then takes the resulting minimum-energy discoveries as new inputs to itself to repeat this procedure recursively - breaking down each best global line approximation into continuously finer, mesoscopic subsections until the finest-granular, microscopic path is resolved.
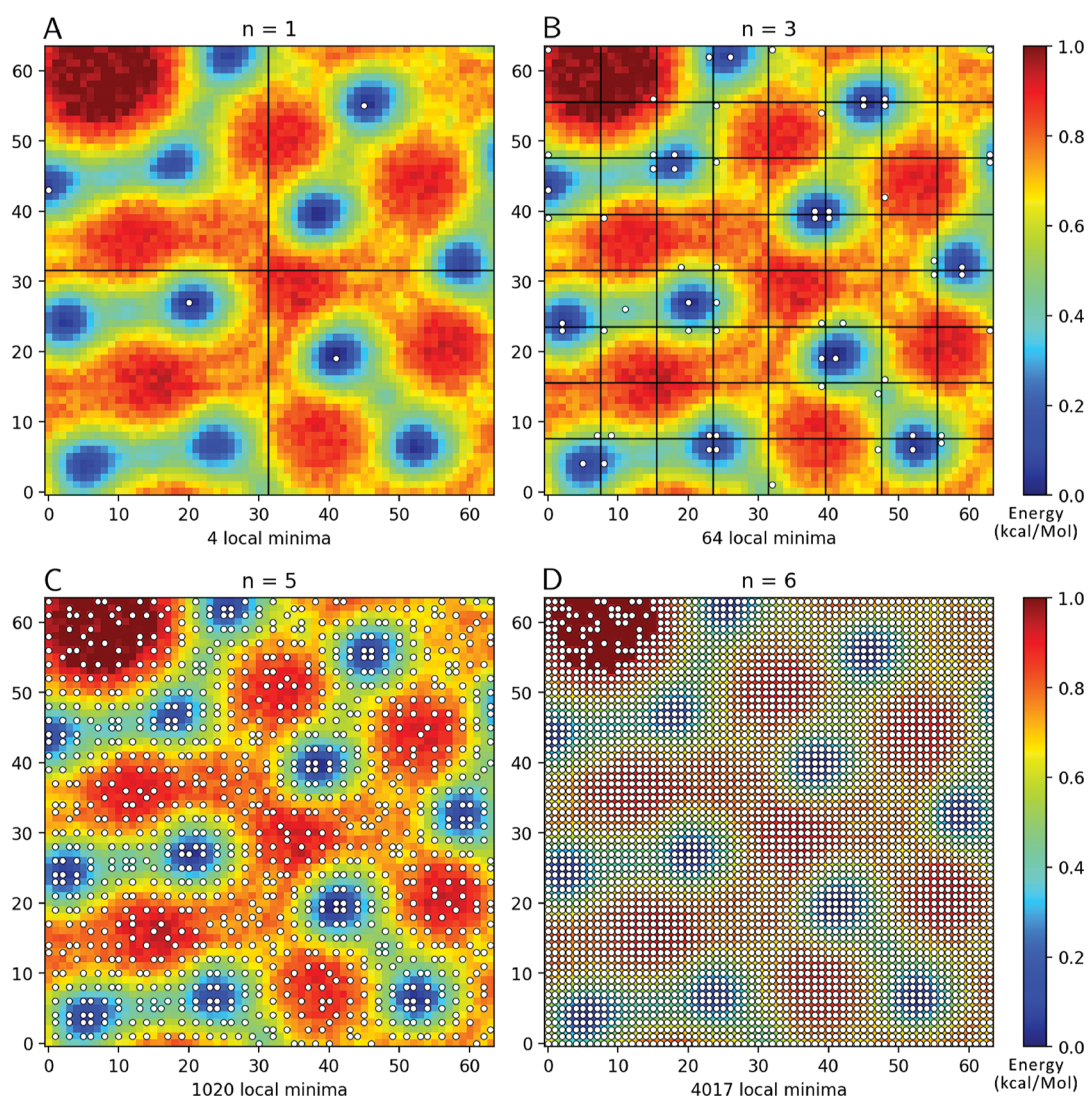
Figure 1. Computationally generated landscape featuring increasing image segmentation depths, with segmenting lines (black) overlaid on n_1 (A) and n_3 (B) for visualization of the division. Local energy-minima are plotted as single white points exclusively occupying each square subdivision. In rare cases of multiple equal-valued minima per segment, only one minimum is selected, with subsequent depths capturing the alternative values. To maximize efficiency, no minima are obtained for segmented regions containing a uniform spread of globally maximum energy levels (as seen in the northwest corner of D). This landscape was created to exemplify the hierarchical scale seen in energy landscapes, whereby additive Gaussian noise was applied across all pixel intensities to represent unavoidable experimental uncertainties during data acquisition and processing.

■ METHODS

POLARIS is a cross-platform, open-source program written in Python 3.x,[40] a graphic user interface built using PyQt5,[41] plots drawn using Matplotlib,[42] and operations performed with the aid of NumPy,[43] Python's Itertools, and Bressenham's Line Algorithm.[44] POLARIS has been designed as a user-friendly tool able to analyze transition pathways through 2D energy landscapes in pursuit of expediting the discovery of their most significant pathways.

By defining spatial subdivisions for isolating minima within the energy landscape (see Section 1: Image Segmentation) and creating the set of all possible permutations between them over a range of increasing line orders (Section 2: Permutational Analysis), POLARIS is able to compare global path approximations, isolate favorable minima, and construct local solutions between them as it methodically implements its optimized outputs as inputs to itself (Section 3: Branching

Recursions). These steps are performed consecutively and will be discussed in this order in the sections below.

As a note, to ensure that this procedure can be performed regardless of data dimensions, the width and height of the original data file are first trimmed to the nearest even pixel when necessary. Next, maximum energy borders are added to the file; expanding one power of 2 above the landscape's dimensions (e.g., for a 70 × 70 dimension landscape, the data is given a border of 29 pixels on each side, such that a $2^7 \times 2^7$ landscape is created, with the extra space filled in uniformly with the highest energy value obtained from the given file). All borders are removed at the end of the computation.

