

RESEARCH ARTICLE

Protein Folding as a Complex Reaction: A Two-Component Potential for the Driving Force of Folding and Its Variation with Folding Scenario

Sergei F. Chekmarev*

Institute of Thermophysics, 630090 Novosibirsk, Russia and Department of Physics, Novosibirsk State University, 630090 Novosibirsk, Russia

* chekmarev@itp.nsc.ru



OPEN ACCESS

Citation: Chekmarev SF (2015) Protein Folding as a Complex Reaction: A Two-Component Potential for the Driving Force of Folding and Its Variation with Folding Scenario. PLoS ONE 10(4): e0121640. doi:10.1371/journal.pone.0121640

Academic Editor: Emanuele Paci, University of Leeds, UNITED KINGDOM

Received: November 25, 2014

Accepted: February 11, 2015

Published: April 7, 2015

Copyright: © 2015 Sergei F. Chekmarev. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: This work was supported by the Russian Science Foundation, grant No. 14-14-00325; <http://www.rscf.ru/>. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The author has declared that no competing interests exist.

Abstract

The Helmholtz decomposition of the vector field of probability fluxes in a two-dimensional space of collective variables makes it possible to introduce a potential for the driving force of protein folding [Chekmarev, J. Chem. Phys. 139 (2013) 145103]. The potential has two components: one component (Φ) is responsible for the source and sink of the folding flow, which represent, respectively, the unfolded and native state of the protein, and the other (Ψ) accounts for the flow vorticity inherently generated at the periphery of the flow field and provides the canalization of the flow between the source and sink. Both components obey Poisson's equations with the corresponding source/sink terms. In the present paper, we consider how the shape of the potential changes depending on the scenario of protein folding. To mimic protein folding dynamics projected onto a two-dimensional space of collective variables, the two-dimensional Müller and Brown potential is employed. Three characteristic scenarios are considered: a single pathway from the unfolded to the native state without intermediates, two parallel pathways without intermediates, and a single pathway with an off-pathway intermediate. To determine the probability fluxes, the hydrodynamic description of the folding reaction is used, in which the first-passage folding is viewed as a steady flow of the representative points of the protein from the unfolded to the native state. We show that despite the possible complexity of the folding process, the Φ -component is simple and universal in shape. The Ψ -component is more complex and reveals characteristic features of the process of folding. The present approach is potentially applicable to other complex reactions, for which the transition from the reactant to the product can be described in a space of two (collective) variables.

Introduction

Protein folding is a significant example of complex reactions, due to entropy effects and possible presence of intermediates and multiple pathways. According to the statistical viewpoint developed in the last decades of the previous century, the folding process is governed by the interplay of the protein potential energy and conformation entropy (see [1–6] for a review). The energy surface projected onto a low dimensional space of variables is funnel-like, because the effective dimension of the conformation space decreases as the protein approaches the native state, which can be explained by a hierarchical organization of the protein structure [7, 8]. The energy and entropy of the protein both decrease from the unfolded states of the protein to its native state but in different manner, so that a free energy barrier is formed that separates the native-like and unfolded states. Depending on the protein type and the conditions under which the reaction takes place, the process of folding can be complicated by the presence of on- and off-pathway intermediates and different pathways. The simplest and widely used approach to take all these factors into account, i.e., the funnel character of the energy landscape, free energy barrier, intermediates and pathways, is to construct a free energy surface (FES) depending on two collective variables [4, 9–11]. One variable is usually chosen to describe how the protein compacts as the contacts between the residues are formed (e.g., the radius of gyration), and the other how the protein conformation is close to the native state (e.g., the fraction of the native contacts). A shortcoming of the FES thus introduced is that it determines only the probabilities for the protein to be in different conformational states but does not account for the order in which these states follow each other, i.e., the local direction in which the protein proceeds. To gain a closer insight into folding dynamics, we have recently introduced a “hydrodynamic” description of the folding process [12]. In this approach, similar to the FESs, the folding process is considered in a reduced space of collective variables, but the calculated folding trajectories are used to determine local flows of transitions (the probability fluxes) in this space. Given the flows, it is possible to construct the vector field and streamlines of the folding flows, similar to how it is done in hydrodynamics [13]. If the process of folding is assumed to mimic physiological conditions, i.e., that the native state is stable and unfolding events are improbable, the folding trajectories are chosen to start at unfolded states of the protein and are terminated upon reaching the native state (“first-passage folding” [14]). In this case, the process of protein folding can be viewed as a steady flow of a folding “fluid” from the unfolded states of the protein to its native state, with the density of the fluid being proportional to the probability for the system to be at the current point of the conformation space.

The hydrodynamic approach has been successfully applied to study folding dynamics of several model proteins—an α -helical hairpin (a lattice model) [12], a SH3 domain [15, 16] and a β -hairpin protein [17] (a C_α -model), and a three-stranded antiparallel β -sheet protein (all-atom simulations) [14, 18]. The primary conclusion drawn from these studies is that the local flow do not generally follow the FESs; even local vortices of the flows can be formed due to partial folding and unfolding of the protein that do not leave clear fingerprints on the FESs [12, 15]. It follows that the FESs do not completely govern the folding dynamics and thus can not be considered to be “true” potentials that determine the driving force of protein folding.

Recently, using a β -hairpin protein [19] as an example, we have shown the Helmholtz decomposition [20] of the vector field of the folding fluxes depending on two collective variables makes it possible to determine the corresponding true potential [17]. For this, the field of the probability fluxes is represented as a sum of divergence-free and curl-free fields. Accordingly, the potential has two components: one component (Φ), which corresponds to the curl-free field, is responsible for the source and sink of the folding flow that represent, respectively, the unfolded states and the native state of the protein, and the other (Ψ), which corresponds to the

divergence-free field, accounts for the flow vorticity generated at the periphery of the flow field and provides the canalization of the flow between the source and sink within the reaction channel.

As such, both the hydrodynamic description of the folding dynamics and the decomposition of the vector field of probability fluxes into divergence-free and curl-free fields involve nothing that is specific of protein folding. Therefore the present approach is potentially applicable to other complex reactions that can be described by two (collective) variables, e.g., unimolecular chemical reactions [21] and conformational transitions in nanoclusters [22]. One recent example of a related approach is the calculation of a reaction path in the framework of the transition-path theory (E and Vanden-Eijnden [23]), in which the flow lines of the probability current of the reactive trajectories were introduced as the lines orthogonal to the isocommitor probability lines. The resulting pictures of the lines obtained in [23] for the model Müller and Brown potential [24] and in our work for folding of a β -hairpin protein [17] are very similar: the Ψ -isolines can be likened to the flow lines, and Φ -isolines to the isocommitor probability lines, although the former, in contrast to the latter, are not mutually orthogonal [17].

The β -hairpin protein we studied in [17] had a very simple folding scenario: a single folding pathway, no intermediates and, correspondingly, two-state kinetics. It is therefore of interest to see how the potential for the driving force changes when the folding scenario becomes more complex, i.e., involves intermediates and additional pathways. A natural approach to this problem would be to consider proteins that have the corresponding complicated scenarios and compare the results to those for the β -hairpin protein [17]. However, such an approach does not seem to be optimal. First, it is difficult to choose the proteins that would have the desired folding scenarios without secondary effects complicating the comparison, e.g., a protein that had parallel folding pathways and no intermediates. Secondly, the chosen proteins should allow the simulations at a reasonable computational cost, because at least the order of thousand of folding trajectories is required to have statistically significant probability fluxes [17]. This restricts the class of possible protein models to simplified, coarse-grained models, such as a C_α -model used in [17]. At the same time, to describe, e.g., off-pathway intermediates, an all-atom approach is generally required, because such intermediates are typically associated with either disulfide bridges (BPTI [25]), or salt-induced states (S6 [26]), or cis prolines (CheY [27]).

Another possible approach is to use model potential energy surfaces that make it possible to generate molecular dynamics trajectories mimicking the protein folding trajectories projected onto a two-dimensional space of collective variables. These trajectories should allow both the calculation of the probability fluxes, which are necessary to determine the potential for the driving force, and the characterization of the process in terms accepted in the protein folding studies, in order to relate the potential to a specific folding scenario. For this purpose, we, similar to [23], employ the two-dimensional Müller and Brown potential [24], which is widely used to test numerical methods for calculation of reaction paths in many-body systems, e.g., [28, 29]. In addition to a simple system (one “folding” pathway without intermediates), we will consider two more complex systems: one has two folding pathways, and the other has an off-pathway intermediate. We will show that in all cases the Φ -component of the potential is simple and universal in shape, while the Ψ -component is more complex and reveals characteristic features of the folding process.

The paper is organized as follows. The next section describes methods: the generation of the PESs and their relation to protein folding, the simulation method, calculating the FESs, a short description of the hydrodynamic approach, and determining the two-component potential with the Helmholtz decomposition of the vector flow field. The subsequent section contains the obtained results and their discussion: for the basic system, a system with two folding

Table 1. Two-well landscape with a single pathway: Parameters of the generating function.

Component	x_0	y_0	A	a	b	c
1	0.34	0.56	-70.0	-30.0	-22.0	-12.0
2	0.75	0.2	-80.0	-40.0	-0.0	-80.0

doi:10.1371/journal.pone.0121640.t001

Table 2. Two-well landscape with two pathways: Parameters of the generating function.

Component	x_0	y_0	A	a	b	c
1	0.35	0.25	-25.0	-30.0	-20.0	-10.0
2	0.65	0.2	-50.0	-50.0	0.0	-100.0
3	0.75	0.45	-25.0	-30.0	-25.0	-10.0
4	0.40	0.65	-37.5	-10.0	0.0	-10.0

doi:10.1371/journal.pone.0121640.t002

pathways, and a system with an off-pathway intermediate. The last section summarizes the results of the work and gives concluding remarks.

