Supporting Information

POLARIS: Path of Least Action Analysis on Energy Landscapes

Evan Seitz[†] and Joachim Frank[*,†,‡]

[†]*Department of Biological Sciences, New York, NY 10032, USA*

[‡]*Department of Biochemistry and Molecular Biophysics, New York, NY 10032, USA*

E-mail: jf2192@cumc.columbia.edu

User Interface

POLARIS supports Comma Separated Spreadsheet (.csv, .txt) inputs in the form of $m \times n$ heat maps. The GUI structure consists of a main window with two tabs entitled 'Coordinates' and 'Settings' (Figure S1, Figure S2). Under the 'Coordinates' tab, an energy landscape can be imported and viewed. Once loaded, between 2 and 10 sets of user coordinates can be chosen, allowing the user to find the one-way path of least action between any two points, a series of points, or creating a cycle of points possibly corresponding to stable, reversible processes.

Advanced settings can be accessed via the 'Settings' tab, where parameters can be set to specify the algorithm's maximum search depth and performance. The 'Transition State Weighting' option can also be enabled to additionally weight POLARIS' comparison of competing lowest-energy paths based on rate limiting steps, as opposed to only the integrated energy along all coordinates in each path (Figure S3, Figure S4 and Figure S5). Multiple processors can also be selected to perform the most time-intensive tasks in parallel. After
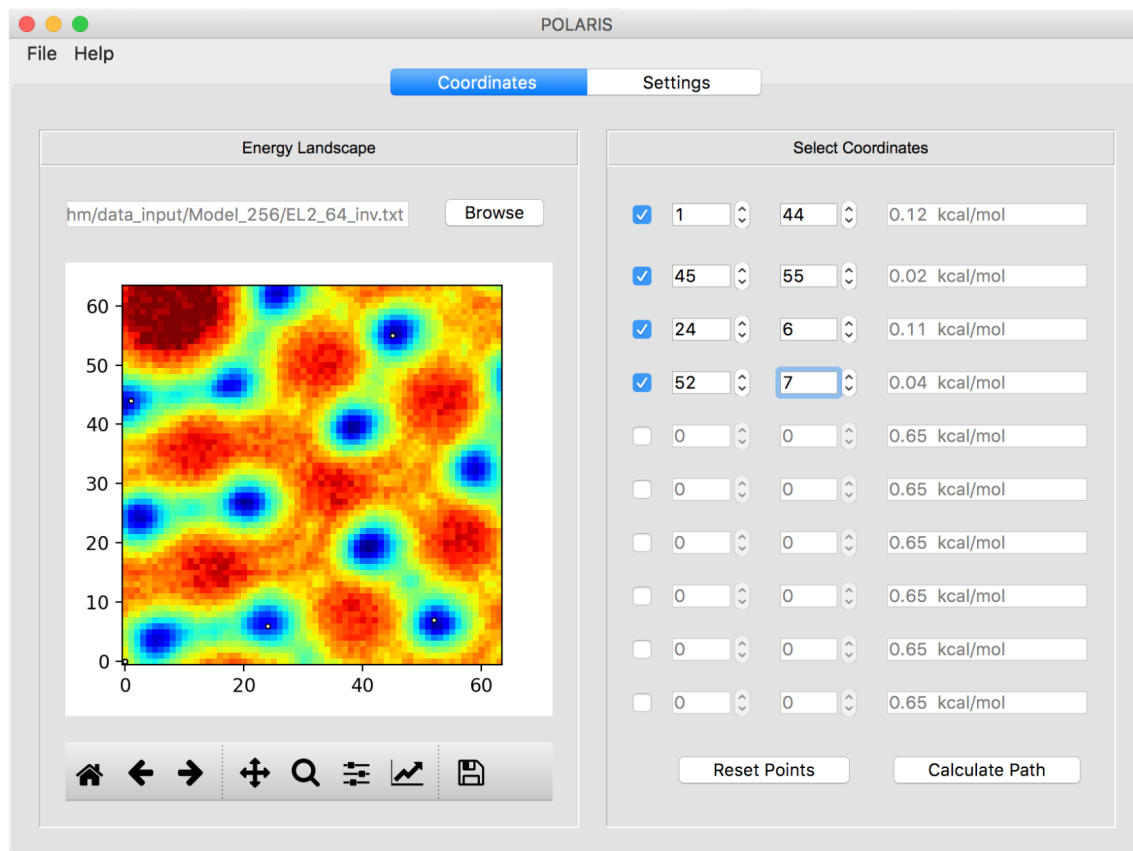
Figure 1: Image from the main page of the POLARIS user interface, allowing users to load a valid data file and add up to 10 intermediate transit locations on the energy landscape as desired.