1. Image Segmentation. Image segmentation is the process of dividing the landscape into a number of equally sized squares and recording the coordinates of the local minimum-energy values within each (Figure 1). The "segmentation depth", n, defines the number of these subdivisions created and, thus, the number of local minimum-energy coordinates recorded. The user defines the
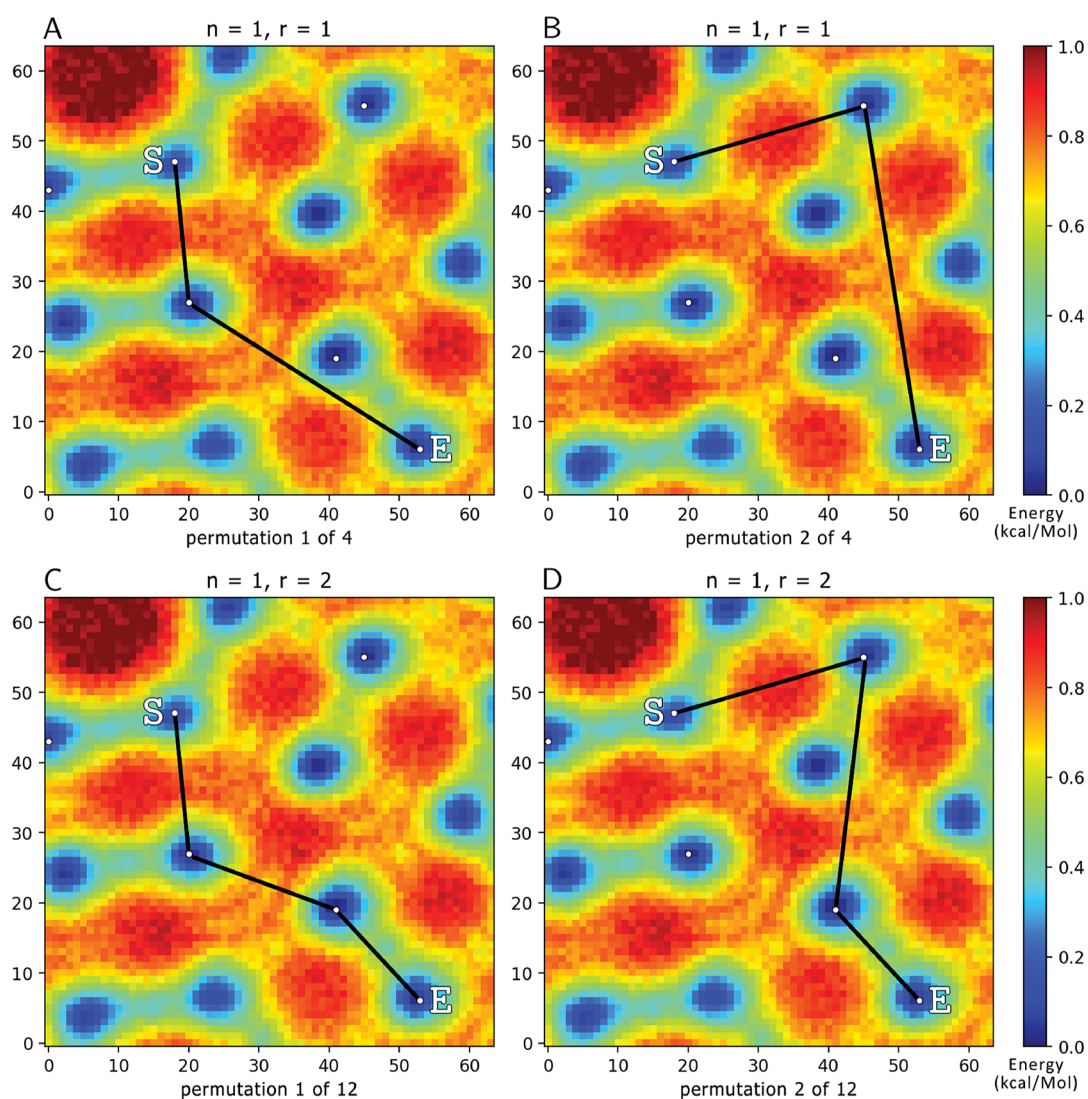
Figure 2. Computationally generated landscape showing permutational order r_1 (A, B) and r_2 (C, D), each assigned n_1. User-defined start and end points are labeled S and E, respectively. For r_1, 4 permutations exist with one unique midpoint each, for which two are shown (A, B). For r_2, 12 permutations exist with two unique midpoints, for which two are shown (C, D). The energies along the full set of these pathway permutations are integrated and compared (e.g., selecting the lowest-energy path of the 4 (A, B) or the lowest-energy path of the 12 (C, D), etc.). The lowest-energy paths obtained from each $\{n_i, r_j\}$ combination are then compared to find the lowest-energy approximation for all given combinations of permutational orders and segmentation depths.

set of segmentation depths, $\{n_i\}$, to use within the parameters tab; e.g., $\{n_7, n_5\}$ or $\{n_7, n_5, n_4, n_3\}$, etc.

For each n_i, $N_i = 4^{n_i}$ image subdivisions are created, with an equivalent number of minimum-energy points stored - such that, for n_1, the coordinates of 4 local energy-minima are stored (Figure 1A). Within this set, the highest allowable value of n_i, n_{max} is defined via that subdivision where further image divisions are impossible (e.g., for a 64 × 64 landscape, n_{max} = $\log_2(64)$ = 6 subdivisions), so that every segmented grid spans an area of exactly 1 pixel (Figure 1D).

As the segmentation depth is increased from n_1 to n_{max}, previously overlooked higher-energy minima (relative to the system's global minima) are geometrically separated from the global minimum values and placed into their own neighboring grids. Since these locations become newly defined local energy-minima at higher depths, this method ensures that all points are eventually considered as potential nodes in the permuta-

tional analysis of the pathway, to be outlined in the next section.

2. Permutational Analysis. Each set of 4^{n_i} minimum-energy coordinates obtained from each n_i during image segmentation is subsequently used as transit options for comparing potential paths across the energy landscape. Here, for each n_i, permutations of straight lines are connected between the set of 4^{n_i} coordinates and the user-defined start point, S, and end point, E (Figure 2). The "permutational order", r, defines how many intermediate transit options can exist between S and E. For example, if there are 4 transit options (via n_1), r_1 would sample all ways of bridging S and E with straight lines using only 1 transit coordinate - a total of 4 permutations are possible. These permutations are created via Itertools, such that for n_1 with minima M_1, M_2, M_3, and M_4, paths $S \rightarrow M_1 \rightarrow E$, $S \rightarrow M_2 \rightarrow E$, $S \rightarrow M_3 \rightarrow E$, and $S \rightarrow M_4 \rightarrow E$ are generated (Figure 2A, Figure 2B).