Methods

Model potential energy surfaces and their relation to protein folding

The Müller and Brown function [24], which we use to determine a two-dimensional (2D) potential energy surface (PES), is written as

$$U(x, y) = \sum_i A_i \exp [a_i(x - x_{0i})^2 + b_i(x - x_{0i})(y - y_{0i}) + c_i(y - y_{0i})^2] \quad (1)$$

where x_{0i} and y_{0i} are the coordinates of the center of i potential well, a_i , b_i and c_i characterize the steepness of the well sides, and A_i determines the depth of the well. The summation in this equation is taken from 2 to 4 depending on the specific form of the PES (Tables 1–3 below). Each point in the (x, y) space in Equation (1) is assumed to represent a conformation of some (imaginary) protein in a 2D space of collective variables; the coordinate x is to characterize the proximity of the protein conformation to the native state, and the coordinate y —the compactness of the protein. The deepest well on the PES is to account for the native-like states, and the shallowest well—for the unfolded states, which include both the completely unfolded (extended) and partly unfolded (semi-compact) conformations. Having this in mind, we will refer to the states of the present model system in the (x, y) space to as protein states, omitting quotation marks.

Two-dimensional PESs are evidently too simple to reproduce the protein folding dynamics in multidimensional conformation space. A principal shortcoming of such surfaces is that the

Table 3. Three-well landscape with an off-pathway intermediate: Parameters of the generating function.

Component	x_0	y_0	A	a	b	c
1	0.34	0.56	-35.0	-30.0	-22.0	-9.0
2	0.75	0.15	-50.0	-40.0	-0.0	-80.0
3	0.75	0.65	-37.0	-25.0	0.0	-80.0

doi:10.1371/journal.pone.0121640.t003

contribution of conformational entropy is unreasonably reduced. In protein folding, when the reaction is projected onto a two-dimensional space of variables, the interplay between the potential energy and entropy leads to a barrier separating the unfolded and native-like states. On the barrier, the energy directs the protein to the native state, because its energy is minimal, but entropy pulls it back, because the number of the unfolded states is much higher than the number of native-like states (in some cases, the entropy of the native state can contribute to its stability [30], however, the overall picture of the process remains unchanged). In a 2D space, the appropriate difference in the number of the unfolded and native-like states is hardly possible to reproduce. Therefore, to mimic folding dynamics, a 2D PES should have a barrier that separates the unfolded and native-like states, although such (energy) barriers are not characteristic of protein folding. In other words, the PES should consist of, at least, two wells, one for the unfolded states and the other for the native-like states. However, if we are interested in the resulting MD trajectories rather than in the mechanism of protein folding, as in the present study, it is, in principle, insignificant what is the nature of the barrier that the system has to overcome—entropic or energetic. What is important here is that the obtained MD trajectories should allow consistent reproduction of the characteristic properties of the folding reaction in a 2D space of collective variables that are essential for the goal of the present study; they are the vector field of the local flows of conformational transitions (probability fluxes)—to calculate the potential for the driving force, and the distribution of the first-passage times and the FES landscape—to relate the results obtained with the given 2D PES to a specific folding scenario.

Simulation method and protocols

The simulations were performed using the constant-temperature MD based on the Langevin equation

$$\frac{d^2\mathbf{r}}{dt^2} + \gamma \frac{d\mathbf{r}}{dt} = -\frac{\partial U}{\partial \mathbf{r}} + \Phi(t) \quad (2)$$

where \mathbf{r} is the radius-vector of the (x, y) point, U is the potential energy of the system given by Equation (1), Φ is a random force generated by the “surroundings” at temperature T , γ is the friction coefficient that introduces viscosity of the surroundings to balance the random force and dissipation, and the system’s mass is set to unity. The random forces have the Gaussian distribution with zero mean and variance $\langle \Phi_i(t)\Phi_j(t + \tau) \rangle = 2\gamma T \delta_{ij} \delta(\tau)$, where the angular brackets denote an ensemble average, the index at Φ stands for the vector component, and δ_{ij} and $\delta(\tau)$ are the Kronecker and Dirac delta functions, respectively. The temperature is measured in the energy units (the Boltzmann constant $k_B = 1$). The equation was numerically integrated using the algorithm of Biswas and Hamann [31] with the time step $\Delta t = 0.001\tau$ and $\gamma = 3/\tau$ (τ is the time scale equal to 1). In each case, 1×10^4 folding trajectories were simulated at a temperature at which the folding kinetics were as desired (two-states or three states).

Free energy surface

To construct the FES, the probability for the system to be at a current point of the (x, y) space, $p(x, y)$, was calculated and then, using the Boltzmann hypothesis, was converted into the free energy

$$G(x, y) = -T \ln p(x, y) \quad (3)$$

The temperature is measured in the Boltzmann constant units.

Probability fluxes

The probability fluxes, determining the local rates of transitions between the points of the (x, y) space, were calculated according to the hydrodynamic description of protein folding [12]. The x -component of the flux $\mathbf{j}(\mathbf{r})$ was calculated as

$$j_x(\mathbf{r}) = \left[\sum_{\mathbf{r}', \mathbf{r}'' (\mathbf{r} \subset \mathbf{r}^*)}^{x''-x'>0} n(\mathbf{r}'', \mathbf{r}') - \sum_{\mathbf{r}', \mathbf{r}'' (\mathbf{r} \subset \mathbf{r}^*)}^{x''-x'<0} n(\mathbf{r}'', \mathbf{r}') \right] / (M \bar{t}_f \Delta y) \tag{4}$$

where M is the total number of simulated trajectories, \bar{t}_f is the mean first-passage time (MFPT), $n(\mathbf{r}'', \mathbf{r}')$ is the number of transitions from state \mathbf{r}' to \mathbf{r}'' , and $\mathbf{r} \subset \mathbf{r}^*$ is a symbolic designation of the condition that the transitions included in the sum have the straight line connecting points \mathbf{r}' to \mathbf{r}'' , which crosses the line $x = \text{const}$ within the segment of the length of Δy centered at the point \mathbf{r} . The y -component of $\mathbf{j}(\mathbf{r})$ is determined in a similar way, except that one selects the transitions crossing the line $y = \text{const}$ within the current segment Δx . The calculations were performed on a grid with discretization $\Delta x = \Delta y = 0.01$.

Potential for the driving force

To determine the potential for the driving force, the Helmholtz decomposition theorem [20] was used, similar to [17]. According to this theorem, any smooth vector field $\mathbf{j}(\mathbf{r})$ can be uniquely represented as

$$\mathbf{j} = \mathbf{j}_{cf} + \mathbf{j}_{df} \tag{5}$$

where \mathbf{j}_{cf} is the curl-free field, and \mathbf{j}_{df} is the divergence-free field, i.e., $\nabla \times \mathbf{j}_{cf} = 0$ and $\nabla \cdot \mathbf{j}_{df} = 0$. Due to the conditions the vectors \mathbf{j}_{cf} and \mathbf{j}_{df} satisfy, they can be written as

$$\mathbf{j}_{cf} = -\frac{\partial \Phi}{\partial x} \mathbf{k}_x - \frac{\partial \Phi}{\partial y} \mathbf{k}_y \tag{6}$$

$$\mathbf{j}_{df} = \frac{\partial \Psi}{\partial y} \mathbf{k}_x - \frac{\partial \Psi}{\partial x} \mathbf{k}_y \tag{7}$$

where $\Phi = \Phi(\mathbf{r})$ is the potential for the curl-free component, $\Psi = \Psi(\mathbf{r})$ is the potential for the divergence-free component, and \mathbf{k}_x and \mathbf{k}_y are the unit vectors for x and y , respectively. Substituting the expressions for \mathbf{j}_{cf} and \mathbf{j}_{df} from Equations (6) and (7) into Equation (5) and regrouping the terms, we have

$$\mathbf{j} = j_x \mathbf{k}_x + j_y \mathbf{k}_y$$

where

$$j_x = -\frac{\partial \Phi}{\partial x} + \frac{\partial \Psi}{\partial y}, \quad j_y = -\frac{\partial \Phi}{\partial y} - \frac{\partial \Psi}{\partial x} \tag{8}$$

The function $\Phi(\mathbf{r})$ and $\Psi(\mathbf{r})$ obey the corresponding Poisson's equations. Taking into account Equation (6) and the condition $\nabla \cdot \mathbf{j}_{df} = 0$, the dot product of the gradient vector and the

probability flux \mathbf{j} determined by Equation (5) gives

$$\Delta\Phi(\mathbf{r}) + q(\mathbf{r}) = 0 \tag{9}$$

where $q(\mathbf{r})$ is the distribution of sources and sinks that is calculated from the simulated probability fluxes $\mathbf{j}(\mathbf{r})$ as

$$q(\mathbf{r}) = \partial j_x / \partial x + \partial j_y / \partial y \tag{10}$$

The function $\Phi(\mathbf{r})$ should satisfy Equation (9) within a region of the \mathbf{r} space in which $q(\mathbf{r}) \neq 0$ and have $\Phi(\mathbf{r}) = 0$ at its boundary (the Dirichlet problem; see, e.g., Sobolev [32]). Similarly, the cross product of the gradient vector and \mathbf{j} , with taking into account Equation (7) and the condition $\nabla \times \mathbf{j}_{cf} = 0$, gives

$$\Delta\Psi(\mathbf{r}) + \omega(\mathbf{r}) = 0 \tag{11}$$

where

$$\omega(\mathbf{r}) = \partial j_y / \partial x - \partial j_x / \partial y \tag{12}$$

is the simulated distribution of vorticity. As previously, the corresponding boundary condition should be satisfied, i.e., it should be $\Psi(\mathbf{r}) = 0$ at the boundary of the region where $\Psi(\mathbf{r})$ is determined.

