SCIENTIFIC REP⚙RTS

Intrinsic Localized Modes in Proteins

Adrien Nicolaï[1,2], Patrice Delarue[2] & Patrick Senet[2]

Protein dynamics is essential for proteins to function. Here we predicted the existence of rare, large nonlinear excitations, termed intrinsic localized modes (ILMs), of the main chain of proteins based on all-atom molecular dynamics simulations of two fast-folder proteins and of a rigid α/β protein at 300 K and at 380 K in solution. These nonlinear excitations arise from the anharmonicity of the protein dynamics. The ILMs were detected by computing the Shannon entropy of the protein main-chain fluctuations. In the non-native state (significantly explored at 380 K), the probability of their excitation was increased by a factor between 9 and 28 for the fast-folder proteins and by a factor 2 for the rigid protein. This enhancement in the non-native state was due to glycine, as demonstrated by simulations in which glycine was mutated to alanine. These ILMs might play a functional role in the flexible regions of proteins and in proteins in a non-native state (i.e. misfolded or unfolded states).

Intrinsic localized modes (ILMs)[1], are members of the large soliton family[2]. They have been predicted and observed in crystals and anti-ferromagnetic materials[3,4] and arise from the anharmonicity of interatomic potentials and from the discreteness of matter at the atomic scale. We expect proteins in solution (water) to subtend ILMs, that is, strongly *localized* waves, because of their well-known anharmonicity[5–11]. Despite the importance of protein dynamics for biological function[12–16], the actual occurence of ILMs in proteins remains to be demonstrated both theoretically and experimentally.

Referring to the anharmonicity of hydrogen bonds between the amide N-H bonds and the carbonyl C = O groups of the protein backbone, Davydov proposed the existence of localized waves in α-helices several decades ago[17]. Using a one-dimensional quantum Hamiltonian, Davydov predicted the self-localization of C = O bond vibrations through their anharmonic coupling with the low-frequency modes of the polypeptide chain[17]. According to his proposal, this localized wave would provide a mechanism to propagate energy within a protein[18]. Experimental evidence of Davydov's solitons in proteins continues to remain elusive[19–22]. The one-dimensional Davydov model is a crude approximation of the structure of a protein. Several authors have attempted to simulate solitons in proteins, including their three-dimensional features, using simplified (coarse-grained) classical[23–25] or quantum dynamic models[26]. These models are very difficult to relate to actual protein dynamics. Protein dynamics span several time-scales in which local displacements of atoms are coupled to more large-scale conformational motions; a full description of these dynamics requires, therefore, an atomistic approach[16].

In recent years, all-atom molecular dynamics (MD) simulations have become a powerful tool that is complementary to experiments to investigate the dynamics of proteins in solution at the atomic scale[27–29]. Here, we used all-atom MD simulations, including realistic interactions between water and amino acids and the full effects of temperature, to theoretically establish the lifetime and statistics of ILMs in proteins.

We predict a new type of fully classical ILMs in proteins: solitons localized in both time and space (similarly to the Peregrine solitons[30,31]). These intermittent ILMs are due to the anharmonicity of the potential energy surface describing the torsional degrees of freedom of the main chain of proteins. The torsional degrees of freedom play an important role in protein dynamics because they govern the low-frequency functional modes of proteins[32–34]. The main-chain torsional angles γ are built from four C^α atoms of consecutive residues in the amino acid sequence[35,36] (Fig. 1). Because the length of the C^α-C^α virtual bond between two consecutive residues is nearly constant, the main-chain conformation is entirely described by the main-chain torsional angles γ and the main-chain bond angles θ (Fig. 1). These coarse-grained angles (CGA) (γ, θ) are part of coarse-grained protein models[37,38] and are used to analyze large conformational changes of proteins and protein folding in all-atom MD simulations[36,39].

[1]Department of Physics, Applied Physics and Astronomy, Rensselaer Polytechnic Institute, Troy, 12180 New York, United States. [2]Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Univ. Bourgogne Franche-Comté, 9 Av. A. Savary, BP 47 870, F-21078 Dijon Cedex, France. Correspondence and requests for materials should be addressed to P.S. (email: psenet@u-bourgogne.fr)
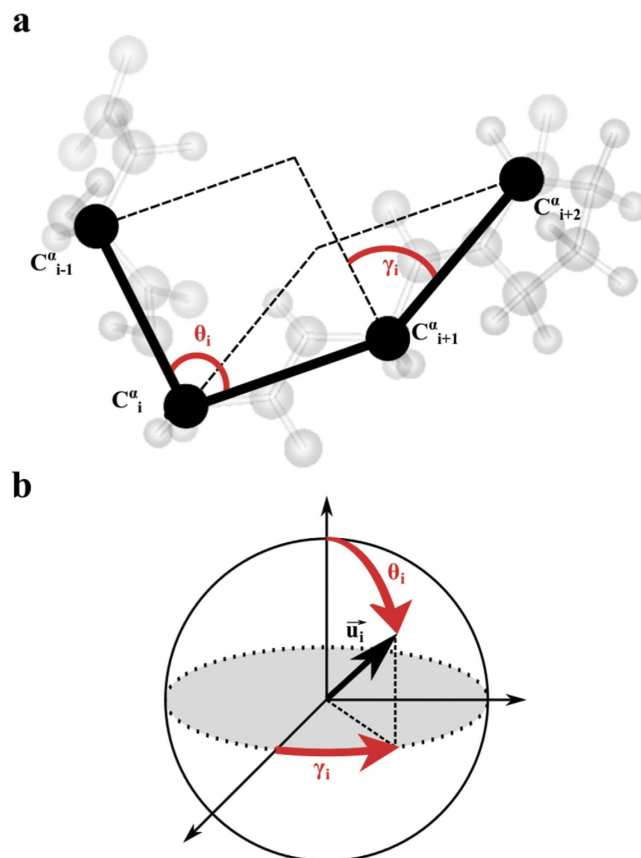
Figure 1. Coarse-grained angles (CGA) and their vector representation. (a) For a residue i, γ_i is the dihedral angle formed by the virtual bonds joining four successive C$^\alpha$ atoms ($i-1$, i, $i+1$ and $i+2$) along the amino acid sequence and θ_i is the bond angle formed by the virtual bonds joining three successive C$^\alpha$ atoms ($i-1$, i and $i+1$) along the amino acid sequence. The first pair of CGA (γ, θ) along the sequence is $(\gamma, \theta)_2$ and the last one is CGA $(\gamma, \theta)_{N-2}$, where N is the total number of residues. The convention for γ angles is the following: each γ angle varies between $-180°$ and $+180°$ with $\gamma = 0°$ being chosen when C$^\alpha_{i-1}$ is cis to C$^\alpha_{i+2}$, and the clockwise rotation of C$^\alpha_{i+1}$-C$^\alpha_{i+2}$ is positive when looking from C$^\alpha_i$ to C$^\alpha_{i+1}$. **(b)** Representation of the CGA pair (γ_i, θ_i) by a unit vector $\mathbf{u_i}$ in spherical coordinates, where γ_i is the azimuth angle, and θ_i is the polar angle.

