RESEARCH ARTICLE

Turbulence in protein folding: Vorticity, scaling and diffusion of probability flows

Vladimir A. Andryushchenko[1,2], Sergei F. Chekmarev[1,2] *

1 Institute of Thermophysics, SB RAS, Novosibirsk, Russia, **2** Department of Physics, Novosibirsk State University, Novosibirsk, Russia

* chekmarev@itp.nsc.ru

Abstract

Recently, when studying folding of a SH3 domain, we discovered that the flows of transitions between protein states can be surprisingly similar to turbulent fluid flows. This similarity was not restricted by a vortex pattern of the flow fields but extended to a spatial correlation of flow fluctuations, resulting, in particular, in the structure functions such as in the Kolmogorov theory of homogeneous and isotropic turbulence. Here, we undertake a detailed analysis of spatial distribution of folding flows and their similarity to turbulent fluid flows. Using molecular dynamics simulations, we study folding of another benchmark system—Trp-cage miniprotein, which has different content of secondary structure elements and mechanism of folding. Calculating the probability fluxes of transitions in a three-dimensional space of collective variables, we have found that similar to the SH3 domain, the structure functions of the second and third orders correspond to the Kolmogorov functions. The spatial distributions of the probability fluxes are self-similar with a fractal dimension, and the fractal index decreases toward the native state, indicating that the flow becomes more turbulent as the native state is approached. We also show that the process of folding can be viewed as Brownian diffusion in the space of probability fluxes. The diffusion coefficient plays a role of the key parameter that defines the structures functions, similar to the rate of dissipation of kinetic energy in hydrodynamic turbulence. The obtained results, first, show that the very complex dynamics of protein folding allows a simple characterization in terms of scaling and diffusion of probability fluxes, and, secondly, they suggest that the turbulence phenomena similar to hydrodynamic turbulence are not specific of folding of a particular protein but are common to protein folding.

Introduction

Protein folding and hydrodynamic turbulence are two challenging problems that attract attention of researchers for years. Turbulent motion of a fluid is a stochastic motion, which arises due to the instability of the fluid flow at large Reynolds numbers, i.e., when the inertia of the fluid motion dominates over viscosity [1–4]. Typically, the turbulent motion appears as a cascade of eddies of various sizes. One example is when large eddies generated by external forces, e.g., by the walls of the pipe through which the fluid flows, disintegrate into smaller eddies

until the latter dissipate due to viscosity (the Richardson cascade [5]). In contrast to the fluid, which is a collection a large number of atoms ($\sim 10^{24}$) and thus can be described in the macroscopic terms such as the average velocity, density, etc., a protein is a system of finite size, which can be as small as of $\sim 10^3$ atoms, and thus requires the description on atomic level. Synthesized on the ribosome as a chain of amino acid residues, a protein folds into a compact functional (native) state. The process of folding is typically very complex, with a variety of folding pathways and metastable states [6–11]. One essential feature of protein folding that makes it similar to hydrodynamic turbulence [1–5] is that the process of folding is inherently a cascade process—in the present case, in the form of sequential rearrangement of the protein structure from an unfolded state to the native state. The cascade nature of the process is also characteristic of the other known types of turbulence—the wave [12], market [13] and superfluid [14] turbulence (see, also, a discussion of the cascades in the latter case [15]).

A detailed analysis of similarity between protein folding and hydrodynamic turbulence becomes possible if, instead of evolution of protein structure in the multidimensional (all-atom) conformational space, we consider probability fluxes of transitions between characteristic states of the protein in a reduced space of collective variables. Such is a recently proposed hydrodynamic description of protein folding [16]. The purpose of that approach was to gain a closer insight into folding dynamics, because typically employed free energy surfaces (FESs) display only the probability for the protein to be in a current state but do not show the direction in which the protein proceeds (folds, unfolds, or dwells in the current state). The process of "first-passage folding", i.e., when the folding trajectories are initiated in a unfolded state of the protein and terminated upon reaching the native state, is of particular interest because it corresponds to physiological conditions when the native state is stable and unfolding events are improbable [17]. Having the probability fluxes, the process of first-passage folding can be viewed as a stationary flow of a "folding fluid" from an unfolded state of the protein to its native state, with the density of the fluid being proportional to the probability for the system to be in the current state. The analysis of the first-passage folding of several model proteins (an α-helical hairpin [16], a SH3 domain [18, 19], and beta3s [17, 20, 21] and 2evq [22] miniproteins) has shown that the folding flows do not generally follow the FESs and typically contain vortices that remind eddies in turbulent flows. To see how the protein folding flows are close to turbulent fluid flows, the folding flows of SH3 domain were characterized in terms accepted in hydrodynamic turbulence [19]. Specifically, there were calculated so called structure functions, which represent velocity space correlation functions [2], or, more exactly, flux space correlation functions, because the folding fluid is highly "compressible" [19]. According to the Kolmogorov theory of isotropic and homogeneous turbulence (K41) [23, 24], the fluctuations of the flow velocities scale with the space increment l as $l^{1/3}$, so that the structure functions of the second and third order vary as $l^{2/3}$ and l, respectively. Very surprisingly, it was found that the corresponding structure functions for folding flows of SH3 domain reveal exactly the same dependence on the increment in the inter-residue contact space [19].

These results for SH3 domain lead to a natural question of how such turbulence phenomena are common to protein folding. To see that, we consider another benchmark system—the Trp-cage miniprotein [25–32], whose secondary structure content and mechanism of folding are essentially different from those for SH3 domain. In particular, the kinetics of Trp-cage folding are single-exponential, while folding kinetics of SH3 domain were double-exponential, and the turbulent flow was observed only for slow folding trajectories [19]. Also, we employ an essentially different approach to study the Trp-cage folding: First, the molecular dynamics (MD) simulations are performed using an all-atom model (CHARMM program [33]), while for the SH3 domain a coarse-grained representation of the protein was used, in which the amino acid residues were considered as monomers placed on positions of C_α-atoms in the

protein chain (C_α-model) [19]. Secondly, the collective variables are determined with a principal component analysis (PCA) method [34], while in the case of SH3 domain they were represented by weakly dependent groups of native contacts, which were considered as "physically" orthogonal variables [19]. We find that despite such a difference between these proteins and their characterization, the structure functions of the second and third orders for the Trp-cage follow the Kolmogorov scaling for isotropic and homogeneous turbulence, similar to those in the case of SH3 domain. Further, we show that the time-rate of change of the variance of folding fluxes is approximately constant in the dominant interval of times, so that it can be considered as a key parameter to characterize folding flows, similar to the rate of energy dissipation in hydrodynamic turbulence. Accordingly, the process of protein folding can be viewed as Brownian diffusion in the space of probability fluxes. Finally, we show that the folding flows are self-similar with a fractal dimension, and the fractal index decreases as the native state is approached.

The paper is organized as follows. The next section briefly describes the protein model, the simulation technique, the methods we used to characterize the folding process, and a general picture of Trp-cage folding (for more details see [32]). The subsequent section presents the results and their discussion. The last section contains concluding remarks.

Folding of Trp-cage miniprotein

System and simulation method

Trp-cage is a 20-residue miniprotein (Asn1-Leu2-Tyr3-Ile4-Gln5-Trp6-Leu7-Lys8-Asp9-Gly10-Gly11-Pro12-Ser13-Ser14-Gly15-Arg16-Pro17-Pro18-Pro19-Ser20; 1L2Y.pdb) [25]. It consists of a N-terminal α-helix (residues 2-8), a 3_{10}-helix (residues 11-14), and a C-terminal polyproline II (PPII) helix (residues 17-19), which form a hydrophobic core with the Trp6 buried in the center (Fig 1). The interactions between Tyr3, Trp6, Gly11, Pro12, Pro18, and Pro19 lead to formation of the Trp-cage fold. To perform MD simulations, the CHARMM program [33] was employed. All heavy atoms and the hydrogen atoms bound to nitrogen or oxygen atoms were considered explicitly; PARAM19 force field [35] and a default cutoff of 7.5Å for the nonbonding interactions were used. To take into account the main effects of the aqueous solvent, a meanfield approximation based on the solvent-accessible surface (SAS) [36] was employed. The simulations were performed with the time step of 2 fs using the Berendsen thermostat (coupling constant of 5 ps) [37]. Although such an approach overestimates folding rates, mostly because of the absence of the friction of protein atoms against the solvent, the relative rates of formation of secondary structural elements are comparable to the values observed in experiment; i.e., α-helices fold in about 1 ns and β-hairpins in about 10 ns [38] compared to experimental values of $\sim 0.1\mu s$ and $\sim 1\mu s$ [39], respectively.