To find the $\Phi(\mathbf{r})$ and $\Psi(\mathbf{r})$, it is convenient to replace the Dirichlet problem with the initial-boundary-value problem (for the other methods of numerical solution of the former, see, e.g., Dorr [33]). In this approach, Equations (9) and (11) are replaced with

$$\partial\Phi(\mathbf{r}, t) / \partial t = \Delta\Phi(\mathbf{r}, t) + q(\mathbf{r}) \tag{13}$$

and

$$\partial\Psi(\mathbf{r}, t) / \partial t = \Delta\Psi(\mathbf{r}, t) + \omega(\mathbf{r}) \tag{14}$$

where t plays a role of time. In each case, starting with some initial distribution, the process is continued until a stationary distribution is achieved, which approximates the solution of the corresponding the Dirichlet problem. In the present paper, we integrated Equations (13) and (14) numerically, using a finite-difference scheme of the first-order approximation for the time derivative and the second-order approximation for the space derivatives. To avoid boundary effects, the region of integration, which was the same for $\Phi(\mathbf{r}, t)$ and $\Psi(\mathbf{r}, t)$, was enlarged on each side by 0.2 in comparison with the region where the simulations were performed. The level of discretization of the \mathbf{r} space was as in Equation (4). At the boundaries of this, enlarged region there was $\Phi(\mathbf{r}, t) = 0$ and $\Psi(\mathbf{r}, t) = 0$ for Equations (13) and (14), respectively, and the initial distributions were taken to be $\Phi(\mathbf{r}, 0) = 0$ and $\Psi(\mathbf{r}, 0) = 0$. The stationary solutions of Equations (13) and (14) were considered to be achieved if the conditions $\{\int [\Delta\Phi(\mathbf{r}, t) / q(\mathbf{r}) + 1]^2 dx dy\}^{1/2} \leq 1 \times 10^{-10}$ and $\{\int [\Delta\Psi(\mathbf{r}, t) / \omega(\mathbf{r}) + 1]^2 dx dy\}^{1/2} \leq 1 \times 10^{-10}$, respectively, were satisfied for the last one third of the total integration time. We note that Equation (8) allows determining the functions $\Phi(\mathbf{r})$ and $\Psi(\mathbf{r})$ directly from the simulated probability fluxes $\mathbf{j}(\mathbf{r})$. For this purpose, the least-squares functional $Q = \int [(j_x + \partial\Phi/\partial x - \partial\Psi/\partial y)^2 + (j_y + \partial\Phi/\partial y + \partial\Psi/\partial x)^2] dx dy$, where j_x and j_y are the simulated probability fluxes, can be minimized with respect to $\Phi(\mathbf{r})$ and $\Psi(\mathbf{r})$ [17]. Both these approaches lead to essentially the same results.

According to Equations (6) and (7), the probability fluxes are proportional to the first-order space derivatives of the potential functions $\Phi(\mathbf{r})$ and $\Psi(\mathbf{r})$, i.e., the fluxes are considered to be drift fluxes. In other words, although the inertia term was present in the Langevin Equation (2), the functions $\Phi(\mathbf{r})$ and $\Psi(\mathbf{r})$ determined from the resulting simulated fluxes $\mathbf{j}(\mathbf{r})$ via

Equations (5)–(7) account for an overdamped motion. The corresponding kinetic equation is the Smoluchowski equation $\partial p/\partial t + \nabla \cdot \mathbf{J} = 0$, where $p(\mathbf{r}, t)$ is the probability density, and \mathbf{J} is the probability current, which can be written as $\mathbf{J} = -D(\mathbf{r})\nabla p(\mathbf{r}, t) + [D(\mathbf{r})/T]\mathbf{F}(\mathbf{r})p(\mathbf{r}, t)$, where $-D(\mathbf{r})\nabla p(\mathbf{r}, t)$ is the diffusion flux, $[D(\mathbf{r})/T]\mathbf{F}(\mathbf{r})p(\mathbf{r}, t)$ is the drift flux, $D(\mathbf{r})$ is the diffusion coefficient, and $\mathbf{F}(\mathbf{r})$ is the driving force (the Boltzmann constant is equal to unity). The zero probability current $\mathbf{J} = 0$ assumes detailed balance, and, correspondingly, a curl-free flow. In this case, the equality $\mathbf{J} = 0$ leads to the stationary equilibrium (Boltzmann) distribution $p(\mathbf{r}) \sim \exp[-G(\mathbf{r})/T]$, where $G(\mathbf{r})$ is the free energy that exerts the driving force $F(\mathbf{r}) = -\nabla G(\mathbf{r})$. However, if the probability current is nonzero, i.e., as in the present case, when the steady flow from the unfolded states to the native state exists, the stationary solution is determined by the condition $\nabla \cdot \mathbf{J} = 0$, so that \mathbf{J} can have a curl component (van Kampen [34]). Such flow is non-equilibrium and is characterized by “irreversible circulation” or “cyclic balance”, which can be considered as a measure of deviation from detailed balance [35–37]. In the present case, the circulating flow is represented by the flux vector \mathbf{j}_{df} and, according to Equation (7), is generated by the Ψ -component of the potential.

We note that the functions $\Phi(\mathbf{r})$ and $\Psi(\mathbf{r})$ in Equations (6) and (7) differ from the conventional potentials in that the driving forces determined by these functions are density weighted. Conventionally, as, e.g., in the Smoluchowski equation above, the probability flux is determined as $\mathbf{j}(\mathbf{r}) \sim \mathbf{F}(\mathbf{r})p(\mathbf{r})$, so that the driving force $\mathbf{F}(\mathbf{r})$, which is equated with the negative gradient of the potential, is proportional to the velocity of motion $\mathbf{v}(\mathbf{r}) = \mathbf{j}(\mathbf{r})/p(\mathbf{r})$. In contrast, Equations (6) and (7) assume that the gradients of the potentials are immediately related to the probability fluxes. The reason why the conventional definition is unsuitable in the present case is that the probability density $p(\mathbf{r})$ varies across the flow field very strongly, i.e., the folding fluid is highly “compressible”. Because of this, as it follows from the equality $\nabla \cdot \mathbf{v}(\mathbf{r}) = \nabla \cdot \mathbf{j}(\mathbf{r})/p(\mathbf{r}) + \mathbf{j}(\mathbf{r}) \cdot \nabla[1/p(\mathbf{r})]$, the velocity field is not divergence-free in a region which is free of flow sources and sinks, in contrast to the field of the probability fluxes, for which $\nabla \cdot \mathbf{j}(\mathbf{r}) = 0$. The last term in the right-hand side of the above equality presents the sources/sinks that arise due to the compressibility of the folding fluid. The appearance of such artificial sources and sinks destroys the natural separation of the folding flow into the divergence- and curl-free components and complicates the interpretation of the $\Phi(\mathbf{r})$ and $\Psi(\mathbf{r})$ functions (see below). Therefore, we determine the potential through the fluxes, according to Equations (6) and (7), but consider the driving force to be density weighted, i.e., $\mathbf{F}(\mathbf{r}) = p(\mathbf{r})\hat{\mathbf{F}}(\mathbf{r})$, where $\hat{\mathbf{F}}(\mathbf{r})$ is the actual driving force.

Results and Discussion

Two-well landscape with a single pathway

This case corresponds to the simplest, two-state kinetics of protein folding, which are characteristic of small proteins [38]. Table 1 gives the values of the parameters determining the potential energy function, Equation (1), and Fig. 1a shows the corresponding PES. The simulations were performed at $T = 3.0$. The native state was associated with the point $x_n = 0.75, y_n = 0.2$. The MD trajectories were initiated in the vicinity of the point $x_u = 0.25, y_u = 0.78$ with a uniform random scattering of the points within $|\Delta_u x| = |\Delta_u y| = 0.1$; these points were intended to mimic the completely unfolded (extended) protein states. The native state was considered to be reached if the deviation from the native state was not larger than $|\Delta_n x| = |\Delta_n y| = 0.01$. The same conditions to choose the initial points (the scattering of the points) and to terminate the trajectories in the native state were used for the other systems.

Fig. 1b shows the FES. It reveals two basins of attraction that are connected by a free energy valley, which plays a role of the reaction channel. One basin, which is located at lower values of

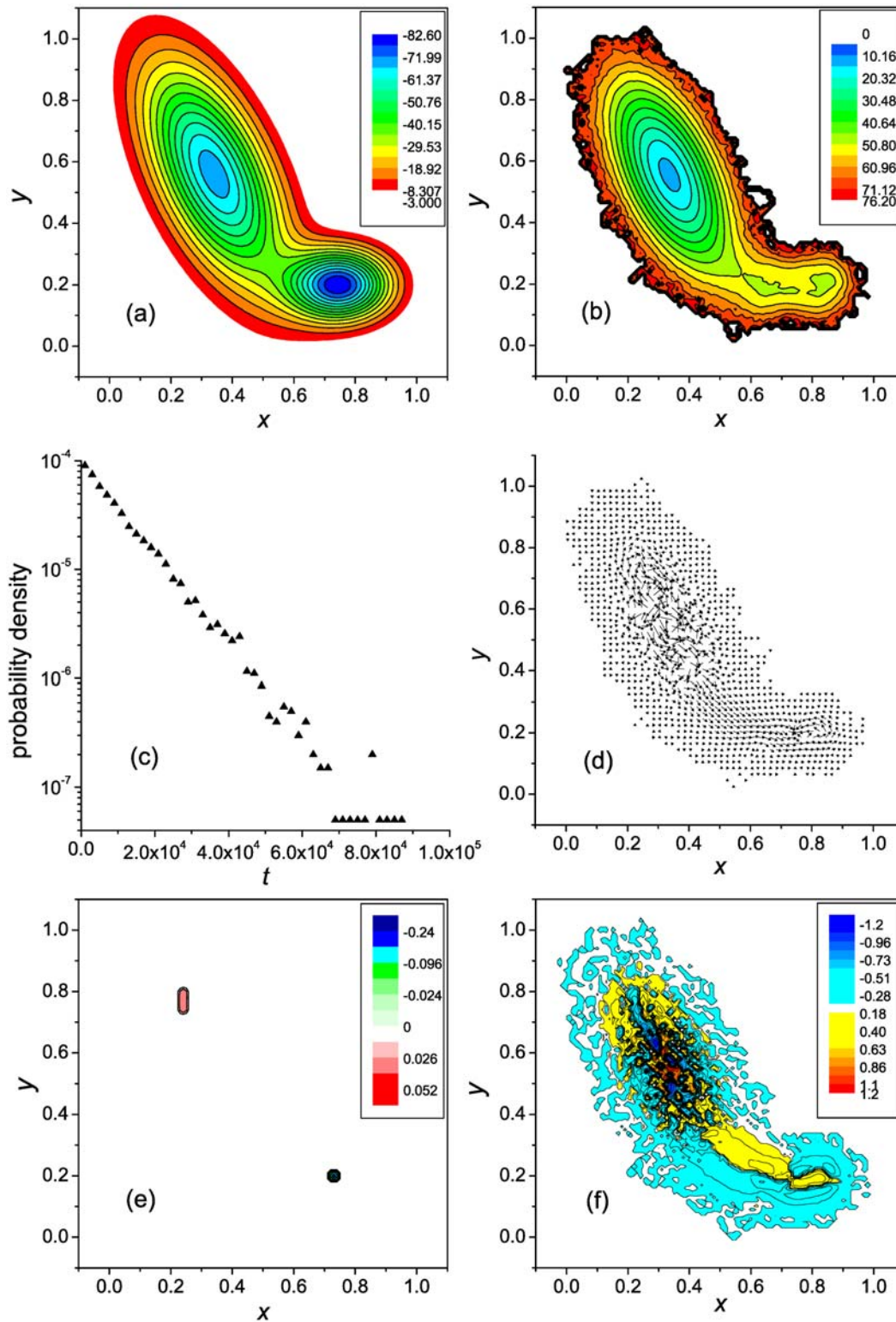


Fig 1. Two-well landscape with a single pathway: General characterization. (a) the potential energy surface, (b) the free energy surface, (c) the first-passage time distribution, (d) the vector flow field, (e) the distribution of the divergence, and (f) the vorticity distribution.