The origin of ILMs found in proteins can be better understood by drawing an analogy between the fluctuations of the protein main chain and those of a simple mechanical system known to substend solitons. The mechanical analog of the protein main chain is a chain of rigid pendulums coupled by harmonic torsional springs and rotating around the same axis[2] [Fig. 2(a)]. At each time, the position of each pendulum i is defined by its angle α_i relative to the vertical. In this mechanical model, anharmonicity arises from the gravitational restoring force, which is proportional to $\sin(\alpha_i)$. This mechanical system may subtend ILMs (sine-Gordon solitons)[2], which interpolate between two rest states (no restoring force) of the system ($\alpha_i = 0$ for all i and $\alpha_i = 2\pi$ for all i). The rest state $\alpha_i = 0$ for all i and the ILM solution of the dynamic equations[2], computed at the maximum of its amplitude ($\alpha = \pi$) for a localization chosen at $i = 10$, are shown in Fig. 2(a). The localized wave is characterized by a well-defined profile of the angle α [Fig. 2(b)] named *kink*. Due to the symmetry of the system, an anti-*kink* solution exists. The combination of the *kink* and anti-*kink* solutions may lead to a localized solution [chosen at $i = 10$ in Fig. 2(b)], that does not propage but does oscillate as a function of time, a so-called sine-Gordon discrete breather[2]. The position of each pendulum can be represented alternatively as a rotating unit vector in a plane: $\mathbf{u_i} = (\cos(\alpha_i), \sin(\alpha_i))$ [Fig. 2(c)]. For each pendulum i, we define the difference between the vector $\mathbf{u_i}$ between the uniform state of the chain [red arrows in Fig. 2(c)] and in its excited state [blue arrows in Fig. 2(c)] by $\Delta\mathbf{u_i}$. The kink [panel (b)] leads to a peak in Δu_i^2 as a function of the pendulum position i [panel (d)] (in the example shown in Fig. 2, the sine-Gordon soliton is centered at $i = 10$). Consequently, the time-dependent fluctuations of the pendulums can be viewed as an ensemble of vectors rotating in a plane [Fig. 2(c)]. Their displacements between different times allow the definition of the localized character of the excitation [Fig. 2(d)].

As for the motion of each pendulum, the dynamics of each torsional angle γ_i of the protein main chain can be represented by a unit vector $\mathbf{u_i} = (\cos(\gamma_i), \sin(\gamma_i))$ rotating in a plane[36,40]. For small fluctuations around their equilibrium orientation, the vectors are harmonically coupled, as are the pendulums in the mechanical model. However, for large angular displacements, anharmonicity arises due to the nonlinearity of the dihedral potential
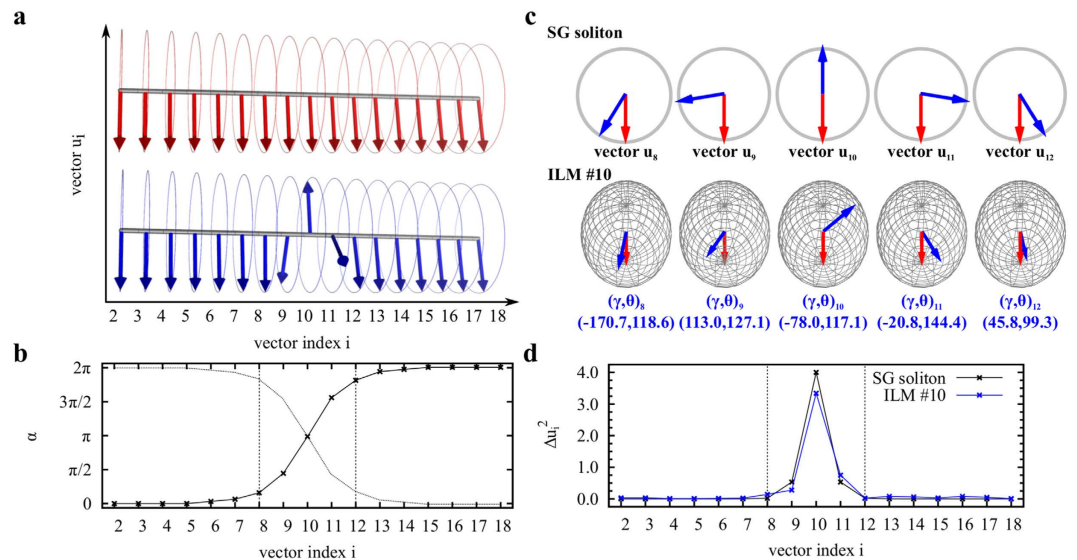
Figure 2. Analogy between solitons of a mechanical model and ILMs computed in proteins. (**a**) System of 17 coupled pendulums in its initial configuration (red) and in its excited state (blue). (**b**) Values of the angle α_i of each pendulum *i* relative to the vertical in the *kink* excited state [blue pendulums in panel (**a**)]. For comparison an anti-*kink* is shown using dashed lines. The dotted lines indicate the region of maximum localization. (**c**) (top) Vectors $\mathbf{u}_i = (\cos(\alpha_i), \sin(\alpha_i))$ computed for the initial (red) and excited (blue) states of panel (**a**) for the region of maximum localization. (bottom) Vectors $\mathbf{u}_i = (\cos(\gamma_i)\sin(\theta_i), \sin(\gamma_i)\sin(\theta_i), \cos(\theta_i))$ extracted from an MD simulation of Trp-cage for an initial state (red) and excited state (blue). For more clarity, each initial vector (in red) has been positioned in the same configuration for all spheres by rotation, and the exact values of the initial state (γ_i, θ_i) are given. (**d**) Amplitude Δu_i^2 of a sine-Gordon soliton (black line) and the ILM extracted from MD simulations (blue line).