The folding trajectories were initiated in unfolded states of the protein and terminated upon reaching the native state. The unfolded states were prepared using the standard CHARMM protocol [33]; i.e., an extended conformation of the protein was first minimized (200 steps of the steepest descent followed by 300 steps of the conjugate gradient algorithm) and then heated to $T = 300K$ and equilibrated for 5×10^3 time steps. A native contact was assumed to be formed if the distance between the C_α-atoms in the residues which are not neighbors in the sequence is less than 6.5Å in all NMR structures [25], which resulted in 35 native contacts. The simulations were conducted for $T = 300K$; at this temperature, the mean first-passage time (MFPT) was minimal and equal to $\approx 36ns$, which is in good agreement with the experimental time (4.1μs [26]), if to take into account that the simulations with implicit solvent overestimate the rate of formation of secondary structure elements by $\approx 10^2$ times.
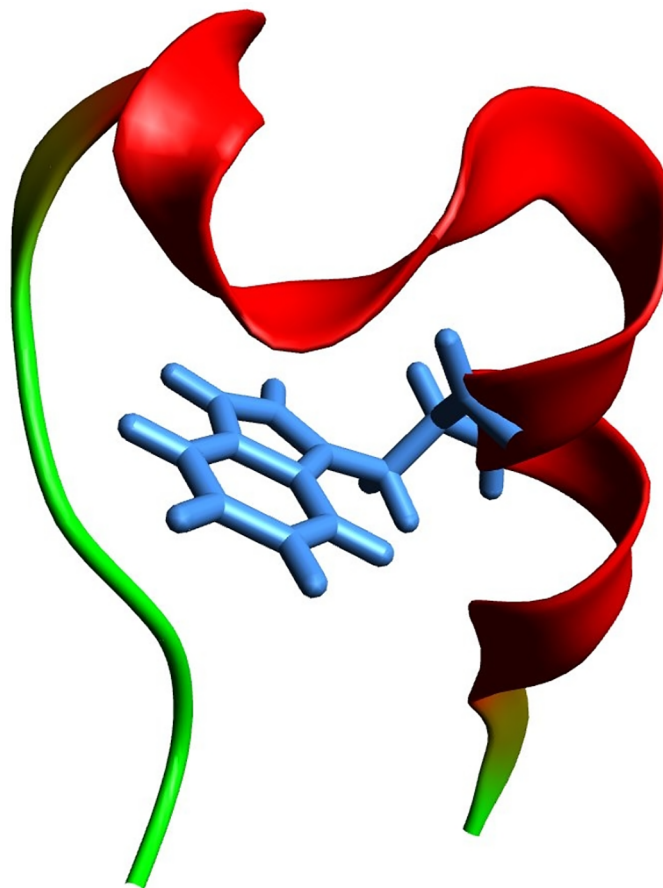
Fig 1. The native structure of the Trp-cage miniprotein (1L2Y.pdb) in a ribbon representation. The Trp6 residue is shown in blue sticks.

There were simulated 100 folding trajectories. Protein conformations ("frames") were stored each 20 ps, which resulted in 229420 conformations in total.

Conformation space and collective variables

To characterize protein conformations, the distances between C_α-atoms in the residues that formed native contacts were used. This space of the distances was then transformed to a space of orthogonal collective variables using the PCA method [34]. It was found that the first three eigenvalues were well separated from the others and captured $\approx 29\%$, $\approx 21\%$ and $\approx 19\%$ of the data variation, which resulted in $\approx 69\%$ of information in total. The eigenvectors corresponding to these modes were chosen to form a three-dimensional (3D) space of collective variables $\mathbf{g} = (g_1, g_2, g_3)$. To determine a two-dimensional (2D) space of variables, $\mathbf{G} = (G_1, G_2)$, the variable G_1 was chosen as g_1, and the variable G_2 was determined as a sum of the second and third eigenvectors weighted according to their eigenvalues, similar to Ref. [22]. Since the collective variables are linear combinations of the original variables (distances), they are measured in the same units as the latter, specifically, in angstroms.

Probability fluxes

The probability fluxes, determining the local rates of transitions between protein states in the **g** space, were calculated according to the hydrodynamic description of protein folding [16]. In the case of the 3D space of collective variables, the g_1-component of the flux **j(g)** was calculated as

$$j_{g_1}(\mathbf{g}) = \left[\sum_{\mathbf{g'},\mathbf{g''}(\mathbf{g}\subset\mathbf{g^*})}^{g_1''-g_1'>0} n(\mathbf{g''},\mathbf{g'}) - \sum_{\mathbf{g'},\mathbf{g''}(\mathbf{g}\subset\mathbf{g^*})}^{g_1''-g_1'<0} n(\mathbf{g''},\mathbf{g'}) \right] / (M\bar{t}_f \Delta g_2 \Delta g_3) \tag{1}$$

where M is the total number of simulated trajectories, $\bar{t}_f$ is the MFPT, $n(\mathbf{g''},\mathbf{g'})$ is the number of transitions from state $\mathbf{g'}$ to $\mathbf{g''}$, and $\mathbf{g} \subset \mathbf{g^*}$ is a symbolic designation of the condition that the transitions included in the sum have the straight line connecting points $\mathbf{g'}$ to $\mathbf{g''}$, which crosses the plane g_1 = const within the elementary cell $\Delta g_2 \times \Delta g_3$ centered at the point **g**. The g_2 and g_3 components of **j(g)** are determined in a similar way, except that one selects the transitions crossing the planes g_2 = const and g_3 = const within the cells $\Delta g_1 \times \Delta g_3$ and $\Delta g_1 \times \Delta g_2$, respectively. In the case of the 2D space, the planes and elementary cells are replaced with the lines and elementary segments along these lines, respectively. The calculations were performed on a grid with discretization $\Delta g_1 = \Delta g_2 = \Delta g_3 = 1$Å. In what follows, distances and times will be measured in angstroms and microseconds, respectively.

Visualization of the streamlines

To visualize the streamlines in the 3D space of variables, $\mathbf{g} = (g_1, g_2, g_3)$, we used passive tracers. Starting from various points of the **g** space, there was solved the equation

$$\frac{d\mathbf{g}}{d\tau} = \mathbf{j}(\mathbf{g}) \tag{2}$$

where **j(g)** is determined by Eq (1), and τ is a parameter ("time"). To calculate intermediate values of **j(g)**, an algorithm of linear interpolation between the neighboring points [40] was used.

In the case of the 2D space of variables, $\mathbf{G} = (G_1, G_2)$, the streamlines can be calculated as the lines corresponding to constant values of the stream function [2]

$$\Psi(G_1, G_2) = \int_{G_2'=0}^{G_2'=G_2} J_{G_1}(G_1, G_2')dG_2' \tag{3}$$

where **J(G)** is the probability flux in the 2D space. Then, two streamlines that satisfy the equations $\Psi(G_1, G_2) = C_1$ and $\Psi(G_1, G_2) = C_2$, where the constant C_1 and C_2 obey the condition $C_2 > C_1$, create a stream tube which contains the $(C_2 - C_1)/\Pi$ fraction of the total flow $\Pi = \int J_{G_1}(G_1, G_2')dG_2'$.