these settings have been decided upon, the user can proceed back to the 'Coordinates' tab and click the 'Calculate Path' button to initiate the path finding algorithm.

Every value of n can be changed in the 'Parameters' section (Figure S2) by first unmarking its corresponding checkbox, if active. As that value of n is altered, the total permutation count will automatically be updated on the right. Adjusting these n parameters for each value of r such that the $P(4^n, r)$ values are all of the same order of magnitude will prevent any extreme rate limiting steps during the computation. Once these parameters have been set, they can be activated by checking each checkbox - thus instructing POLARIS to use those specific parameters within its search.
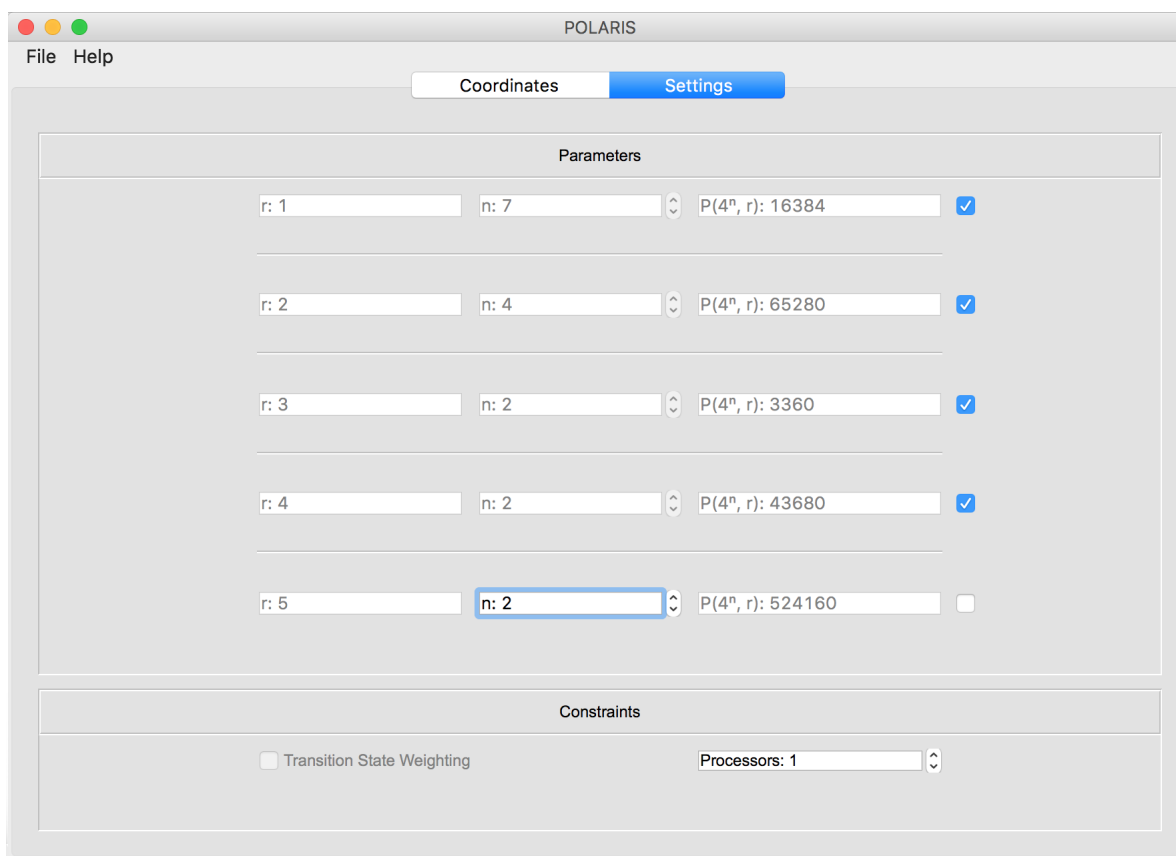