energy surface. As for the pendulums, the orientation of each vector $\mathbf{u_i}$ corresponding at $\gamma_i = 0$ and $\gamma_i = 2\pi$ are equivalent, and solitons similar to discrete breathers of sine-Gordon type are expected in proteins by analogy.

More generally, the fluctuations of each pair of CGAs in the protein can be represented by a unit vector:

$$\mathbf{u}_i = (\cos(\gamma_i)\sin(\theta_i), \ \sin(\gamma_i)\sin(\theta_i), \ \cos(\theta_i)),$$

with one end fixed and the other describing a stochastic path along the surface of a sphere [Fig. 1(b)]. An example of the vectors $\mathbf{u_i}$ built on (γ_i, θ_i) (Fig. 1) computed from an all-atom MD trajectory of Trp-cage protein[41] (see the Results section) are shown for a uniform state (red arrows) and an excited state (blue arrows) occuring 1 ps later at $i = 10$ [Fig. 2(c)]. The localized character of this particular excitation can be seen in Fig. 2(d), where the Δu_i^2 values are compared for the pendulum model and the protein. As seen in Fig. 2(c), the vector displacements are primarily due to the fluctuations of the γ torsional angles. As shown in the Results section, the typical ILM shown in Fig. 2(d) is a rare event in MD simulations of proteins and is localized in both time and space.

In the present work, we present evidence of ILMs of the soliton type [as in Fig. 2(d)] in the spontaneous, unbiased fluctuations of the main chain of model proteins at different temperatures using all-atom MD in explicit solvent (water). We predict the existence, statistics, and biophysical properties of ILMs in proteins and their relation with the protein free-energy landscape. The particular questions we address are as follows: what is the probability of finding an ILM occurring spontaneously in the native (folded) and non-native (misfolded or unfolded) states of a protein in solution? How do the ILMs depend on the secondary structures of the protein and on its chemical composition?

Results

Evidence of ILMs in unbiased MD simulations. As the loss of rigidity due to the unfolding of a protein increases the anharmonicity of its free-energy landscape, we investigated the dynamics of two ultrafast-folder proteins, Trp-cage[41,42] and the chicken villin headpiece fragment HP-36[43,44], above their folding temperature. These proteins were chosen because they have been extensively studied using MD simulations and experiments and because unfolding events can be reproduced by unbiased all-atom MD simulations in explicit solvent within a reasonable computational time[45–48]. Trp-cage is a 20-residue protein designed to aid in understanding protein folding mechanisms consisting of one α-helix and one (3/10)-helix[41]. HP-36 is a 36-residue protein corresponding to the C terminus of the 76-residue chicken villin headpiece domain[43,44]. It consists of three α-helices. Because of their small size and fast kinetics, Trp-cage and HP-36 have become typical model proteins for MD simulations of protein folding[45–48]. For comparison, we analyzed also the dynamics of a rigid 46-residue α/β model protein (VA3)[40]. Because of the presence of three disulfide bonds (namely 3–40, 4–32, and 16–26), VA3 remains folded in all MD trajectories[40] while exploring the non-native state at 380 K.

	$T = 380\,\mathrm{K}$					$T = 300\,\mathrm{K}$	
	run 1	run 2	run 3	mutant$^+$	mutant*	run 1	run 2
N_{ILM}	251	302	199	321	58	71	17
N_{ILM}/N_t (in %)	0.0502	0.0604	0.0398	0.0642	0.0116	0.0142	0.085
$i = 9$ (LYS-ASP-**GLY***-**GLY***)	95	92	42	167	2	—	—
$i = 10$ (ASP-GLY*-**GLY***-PRO)	93	110	59	113	7	2	16
$i = 14$ (SER-SER-**GLY**$^{+,*}$-ARG)	33	63	41	2	7	—	—
Solitons $i = 9$, 10 and 14	221	265	142	282	16	2	16
$i = 18$ (PRO-PRO-PRO-SER)	24	29	46	25	26	62	—

Table 1. Number N_{ILM} of ILMs detected in MD simulations of Trp-cage at $T = 380\,\mathrm{K}$ and $300\,\mathrm{K}$. Localization along the amino acid sequence and statistics of the most frequent ILMs ($i = 9$, 10, 14 and 18) are given. For each mutant, each residue replaced by ALA is marked by a symbol. All MD runs were of a 500 ns duration except run 2, at $T = 300\,\mathrm{K}$ (20 ns, see the Methods section).