For each user-defined value of n_i, a partnering value for the permutational order, r_j, must also be chosen, e.g., $\{(n_7, r_1), (n_5, r_2)\}$, etc. Here, r_j can range from r_0 to r_{max}, with r_0 representing the line directly connecting S to E (with no midpoints in between). The value of r_{max} is restricted within the user interface to 5, with higher values deemed counterproductive to the aim of this interpolative procedure. Within this formulation, the standard permutation notation $^{N_i}P_{r_j}$ represents a unique combination of some n_i and r_j, with $N_i = 4^{n_i}$. As an example, for n_2 with r_2, one such path $S \rightarrow M_1 \rightarrow M_{16} \rightarrow E$ is created among 239 other permutations, via $N_2 = 4^{n_2} = 16$ and $^{16}P_2 = \frac{16!}{(16-2)!} = 240$.

Within each member of a given $\{n_i, r_j\}$ permutational pool (e.g., $\{^{4^1}P_1, ^{4^1}P_2, ...\}$, as seen in Figure 2), each set of ordered transit points is connected by straight lines, drawn using a variant of Bressenham's Line Algorithm modified for energy awareness. These path approximations act to gather overhead awareness across different regions in the energy landscape. The total energies across the paths formed by each of these permutations is then independently integrated, with only the transit points from the minimum-energy permutation stored for that $\{n_i, r_j\}$. This process is then repeated for all combinations of $\{n_i, r_j\}$, ultimately only storing the set of transit points belonging to the lowest-energy path approximation discovered between the initial S and E. For example, if parameters $\{(n_7, r_1), (n_5, r_2)\}$ are chosen, $16,384 + 1,047,552 = 1,063,936$ paths are generated - with only the original list of minimum-energy coordinates pertaining to the path of overall lowest-energy stored. These minimum-energy coordinates are then introduced as inputs for the subsequent computations.

3. Branching Recursion. POLARIS next uses the minimum-energy nodes discovered in the previously performed permutational analysis recursively as inputs to itself (replacing the initial user-defined inputs S and E with the set of these newly discovered intermediate, minimum-energy nodes). First, a for-loop is created between each pair of intermediate nodes, such that if the pathway containing points $S \rightarrow M_1 \rightarrow M_2 \rightarrow E$ were found as a minimum among all other permutations in the previously described steps, a loop containing $S \rightarrow M_1$, $M_1 \rightarrow M_2$, and $M_2 \rightarrow E$ would emerge.

Within this loop, POLARIS performs branching recursion (Figure 3), repeating all of these aforementioned procedures on its newly obtained outputs. From here, the permutational steps are repeated for each new set of start and end points discovered - proceeding recursively down each inner branch, until two output points having lowest-line approximation r_0 are found within each one of its individual *leaves* (defining the maximum extent of each *branch*). Upon encountering each leaf-break, the coordinate of that leaf is globally recorded. This process continues with the algorithm navigating throughout its recursive hierarchy until the path of least action is filled in completely with a set of coordinates spanning from the initial, user-defined start point to the initial, user-defined end point.

This process is akin to performing interpolation on the space between each start and end point within each recursion, whereby two anchors are known, and all points in between unknown. Those unknown points are then iteratively discovered via a complete set of the previously described permutational comparisons. Ultimately, these recursive steps break down the best global line approximation into continuously finer subsections within each recursion until the
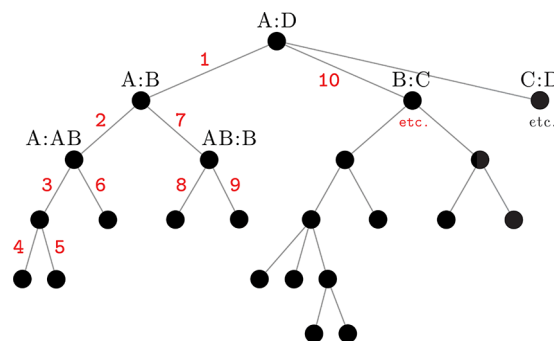

Figure 3. (In the following, we use the notation POLARIS-(input1, input2) to describe one complete cycle of Image Segmentation and Permutational Analysis, starting with points input1 and input2. If the minimum-line outputs from the initial seed, POLARIS(A, D), include points B, C, and D, POLARIS(A, B), POLARIS(B, C), and POLARIS(C, D) are subsequently called - with each one forming a new branch across the next-highest tier in the hierarchy. Red numbers indicate the order in which POLARIS navigates its recursive hierarchy. Broken leaves represent pairs of data points where further segmentation is futile; i.e., no lower energy line exists between them (occurring exclusively when r_0 is returned as the lowest-energy approximation).

finest-granular path is resolved - giving rise to both global and local awareness of the landscape.

To remove unnecessary permutational calculations within the branching recursion hierarchy and, thus, shorten the computational effort, once a set of $\{n_i, r_j\}$ parameters is chosen, only those combinations within the set that are spatially viable for the current S and E coordinates are computed. Thus, if the Euclidean distance between one set of S and E coordinates in the branch is smaller than the distance covered by the minimal spanning line of a specifically chosen combination of n and r, the permutational analysis for that n and r combination will be skipped.

4. Pathway Pruning. As an artifact of Bresenham's Line Algorithm for drawing straight line approximations, there remain a multitude of ways to alias a diagonal line between the same two sets of coordinates (e.g., two different permutations can emerge for the same straight line when drawn $S \rightarrow E$ versus $E \rightarrow S$). To work around this problem, pruning techniques have been implemented on the final set of coordinates of the complete least action trajectory. These include finite, local perturbations along each coordinate in the completed path. So long as these perturbations preserve the continuity of the overall path (no breaks), every point is iteratively sent to occupy its set of unrestricted neighboring pixels and reshuffle into those coordinates that ultimately minimize the pathway's global energy.

As a note, any coordinates having zero-energy are never pruned from the pathway, safeguarded during this process for their significance in the energy landscape. This ensures that the overall line may still pass through these locations - including situations where the zero-energy point is neighbored by three other points in the path. Within this exception, it follows that no energy contribution is added to the path of least action.