doi:10.1371/journal.pone.0121640.g001

x , can be associated with unfolded conformations. For the most part, these conformations should be considered as partly unfolded (semi-compact), because extended conformations rapidly transform into semi-compact ones [11]. The other basin corresponds to native-like conformations. Since the trajectories were terminated upon reaching the native state, this basin is not so well formed as it would be under the equilibrium folding conditions, when the protein resides in the native basin until it unfolds (cf., e.g., the FESs for the first-passage and equilibrium folding of the three-stranded β -sheet protein [14]). At the same time, the free energy barrier between the unfolded and native-like states is rather well formed ($x \approx 0.57$).

The distribution of the first passage times is depicted in Fig. 1c. The distribution is close to single-exponential, evidencing that the kinetics are two-state. The MFPT \bar{t}_f is equal to $\approx 1 \times 10^4$.

Fig. 1d shows the vector field of the probability fluxes. For the illustration purpose, the lengths of the vectors are rescaled as $\mathbf{j}_{\text{sc1}} = q(\mathbf{j}/j)j^{1/4}$, where \mathbf{j} is the original vector determined by Equation (4), j is its length, and q is the factor equal to 0.125. As is seen, the complex motion of the system that exists in the basin associated with unfolded states changes to a regular motion in the transition region between this and native basins, i.e., in the region which contains the free energy barrier. In this respect, the present vector flow field is similar to what was observed for an α -helical hairpin, which also had a single folding pathway and two-state kinetics [12]. A vortex observed in the native basin is mainly due to the strong condition we used to terminate the trajectories ($|\Delta_n x| = |\Delta_n y| = 0.01$). Because of it, the system had to explore a large portion of the native-like states until the native state was found with the required accuracy. For actual protein folding, when the multidimensional conformation space is projected on a two-dimensional space of collective variables, this effect is not present [12, 17]. If the values of $\Delta_n x$ and $\Delta_n y$ were taken to be comparable to the size of the native basin ($|\Delta_n x| = |\Delta_n y| \sim 0.1$), the vortex was not formed. Instead, a region around the native state was created, into which the trajectories did not penetrate, i.e., the trajectories ended at the boundary of this region. We also note that since every trajectory initiated in an unfolded state reached the native state, and was terminated in it, the total flow from the unfolded states to the native state $G(x) = \int j_x(x, y) dy$ is constant; it is equal to the reciprocal MFPT $1/\bar{t}_f$ [see Equation (4)].

Fig. 1e and 1f show, respectively, the distributions of the flow divergence $q(\mathbf{r})$ and its vorticity $\omega(\mathbf{r})$, which are determined by Equations (10) and (12). Similar to the probability fluxes (Fig. 1d), for the illustration purpose, the values of $\omega(\mathbf{r})$ are rescaled as $\omega_{\text{sc1}} = (\omega/|\omega|)|\omega|^{1/4}$. Fig. 1e shows that except for the local regions where the trajectories were initiated and terminated, i.e., the source and sink regions, the flow is divergence-free. The vorticity, in contrast, is nonzero across the entire flow field (Fig. 1f). First of all, it is seen that the motion of the system in the basin for the unfolded states is not completely disordered—a set of the vortices of opposite sign exists there (the negative vorticity corresponds to the clock-wise motion, and the positive vorticity to the counterclockwise motion). Secondly, the vorticity has different signs on different sides of the reaction channel, because the intensity of the flow decreases toward both sides of the channel (Fig. 1d). As has been mentioned in [17], this decrease of the flow intensity toward the channel sides is a natural phenomenon, because the lower the probability to visit some region of the conformation space (in the present case, a side of the channel), the smaller the flows in this region; similar nonuniform distributions of the flows were previously observed for an α -helical hairpin [12], SH3 domain [15], and β -hairpin [17]. Therefore, the vorticity of this type is an intrinsic property of the folding dynamics.

The calculated functions $\Phi(\mathbf{r})$ and $\Psi(\mathbf{r})$ are shown in Fig. 2a and 2b, respectively. In agreement with the Helmholtz decomposition [20], $\Phi(\mathbf{r})$ accounts for the source and sink of the folding flow, and $\Psi(\mathbf{r})$ for the vorticity effects. The surface for the Φ -component is simple; it is

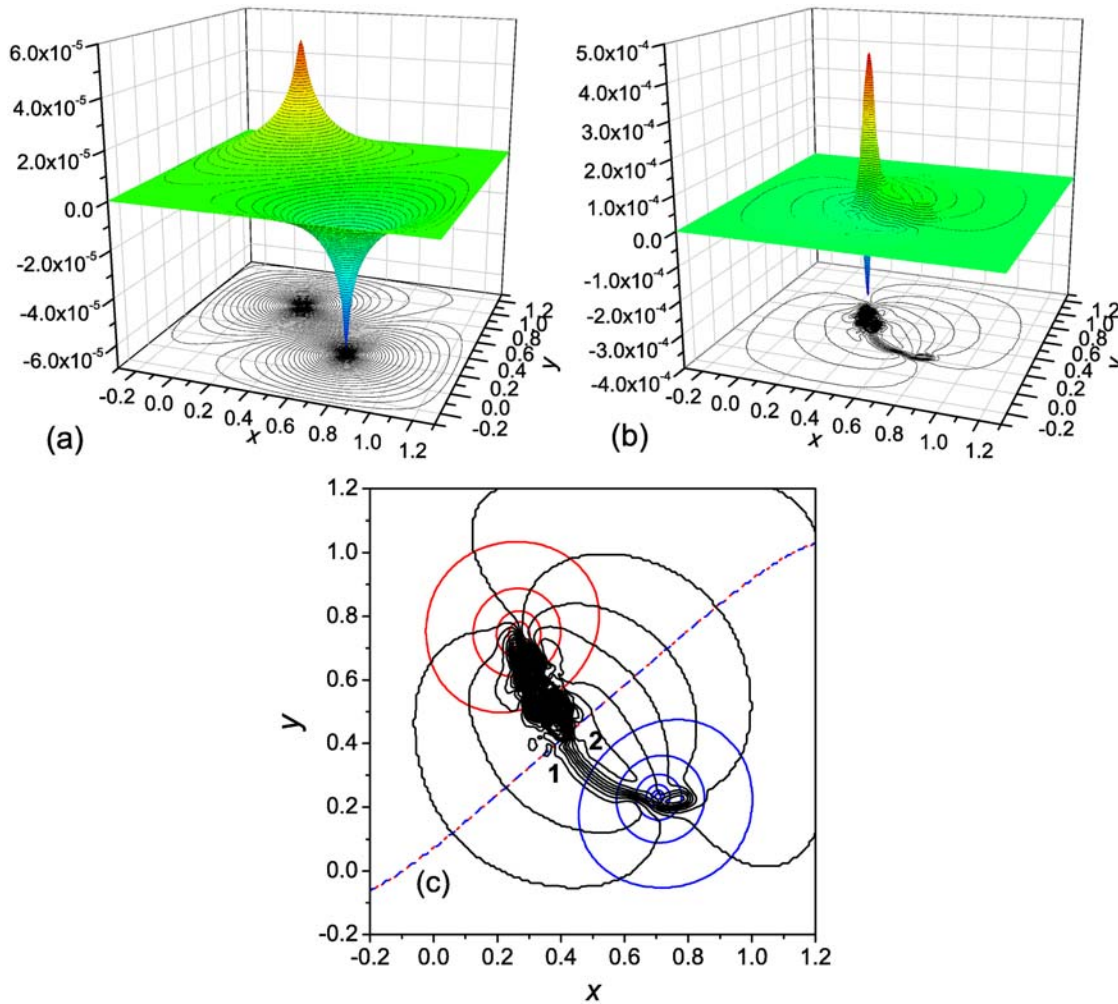


Fig 2. Two-well landscape with a single pathway: The potential for the driving force. (a) the Φ -component of the potential, (b) the Ψ -component, and (c) the isolines $\Phi(\mathbf{r}) = \text{const}$ (blue and red lines are, respectively, for the negative and positive values, and the blue-red dashed line is for the zero value), and $\Psi(\mathbf{r}) = \text{const}$ (black lines). In panel c the $\Phi(\mathbf{r}) = \text{const}$ and $\Psi(\mathbf{r}) = \text{const}$ isolines are shown with interval of $0.1/t_r \approx 1 \times 10^{-8}$ (see the text). Characteristic values of $\Psi(\mathbf{r})$ in the reaction channel: in region 1 $\Psi(\mathbf{r}) \approx -2 \times 10^{-5}$, and in region 2 $\Psi(\mathbf{r}) \approx 5 \times 10^{-5}$.