General picture of Trp-cage folding

As has been indicated in the Introduction, the process of first-passage folding, which we consider in the present paper, can be viewed as a stationary flow of a folding fluid from an unfolded state of the protein to its native state. Fig 2a and 2b show the general picture of the flow field—the vector flow field and the folding trajectories in the form of passive tracers, respectively. The common understanding of the process of folding of Trp-cage is that it can fold trough one of two (or through both) characteristic folding pathways [28–32]: in one pathway (I), the collapse of the hydrophobic core precedes the formation of the α-helix, and in the
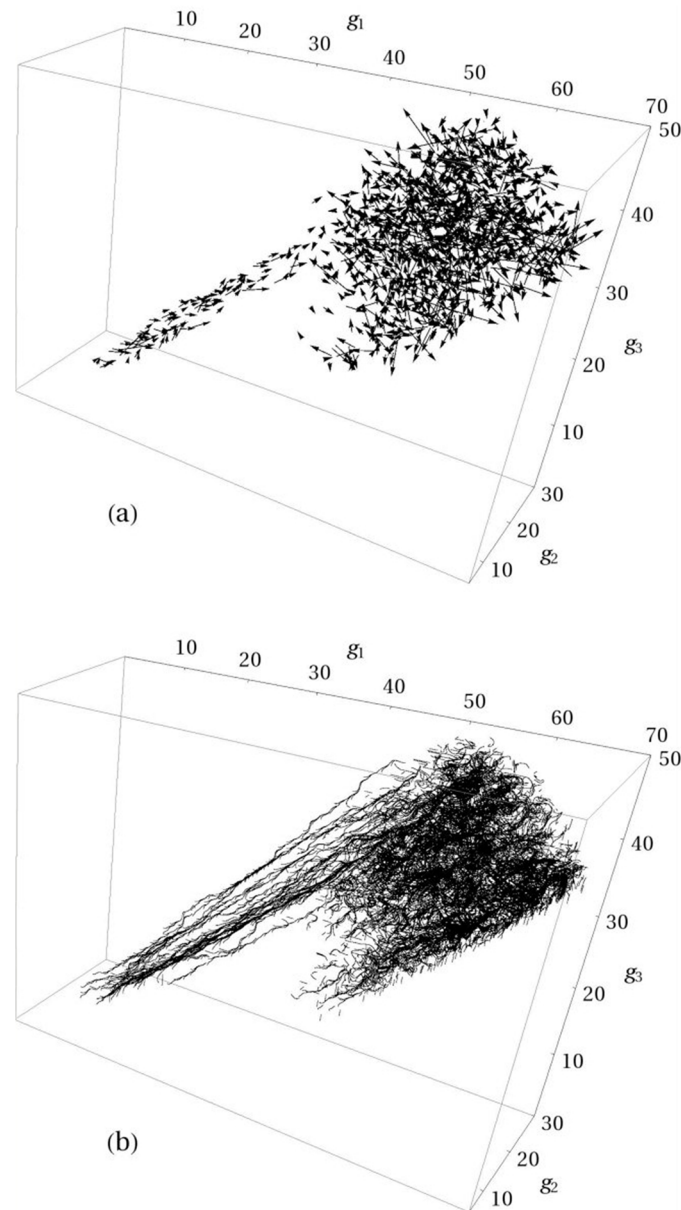
Fig 2. Three-dimensional flow field of the Trp-cage folding. Panel (**a**) shows the vector flow field, and panel (**b**) depicts the folding trajectories in the form of passive tracers (for illustration purpose, twenty randomly selected trajectories were chosen). Folding trajectories are initiated in the region of unfolded states ($g_1 \approx 12.0$, $g_2 \approx 14.0$, $g_3 \approx 7.0$) and terminated in the native state ($g_1 \approx 65.5$, $g_2 \approx 14.8$, $g_3 \approx 38.1$).

other pathway (II), the α-helix forms first. Fig 3a and 3b show the streamlines of the folding flow superimposed, respectively, on the FES and the distribution of flow vorticity. To make the vortex picture of the flow field clearer, Fig 3c also presents the folding trajectories in the form of passive tracers. The free energy was calculated as

$$F(\mathbf{G}) = -k_{\mathrm{B}}T \ln p(\mathbf{G}) \tag{4}$$

where $p(\mathbf{G})$ is the probability for the system to be found at the point $\mathbf{G} = (G_1, G_2)$ and k_B is the
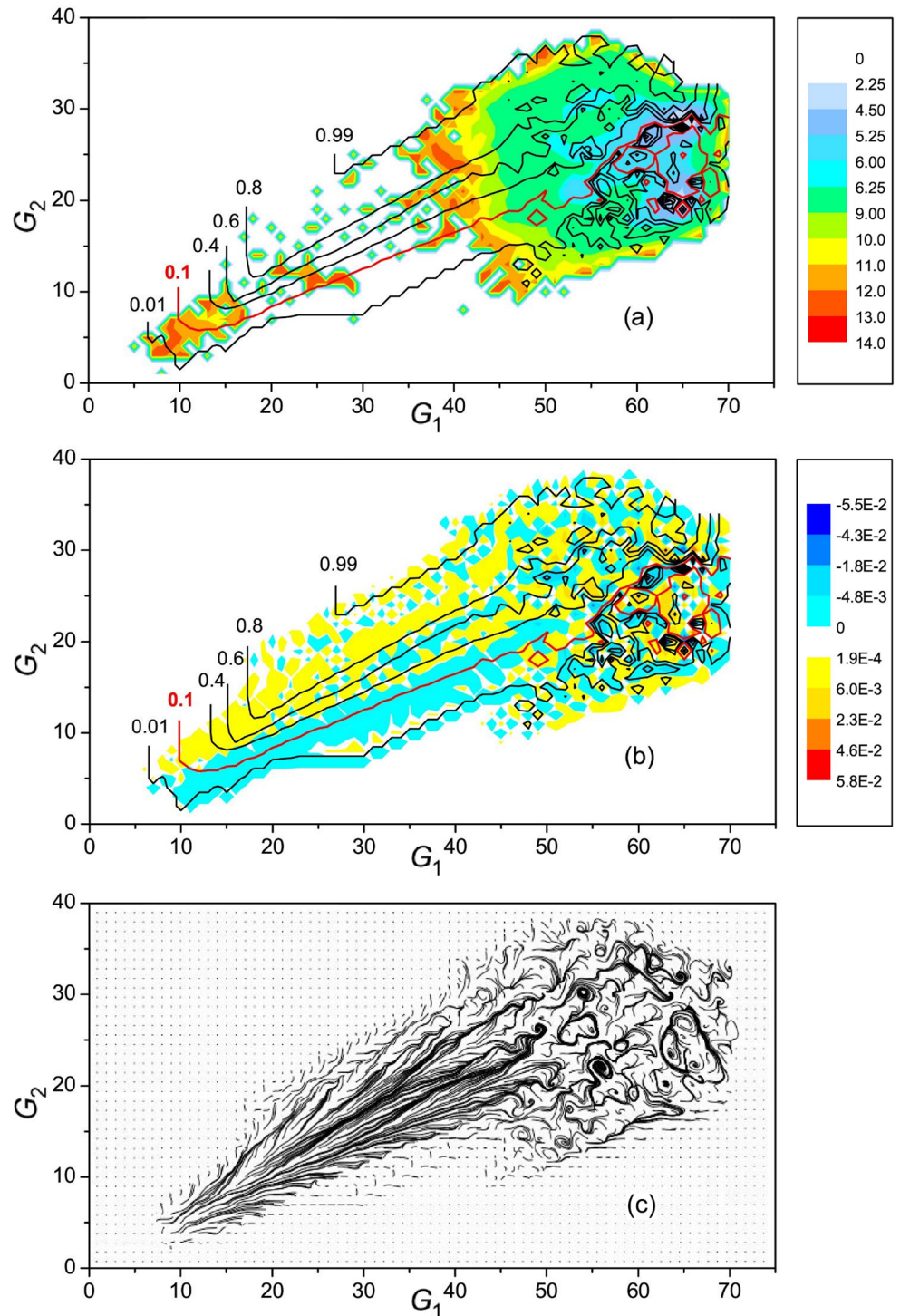
Fig 3. Two-dimensional flow field. Streamlines of the folding flows superimposed on (**a**) the free energy surface and (**b**) the vorticity distribution, and (**c**) folding trajectories in the form of passive tracers. The negative vorticity corresponds to a clockwise motion, and the positive vorticity to an anti-clockwise motion. Figures at the streamlines denote the fractions of the total folding flow restricted by the current streamlines. The lowest stream tube (up to 0.1 traction of the total flow) represents pathway II, and the other stream tubes correspond to pathway I. The color scale bars at panels (**a**) and (**b**) show, respectively, the levels of the free energy and vorticity.

https://doi.org/10.1371/journal.pone.0188659.g003

Boltzmann constant, and the vorticity was calculated as

$$\omega(\mathbf{G}) = \partial J_{G_2}/\partial G_1 - \partial J_{G_1}/\partial G_2 \qquad (5)$$