■ **RESULTS**

An experimentally generated free-energy landscape was used for testing POLARIS' performance (Figure 4) - previously constructed outside of this study through application of the ManifoldEM techniques to a set of ribosomal cryo-EM data.[19]
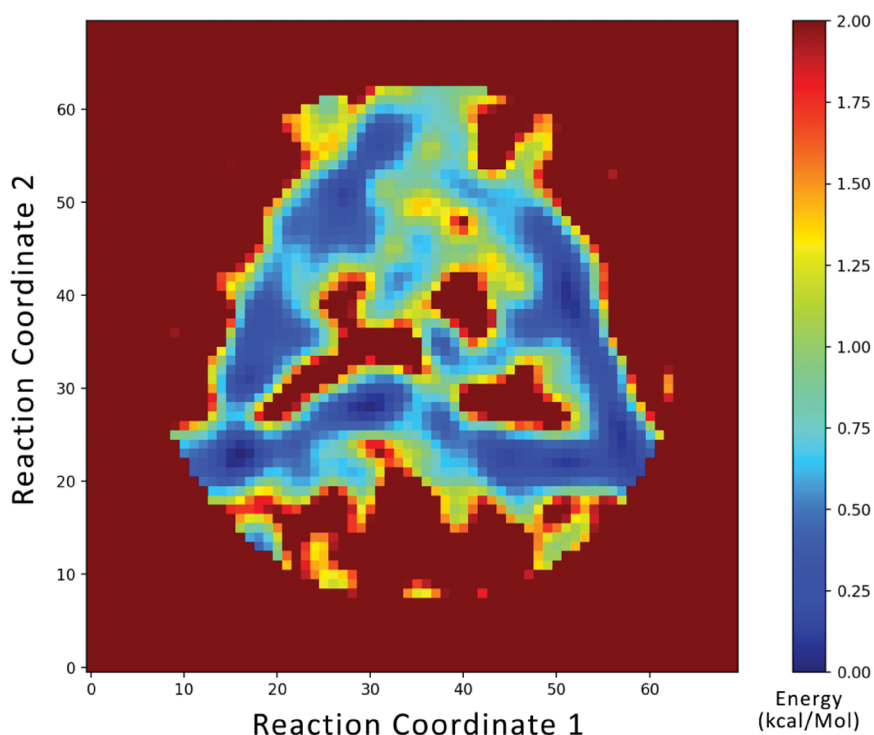
Figure 4. Energy landscape for the reaction coordinates of the naked ribosome during translation elongation. Data taken from ribosomal reaction coordinates determined from cryo-EM data by ManifoldEM.[19]
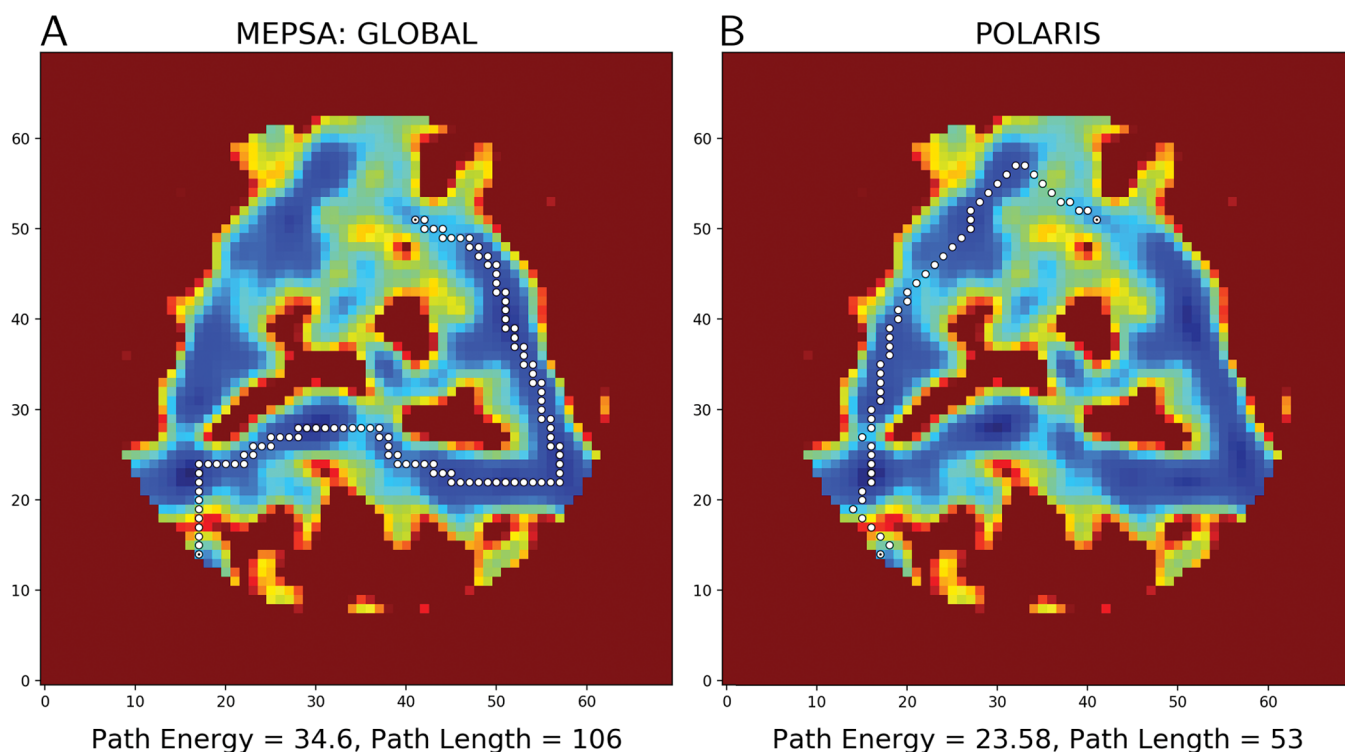

Figure 5. Comparison between MEPSA's "GLOBAL" algorithm (A) and POLARIS (B). POLARIS found a radically different global path than MEPSA, returning a path of integrated energy 11.0 kcal/mol less with a 53-point difference in path lengths. For the POLARIS run, the option "Transition State Weighting" was used with search depths $\{r_1, n_7\}$, $\{r_2, n_5\}$, and $\{r_4, n_3\}$, taking approximately 6 min using multiprocessing with 4 1.2 GHz processors and 8 GB of memory.