doi:10.1371/journal.pone.0121640.g002

characterized by two peaks of different sign—one (positive) peak corresponds to the source of the folding flow, and the other (negative) peak represents the sink of the flow. Poisson's Equation (9), in which $q(\mathbf{r}) > 0$ in the source region, $q(\mathbf{r}) < 0$ in the sink region, and $q(\mathbf{r}) = 0$ in the rest of the \mathbf{r} space, can be viewed as an equation that describes the stationary diffusion of some substance of density $\Phi(\mathbf{r})$ between the source and sink. It is evident that the substance ejected from the source will spread over the surface in all directions. Therefore, to canalize the flow between the source to sink, an addition force is required. This force is produced by the Ψ -component of the potential. Two properties of the $\Psi(\mathbf{r})$ surface are noteworthy. The first one is that in the region between the source and sink, where the $\Phi(\mathbf{r})$ surface is relatively flat, two parallel ridges of different sign are formed that connect the source and sink. More clearly, it is seen from Fig. 2c, where characteristic values of the Ψ -component in the regions corresponding to the ridges are indicated: the Ψ -component decreases from the periphery of the flow field, where it is zero (Fig. 2b), to approximately -2×10^{-5} in region 1 (the first ridge) and then

increases to 5×10^{-5} in region 2 (the second ridge), with a subsequent decrease to zero at the opposite side of the flow field. The second property of the $\Psi(\mathbf{r})$ surface is the presence of (two) peaks of different sign in the region for the unfolded states, which are due to the local vortices formed in this region (Fig. 1f). Such vortices were observed for the α -helical hairpin [12], SH3 domain [15] and three-stranded β -sheet protein [14], as a result of partial folding and unfolding of the protein in the attempts to overcome the barrier to the native state. It should be noted that in contrast to the presence of parallel ridges of different sign in $\Psi(\mathbf{r})$, which is typical of the folding process [17], the appearance of the peaks in $\Psi(\mathbf{r})$ depends on the value of the free energy barrier between the unfolded and the native states. If the barrier is large enough, the system dwells in the basin for the unfolded states performing a circulating motion, as in the present case; then the peaks are formed. However, if the barrier is low, so that the system easily overcomes it, the peaks are absent [17].

Fig. 2c depicts the equipotential lines $\Phi(\mathbf{r}) = \text{const}$ and $\Psi(\mathbf{r}) = \text{const}$. In each family, the lines are shown with the interval of $0.1/\bar{t}_f$. In the intermediate region between the source and sink, where the $\Phi(\mathbf{r})$ surface is relatively flat, the $\Psi(\mathbf{r})$ isolines are directed along the probability fluxes [17], i.e., each of them presents a streamline of the flow. Every two adjacent streamlines form a stream tube, so that the (ten) stream tubes between the $\Psi(\mathbf{r})$ isolines in this region contain the total flow from the source to the sink [see Equation (4)]. As has been mentioned in the Introduction, in contrast to [23], in which the probability current lines (streamlines) were introduced as the lines orthogonal to the isocommittor probability lines, the sets of the $\Phi(\mathbf{r}) = \text{const}$ and $\Psi(\mathbf{r}) = \text{const}$ lines are generally not mutually orthogonal because the folding flow is not curl-free; the orthogonality of these lines would require the flow to be both divergence- and curl-free [13] (see the discussion of this issue in [17]).

As has been previously indicated, the velocity field $\mathbf{v}(\mathbf{r}) = \mathbf{j}(\mathbf{r})/p(\mathbf{r})$ for a compressible folding fluid, i.e., for the probability density $p(\mathbf{r}) \neq \text{const}$, is not divergence-free. The calculations show that if the velocities are used instead of the fluxes in Equations (6) and (7), the Φ and Ψ components of the potential presented in Fig. 2 change very considerably. The main reason is that the probability density $p(\mathbf{r})$, which has a large value in the basin for the unfolded states, drastically decreases toward the native state. As a result, the velocity of the flow $\mathbf{v}(\mathbf{r})$ increases, and the effective source of the flow, determined by the velocity, shifts toward the flow sink. Correspondingly, the positive peak in $\Phi(\mathbf{r})$ shifts toward the negative peak, and the ridges between the source and sink regions in $\Psi(\mathbf{r})$ become considerably shorter. In other words, both $\Phi(\mathbf{r})$ and $\Psi(\mathbf{r})$ surfaces lose their direct connection with the physical picture of the process.

Two-well landscape with two pathways

One complication of the folding scenario described in the previous section is the presence of multiple folding pathways [39, 40]. We will consider the simplest case when there are two independent pathways, as, e.g., in the B domain of protein A due to the symmetrical backbone structure of the protein (a three-helix bundle) [41, 42], or in the fyn SH3 domain, where the fast folding trajectories are organized in two essentially independent routes due to the reverse order of formation of the contacts between the β_1 and β_5 strands and the RT-loop and the β_4 strand [15]. The values of the parameters determining the potential energy function in the present case are given in Table 2, and the corresponding PES is shown in Fig. 3a. The temperature at which the simulations were performed was $T = 1.5$. The native state was associated with the point $x_n = 0.65$, $y_n = 0.2$, and the MD trajectories were initiated in the vicinity of the point $x_u = 0.35$, $y_u = 0.8$. The MFPT is $\bar{t}_f \approx 2.2 \times 10^3$. Figs. 3b-3f and 4a-4c show the simulation results, akin to the corresponding panels of Figs. 1 and 2.

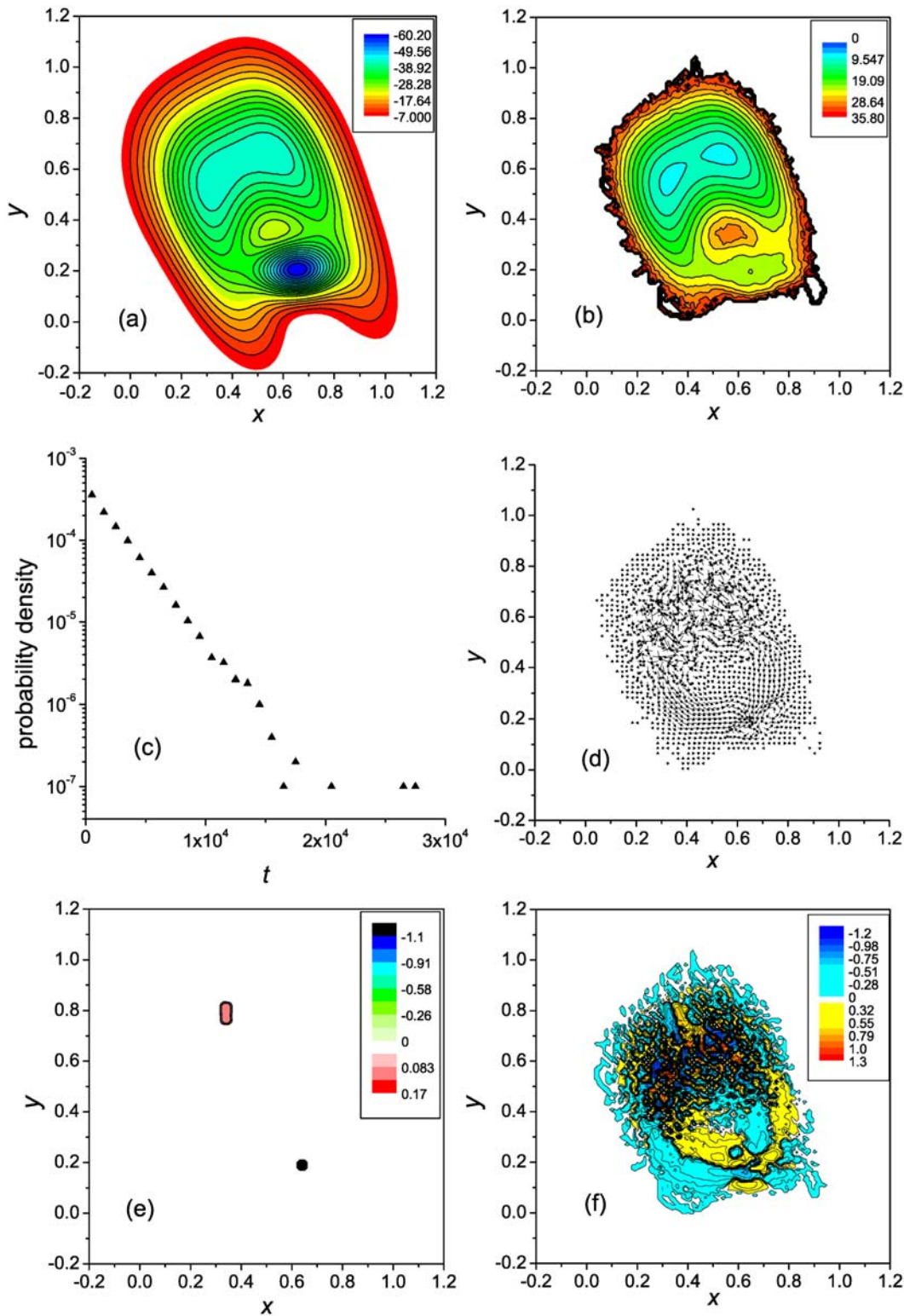


Fig 3. Two-well landscape with two pathway: General characterization. Notations are as in Fig. 1.