Figure 2: Image from the 'Settings' tab of the POLARIS user interface, allowing users to set parameters and constraints as desired. Note that each n_j has been chosen such that the total number of permutations for each $\{^N P_r\}$ combination are of approximately the same order of magnitude, and initially chosen as to be in line with the range provided by n_{max} (16,384 total permutations). As this order of magnitude is increased, so too will the computation time.

Upon completion of the back-end algorithm, landscape-path plots (Figure S6) and transition state diagrams (Figure S7) are saved automatically as .png images. Additionally, coordinates of the minimum-energy path and its respective energies are automatically generated within plain text files in the form of three column lists (x, y, $z \in$ energy). These files also contain the total integrated energy, overall length of each trajectory, user-defined parameters and elapsed computation time in their header.
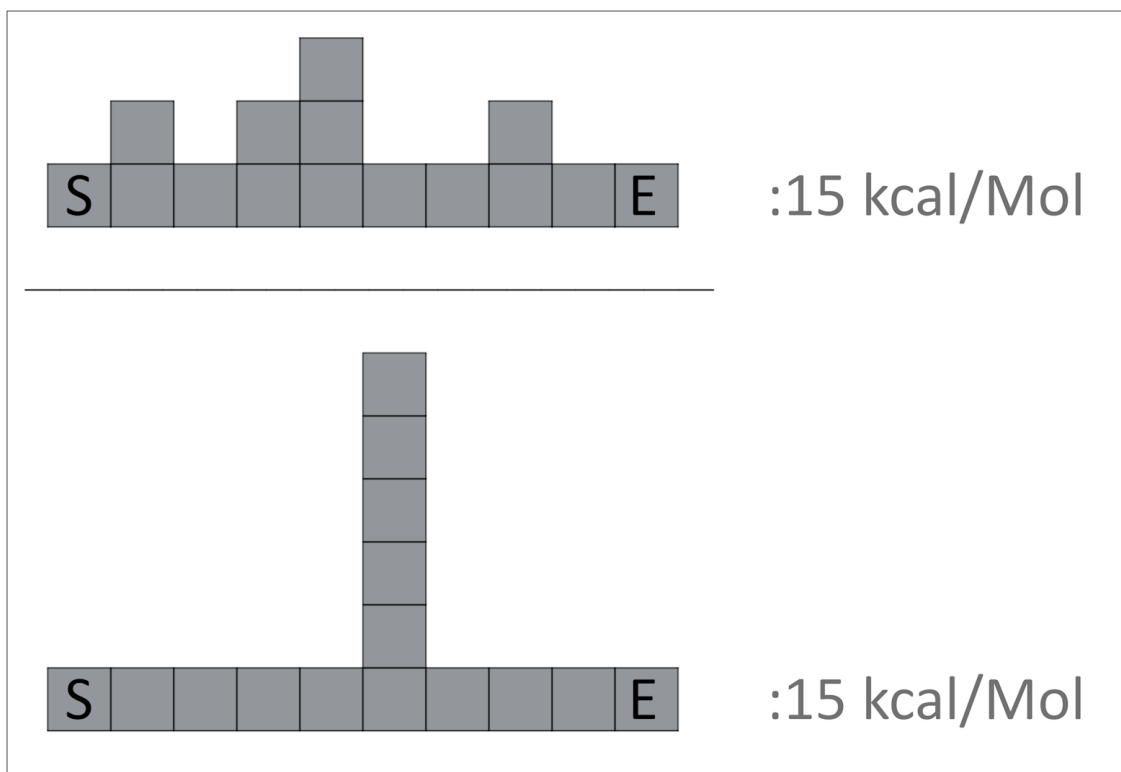