The ribosome has long been described as a "thermal ratchet machine",[45] whereby random energetic perturbations lead to large-scale shifts across its available configurations. The landscape obtained represents the conformational state space available to the naked ribosome (i.e., the ribosome in the absence of its functional ligands), with the "reaction coordinates" corresponding to the two highest-ranking, orthogonal factors for motion. Based on the manifold analysis,
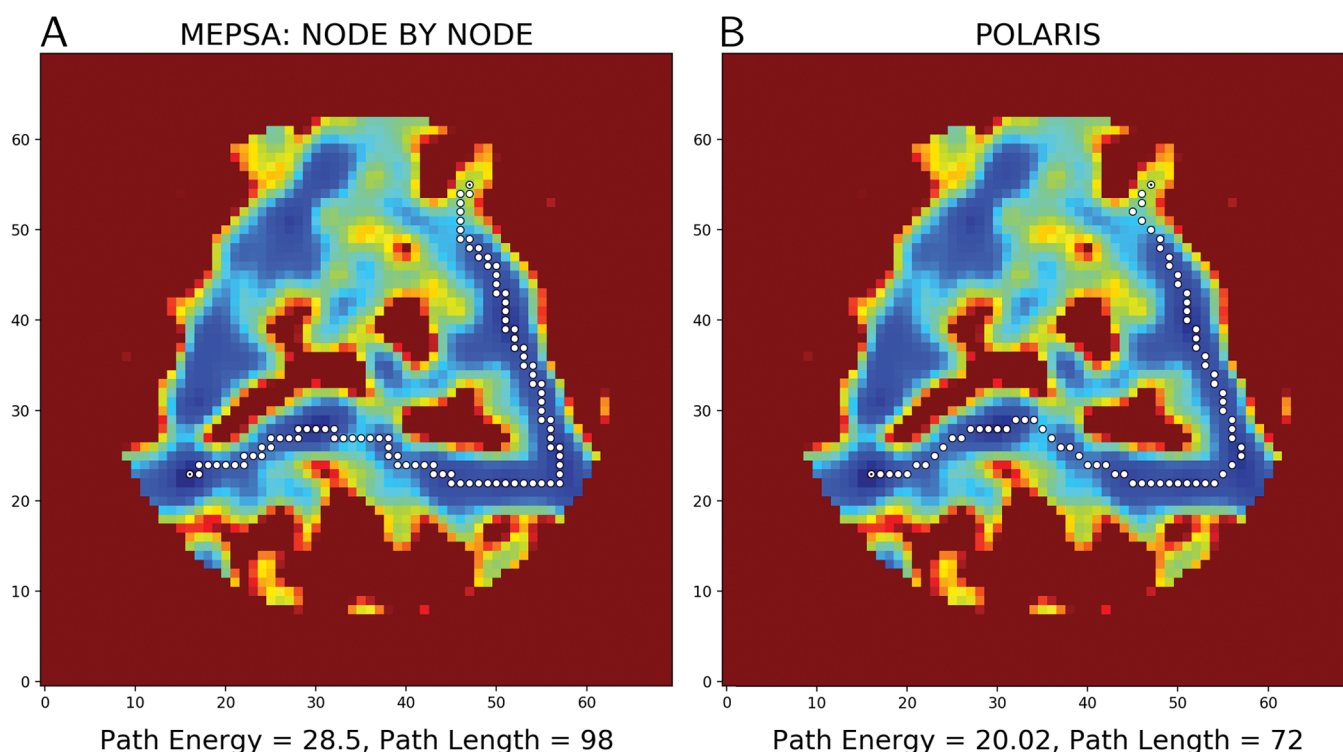
Figure 6. Comparison between MEPSA's "NODE BY NODE" algorithm (A) and POLARIS (B). While both algorithms again found the same approximate global path, POLARIS returned a path of integrated energy 8.4 kcal/mol less with a 26-point difference in path lengths. For the POLARIS run, "Transition State Weighting" was used with search depths $\{r_1, n_7\}$, $\{r_2, n_5\}$, and $\{r_3, n_3\}$, taking approximately 5 min using multiprocessing with 4 1.2 GHz processors and 8 GB of memory.

these orthogonal factors are thought to represent the ribosome's leading, human-defined conformational motions during translocation - encompassing a composite of large-scale movements of its two subunits,[19,46] including an intersubunit rotation and a swivel motion of the small-subunit head. Analysis of 3D reconstructions along the naked ribosome's trajectory revealed that in the thermal bath it undergoes conformational changes akin to those observed for factor- and GTP hydrolysis-driven translating ribosomes - clearly demonstrating the macromolecule's intrinsically flexible architecture.[3]

Conformations reported by cryo-EM reveal that each "state" of the ribosome is actually an ensemble of structurally similar configurations clustered within a specific minimum-region on the energy landscape.[3] At the same time, the rugged, higher-energy hills correspond to the less-favorable states of the macromolecular complex as it travels from one minimum to the next. Given such a system, to properly delineate the ribosome's key mechanisms of translation control, POLARIS was used to determine the most likely transitional pathways between different selections of energy basins.

Comparison of Results from POLARIS and MEPSA. For validation, POLARIS' results were compared to those obtained with a published energy landscape analysis algorithm entitled MEPSA (Minimum Energy Pathway Analysis),[30] which uses an approach similar to Dijkstra's algorithm,[24] with small differences in the sampling and trace-back. The metric used to compare the two algorithms is the final integrated energy from a single continuous list of coordinates spanning between any two user-defined points.

The MEPSA tool allows two options for algorithmic comparison: the self-defined, less accurate "GLOBAL" option (capable of accepting arbitrary user start and end inputs as well as predefined anchor points) and the more accurate option "NODE BY NODE" (accepting only predefined global minima as user inputs). For the first comparison, the "GLOBAL" approach was chosen to compare each program's output given any set of user-defined start and end coordinates, here using `(41, 51)` and `(17, 14)`, respectively, on the ribosome energy landscape (Figure 4). Chosen for the complexity of the regions in between, this trajectory includes opportunities for traversing many narrow, branching low-energy pathways throughout the center of the landscape, as well as a final leap across a mountain range of highest-energy coordinates into a pool of minimum-energy.