doi:10.1371/journal.pone.0121640.g003

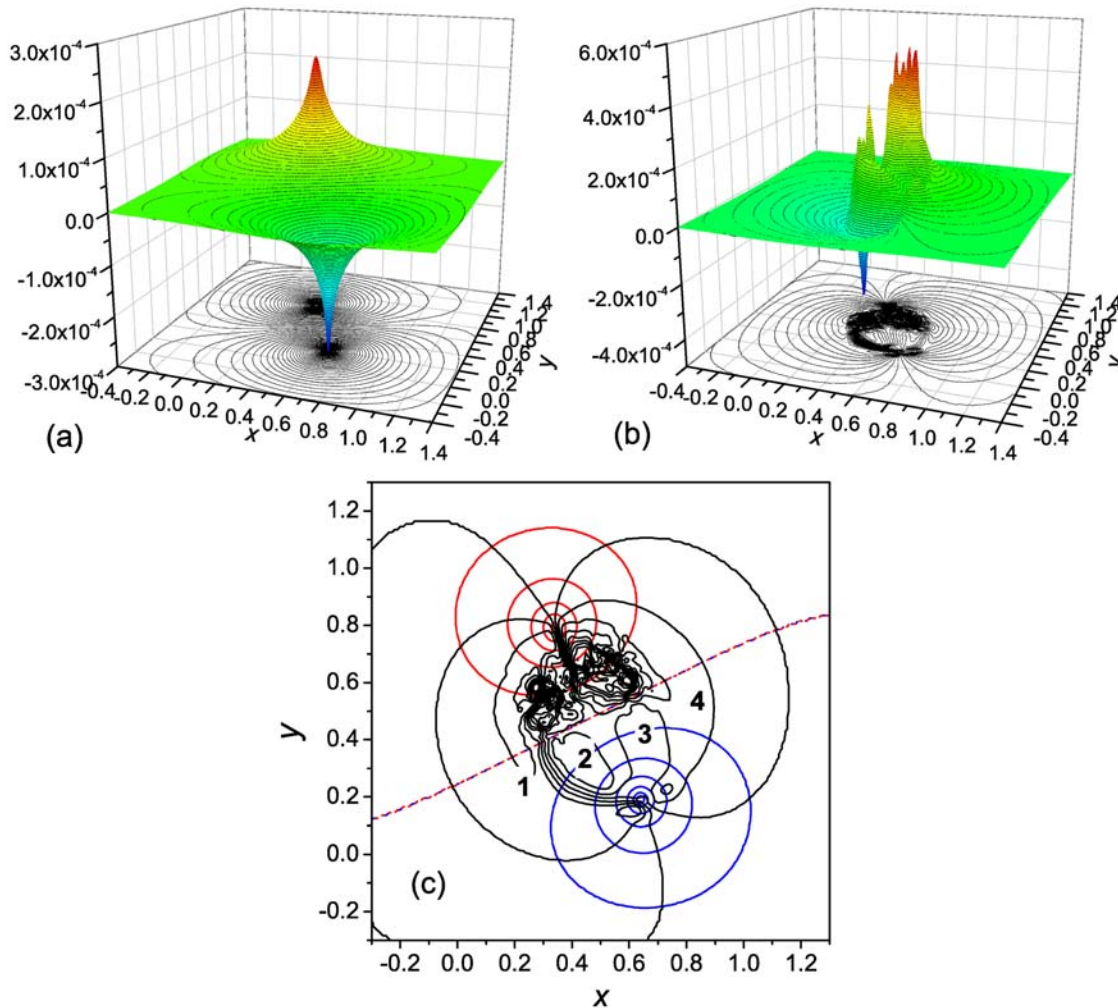


Fig 4. Two-well landscape with two pathway: The potential for the driving force. General notations are as in Fig. 2. In panel c the $\Phi(r) = \text{const}$ and $\Psi(r) = \text{const}$ isolines are shown with interval of $0.1/\bar{t}_f \approx 4.5 \times 10^{-5}$. Characteristic values of $\Psi(r)$ in the reaction channel: $\Psi(r) \approx -1 \times 10^{-4}$ (region 1), $\Psi(r) \approx 1.3 \times 10^{-4}$ (region 2), $\Psi(r) \approx 3.8 \times 10^{-5}$ (region 3), and $\Psi(r) \approx 8.2 \times 10^{-5}$ (region 4).

doi:10.1371/journal.pone.0121640.g004

According to Fig. 3a and 3b, the right-hand pathway from the unfolded to the native state is more effective (see also Fig. 3d). Since the pathways are independent, the kinetics remain two-state [43] (Fig. 3c), as it were in the case of a single pathway (Fig. 1c). Also, similar to the single pathway (Fig. 1f), each of the pathways is characterized by the change of the flow vorticity from negative to positive across the corresponding reaction channel. Accordingly, two pairs of parallel ridges are formed that connect the source and sink regions (Fig. 4b). Some details are given in Fig. 4c. The function $\Psi(r)$ first decreases from zero at the periphery of the flow field (Fig. 4b) to approximately -1.4×10^{-4} (region 1) and then increases to 1.3×10^{-4} (region 2), which represents the first pair of ridges. This is followed by formation of the second pair of ridges, in which $\Psi(r)$ decreases to 3.8×10^{-5} (region 3) and then increases to 8.2×10^{-5} (region 4). The peaks in the Ψ -component are of the same origin as for the single pathway PES (Fig. 2b). Despite the dynamics of the system in the present case, and the $\Psi(r)$ surface as well, are more complex than in the case of the single pathway, the Φ -component (Fig. 4a) is of the same shape as in the latter case.

Three-well landscape with an off-pathway intermediate

Another typical complication of the basic folding scenario described previously is the presence of off-pathway intermediates, which lead to a deviation from two-state kinetics [11]. When comparable with the native state in the life-time, such intermediates can play a role of “latent” states [44]. Table 3 gives the parameters of the potential energy function we used to construct the model PES in this case, and Fig. 5a depicts the PES. The simulations were performed at $T = 1.25$. The native state was associated with the point $x_n = 0.75, y_n = 0.15$, and the MD trajectories were initiated in the vicinity of the point $x_u = 0.25, y_u = 0.8$. The off-pathway intermediate is presented by a basin centered at $x = 0.75, y = 0.65$. The MFPT is $\bar{t}_f \approx 2.5 \times 10^4$. The simulation results are shown in Figs. 5b-5f and 6a-6c.

In accord to the FES (Fig. 5b), Fig. 5c indicates that the kinetics are double-exponential; the MFPTs for the fast and slow trajectories are equal to approximately 1.5×10^4 and 1.5×10^5 , respectively. In the fast trajectories, the system goes from the unfolded states to the native state directly, while in the slow trajectories, it visits the intermediate (possibly, several times) before proceeding to the native state. The directed motion toward the native state is accompanied by a gradual change of the flow vorticity across the reaction channel (Fig. 5f), similar to what was observed for the systems considered in the previous sections. Accordingly, a pair of parallel ridges is formed that connect the source (unfolded states) and the sink (native state) of the flow (cf. Figs. 6b and 5d). At the same time, the reaction channel between the unfolded states and the intermediate does not possess these properties, because the flow toward the intermediate is compensated by the backward flow. Correspondingly, the intermediate presents neither a source nor sink of the flow, which is reflected in the distribution of divergence (Fig. 5e) and the Φ -component (Fig. 6a), and the $\Psi(\mathbf{r})$ surface does not have ridges connecting the unfolded states and the intermediate (Fig. 6b). Also note that a set of vortices is formed in the basin associated with the intermediate (Fig. 5f), similar to the basins for the unfolded states in this (Fig. 5f) and previous systems (Figs. 1f and 3f). These vortices lead to formation of a set of peaks of positive and negative sign that the Ψ -component has in the intermediate basin.

Conclusions

Using model potential energy surfaces generated with the 2D Müller and Brown potential [24], three characteristic scenarios of protein folding have been considered: a single pathway from the unfolded to the native state without intermediates (two-state kinetics, a basic scenario), two parallel pathways without intermediates (two-state kinetics), and a single pathway with an off-pathway intermediate (three-state kinetics). It has been shown that despite a reduced contribution of conformation entropy, the 2D surfaces enable consistent reproduction of the characteristic properties of the folding reaction in a 2D space of collective variables that are essential for the goal of the present study, i.e., the vector field of the probability fluxes, which is used to calculate the potential for the driving force, and the distribution of the first-passage times and the FES landscape, which are necessary to relate the results obtained with the 2D PES to a specific folding scenario. To calculate the probability fluxes, the hydrodynamic description of the folding reaction [12] was employed. In it, the process of the first-passage folding is viewed as a steady flow of the representative points of the system from the unfolded to the native state. One essential property of the hydrodynamic approach is that the Helmholtz decomposition [20] of the vector field of the probability fluxes makes it possible to introduce the potential for the driving forces of folding [17]. This potential has two components, $\Phi(\mathbf{r})$ and $\Psi(\mathbf{r})$; the Φ -component accounts for the sources and sinks of the folding flows, and the Ψ -component for the flow vorticity effects. Both components obey Poisson's equations; in the equation for $\Phi(\mathbf{r})$,