Three all-atom MD trajectories with different initial conditions (run 1, run 2 and run 3) at $T = 380\,\mathrm{K}$ and one MD trajectory at $T = 300\,\mathrm{K}$ each of a duration of 500 ns were conducted for Trp-cage, and two all-atom MD trajectories with different initial conditions (run 1 and run 2) at $T = 380\,\mathrm{K}$ and one MD trajectory a $T = 300\,\mathrm{K}$ each of a duration of 500 ns were conducted for HP-36 in explicit water (see the Methods section). In addition, one all-atom MD trajectory at 300 K and one at 380 K each of a duration of 500 ns were performed for VA3 (see the Methods section). The coordinates of the proteins were recorded every ps. Each MD run, therefore, represents 500,001 snapshots, from which the vector $\mathbf{u}_i$ associated with each pair (γ_i, θ_i) (Fig. 1) was computed. The fluctuations of the protein main chain between two consecutive snapshots were represented by the sequence of the displacements $\Delta\mathbf{u}_i(t) = \mathbf{u}_i(t) - \mathbf{u}_i(t-1)$ for all $i = 2$ to $N - 2$ and $t \leq 1$. The degree of localization of these fluctuations was measured by the normalized Shannon entropy S computed from the square displacements $\Delta u_i(t)^2$ along the sequence (see the Methods section). An excitation localized on a single pair of CGAs corresponds to $S = 0$ (minimum entropy, strongly localized fluctuations), and an excitation uniformly distributed on all CGAs corresponds to $S = 1$ (maximum entropy, delocalized fluctuations). The calculation of $S(t)$ is a systematic means to detect rare large localized excitations in MD trajectories. The ILMs are defined here by the excitations for which $S \leq 0.5$ (see the Methods section for the choice of this cutoff value). For example, the ILM shown in Fig. 2(d), which is the excitation that has the largest amplitude $\Delta u_{10}^2 = 3.34$ in the MD run 1 ($T = 380\,\mathrm{K}$) of Trp-cage, had a value of $S = 0.47$. The results for the Trp-cage protein are discussed next, and similar results for additional MD runs of Trp-cage, HP-36 and VA3 are shown in the Supplementary Information.

ILMs ($S \leq 0.5$) are rare events. For example, in the MD run 1 of Trp-cage, only 251 ILMs were found (Table 1), which represents 0.05% of the total number of main-chain fluctuations recorded over 500 ns. The probability of observing ILMs was similar in the other MD runs of Trp-cage (Table 1). In all MD runs of Trp-cage, the most frequent ILMs were located at the same specific positions along the amino acid sequence [Fig. 3(a)]: $i = 9$, 10, 14 and 18. The excitations at $i = 9$, 10 and 14 are typical solitons of sine-Gordon type [compare Fig. 2(d) to Fig. 3(b)]. The largest amplitudes of ILMs were found for these three sites, with $\Delta u^2 = 3.0$, 3.3 and 3.1 for a soliton centered at $i = 9$, 10 and 14, respectively. Most of the solitons (80%) centered at $i = 9$, 10 and 14 with the largest amplitudes ($\Delta u^2 > 2.0$) corresponded to cis-trans or trans-cis transitions of the four C$^\alpha$ segments [typical examples are shown in Fig. 3(c)].

The solitons localized at $i = 9$, 10 and 14 [shown in Fig. 3] shared a common feature: they all had a glycine (GLY) residue at $i + 1$ in the amino acid sequence (in bold font in Table 1). As shown in SI (Supplementary Tables 1 and 2), similar results were found for HP-36 and VA3 proteins: solitons localized at $i = 51$ and $i = 73$ in HP-36 have a GLY residue at $i = 52$ and $i = 74$, respectively, and solitons localized at $i = 36$ in VA3, have a GLY residue at $i = 37$. In the Trp-cage, the most frequent soliton ($i = 10$) corresponded to a rotation around a virtual bond formed by two GLYs (GLY10-GLY11) (Table 1). Because of its small side-chain (H atom), GLY can adopt a larger set of conformations in a polypeptide chain, which may explain why the highest probabilities of ILMs are observed at $i = 9$, 10 and 14. To test this hypothesis, we ran an MD trajectory for the mutant Trp-cage G15A at $T = 380\,\mathrm{K}$. The number of ILMs of soliton type at $i = 14$ decreases by an order of magnitude (Table 1). To further test the role of GLYs, we ran a MD trajectory for the triple mutant Trp-cage G10A-G11A-G15A at $T = 380\,\mathrm{K}$. The number of ILMs of soliton type at $i = 9$, 10 and 14 decreased drastically (Table 1). This observation reveals, for the first time, the role of GLY residues in the localization and the probability of the appearance of ILMs of the soliton type.

The ILMs located at the C-termimus of the chain ($i = 18$) are *not* similar to sine-Gordon solitons (Supplementary Figure 1) but are localized excitations that also exist in a harmonic chain with free ends and are due to the broken symmetry of the chain at its extremities[49]. These ILMs do not depend on the presence of GLY residues, and their probability is similar for Trp-cage and its mutants (Table 1).

A few ILMs ($S \leq 0.5$), all with small amplitudes ($0.4 < \Delta u^2 < 1.9$), were also observed very rarely (not more than six times in 500,001 snapshots) at $i = 2$, 7, 8, 11, 12, 13, 15, 16 and 17 in the MD runs of Trp-cage (Supplementary Table 3). Except at $i = 2$, which corresponds to a mode located at the N-terminus (Supplementary Figure 1), all of these ILMs were similar to sine-Gordon solitons.