The streamlines, which divide the total folding flow from the unfolded to the native state into stream tubes, show that approximately 90% of the flow follow pathway I, in agreement with the previous MD simulation studies at $T = 300K$ [28, 31]. The flow in this pathway is well directed to the native state and filled with small vortices which do not effect the general directions of the flow. In contrast, the flow in pathway II, which accounts just for 10% of the total flow, is much more complex. In particular, it contains a set of relatively large opposite-directed vortices in the region adjacent to the native state. As the previous study has shown [32], the clockwise vortices surrounding the group of anti-clockwise vortices that is centered at $G_1 \approx 63$ and $G_2 \approx 22$ form a large clockwise vortex. It is created due to a repeated partial unfolding of native-like conformations to the conformations that have a partly unformed α-helix and broken alignment of the α- and PPII-helices, which is followed by the return of the protein to a native-like state. The smaller, opposite-directed vortices within this, larger vortex, correspond to less significant changes in the protein structure; here, the rearrangements are mostly restricted to a partial forming/unforming the α-helix. The present complexity of the folding flows in pathway II does not lead to a considerable deviation from two-state kinetics; the distribution of the first-passage times remains essentially single-exponential (Fig 4). We note that the appearance of vortices in the flow field is not surprising [21, 41] because the condition of stationary flow (in the present case, from the source to the sink) $\Delta \cdot \mathbf{J} = 0$ does not rule out the presence of a curl-component in $\mathbf{J}$ [42]. Such whirling flows are characterized by "irreversible circulation" or "cyclic balance", which determine the degree of deviation from detailed balance [43–45].
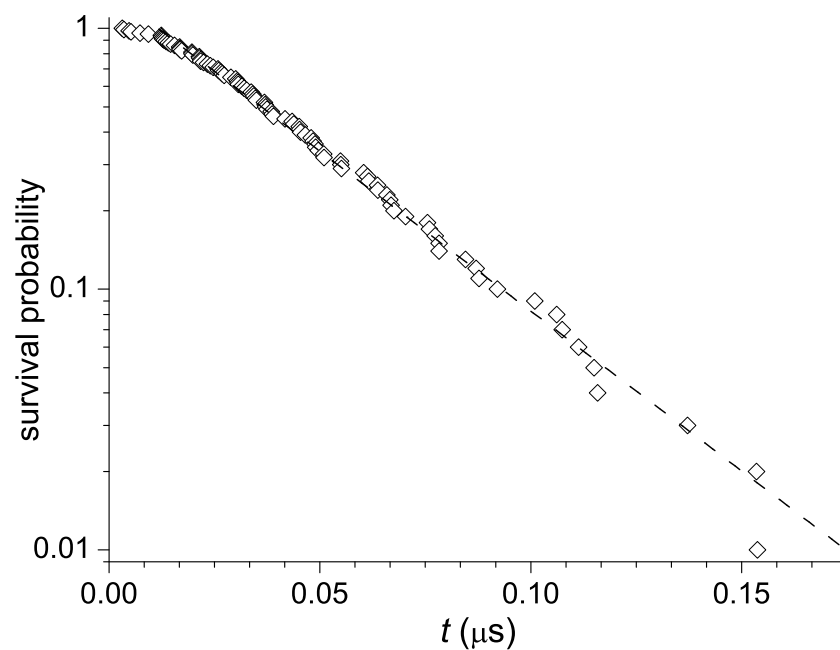
Fig 4. Cumulative distribution of the first-passage times. The labels correspond to the simulation results, and the dashed line is the best-fit exponential approximation with the waiting time $\approx 0.036\mu s$.

Results and discussion

In contrast to classical hydrodynamic turbulence, which considers an incompressible fluid [1–4], the folding fluid is highly "compressible" because the probability for the system to visit different points of conformation space, which plays a role of the density of the folding fluid, varies by several order of magnitude (see Eq (4) and Fig 3a). Therefore, to characterize the folding flow field, the probability fluxes are more suitable than the velocities [19]. This is related not only to turbulence phenomena but also to a general description of the folding flows. In particular, according to the Helmholtz decomposition theorem, a natural separation of the folding flow field into a curl-free and divergence-free vector fields is allowed, which results in a two-component potential of the driving force of protein folding [22].

We start with the study of the space distribution of folding fluxes depending on the scale of spatial coarse-graining. As has been previously shown for SH3 domain [19] and beta3s mini-protein [20], although the folding flow field is far from uniform (Fig 2), the distribution of folding flows possesses a well pronounced property of self-similarity. To see if the fluxes for Trp-cage are also self-similar, and to determine the self-similarity index, we calculated the function $G_k(L) = < |J_{g_k,L}|/\bar{j}_{g_k} >$, where $|J_{g_k,L}|$ is the absolute value of g_k component of the flow through the square of linear size L, $\bar{j}_{g_k} = (\sum_1^M j_{g_k,i}^2/M)^{1/2}$ is the average flux in g_k-direction through the elementary square, M is the number of elementary squares the square of size L covers, and the angular brackets denote the averaging over the g_k = const cross-sections of the $\mathbf{g} = (g_1, g_2, g_3)$ space. The linear size L is measured in units of the elementary square linear size equal to 1Å. The maximum value of L was chosen to be not larger than 5Å, because the flow field is very narrow in the g_2 direction; it varies from $\approx$ 5Å at small values of g_1 to $\approx$ 20Å at large g_1 values (Fig 2). Fig 5a–5c presents the results. In each panel, the values of $G_k(L)$ are shown for regions of conformation space that gradually shift from the unfolded to the native state along the g_1 coordinate. Specifically, the triangles-up correspond to $10 < g_1 \leq 30$, triangles-down to $30 < g_1 \leq 50$, and circles to $50 < g_1 \leq 70$. The lines show the corresponding best-fits of the data to the equation $G_k(L) \sim L^{D_k}$. It is seen that for all directions ($k = 1, 2, 3$), the flow space distributions are self-similar, and the values of D_k vary between approximately 0.7 and 1.4, i.e., the distributions are fractal [46]. Also, as the native state is approached, the fractal index decreases, indicating that the flow deviates from a uniform flow, for which $D = 2$, more and more. These results are in line with the previous studies of folding of SH3 domain [19] and beta3s miniprotein [17], where D_k decreased from $\approx$ 1.5 to $\approx$ 0.7 toward the native state.

Let us now turn to the structure functions. Specifically, we consider the conventional longitudinal functions [1–4, 23, 24], in which the increment of the flux between two points is projected on the line connecting these points. The second-order structure function is defined as

$$C_{ll}(l) = \langle \delta j_{||}(l)^2 \rangle \tag{6}$$

and the third-order function as

$$C_{lll}(l) = \langle \delta j_{||}(l)^3 \rangle \tag{7}$$

Here

$$\delta j_{||}(l) = [\mathbf{j}(\mathbf{g} + \mathbf{l}) - \mathbf{j}(\mathbf{g})] \cdot \mathbf{l}/l \tag{8}$$

where $\mathbf{l}$ is the increment in the $\mathbf{g}$ space, and the angular brackets denote ensemble averages. Fig 6a and 6b shows the calculated structure functions. It is seen that there is a range of space increments, approximately $30 < l < 55$, where the functions scale with l as the Kolmogorov (K41) theory for isotropic and homogeneous turbulence [23, 24] predicts for the inertial
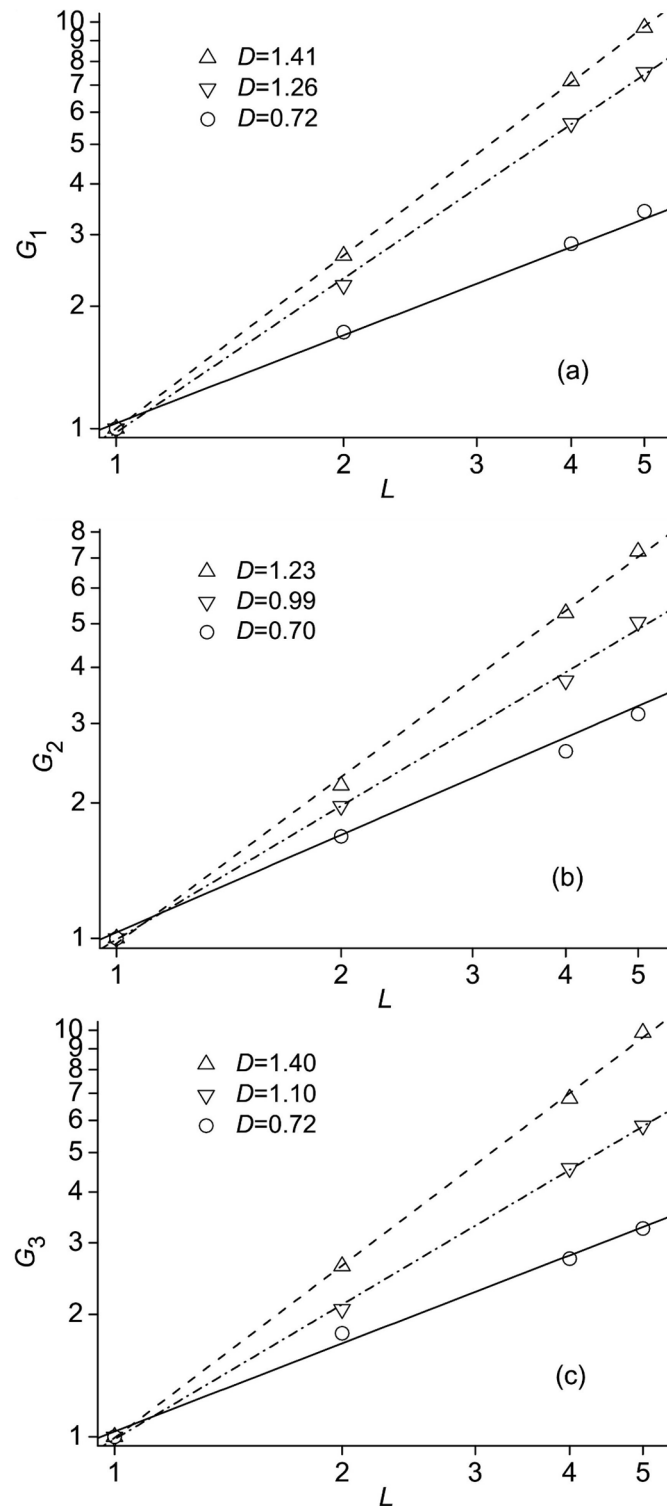
Fig 5. The functions $G_k(L)$ representing the k-components of the flow depending on the scale of coarse-graining L. Panels **(a)**, **(b)** and **(c)** are for $k = 1$, 2 and 3, respectively. Triangles-up correspond to the region of conformation space $10 < g_1 \leq 30$, triangles-down to $30 < g_1 \leq 50$, and circles to $50 < g_1 \leq 70$. The lines show the best-fits of the data to the equation $G_k(L) \sim L^{D_k}$.