Figure 3: Schematic to demonstrate the need for a proper weighting function when comparing possible paths in the energy landscape. Two unique 1D paths are shown here (top and bottom) with unique start-point (S) and end-point (E), while covering a total of 10 coordinates each of varying energies. Each path has the same net integrated energy (15 units, with 1 unit per block in this simplification). Even though these two paths are energetically degenerate, transition state theory would claim that the bottom path is much less favorable due to its singular energy-spike (representing the path's 'activation energy') - having a much lower probability of being surpassed than any of the individual (smaller) peaks in the top path. POLARIS' 'Transition State Weighting' function incorporates both of these two properties (net integrated energy and relative height of energy-peaks) when assessing and comparing all possible paths.
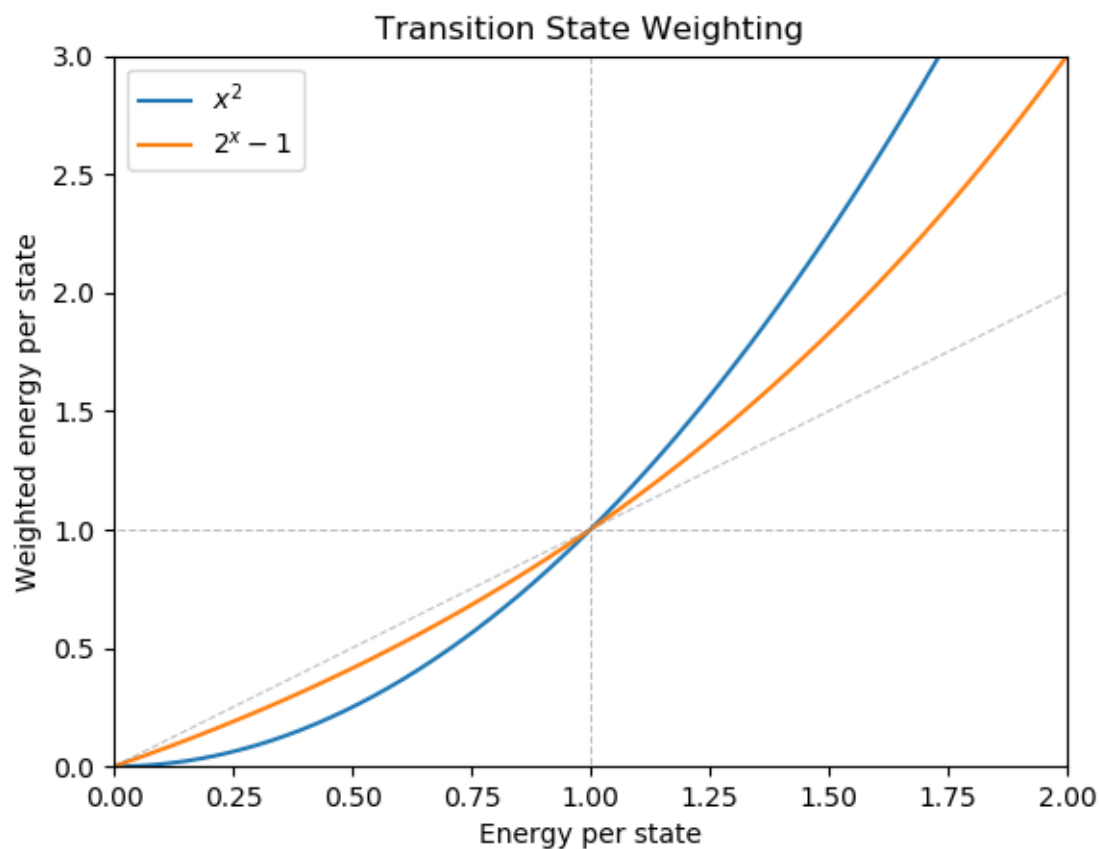
Figure 4: Two possible weighting functions are plotted for comparing pathways via the 'Transition State Weighting' option. Each function aims to make higher energy coordinates increasingly less favorable, while weighting the regime of lower energy coordinates (in the region leading up to 1 kcal/mol) more favorably. The outcomes of these weights are demonstrated in Figure S7.