Free-energy landscape and ILMs.
The probability of observing an ILM at a given time is smaller if the protein is in its native state (rigid, folded state) than in a non-native state (flexible, misfolded or unfolded states).
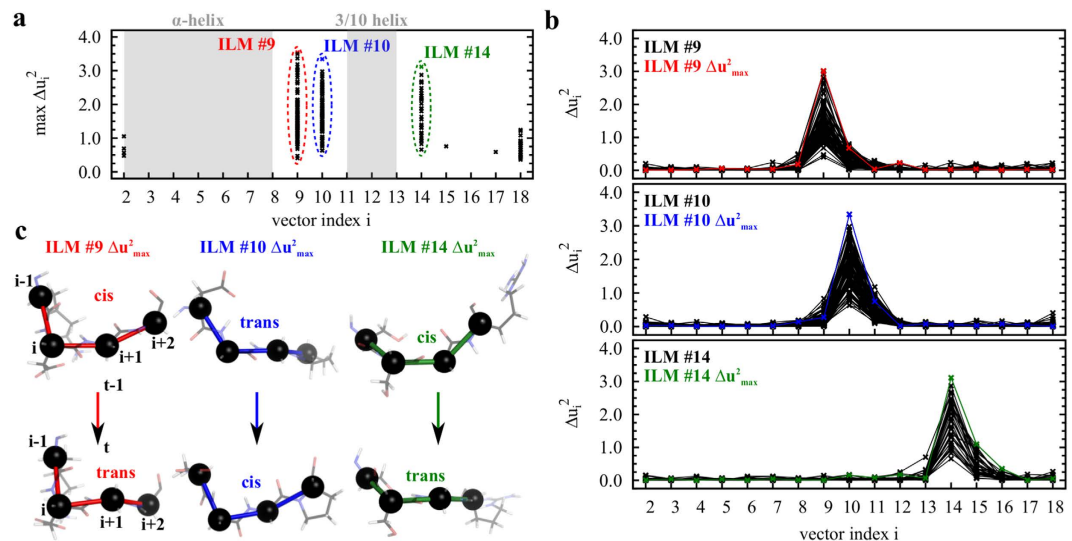
Figure 3. Intrinsic localized modes (ILMs) for the Trp-cage protein at $T = 380$ K. Results of the MD run 1 are shown. (**a**) Maximum amplitudes Δu^2 of ILMs as a function of the residue index i. (**b**) Amplitudes Δu^2 of ILMs $i = 9$, 10 and 14 as a function of the vector index. The ILMs showing the largest amplitude are highlighted in red (ILM #9), blue (ILM #10) and green lines (ILM #14). (**c**) Typical cis-trans and trans-cis transitions observed in ILMs #9, #10 and #14 of maximum amplitude.

The native state of Trp-cage is defined by the ensemble of the most probable conformations explored at $T = 300$ K. The native state is better represented as basins in the free-energy landscape of the protein[6,16].

We represented the free-energy landscape of the main chain of Trp-cage by the sequence of the effective free-energy maps $V(\gamma, \theta)_n$ computed from the probability densities of each pair of CGA $(\gamma, \theta)_n$ in the MD trajectories (see the Methods section and Supplementary Figures 2 and 3). The sequence of $V(\gamma, \theta)_n$ has proven useful in describing protein folding[50], conformation dynamics[40] and allosteric communication[51]. For each $V(\gamma, \theta)_n$, we defined the native basin as the region of the $(\gamma, \theta)_n$ space within 3 k_BT from the minimum of $V(\gamma, \theta)_n$ at $T = 300$ K (Supplementary Figure 2). The 3 k_BT cutoff ensures that all of the experimental structures of Trp-cage observed by NMR at 282 K[41] correspond to (γ, θ) angles located in the native basins (Supplementary Figure 4).