https://doi.org/10.1371/journal.pone.0188659.g005
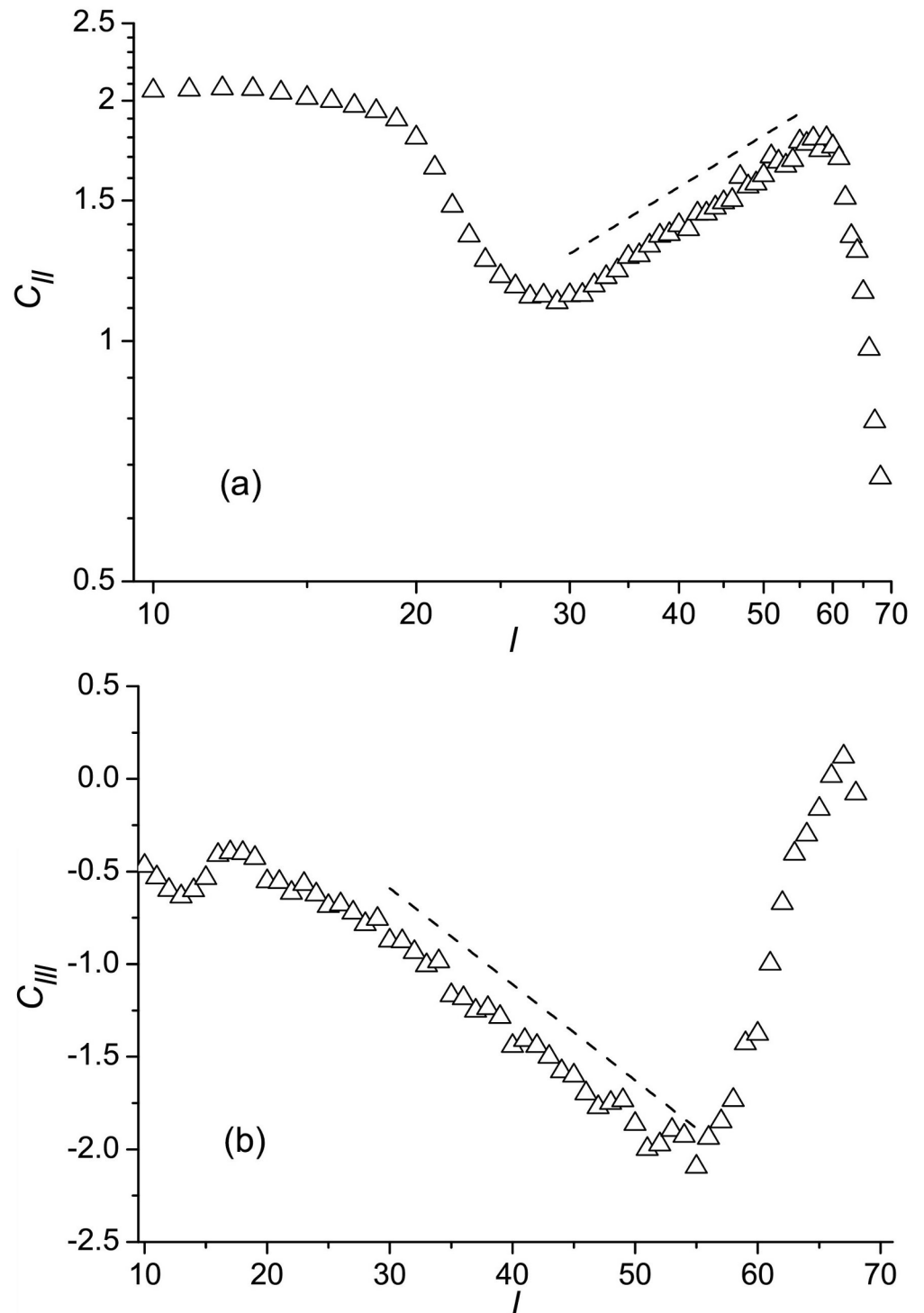
Fig 6. Structure functions of (a) the second and (b) third orders. The dished lines in panels (**a**) and (**b**) represent the functions $C_{II} \sim l^{2/3}$ and $C_{III} \sim l$, respectively.

https://doi.org/10.1371/journal.pone.0188659.g006

interval of scales, i.e., $C_{ll}(l) \sim l^{2/3}$ and $C_{lll}(l) \sim l$. The lower bound of this range is considerably larger than the characteristic distance on which the inter-residue contacts form and break (the nonbonding interaction cutoff is 7.5Å), and the upper bound is smaller than the length of the unfolded protein chain (≈ 70Å), which determines the overall size of the flow field (Fig 2). Therefore, similar to the inertial interval of scales in hydrodynamic turbulence, the only distance on which the flow increments essentially depend within the present range is the current space increment l.

According to the K41, the time-rate of change of the kinetic energy of fluctuations per unit mass $\epsilon_{hd} = de/dt$ is finite and constant in the inertial interval of scales, and thus it plays a role of the key parameter that determines the behavior of flow fluctuations on these scales. Then, the dimension analysis gives $\delta v \sim (\epsilon_{hd} l)^{1/3}$ [23, 24]. The kinetic energy of flow fluctuations per unit mass is $e = (\sum m v_i^2/2)/M$, where m is the molecular mass, v_i is the fluid velocity in i point of the space volume the fluid fills, and M is the mass of the fluid. It can be rewritten as $e = (\sum m v_i^2/2)/(mnV) = (\sum v_i^2/2)/(nV) = n^{-3}(\sum j_i^2/2)/V = n^{-3}\sigma^2/2$, where V is the space volume, n is the (numerical) density of the fluid, $j_i = nv_i$ is the fluid flux, and σ^2 is the variance of the fluxes per unit volume. This suggests that in the case of protein folding, or more generally, in the case of compressible fluid, the time-rate of change of the variance of fluxes per unit volume $\epsilon_{pf} = d\sigma_{pf}^2/dt$, where $\sigma_{pf}^2 = \langle [\mathbf{j}(\mathbf{g}) - \langle \mathbf{j}(\mathbf{g}) \rangle]^2/V \rangle$, plays a role of the key parameter, similar to ϵ_{hd} in hydrodynamic turbulence. Accordingly, the relation $\delta v \sim (\epsilon_{hd} l)^{1/3}$ transforms into $\delta j \sim (\epsilon_{pf} l)^{1/3}$, indicating that the flux distribution is self-similar with respect to the space increment. To perform its function, ϵ_{pf} should be constant. To see if this is true, we calculated the time-dependent variance of the fluxes

$$\Delta \sigma_{pf}^2(\Delta t) = \langle \sigma_{pf}^2(t + \Delta t) - \sigma_{pf}^2(t) \rangle_t \quad (9)$$

where

$$\sigma_{pf}^2(\tau) = \langle \{\mathbf{j}[\mathbf{g}(\tau)] - \langle \mathbf{j}[\mathbf{g}(\tau)] \rangle_{\mathbf{g}} \}^2 \rangle_{\mathbf{g}}/V(\tau) \quad (10)$$