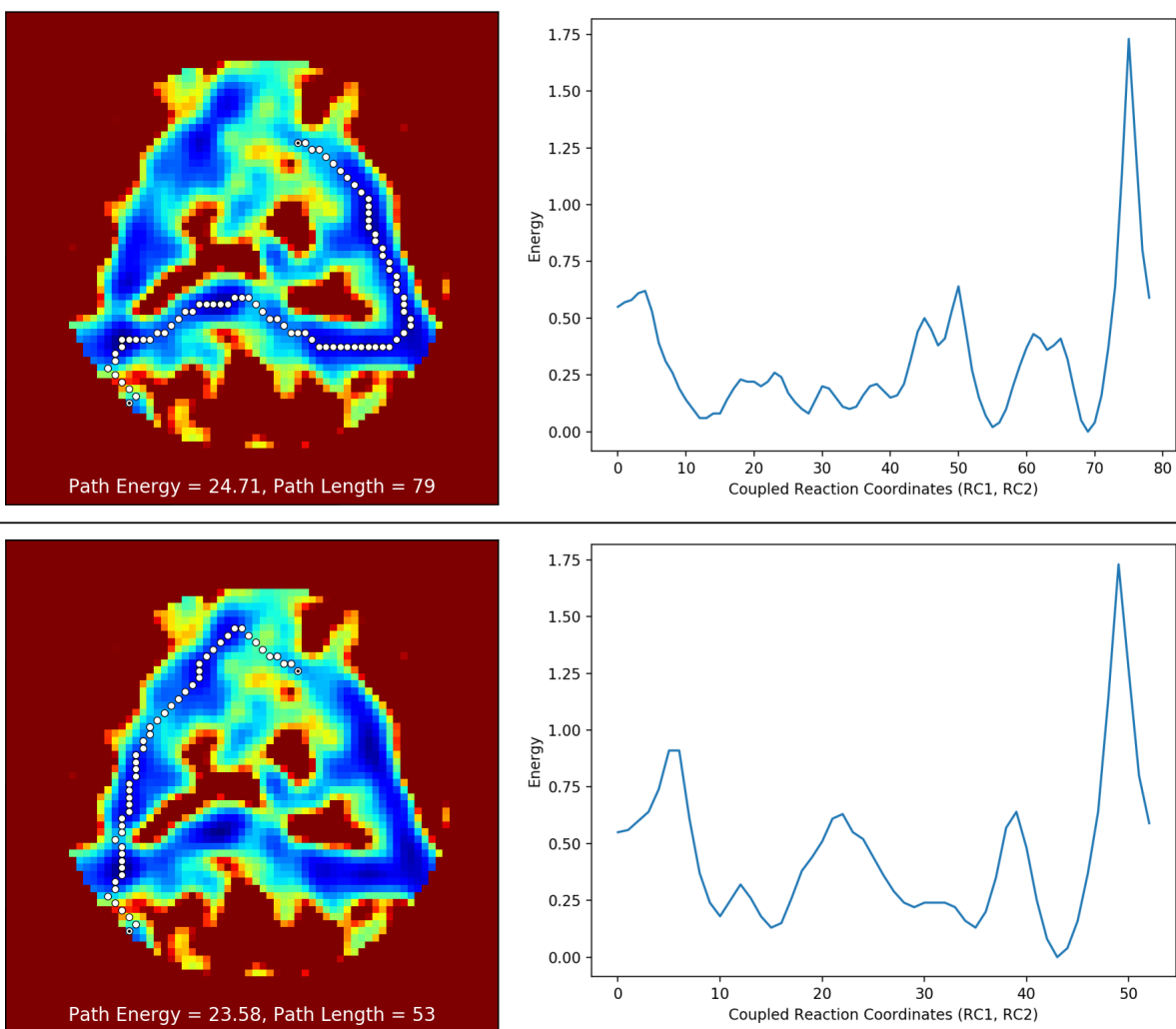
Figure 5: Further examination of the two weighting functions shown in Figure S4, with the least-energy path output discovered by weighting with x^2 shown on top and $2^x - 1$ on bottom. Although these two weighting functions are almost identical, each discovered drastically different routes through the energy landscape. Even so, the top path only has 1.13 kcal/mol more total integrated energy than the bottom path, while also including 26 more transit points. Such a drastic global difference emerging under such subtle underlying changes demonstrates the delicate