The net energy of the least-energy pathway found by MEPSA was 34.6 kcal/mol with a length of 106 points, containing many unnecessary regions where the algorithm appeared to draw jagged lines between its predefined nodes (Figure 5A). For the same points, POLARIS returned a globally different 53-point 23.6 kcal/mol path (Figure 5B), showing a difference of 11.0 kcal/mol between the two algorithms in favor of POLARIS. To place this energy difference within biological context, under standard conditions, the kinetic energy released by the hydrolysis of a single molecule of GTP is approximately 7 kcal/mol.[47]

For the "NODE BY NODE" comparison, MEPSA's node 2 `(16, 23)` and node 41 `(47, 55)` were selected from the set of MEPSA's predefined nodes based on their proximal similarity to the arbitrary points selected in the "GLOBAL" comparison above. The pathway found by MEPSA returned an integrated energy of 28.5 kcal/mol with 98 points (Figure 6A). POLARIS used the same nodes 2 and 41 as user inputs and identified a 72-point pathway having an integrated energy of 20.0 kcal/mol (Figure 6B). The total difference between the
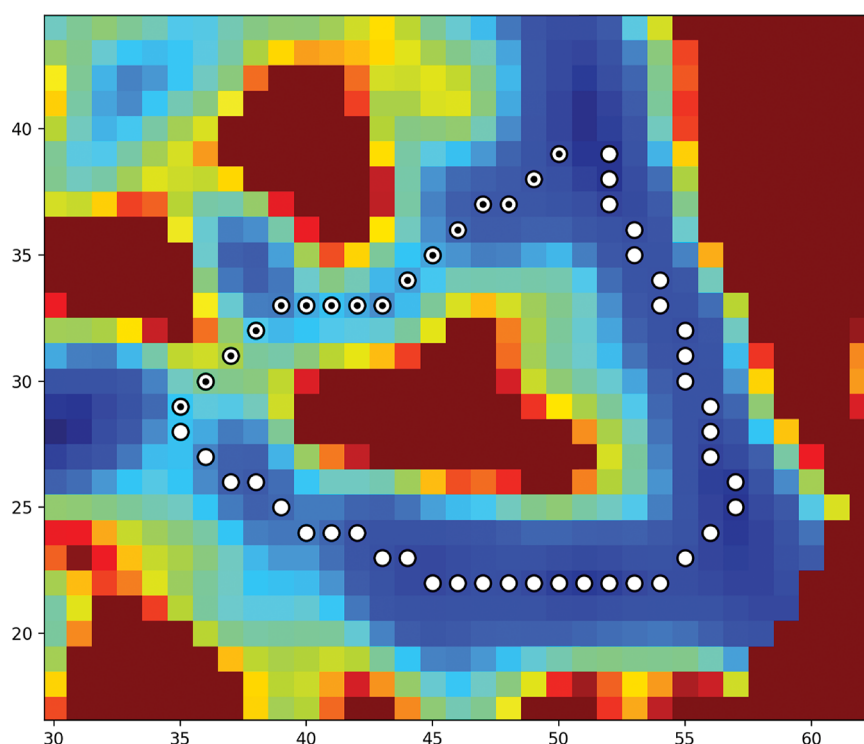
Figure 7. Comparison of two approximately degenerate paths found via POLARIS. To create each path, midpoints were placed alongside the initial user-defined start and end points - forcing POLARIS' exploration of both routes independently on separate runs. These two diverging paths go through the center (black points) and southeast (white points) regions of the landscape and contain path energies 9.0 kcal/mol (with 16 points) and 8.5 kcal/mol (with 37 points), respectively. With only a difference of 0.5 kcal/mol between them, it is possible that such a degenerate least-energy bifurcation could represent novel reaction mechanics that allow for flexibility in macromolecular processes both spatially and temporally. For example, the bifurcation seen above may represent a shortcut in the ribosomal work cycle (elongation) that only becomes available under specific buffer or temperature conditions - allowing the ribosome to modulate its reaction rates based on fluctuating environmental signals.

two minimum-energy approximations found by the two algorithms was 8.5 kcal/mol, again favoring POLARIS.

DISCUSSION

In both comparisons with MEPSA, POLARIS' outputs gave a substantially lower energy path while providing higher flexibility in the choice of user-defined start and end points, as well as in the number of optional user-defined transit locations. MEPSA's more flexible "GLOBAL" analysis performed most poorly in comparison to POLARIS, with a different global route and a greater difference in path energies and number of path-points. While the final paths from MEPSA's "NODE BY NODE" analysis and POLARIS followed the same broad regions throughout the landscape, the MEPSA algorithm appeared to favor the production of jagged segments and deviated from POLARIS' minimum path within regions of its locally defined nodes (such as in the southeast corner of the landscape).

As a further explanation for this large difference in energy and number of path points, MEPSA grows its trajectory in strictly cardinal directions (N, E, S, W), while POLARIS makes allowances for both cardinal and ordinal directivity (N, NE, E, SE, S, SW, W, NW) whenever locally required. Since the energy landscape represents a molecule's movement as a combination of its two reaction coordinates, any combination of the two must be accounted for - including cardinal (where only one coordinate is altered at a time) and ordinal (where both coordinates change simultaneously). In theory, this

principle should hold for all reaction coordinates, regardless of context.

As for computation time, it should be noted that MEPSA generated the above paths within seconds. However, MEPSA's self-defined "global minimum" solutions were undershot considerably by POLARIS' techniques. While POLARIS is capable of similar speeds at drastically lowered segmentation depth and permutational order, the lowest-energy trajectories found (shown here) required more computation time (5−10 min, with 4 1.2 GHz processors and 8 GB of memory). Thus, when weighting accuracy over timing, POLARIS' methods seem considerably more fit. Although these differences may seem trivial from a macroscopic view, such exactitude is essential on the microscopic level for accurately calculating biologically relevant reaction rates via downstream algorithms. In the discussion to follow, POLARIS will be analyzed in terms of three features, "Completeness", "Accuracy", and "Complexity".

Completeness. The greater the number of permutations that POLARIS is allowed to iterate through and compare against, the more accurate its computations will be in obtaining the minimum-energy path. However, since the purpose of POLARIS is to limit this exhaustive search, it instead seeks to iteratively find the set of optimal approximations by use of feasible parameters (user-defined $\{^{N}P_r\}$ combinations) between any two successive recursions, thus performing a series of computationally inexpensive operations that ultimately give rise to the same output as would one computationally exhaustive approach. Because of this, limits must be drawn

on the total number of permutations allowed for the program to compute and compare between.