is the variance of the fluxes per unit volume at time τ, Δt is the time increment, $\mathbf{j}[\mathbf{g}(\tau)]$ is the space distribution of fluxes at time τ, $V(\tau)$ is the volume of the $\mathbf{g}$-space the system occupies at time τ, and the angular brackets denote ensemble averages over time and conformation space, which are indicated, respectively, by indices t and $\mathbf{g}$ at the brackets. The calculations presented in Fig 7 show that for the dominant interval of times, where statistics are not too poor (specifically at $\Delta t < 0.11\mu$s, which covers $\approx 95\%$ of folding trajectories; see Fig 4), $\Delta \sigma_{pf}^2(\Delta t)$ changes with Δt essentially linearly. We thus find that the time-rate of change of σ_{pf}^2 is approximately constant in the course of Trp-cage folding, so that the quantity $\epsilon_{pf} = d\sigma_{pf}^2/dt$ can be considered as a key parameter for the folding process, similar to the time-rate of change of the kinetic energy per unit mass in hydrodynamic turbulence ϵ_{hd}. Accordingly, the structure functions $C_{ll}(l)$ [Eq (6)] and $C_{lll}(l)$ [Eq (7)] are written as

$$C_{ll}(l) \sim (\epsilon_{pf} l)^{2/3} \quad (11)$$

and

$$C_{lll}(l) \sim \epsilon_{pf} l \quad (12)$$

in agreement with their scaling in Fig 6. The Fourier transform of $C_{ll}(l)$ gives the "variance spectrum" $\Sigma_k \sim \epsilon_{pf}^{2/3} k^{-5/3}$, where k is the wave number, which is similar to the famous Kolmogorov spectrum $E_k \sim \epsilon_{hd}^{2/3} k^{-5/3}$ for the energy cascade in hydrodynamic turbulence [23, 24].
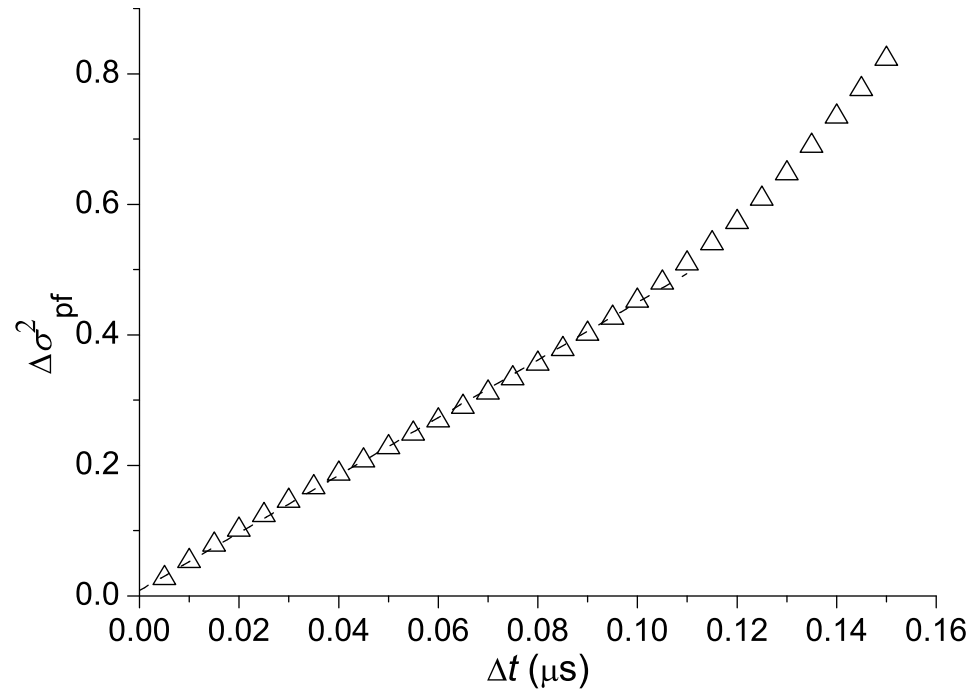
Fig 7. Time-dependent variance of the probability fluxes. The dashed line is the best-fit of the data for $\Delta t < 0.11\mu s$ to a linear equation. To have $\Delta \sigma_{pf}^2$ in the same value scale with the structure functions, the volume $V(\tau)$ in Eq (10) was taken as a fraction of the maximum volume ($\approx 12.2 \times 10^3 \text{ Å}^3$).

https://doi.org/10.1371/journal.pone.0188659.g007

The linear change of σ_{pf}^2 with time (Fig 7) suggests that the process of protein folding can be considered as Brownian diffusion in the space of folding fluxes $\mathbf{j}(\mathbf{g})$ against the drift flow $\langle \mathbf{j}[\mathbf{g}(t)] \rangle_{\mathbf{g}}$. The diffusion coefficient is determined as $(d\sigma_{pf}^2/dt)/6$ (e.g., [47]), or $\epsilon_{pf}/6$. Accordingly, the above discussed condition of the constant rate of change of the variance of folding fluxes, which underlies the observed flux scaling, can be restated in more general terms, i.e., as a requirement that the folding fluxes should represent Brownian diffusion with the diffusion coefficient equal to $\epsilon_{pf}/6$.

The third-order structure function $C_{III}(l)$ in Fig 6b is negative. In hydrodynamic turbulence, this corresponds to a direct (Richardson [5]) cascade of eddies, in which large-scale eddies generated by outer forces disintegrate into smaller eddies until the latter dissipate due to viscosity. In more general terms, the negative value of the $C_{III}(l)$ can be associated with the transition from a well-organized (large scale) motion to a stochastic (small scale) motion, as schematically illustrated in Fig 8. As can be seen from this figure, irrespective of whether the initial point is taken in the region of well-directed flow and the terminal point is chosen in the stochastic flow region, or vise versa, the "longitudinal" increment of the flow $\delta j_{||}(l)$ given by Eq (8) will be negative, and, thus the function $C_{III}(l)$ will also be negative [Eq (7)].