nature of these landscapes. While the latter function has been found to give better results (and is enabled when the 'Transition State Weighting' option is chosen), these differences highlight the importance of setting transit locations in the user interface to best guide the path along regions of highest interest.
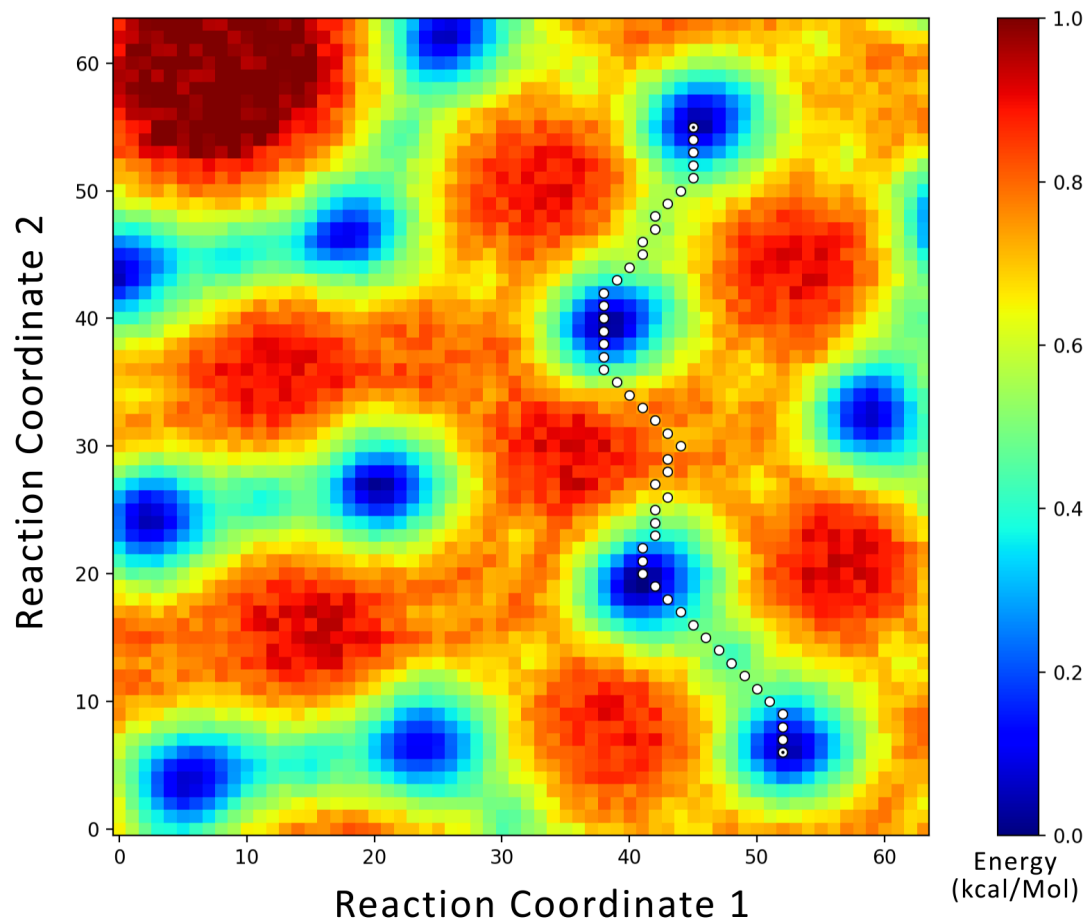
Figure 6: Example output of the least action path for an exemplary computationally-generated landscape.
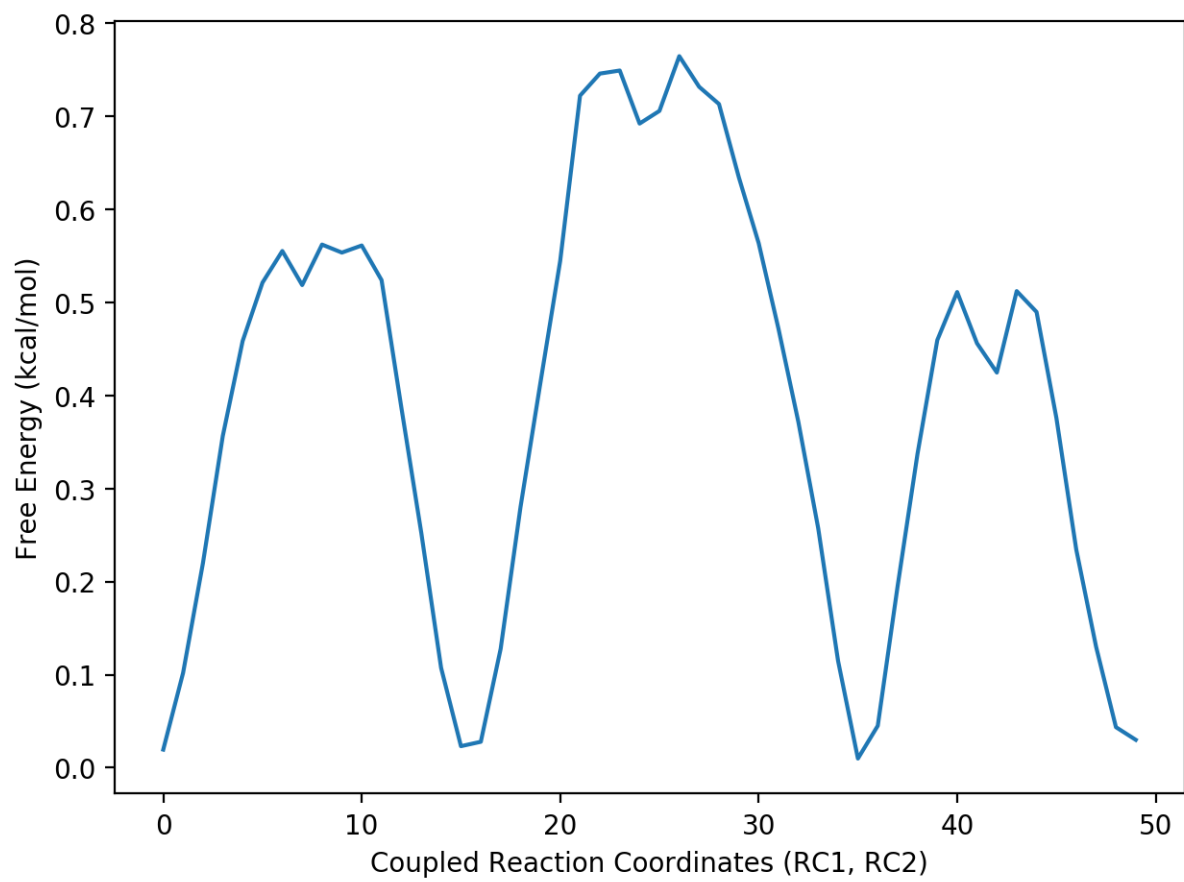
Figure 7: Example transition state diagram for the path seen in Figure S6, with energies plotted against the set of coupled coordinates (RC1, RC2).