This total permutation limit is set by the user under the "Parameters" section found within the "Settings" tab (Figure S2). Here, combinations of n and r can be chosen for each $\{r_{i=1}^{5}\}$, with the total number of permutations accessible from that combination calculated to the right (as given within the adjacent $P(4^n, r)$ entry box).

For r_1, the maximum image subdivision depth, n_{max}, is automatically chosen from the given data file. At this depth, every value on the landscape is represented as a possible minimum node at least once, with this parameter thus comparing between all permutations of single points between each new S and E varying across all recursions. As n_{max} contains all other minima options within it (obtained individually from n_1 to n_{max-1}), it is redundant to compute any other combination less than this value for r_1 (with equivalent logic holding for all other r_i).

Accuracy. Uncertainty is first introduced in the choice of the overhead metric and its ability to realistically encapsulate the dynamics of the system being explored. For example, for the potential energy surface of a chemical reaction, a reaction could be approximated either by the path having the lowest integrated energy or by the path having to surmount the lowest maximum peak or "activation energy" (Figure S3). To provide flexibility in this overhead metric, POLARIS offers the "Transition State Weighting" constraint, which can be enabled to weight the comparison of competing lowest-energy paths based on their rate-limiting step (point of maximal energy through which that path passes) instead of by just the net integrated energy along that path (Figure S4). When this feature is activated, POLARIS weights all energies across the landscape based on a power function - keeping lower energies approximately untouched while making higher-energy coordinates increasingly more unfavorable. When bifurcation opportunities exist within a landscape (as in the case of the central, branching region of Figure 4), significant uncertainty can be introduced for the path of least action. This is especially relevant when such bifurcations are approximately degenerate - leading to two distinct paths of almost equivalent energy spanning radically different regions of the landscape.

During pathway comparisons on the example ribosome landscape, POLARIS isolated one such nearly equivalent-energy bifurcation - dividing the ribosome's most probable sequence of configurations into two discrete sets separated by a high-energy island (Figure 7). From an accuracy perspective, this instance illustrates the reliance on POLARIS' user parameters in constructing its least-energy path, as well as the importance of the user experimenting with these parameters to achieve optimum results (Figure S5).

As an aside, an exploration of such bifurcations can be instrumental for elucidation of a given system's underlying dynamics. For example, the user can supply defined midpoints between the set of start and end coordinates to resolve a bifurcation (e.g., by placing one anchor at a minimum point in the middle of the landscape for one run, followed by another anchor instead at a minimum in the southeast corner for the next), with the added ability to then compare each least-energy approximation after completion of both runs.

Finally, as a note on scalability, it is anticipated that POLARIS will discover slightly different paths of least action given different "binnings" of the same energy landscape (e.g., if a 70 × 70 pixel landscape is compressed into a lower resolution

image). However, these results should only be globally different when relatively close bifurcation possibilities also exist within that landscape along the regions of the intended path (as is similarly described above).

Complexity. In the application, the number of contending lowest-energy paths becomes very large as the size and topological complexity of the landscape increases - making an exhaustive computational search for the path of least action infeasible toward these limits. Large and highly complex maps should therefore be used with this limitation in mind.

The range of available permutations to search through is proportional to the size of the data file (as seen via the available combinations of n and r). As increasingly larger combinations of $\{n_i, r_j\}$ are chosen, the time for each search will also increase, as defined by the number of available permutations (i.e., $\{^N P_r\}$). Thus, the time required by each run is ultimately governed by the algorithm's rate limiting step, `itertools.permutations`, via $O(^N P_r)$.

In its current state, POLARIS is best applied to energy landscapes of biologically relevant size, at dimensions similar to the scale of reaction coordinates seen within chemistry and biophysics,[19,48] (i.e., a 70 × 70 dimension landscape). To minimize computational efforts for landscapes larger than these dimensions, pixel values could be binned beforehand or masked out to only those regions of interest - both of which will be supplied as future options within the POLARIS user interface.

■ ASSOCIATED CONTENT

ⓈI **Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.9b01108.

> Figure S1, coordinates tab of POLARIS user interface; Figure S2, settings tab of POLARIS user interface; Figure S3, schematic of energy landscape weighting functions; Figure S4, graph comparing weighting functions; Figure S5, output comparison from weighting functions; Figure S6, exemplary output for path of least action; Figure S7, exemplary transition state diagram for path of least action (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Joachim Frank — *Department of Biological Sciences and Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, United States;* ⓞ orcid.org/0000-0001-5449-6943; Email: jf2192@cumc.columbia.edu

Author

Evan Seitz — *Department of Biological Sciences, Columbia University, New York, New York 10032, United States;* ⓞ orcid.org/0000-0001-6844-2444

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.9b01108

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Wales, D. *Energy Landscapes*; Cambridge University Press: 2003; DOI: 10.1017/CBO9780511721724.

(2) Errol, L. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, 2nd ed.; Springer: 2011; p 21, DOI: 10.1007/978-90-481-3862-3.

(3) Whitford, P. C.; Altman, R. B.; Geggier, P.; Terry, D. S.; Munro, J. B.; Onuchic, J. N.; Spahn, C. M. T.; Sanbonmatsu, K. Y.; Blanchard, S. C. *Dynamic views of ribosome function: Energy landscapes and ensembles*; Springer: Vienna, 2011; DOI: 10.1007/978-3-7091-0215-2_24.

(4) Munro, J.; Sanbonmatsu, K.; Spahn, C.; Blanchard, S. Navigating the ribosome's metastable energy landscape. *Trends Biochem. Sci.* **2009**, *34* (8), 390−400.

(5) Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. Biomolecular Dynamics: Order-Disorder Transitions and Energy Landscapes. *Rep. Prog. Phys.* **2012**, *75*, 076601.

(6) Weber, D.; Bellinger, D.; Engels, B. *Methods Enzymol.* **2016**, *578*, 145−167.

(7) Wales, D.; Scheraga, H. *Science* **1999**, *285*, 1368−1372.

(8) Schramm, V. Enzymatic Transition States and Transition State Analog Design. *Annu. Rev. Biochem.* **1998**, *67*, 693−720.

(9) Truhlar, D.; Garrett, B.; Klippenstein, S. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, *100*, 12771−12800.

(10) Pappu, R.; Hart, R.; Ponder, J. Analysis and Application of Potential Energy Smoothing and Search Methods for Global Optimization. *J. Phys. Chem. B* **1998**, *102*, 9725−9742.