Both the function $G_k(L)$ (Fig 5a–5c) and the structure functions $C_{II}(l)$ and $C_{III}(l)$ (Fig 6a and 6b) reveal that the folding flows are self-similar. At the same time, their self-similarities are different in that the $G_k(L)$ displays a "transversal" self-similarity of the flow distributions, and the structure functions show a "longitudinal" self-similarity. It is thus of interest to see how, and if, the "transversal" and "longitudinal" self-similarities are consistent with each other. Since the flow through a region of size L scales with L as $J(L) \sim L^D$ (Fig 5), and the total volume V remains the same at different L, $\epsilon_{pf}(L) = d\sigma_{pf}^2(L)/dt \sim J(L)^2/T \sim L^{2D}/T^3$, where T stands for
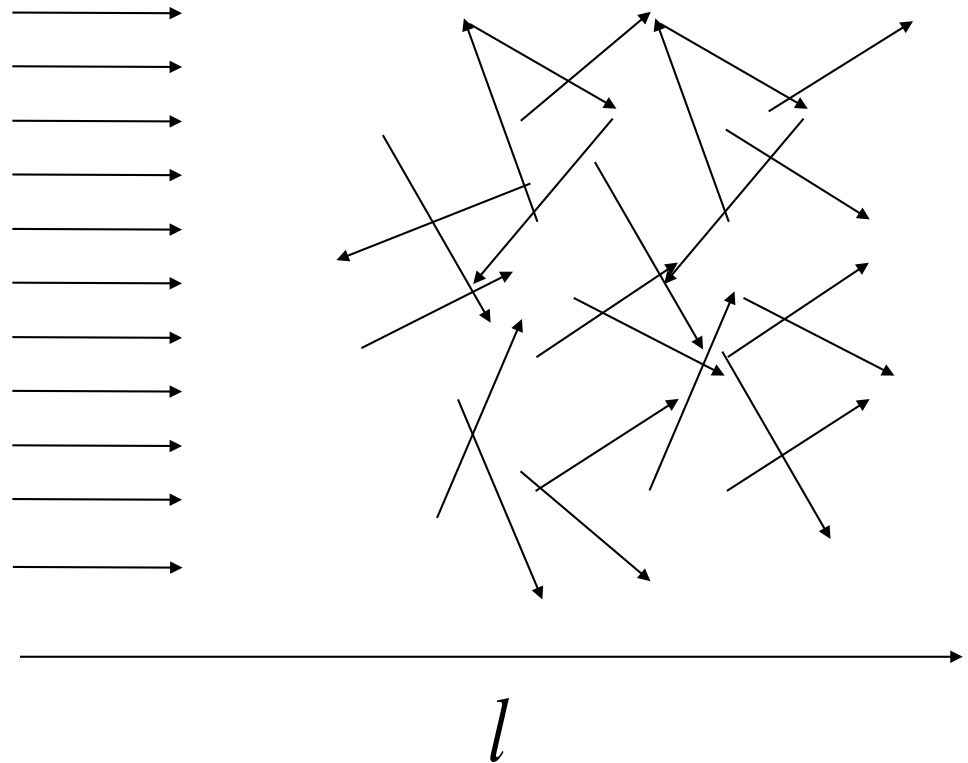
Fig 8. Schematic representation of the transition from a directed flow to a stochastic flow.
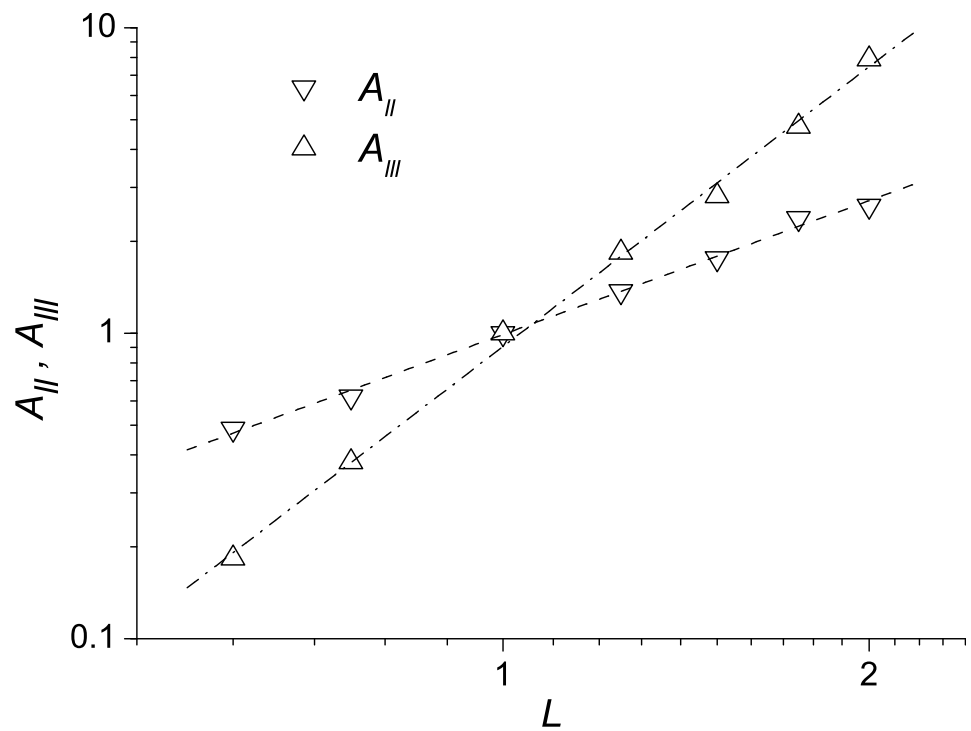
Fig 9. Prefactors A_{II} and A_{III} determining the structure function scaling with the coarse-graining length L. Triangles-down and triangles-up are for the structures of the second and third orders, respectively. The dashed and dash-dot lines represent the best-fits of A_{II} and A_{III} to the equations $A_{II} \sim L^{D_{II}}$ and $A_{III} \sim L^{D_{III}}$ ($D_{II} \approx 1.45$ and $D_{III} \approx 3.0$).

time. Then, according to Eqs ([11](#)) and ([12](#)), the second-order structure function should scale with L as $C_{II}(l, L) = A_{II}(L)C_{II}(l, L_0)$, where $A_{II}(L) \sim L^{2D}$, and the third-order structure function as $C_{III}(l, L) = A_{III}(L)C_{III}(l, L_0)$, where $A_{III}(L) \sim L^{3D}$ ($L_0 = 1$Å). The calculated relations ([Fig 9](#)) show that the exponents in these equations are $D \approx 0.73$ and $D \approx 1$ for the second- and third-order structure functions, respectively, which are within the range of variation of the fractal index in [Fig 5a–5c](#) ($D = 0.7$–1.4). Also, these values of D correspond better to the region adjacent to the native state, where folding flow is more turbulent.

Conclusions

Turbulent behavior of protein folding flows was first observed when folding of a SH3 domain was studied [19]. Most surprising was that the folding fluxes in the space of collective variables scaled with the space increment similar to the fluid velocities in the Kolmogorov (K41) theory of isotropic and homogeneous turbulence [23, 24]. In the present paper, to see whether such similarity between folding flows and turbulent fluid flows is specific of the SH3 domain or may be common to proteins, we consider another benchmark system—Trp-cage miniprotein. We have studied its folding in detail recently [32] and found the results in good general agreement with the previous works [25–31]. The Trp-cage miniprotein differs from the SH3 domain essentially, both in the structure and mechanism of folding. In particular, kinetics of Trp-cage folding are single-exponential, while for SH3 domain we had double-exponential kinetics, and turbulence was observed only for slow folding trajectories [19]. Further, the approaches to simulate and characterize the folding process are different. The simulations of Trp-cage folding are performed using an all-atom model (CHARMM program [33]), while for the SH3 domain a coarse-grained representation of the protein was used in the form of C_α-model [19]. Also, in the present case, the collective variables are determined with a PCA method, whereas in the case of SH3 domain they were represented by weakly dependent groups of native contacts [19]. Despite such a considerable difference between the SH3 domain and Trp-cage miniprotein cases, we have found that the structure functions of the second and third orders for the Trp-cage folding follow the Kolmogorov scaling similar to what was observed for the SH3 domain, i.e., $C_{II}(l) \sim l^{2/3}$ and $C_{III}(l) \sim l$, where l is the increment in the space of collective variables. In contrast to classical hydrodynamic turbulence, which considers an incompressible fluid, and thus uses fluid velocities to characterize the flow, we employ flow fluxes because folding fluid is very compressible. In this characterization, the variance of folding fluxes per unit volume $\sigma_{pf}^2(\mathbf{g})$, where $\mathbf{g}$ is the point in the three dimensional space of collective variables, plays a role of the kinetic energy of fluctuation per unit mass in hydrodynamic turbulence. The calculation of σ_{pf}^2 as a function of time has shown that it varies with time essentially linearly, so that the quantity $\epsilon_{pf} = d\sigma_{pf}^2/dt$ represents the key parameter to characterize the folding flows, similar to the time-rate of change of the kinetic energy per unit mass in hydrodynamic turbulence. In more general terms, the process of protein folding in the space of probability fluxes represents Brownian diffusion (against the drift flow) with the diffusion coefficient equal to $\epsilon_{pf}/6$. The analysis of the probability flux distribution scaling with the size of coarse-graining of the conformational space has also shown that the distributions are self-similar with a fractal dimension, and the fractal index decreases toward the native state, indicating that the flow becomes more turbulent as the native state is approached.

The obtained results, first, show that the very complex dynamics of protein folding allows a simple characterization in terms of scaling and diffusion of probability fluxes, and, secondly, they suggest that the turbulence phenomena similar to hydrodynamic turbulence are not specific of folding of a particular protein but are common to protein folding.

Author Contributions

Conceptualization: Sergei F. Chekmarev.

Data curation: Vladimir A. Andryushchenko.

Formal analysis: Vladimir A. Andryushchenko, Sergei F. Chekmarev.

Funding acquisition: Sergei F. Chekmarev.

Investigation: Vladimir A. Andryushchenko.

Methodology: Vladimir A. Andryushchenko, Sergei F. Chekmarev.

Project administration: Sergei F. Chekmarev.

Software: Vladimir A. Andryushchenko.

Validation: Vladimir A. Andryushchenko.

Visualization: Vladimir A. Andryushchenko.

Writing – original draft: Sergei F. Chekmarev.

Writing – review & editing: Sergei F. Chekmarev.