To quantify the native character (*NC*) of the protein as a function of time in MD trajectories at $T = 380$ K, we counted the % of CGA remaining within their native basins, as illustrated in Fig. 4(a), for a typical trajectory. At $T = 380$ K, Trp-cage partially folds/unfolds during MD runs, *i.e.*, explores non-native states far from the native basins of the free-energy maps, as shown for selected CGA pairs in Fig. 4(b) and for all of the CGAs in Supplementary Figure 3. Typical structures of the protein in native and non-native states are shown in Fig. 4(c) and in the Supplementary Figure 5 for another MD trajectory (run 2) for comparison. As shown in Fig. 4(a), a sharp transition occurs between the initial native state of Trp-cage, which lasts for approximately 100 ns, to a non-native state in which the molecule remains until the end of the trajectory [Fig. 4(a)]. Results similar to those presented in Fig. 4(a) were found for all the MD trajectories of Trp-cage and HP-36 at $T = 380$ K (with different sequences of folding and unfolding events depending on the initial conditions of the MD runs, see Supplementary Figures 5 and 6). As clearly shown in Fig. 4 [bottom of panel (a)], the number of ILMs was larger in the non-native portion of the trajectory than in its native portion. The same results were observed for all MD runs of Trp-cage, HP-36 and VA3 (Supplementary Figures 5 and 6) and were quantified by computing the probability of observing a soliton in the native portion and in the non-native portion of the trajectory (Table 2). For the fast-folder proteins examined in the present work, Table 2 demonstrates that the probability of observing a soliton is larger by a factor varying between approximately 9 and 28 in the non-native portion of the trajectory compared to the native portion at $T = 380$ K. The probability of observing a soliton is only about twice larger in the non-native portion of the MD trajectory of VA3 than in the native portion of the trajectory at $T = 380$ K (Table 2). Because of its three disulfide bridges, VA3 did not unfold in the MD run and explored a non-native state with a relative high average of *NC* ($<NC>_{t_{nn}} = 75\%$, see Supplementary Figure 6). In reference to Table 2, it is worth noting that the probability of observing a soliton in the non-native portion of the trajectories at $T = 300$ K is difficult to evaluate accurately, as the time spent in a non-native state is extremely small.
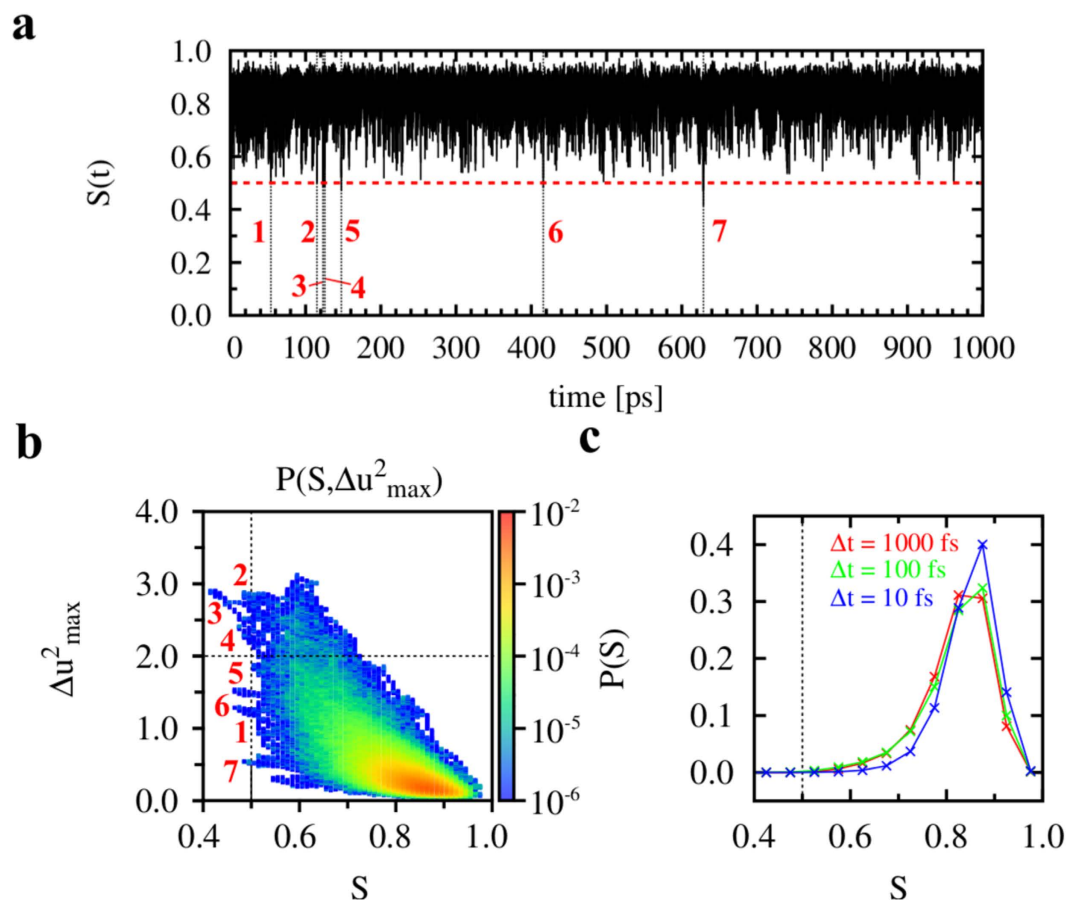
Statistics of ILMs at different time resolutions. To improve the statistics of ILMs, we ran eight additional short trajectories of 1 ns duration using initial non-native structures of Trp-cage extracted at different times from MD run 1 at $T = 380$ K and recording a snapshot every fs. The displacements of the main chain, $\Delta \mathbf{u}_i$, were first computed from these trajectories by using a sliding window of $\Delta t = 1$ ps shifted every fs. That is, the time-scale of the displacements Δu_i was identical to that discussed in the previous sections. The ILMs of the soliton type similar to those reported in Fig. 3(b) were detected by computing $S(t)$ every fs [see Fig. 5(a)]. Seven solitons were detected [Fig. 5(a)] at $i = 10$ (solitons 1 to 5); $i = 3$ (soliton 6) and $i = 14$ (soliton 7) (not shown). Soliton 3 had the largest amplitude and was the most localized [Fig. 5(b)]. The probability of observing solitons in these short
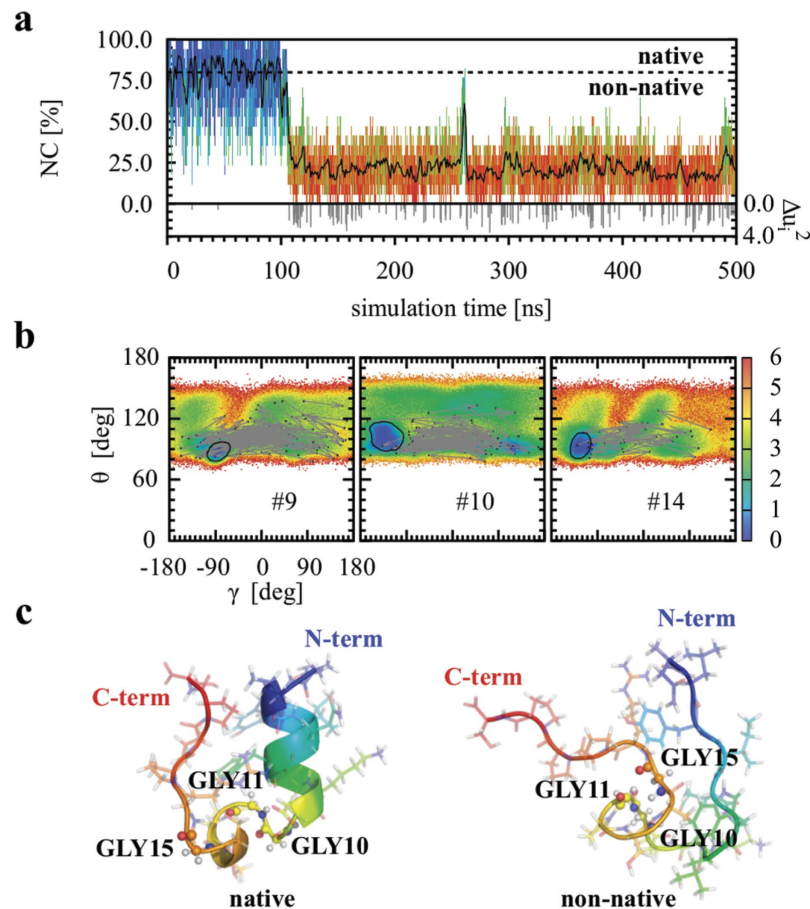
Figure 4. Relation between ILMs and the native/non-native character of the Trp cage protein (run 1 is shown). (**a**) Native character (*NC*) (in %) as a function of simulation time. The color line using the BGR palette (from blue, which corresponds to $NC = 100\%$, green, which corresponds to $NC = 50\%$, to red, which corresponds to $NC = 0\%$) represents the NC computed every 1 ps, and the black line represents the mobile average of the NC computed in a time window of 1 ns. Gray impulses shown on the time axis of the lower graph represent the occurrence of an ILM ($S \leq 0.5$) along the MD trajectory. (**b**) Effective free-energy map $V(\gamma, \theta)_n$ (kbT units) for $(\gamma, \theta)_9$ (left panel), $(\gamma, \theta)_{10}$ (middle panel) and $(\gamma, \theta)_{14}$ (right panel). Black lines represent free-energy isolines $V - V_{min} = 3\,k_B T$ computed from the effective free-energy maps $V(\gamma, \theta)_n$ at $T = 300\,K$ (Supplementary Figure 2). Vectors in gray lines represent the displacements on the map for all the corresponding ILMs. (**c**) Structures of the Trp cage in typical native (left panel) and non-native (right panel) conformations. Structures are shown in a cartoon plus stick representation. GLY residues are highlighted in ball and stick representations. The color code corresponds to an RGB palette, from N-terminus (residue 1) to C-terminus (residue 20).