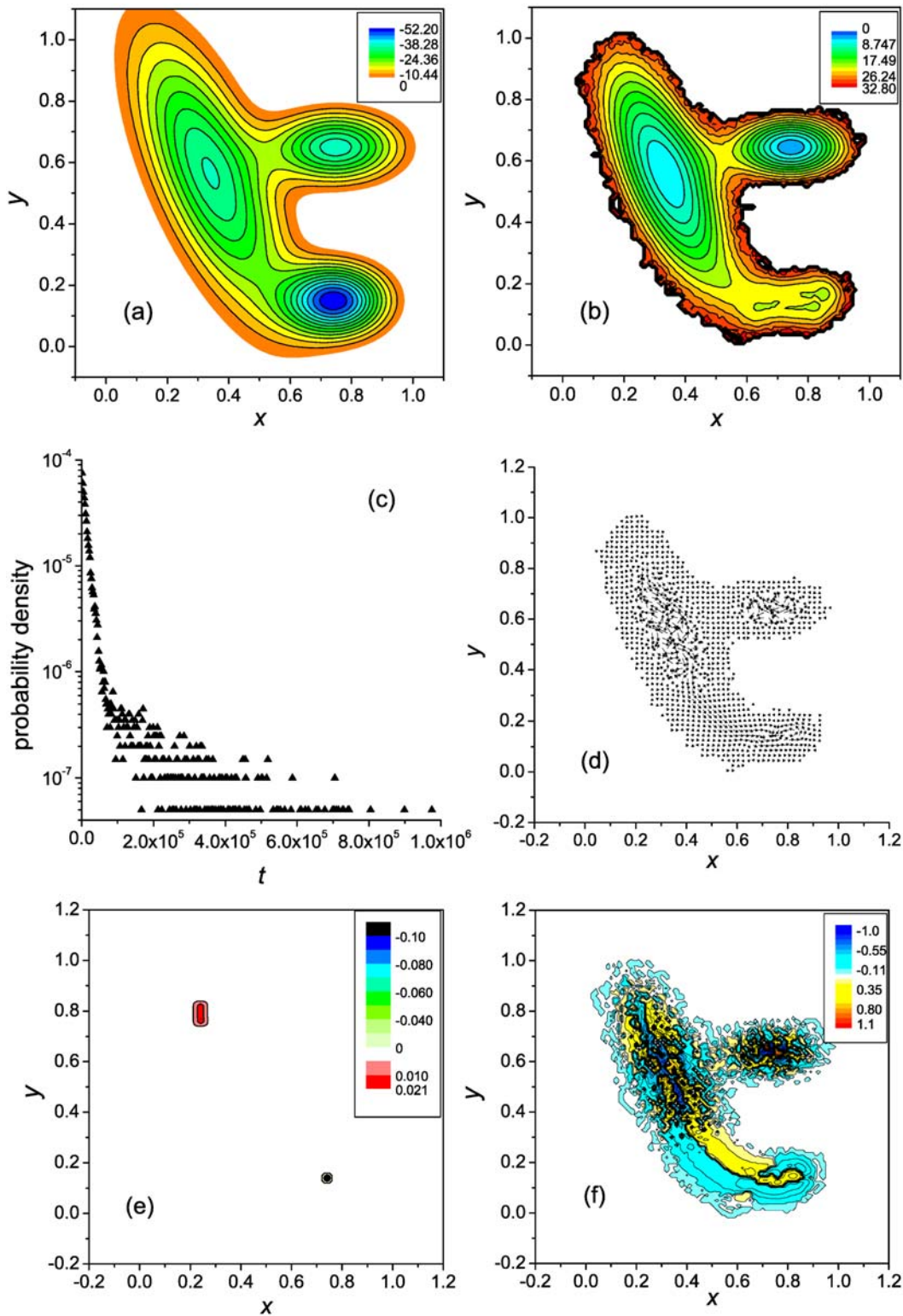


Fig 5. Three-well landscape with an off-pathway state: General characterization. Notations are as in Fig. 1.

doi:10.1371/journal.pone.0121640.g005

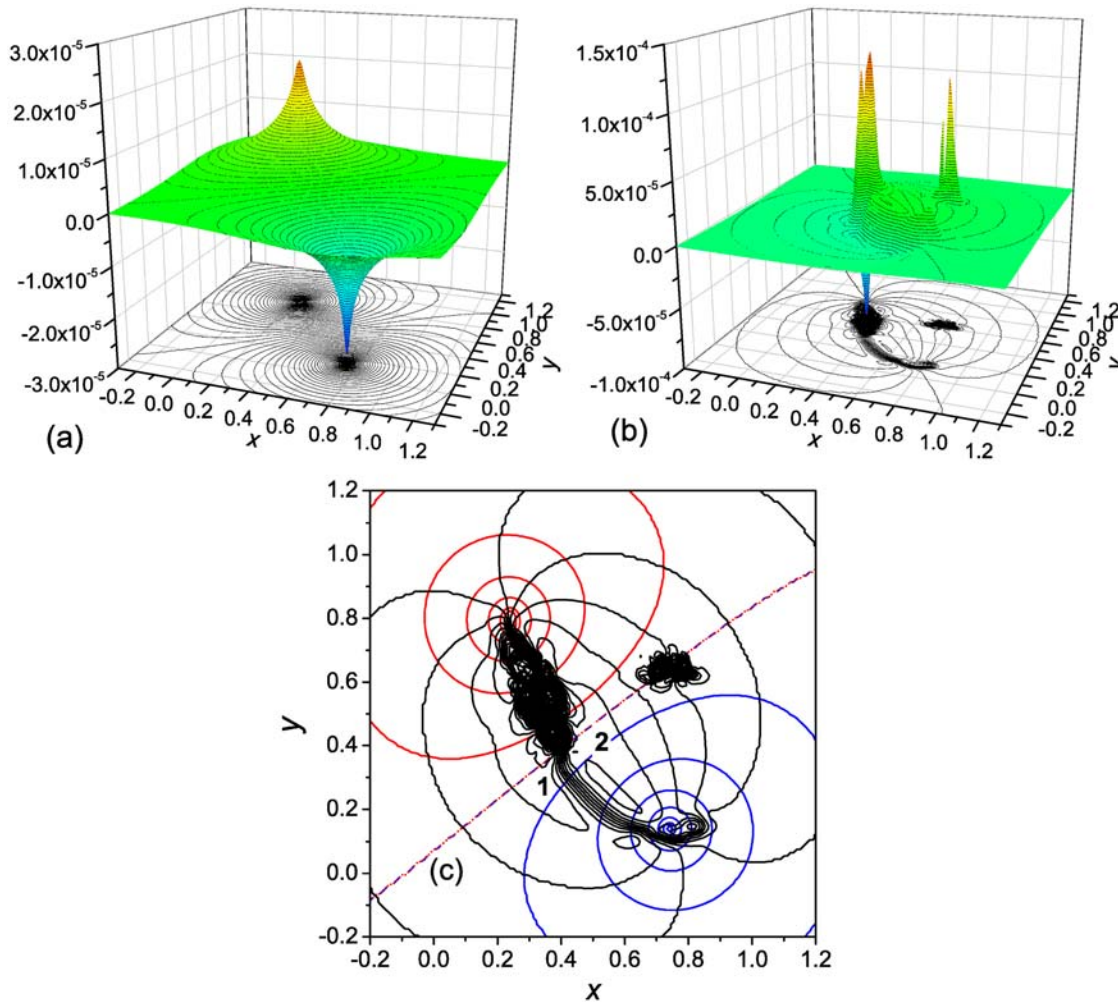


Fig 6. Three-well landscape with an off-pathway state: The potential for the driving force. General notations are as in Fig. 2. In panel c the $\Phi(\mathbf{r}) = \text{const}$ and $\Psi(\mathbf{r}) = \text{const}$ isolines are shown with interval of $0.1/\bar{t}_f \approx 4 \times 10^{-6}$. Characteristic values of $\Psi(\mathbf{r})$ in the reaction channel: in region 1 $\Psi(\mathbf{r}) \approx -1 \times 10^{-5}$, and in region 2 $\Psi(\mathbf{r}) \approx 2 \times 10^{-5}$.

doi:10.1371/journal.pone.0121640.g006

the source/sink term represents the distribution of the flow sources and sinks, and in the equation for $\Psi(\mathbf{r})$, the distribution of the flow vorticity.

The consideration of the above mentioned scenarios has led us to the following conclusions: Despite all possible complexity of the folding process, the Φ -component of the potential is simple and universal in shape. It has two peaks of different sign; one (positive) peak corresponds to the source of the folding flow, representing the unfolded states in which the folding trajectories are initiated, and the other (negative) peak corresponds to the sink of the flow, representing the native state, where the trajectories are terminated. The Ψ -component is more complex and depends on the specificity of the folding process. The most notable property of the $\Psi(\mathbf{r})$ surface is the presence of parallel ridges of different sign that connect the source and sink of the flow; if there are several independent pathways, the corresponding pair of such ridges is formed for each of them. These ridges are the result of the flow vorticity, which gradually changes across the free energy valley connecting the unfolded and native states (the reaction channel) and has different signs at the sides of the valley, because the folding flow weakens toward both sides.

The components Φ and Ψ act in concord, i.e., in the region between the source and sink they are such that the fluxes generated by these components contribute to the directed flow from the source to the sink, while in the regions behind the source and sink they compensate each other. In the region between the source and sink, where the $\Phi(\mathbf{r})$ surface is relatively flat, the Ψ -component plays a primary role, providing the canalization of the flow within the reaction channel. Another, although not so characteristic [17], property of the $\Psi(\mathbf{r})$ surface is the possible presence of peaks of different sign in the regions where the system dwells performing a circulating motion, as it happens in the basins for the unfolded states and intermediates. The formation of these peaks depends on the heights of the barriers to leave these basins, i.e., if the system leaves the basin easily, the peaks are not formed.

For a system with many degrees of freedom, the present potential reflects the complexity of the reaction only in the part that is captured by two collective variables. For protein folding, the practice of application of the principal component analysis and other methods of dimensionality reduction (for a review, see, e.g., [18]) suggests that the general picture of the reaction, i.e., the dominant folding pathways and long-living intermediates, will likely be captured in a two-variable description. At the same time, some specific features of the reaction, which require a more detailed characterization of the process, may remain hidden, e.g., the conformational heterogeneity of the denatured state [45, 46] and the transitions between intermediate conformations [47]. In these cases, another pair of collective variables, which reveal the corresponding elements of the reaction, can be used to see how the potential changes, similar as the FESs for different pairs of the variables are constructed to make the folding picture clearer [47, 48].

The basic components of the present approach are the calculation of the vector field of the probability fluxes in a two-dimensional space of orthogonal variables and its subsequent decomposition into divergence-free and curl-free fields. None of them is specific of protein folding. Therefore, the approach is potentially applicable to any complex reaction, for which the transition from the reactant to the product can be described by two (collective) variables.

Author Contributions

Conceived and designed the experiments: SC. Performed the experiments: SC. Analyzed the data: SC. Contributed reagents/materials/analysis tools: SC. Wrote the manuscript: SC.

References

1. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: The energy landscape perspective. *Annu Rev Phys Chem.* 1997; 48: 545–600. PMID: [9348663](#)
2. Dobson CM, Sali A, Karplus M. Protein folding: A perspective from theory and experiment. *Angew Chem Int Edit.* 1998; 37: 868–893. doi: [10.1002/\(SICI\)1521-3773\(19980420\)37:7%3C868::AID-ANIE868%3E3.3.CO;2-8](#)
3. Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci.* 2000; 25: 331–339. doi: [10.1016/S0968-0004\(00\)01610-8](#) PMID: [10871884](#)
4. Shea JE, Brooks CL III. From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annu Rev Phys Chem.* 2001; 52: 499–535. doi: [10.1146/annurev.physchem.52.1.499](#) PMID: [11326073](#)
5. Dill KA, Ozkan SB, Shell MS, Weikl TR. The protein folding problem. *Ann Rev Biophys.* 2008; 37: 289–316. doi: [10.1146/annurev.biophys.37.092707.153558](#)
6. Dill KA, MacCallum JL. The protein-folding problem, 50 years on. *Science* 2012; 338: 1042–1046.
7. McLeish T. Protein folding in high-dimensional spaces: Hypergutters and the role of nonnative interactions. *Biophys J.* 2005; 88: 172–183.
8. Rimratchada S, McLeish TC, Radford SE, Paci E. The role of high-dimensional diffusive search, stabilization, and frustration in protein folding. *Biophys J.* 2014; 106: 1729–1740.