References

1. Monin AS, Yaglom AM. Statistical fluid mechanics. Pt.2. Cambridge: MIT Press; 1975.

2. Landau LD, Lifshitz EM. Fluid mechanics. New York: Pergamon; 1987.

3. Frisch U. Turbulence: The Legacy of A. N. Kolmogorov. Cambridge University Press; 1995.

4. Lesieur M. Turbulence in fluids. Fluid mechanics and its applications. Dordrecht: Springer; 2008.

5. Richardson LF. Weather prediction by numerical process. Cambridge University Press; 1922.

6. Dinner AR, Šali A, Smith LJ, Dobson CM, Karplus M. Understanding protein folding via free-energy surfaces from theory and experiment. Trends Biochem Sci. 2000 Jul; 25(7):331–339. https://doi.org/10.1016/S0968-0004(00)01610-8 PMID: 10871884

7. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem. 1997 Oct; 48(1):545–600. https://doi.org/10.1146/annurev.physchem.48.1.545 PMID: 9348663

8. Dobson CM, Šali A, Karplus M. Protein folding: a perspective from theory and experiment. Angew Chem Int Ed. 1998 Apr; 37(7):868–893. https://doi.org/10.1002/(SICI)1521-3773(19980420)37:7%3C868::AID-ANIE868%3E3.0.CO;2-H

9. Shea J-E, 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 Oct; 53(1):499–535. https://doi.org/10.1146/annurev.physchem.52.1.499

10. Lindorff-Larsen K, Røgen P, Paci E, Vendruscolo M, Dobson CM. Protein folding and the organization of the protein topology universe. Trends Biochem Sci. 2005 Jan; 30(1):13–19. https://doi.org/10.1016/j.tibs.2004.11.008 PMID: 15653321

11. Dill KA, Ozkan SB, Shell MS, Weikl TR. The protein folding problem. Annu Rev Biophys. 2008 Jun; 37:289–316. https://doi.org/10.1146/annurev.biophys.37.092707.153558 PMID: 18573083

12. Zakharov VE, L'vov VS, Falkovich GE. Kolmogorov spectra of turbulence I. Wave turbulence. Berlin: Springer-Verlag; 1992.

13. Ghashghaie S, Breymann W, Peinke J, Talkner P, Dodge Y. Turbulent cascades in foreign exchange markets. Nature. 1996 Jun; 381(6585):767–770. https://doi.org/10.1038/381767a0

14. Nemirovskii SKQuantum turbulence: Theoretical and numerical problems. Phys Rep. 2013 Mar; 524(3):85–202. https://doi.org/10.1016/j.physrep.2012.10.005

15. Barenghi CF, Sergeev YA, Baggaley AW. Regimes of turbulence without an energy cascade. Sci Rep. 2016 Oct; 6:35701. https://doi.org/10.1038/srep35701 PMID: 27761005

16. Chekmarev SF, Palyanov AYu, Karplus M. Hydrodynamic description of protein folding. Phys Rev Lett. 2008 Jan; 100(1):018107. https://doi.org/10.1103/PhysRevLett.100.018107 PMID: 18232827

17. 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 Apr; 118(16):4287–4299. https://doi.org/10.1021/jp412729r PMID: 24669953

18. Kalgin IV, Karplus M, Chekmarev SF. Folding of a SH3 domain: Standard and "hydrodynamic" analyses. J Phys Chem B. 2009 Sep; 113(38):12759–12772. https://doi.org/10.1021/jp903325z PMID: 19711956

19. Kalgin IV, Chekmarev SF. Turbulent phenomena in protein folding. Phys Rev E. 2011 Jan; 83(1): 011920. https://doi.org/10.1103/PhysRevE.83.011920

20. Kalgin IV, Caflisch 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 Apr; 117(20):6092–6105. https://doi.org/10.1021/jp401742y PMID: 23621790

21. Kalgin IV, Chekmarev SF. Folding of a β-Sheet Miniprotein: Probability Fluxes, Streamlines, and the Potential for the Driving Force. J Phys Chem B. 2015 Jan; 119(4):1380–1387. https://doi.org/10.1021/jp5112795 PMID: 25544646

22. Chekmarev SF. Protein folding: Complex potential for the driving force in a two-dimensional space of collective variables. J Chem Phys. 2013 Oct; 139(14):145103. https://doi.org/10.1063/1.4824133 PMID: 24116649

23. Kolmogorov AN. The local structure of turbulence in incompressible viscous fluids at very large Reynolds numbers. Dokl Akad Nauk SSSR. 1941; 30:299–303. Reprinted in Proc R Soc Lond A. 1991 Jul;434:9–13.

24. Kolmogorov AN. Dissipation of energy in locally isotropic turbulence. Dokl Akad Nauk SSSR. 1941; 32:19–21. Reprinted in Proc R Soc Lond A. 1991 Jul;434:15–17.

25. Neidigh JW, Fesinmeyer RM, Andersen NH. Designing a 20-residue protein. Nat Struct Mol Biol. 2002 Apr; 9(6):425–430. https://doi.org/10.1038/nsb798

26. Qiu L, Pabit SA, Roitberg AE, Hagen SJ. Smaller and faster: The 20-residue Trp-cage protein folds in 4 μs. J Am Chem Soc. 2002 Nov; 124(44):12952–12953. https://doi.org/10.1021/ja0279141 PMID: 12405814

27. Zhou R. Trp-cage: folding free energy landscape in explicit water. Proc Natl Acad Sci USA. 2003 Nov; 100(23):13280–13285. https://doi.org/10.1073/pnas.2233312100 PMID: 14581616

28. Chowdhury S, Lee MC, Duan Y. Characterizing the rate-limiting step of Trp-cage folding by all-atom molecular dynamics simulations. J Phys Chem B. 2004 Sep; 108(36):13855–13865. https://doi.org/10.1021/jp0478920

29. Juraszek J, Bolhuis P. Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. Proc Natl Acad Sci USA. 2006 Oct; 103(43):15859–15864. https://doi.org/10.1073/pnas.0606692103 PMID: 17035504

30. Zheng W, Gallicchio E, Deng N, Andrec M, Levy RM. Kinetic network study of the diversity and temperature dependence of Trp-cage folding pathways: Combining transition path theory with stochastic simulations. J Phys Chem B. 2011 Feb; 115(6):1512–1523. https://doi.org/10.1021/jp1089596 PMID: 21254767

31. Han W, Schulten K. Characterization of folding mechanisms of Trp-cage and WW-domain by network analysis of simulations with a hybrid-resolution model. J Phys Chem B. 2013 Oct; 117(42): 13367–13377. https://doi.org/10.1021/jp404331d PMID: 23915394

32. Andryushchenko VA, Chekmarev SF. A hydrodynamic view of the first-passage folding of Trp-cage miniprotein. Eur Biophys J. 2016 Apr; 45(3):229–243. https://doi.org/10.1007/s00249-015-1089-7 PMID: 26559408

33. Brooks BR, Brooks CL III, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. J Comput Chem. 2009 Jul; 30(10):1545–1614. https://doi.org/10.1002/jcc.21287 PMID: 19444816

34. Jolliffe I. Principal component analysis. New York: Springer Verlag; 2002.

35. Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular systems. J Chem Phys. 1996 Aug; 105(5):1902–1921. https://doi.org/10.1063/1.472061

36. Ferrara P, Apostolakis J, Caflisch A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. Proteins: Struct Funct Bioinf. 2002 Jan; 46(1):24–33. https://doi.org/10.1002/prot.10001

37. Ryckaert J-P, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977 Mar; 23(3):327–341. https://doi.org/10.1016/0021-9991(77)90098-5

38. Ferrara P, Apostolakis J, Caflisch A. Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. J Phys Chem B. 2000 May; 104(20):5000–5010. https://doi.org/10.1021/jp994157t

39. Eaton WA, Muñoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, et al. Fast kinetics and mechanisms in protein folding. Annu Rev Biophys Biomol Struct. 2000 Jun; 29(1):327–359. https://doi.org/10.1146/annurev.biophys.29.1.327 PMID: 10940252

40. Darmofal DL, Haimes R. An analysis of 3D particle path integration algorithms. J Comput Phys. 1996 Jan; 123(1):182–195. https://doi.org/10.1006/jcph.1996.0015

41. Chekmarev SF. Protein folding as a complex reaction: a two-component potential for the driving force of folding and its variation with folding scenario. PLoS ONE. 2015 Apr; 10(4):e0121640. https://doi.org/10.1371/journal.pone.0121640 PMID: 25848943

42. Van Kampen NG. Stochastic processes in physics and chemistry. Vol. 1. Amsterdam: North-Holland; 1992.

43. Tomita K, Tomita H. Irreversible circulation of fluctuation. Prog Theor Phys. 1974 Dec; 51(6): 1731–1749. https://doi.org/10.1143/PTP.51.1731

44. Graham R. Covariant formulation of non-equilibrium statistical thermodynamics. Z Phys B. 1977 Dec; 26(4):397–405. https://doi.org/10.1007/BF01570750

45. Eyink GL, Lebowitz JL, Spohn H. Hydrodynamics and fluctuations outside of local equilibrium: driven diffusive systems. J Stat Phys. 1996 May; 83(3–4):385–472. https://doi.org/10.1007/BF02183738

46. Moon FC. Chaotic and fractal dynamics. New York: Wiley; 1992.

47. Gardiner CW. Handbook of stochastic methods: For physics, chemistry and the natural sciences. Berlin: Springer; 1983.