(11) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **1991**, *254*, 1598−1603.

(12) Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature (London, U. K.)* **2007**, *450*, 964−972.

(13) Smock, R. G.; Gierasch, L. M. Sending signals dynamically. *Science* **2009**, *324* (5924), 198−203.

(14) Frank, J. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*; Oxford University Press: Oxford, NY, 2006; DOI: 10.1093/acprof:oso/9780195182187.001.0001.

(15) Sigworth, F. J. Principles of cryo-EM single-particle image processing. *Microscopy* **2016**, *65*, 57−67.

(16) Frank, J. Generalized single-particle cryo-EM: a historical perspective. *Microscopy* **2016**, *65*, 3−8.

(17) Frank, J. Single-Particle Reconstruction - Story in a Sample (Nobel Lecture); 2017; DOI: 10.1002/anie.201802770.

(18) Fischer, N.; Konevega, A. L.; Wintermeyer, W.; Rodnina, M. V.; Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature (London, U. K.)* **2010**, *466*, 329−333.

(19) Dashti, A.; Schwander, P.; Langlois, R.; Fung, R.; Li, W.; Hosseinizadeh, A.; Liao, H. Y.; Pallesen, J.; Sharma, G.; Stupina, V. A.; Simon, A. E.; Dinman, J. D.; Frank, J.; Ourmazd, A. Trajectories of the ribosome as a Brownian nanomachine. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 17492−17497.

(20) Chen, B.; Frank, J. Two promising future developments of cryo-EM: capturing short-lived states and mapping a continuum of states of a macromolecule. *Microscopy* **2016**, *65*, 69−79.

(21) Frank, J.; Ourmazd, A. Continuous Changes in Structure Mapped by Manifold Embedding of Single-Particle Data in Cryo-EM. *Methods (Amsterdam, Neth.)* **2016**, *100*, 61−67.

(22) Maupertuis, P. L. Accord de differentes loix de la nature qui avoient jusqu'ici paru incompatibles; Wikisource, 1744.https://fr.wikisource.org/wiki/Accord_de_diff%C3%A9rentes_loix_de_la_nature_qui_avoient_jusqu%E2%80%99ici_paru_incompatibles (accessed Feb 4, 2020).

(23) Euler, L. Reflexions sur quelques loix generales de la nature; *Memoires de l'academie des sciences de Berlin*; 1748.

(24) Dijkstra, E. W. *Communication with an Automatic Computer*; 1959.

(25) Cherkassky, B. V.; Goldberg, A. V.; Radzik, T. Shortest paths algorithms: theory and experimental evaluation. *Mathematical Programming* **1996**, *73*, 129−174.

(26) Hart, P. E.; Nilsson, N. J.; Raphael, B. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics* **1968**, *4*, 100−107.

(27) Floyd, R. Algorithm 97: Shortest Path. *Commun. ACM* **1962**, *5*, 345.

(28) Milanfar, P. A Tour of Modern Image Filtering: New Insights and Methods, Both Practical and Theoretical. *IEEE Signal Processing Magazine* **2013**, *30*, 106−128.

(29) Cheung, G.; Magli, E.; Tanaka, Y.; Ng, M. K. Graph Spectral Image Processing. *Proc. IEEE* **2018**, *106*, 907−930.

(30) Marcos-Alcalde, I.; Setoain, J.; Mendieta-Moreno, J. I.; Mendieta, J.; Gómez-Puertas, P. MEPSA: minimum energy pathway analysis for energy landscapes. *Bioinformatics* **2015**, *31*, 3853−3855.

(31) Pratt, L. A statistical method for identifying stransition states in high dimensional problems. *J. Chem. Phys.* **1986**, *85*, 5045.

(32) Elber, R.; Karplus, M. A method for determining reaction paths in large molecules: application to myoglobin. *Chem. Phys. Lett.* **1987**, *139*, 375−380.

(33) Sheppard, D.; Henkelman, G.; Terrell, R. Optimization methods for finding minimum energy paths. *J. Chem. Phys.* **2008**, *128*, 128−137.

(34) Henkelman, G. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901.

(35) Trygubenko, S.; Wales, D. A doubly nudged elastic band method for finding transition states. *J. Chem. Phys.* **2004**, *120*, 2082.

(36) E, W.; Vanden-Eijnden, E.; Ren, W. String method for the study of rare events. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2002**, *66*, 052301.

(37) E, W. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.* **2007**, *126*, 164103.

(38) Seitz, E. Repository for POLARIS (bioRxiv Release). *Zenodo* 2020; https://zenodo.org/record/3612185#.XjnV9WhKiUk (accessed Feb 4, 2020).

(39) Botea, A.; Muller, M.; Schaeffer, J. Near optimal hierarchical path-finding. *Journal of Game Development* **2004**, *1*, 7−28.

(40) Rossum, G. *Python Reference Manual*; 1995.

(41) PyQT, *PyQt Reference Guide*; 2012. https://het.as.utexas.edu/HET/Software/PyQt/index.html (accessed Feb 4, 2020).

(42) Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90−95.

(43) Oliphant, T. E. *Guide to NumPy*, 2nd ed.; CreateSpace Independent Publishing Platform: USA, 2015.

(44) Kuzmin, Y. P. Bresenham's Line Generation Algorithm with Built-in Clipping. *Comput. Graph. Forum* **1995**, *14*, 275−280.

(45) Spirin, A. The Ribosome as a Conveying Thermal Ratchet Machine. *J. Biol. Chem.* **2009**, *284*, 21103−21119.

(46) Whitford, P. C.; Blanchard, S. C.; Cate, J. H. D.; Sanbonmatsu, K. Y. Connecting the Kinetics and Energy Landscape of tRNA Translocation on the Ribosome. *PLoS Comput. Biol.* **2013**, *9*, e1003003.

(47) Aqvist, J.; Kamerlin, S. Exceptionally large entropy contributions enable the high rates of GTP hydrolysis on the ribosome. *Sci. Rep.* **2015**, *5*, 15817.

(48) Dashti, A.; Hail, B. D.; Mashayekhi, G.; Schwander, P.; des Georges, A.; Frank, J.; Ourmazd, A. *Functional Pathways of Biomolecules Retrieved from Single-particle Snapshots*. 2018, DOI: 10.1101/291922 https://www.biorxiv.org/content/10.1101/291922v1 (accessed Feb 4, 2020).