9. Socci ND, Onuchic JN, Wolynes PG. Protein folding mechanisms and the multidimensional folding funnel. *Proteins* 1998; 32: 136–158. doi: [10.1002/\(SICI\)1097-0134\(19980801\)32:2%3C136::AID-PROT2%3E3.3.CO;2-5](https://doi.org/10.1002/(SICI)1097-0134(19980801)32:2%3C136::AID-PROT2%3E3.3.CO;2-5) PMID: [9714155](https://pubmed.ncbi.nlm.nih.gov/9714155/)
10. Boczek EM, Brooks CL III. First-principles calculation of the folding free energy of a three-helix bundle protein. *Science* 1995; 269: 393–396. doi: [10.1126/science.7618103](https://doi.org/10.1126/science.7618103) PMID: [7618103](https://pubmed.ncbi.nlm.nih.gov/7618103/)
11. Chekmarev SF, Krivov SV, Karplus M. Folding time distributions as an approach to protein folding kinetics. *J Phys Chem B* 2005; 109: 5312–5330.
12. Chekmarev SF, Palyanov AY, Karplus M. Hydrodynamic description of protein folding. *Phys Rev Lett*. 2008; 100: 018107. doi: [10.1103/PhysRevLett.100.018107](https://doi.org/10.1103/PhysRevLett.100.018107) PMID: [18232827](https://pubmed.ncbi.nlm.nih.gov/18232827/)
13. Landau LD, Lifshitz EM. *Fluid Mechanics*. Oxford: Pergamon Press; 1987.
14. Kalgin IV, Chekmarev SF, Karplus M. First passage analysis of the folding of a β -sheet miniprotein: Is it more realistic than the standard equilibrium approach? *J Phys Chem B* 2014; 118: 4287–4299.
15. Kalgin IV, Karplus M, Chekmarev SF. Folding of a SH3 domain: Standard and “hydrodynamic” analyses. *J Phys Chem B* 2009; 113: 12759–12772.
16. Kalgin IV, Chekmarev SF. Turbulent phenomena in protein folding. *Phys Rev E* 2011; 83: 011920. doi: [10.1103/PhysRevE.83.011920](https://doi.org/10.1103/PhysRevE.83.011920)
17. Chekmarev SF. Protein folding: Complex potential for the driving force in a two-dimensional space of collective variables. *J Chem Phys*. 2013; 139: 145103.
18. Kalgin IV, Caffisch A, Chekmarev SF, Karplus M. New insights into the folding of a β -sheet miniprotein in a reduced space of collective hydrogen bond variables: Application to a hydrodynamic analysis of the folding flow. *J Phys Chem B* 2013; 117: 6092–6105.
19. Andersen NH, Olsen KA, Fesinmeyer RM, Tan X, Hudson FM, Eidenschink LA, et al. Minimization and optimization of designed β -hairpin folds. *J Am Chem Soc*. 2006; 128: 6101–6110 doi: [10.1021/ja054971w](https://doi.org/10.1021/ja054971w) PMID: [16669679](https://pubmed.ncbi.nlm.nih.gov/16669679/)
20. Weber HJ, Arfken GB. *Essential Mathematical Methods for Physicists*. San Diego: Elsevier; 2004.
21. Forst W. *Theory of Unimolecular Reactions*. New York: Academic Press; 1973.
22. Wales DJ. *Energy landscapes: Applications to clusters, biomolecules and glasses*. Cambridge: University Press; 2003.
23. E W, Vanden-Eijnden E. Transition-path theory and path-finding algorithms for the study of rare events. *Annu Rev Phys Chem*. 2010; 61: 391–420. doi: [10.1146/annurev.physchem.040808.090412](https://doi.org/10.1146/annurev.physchem.040808.090412) PMID: [18999998](https://pubmed.ncbi.nlm.nih.gov/18999998/)
24. Müller K, Brown LD. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor Chim Acta* 1979; 53: 75–93. doi: [10.1007/BF00547608](https://doi.org/10.1007/BF00547608)
25. Weissman JS, Kim PS. Reexamination of the folding of BPTI: Predominance of native intermediates. *Science* 1991; 253: 1386–1393. doi: [10.1126/science.1716783](https://doi.org/10.1126/science.1716783) PMID: [1716783](https://pubmed.ncbi.nlm.nih.gov/1716783/)
26. Otzen DE, Oliveberg M. Salt-induced detour through compact regions of the protein folding landscape. *Proc Natl Acad Sci USA* 1999; 96: 11746–11751. doi: [10.1073/pnas.96.21.11746](https://doi.org/10.1073/pnas.96.21.11746) PMID: [10518521](https://pubmed.ncbi.nlm.nih.gov/10518521/)
27. Munoz V, Lopez EM, Jager M, Serrano L. Kinetic characterization of the chemotactic protein from *Escherichia coli*, CheY. Kinetic analysis of the inverse hydrophobic effect. *Biochemistry* 1994; 33: 5858–5866. doi: [10.1021/bi00185a025](https://doi.org/10.1021/bi00185a025) PMID: [8180214](https://pubmed.ncbi.nlm.nih.gov/8180214/)
28. Olender R, Elber R. Calculation of classical trajectories with a very large time step: Formalism and numerical examples. *J Chem Phys*. 1996; 105: 9299–9315. doi: [10.1063/1.472727](https://doi.org/10.1063/1.472727)
29. Jas GS, Hegefeld WA, Májek P, Kuczera K, Elber R. (2012) Experiments and comprehensive simulations of the formation of a helical turn. *J Phys Chem B* 2012; 116: 6598–6610. doi: [10.1021/jp211645s](https://doi.org/10.1021/jp211645s) PMID: [22335541](https://pubmed.ncbi.nlm.nih.gov/22335541/)
30. Dagan S, Hagai T, Gavrilov Y, Kapon R, Levy Y, Reich Z. Stabilization of a protein conferred by an increase in folded state entropy. *Proc Natl Acad Sci USA* 2013; 110: 10628–10633. doi: [10.1073/pnas.1302284110](https://doi.org/10.1073/pnas.1302284110) PMID: [23754389](https://pubmed.ncbi.nlm.nih.gov/23754389/)
31. Biswas R, Hamann D. Simulated annealing of silicon atom clusters in Langevin molecular dynamics. *Phys Rev B* 1986; 34: 895–901. doi: [10.1103/PhysRevB.34.895](https://doi.org/10.1103/PhysRevB.34.895)
32. Sobolev S. *Partial Differential Equations of Mathematical Physics*. Oxford: Pergamon Press; 1964.
33. Dorr FW. The direct solution of the discrete poisson equation on a rectangle. *SIAM Rev*. 1970; 12: 248–263. doi: [10.1137/1012045](https://doi.org/10.1137/1012045)
34. Van Kampen NG. *Stochastic Processes in Physics and Chemistry*, volume 1. Amsterdam: North-Holland; 1992.
35. Tomita K, Tomita H. Irreversible circulation of fluctuation. *Prog Theor Phys*. 1974; 51: 1731–1749. doi: [10.1143/PTP.51.1731](https://doi.org/10.1143/PTP.51.1731)

36. Graham R. Covariant formulation of non-equilibrium statistical thermodynamics. *Z Phys B Con Mat.* 1977; 26: 397–405. doi: [10.1007/BF01570750](https://doi.org/10.1007/BF01570750)
37. Eyink GL, Lebowitz JL, Spohn H. Hydrodynamics and fluctuations outside of local equilibrium: Driven diffusive systems. *J Stat Phys.* 1996; 83: 385–472.
38. Jackson SE. How do small single-domain proteins fold? *Fold Des.* 1998; 3: R81–R91.
39. Levinthal C. How to fold graciously. In: Debrunner P, Tsibris JCM, Münck E, editors. *Mossbauer Spectroscopy in Biological Systems, Proceedings of a Meeting held at Allerton House, Monticello.* Urbana, Illinois: University of Illinois Press; 1969. pp. 22–24.
40. Wang J, Onuchic J, Wolynes P. Statistics of kinetic pathways on biased rough energy landscapes with applications to protein folding. *Phys Rev Lett.* 1996; 76: 4861. doi: [10.1103/PhysRevLett.76.4861](https://doi.org/10.1103/PhysRevLett.76.4861) PMID: [10061399](https://pubmed.ncbi.nlm.nih.gov/10061399/)
41. Itoh K, Sasai M. Flexibly varying folding mechanism of a nearly symmetrical protein: B domain of protein A. *Proc Natl Acad Sci USA* 2006; 103: 7298–7303.
42. Sato S, Fersht AR. Searching for multiple folding pathways of a nearly symmetrical protein: Temperature dependent Φ -value analysis of the B domain of protein A. *J Mol Biol.* 2007; 372: 254–267.
43. Levy RM, Dai W, Deng NJ, Makarov DE. How long does it take to equilibrate the unfolded state of a protein? *Protein Sci.* 2013; 22: 1459–1465.
44. Palyanov AY, Krivov SV, Karplus M, Chekmarev SF. A lattice protein with an amyloidogenic latent state: Stability and folding kinetics. *J Phys Chem B* 2007; 111: 2675–2687.
45. Krivov SV, Karplus M. Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc Natl Acad Sci USA* 2004; 101: 14766–14770. doi: [10.1073/pnas.0406234101](https://doi.org/10.1073/pnas.0406234101) PMID: [15466711](https://pubmed.ncbi.nlm.nih.gov/15466711/)
46. Caflisch A. Network and graph analyses of folding free energy surfaces. *Curr Opin Struc Biol.* 2006; 16: 71–78. doi: [10.1016/j.sbi.2006.01.002](https://doi.org/10.1016/j.sbi.2006.01.002)
47. Lange OF, Grubmüller H. Full correlation analysis of conformational protein dynamics. *Proteins* 2008; 70: 1294–1312. doi: [10.1002/prot.21618](https://doi.org/10.1002/prot.21618) PMID: [17876828](https://pubmed.ncbi.nlm.nih.gov/17876828/)
48. Zheng W, Qi B, Rohrdanz MA, Caflisch A, Dinner AR, Clementi C. Delineation of folding pathways of a β -sheet miniprotein. *J Phys Chem B* 2011; 115:13065–13074.