Protein	T [K] (run)	t_n [ns]	t_{nn} [ns]	$<NC>_{t_{nn}}$	N_n	N_{nn}	P_n	P_{nn}	P_{nn}/P_n
Trp cage	300(1)	487	13	—	9	—	0.018	—	—
	380(1)	57	443	27%	1	222	0.018	0.501	27.8
	380(2)	11	489	26%	—	270	—	0.552	—
	380(3)	71	429	32%	1	149	0.014	0.347	24.6
HP-36	300(1)	478	22	—	28	3	0.059	0.136	2.3
	380(1)	64	436	58%	1	83	0.016	0.190	11.9
	380(2)	185	315	66%	4	58	0.022	0.184	8.4
VA3	300(1)	500	—	—	33	—	0.066	—	—
	380(1)	245	255	74%	12	22	0.049	0.086	1.8

Table 2. Statistics of ILMs of the soliton type in the native and non-native states computed from MD simulations of Trp-cage, chicken villin headpiece fragment HP-36, and VA3 (MD trajectories of a duration of 500 ns are presented). The subscripts *n* and *nn* define the native and non-native states of the protein, respectively. ILMs at the N-terminus (localized at $i = 2$) and C-terminus (localized at $i = 18$, 34 and 44 for Trp-cage, HP-36 and VA3, respectively) were not considered.

**Figure 5. Statistical analysis of ILMs at different time resolutions.** (**a**) Shannon entropy S as a function of simulation time computed every 1 fs from a typical 1 ns MD run of Trp-cage at $T = 380$ K. The black dashed lines indicate ILMs ($S \leq 0.5$), and the red dashed line indicates the criterion $S = 0.5$. (**b**) 2-D probability distribution function $P(S, \Delta u_{max}^2)$ computed from the same MD run of Trp-cage at 380 K as in panel a. The horizontal and vertical black dashed lines indicate the limits $\Delta u_{max}^2 = 2.0$ and $S = 0.5$, respectively. (**c**) Probability distribution function $P(S)$ computed using different time resolution using the same trajectory as in panels a and b.

trajectories was similar to that reported in Table 1. If the snapshots were recorded only every ps (as in the 500 ns MD runs 1, 2 and 3), only soliton 3 (located at $i = 10$) would have been detected. Interestingly, the probability density of the entropy $P(S)$ was fairly independent of the time-scale used to compute the main-chain displacements Δu_i [see Fig. 5(c) for $\Delta t = 100$ fs and $\Delta t = 10$ fs]. The ILMs ($S \leq 0.5$) were always found in the tails of $P(S)$ (rare events).

Soliton dynamics at a femtosecond time resolution. As an example, we analyzed the dynamics of soliton 3 (located at $i = 10$) detected in the trajectory shown in Fig. 5(a), for which snapshots were recorded every fs. The time $t = 0$ was defined as the time at which the soliton was detected, with $\Delta t = 1$ ps in the trajectory analyzed in Fig. 5(a). The displacement $\Delta \mathbf{u}_{10}(t)$ was computed at each fs by summing the displacements $\Delta \mathbf{u}_{10}$ every fs from $t = -5$ ps to $t = 5$ ps (by assuming $\Delta \mathbf{u}_{10}(t = -5\text{ps}) \equiv 0$). The values of $\Delta u_{10}(t)^2$ shown in Fig. 6(a), demonstrated that the ILM is strongly localized between $t = -1$ ps and $t = 2.5$ ps. A careful analysis of the variation of the $S(t)$ [Fig. 6(b)] shows that the life time of the soliton (*i.e.*, the time for which $S(t)$ remains lower or close to 0.5) is only 400 fs (from $t = -300$ fs to $t = 100$ fs). This extreme time-space localization corresponds to a jump in the γ_{10} free-energy profile from the top of a barrier ($\gamma_{10} = -30°$) to a small metastable state ($\gamma_{10} = 60°$) [Fig. 6(c)]. As already noted in the introduction and shown in Fig. 2, the ILMs of the soliton-type correspond mainly to a large variation of the dihedral angle γ, as seen for soliton 3 in Fig. 6(c).

Discussion

In the present all-atom MD study, ILMs were successfully detected in the spontaneous thermal fluctuations of the main-chain of proteins. We found that ILMs of soliton type are short-life events. In the example detailed in Fig. 6, the lifetime is ~400 fs ($S \leq 0.5$) [it would be 1,500 fs if we adopted a less strict definition of the spatial localization of the main-chain deformation ($S \leq 0.6$)]. These sub-picosecond localized excitations may be accessible for experimental investigation using 2D-IR spectroscopy, as demonstrated for the fast dynamics of the Ramachandran dihedral angular fluctuations of a three-peptide chain[52]. The probability of an ILM is approximately 0.05% in MD simulations. As rare events, these ILMs do not contribute significantly to the partition function of the system;
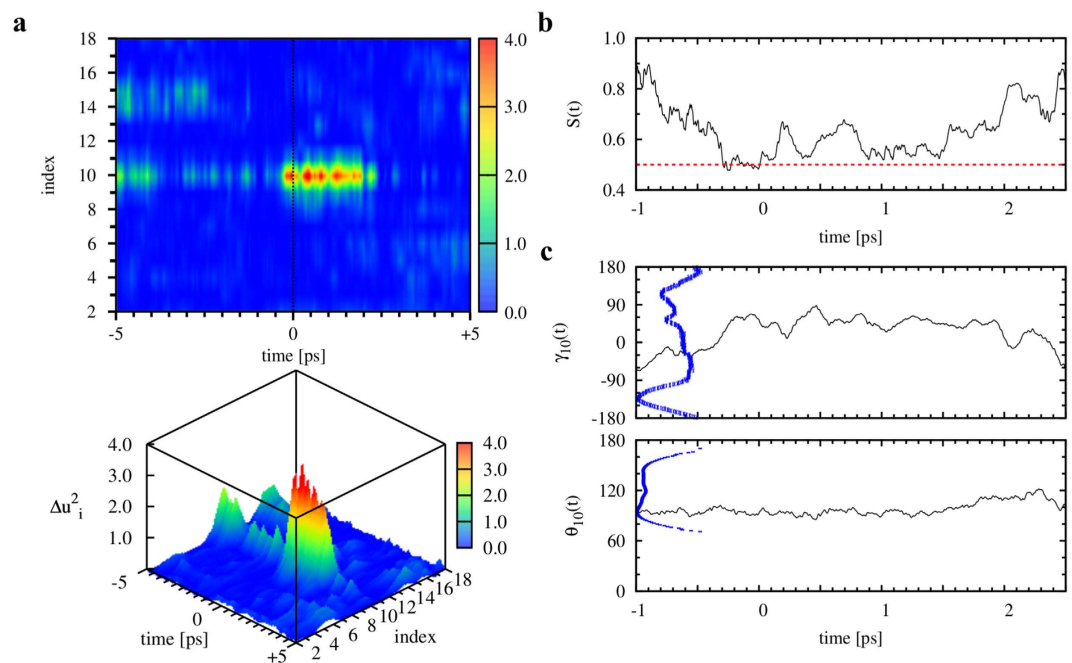
Figure 6. Typical time evolution of an ILM of soliton type in MD trajectories of Trp-cage of 1 ns duration for which the coordinates were recorded every fs. The time $t = 0$ corresponds to the detection of soliton 3 in the MD trajectory shown in Fig. 5. (**a**) Two-dimensional (up) and three-dimensional (bottom) color maps of Δu_{10}^2 in a window of 10 ps. (**b**) Corresponding variation of $S(t)$ between $t = -1$ ps and $t = 2.5$ ps. (**c**) Variations of $(\gamma, \theta)_{10}$ (thin black lines) in the time window represented in (**b**). The free-energy profiles $V(\gamma_{10})$ and $V(\theta_{10})$ are also shown in thick blue lines. A movie of the time evolution of the soliton shown is available in the SI.

they are a priori negligible from a thermodynamic equilibrium point of view. However, the sine-Gordon localized excitations may help to cross an activation barrier as illustrated in Fig. 6(c) and may be important from a kinetic point of view. Intermittent events often do play a role in biological function, for example, in conformational gating of a ligand in enzymes[53].

Solitons were previously predicted to occur in stiff regions of a protein (in an α-helix or at catalytic sites) using simplified models of a polypeptide chain[23]. The opposite is found here by using all-atom MD simulations of two α-helical fast-folder proteins and a rigid protein. We found ILMs in the *flexible* regions of the proteins (loops and N- and C-termini), as illustrated in Fig. 3(a). These ILMs are different from the usual classical discrete breathers as they are strongly localized in time. Solitons with the highest amplitudes mainly corresponded to (incomplete) cis-trans or trans-cis transitions [Fig. 3(c)], which are more probable in flexible segments of a protein main chain. The flexibility governs the probability of observing a soliton: the probability of observing an ILM of the soliton type [Fig. 2(d)] is enhanced in the non-native state compared to that in the native state and in protein segments containing GLY residues. This result is expected because the main chain experiences more of the anharmonic portion of the free-energy landscape in large fluctuations than it does in small fluctuations. The non-native state was significantly explored at $T = 380$ K. To confirm that the solitons occur more frequently in our simulations because the protein is unstructured (and is thus more flexible) and not because the temperature is increased, we ran an additional 20 ns trajectory at $T = 300$ K by selecting one of the most unfolded structures of Trp-cage as the initial structure of the new trajectory, as we did previously[39]. The probability of observing solitons was enhanced by a factor of approximately 6 compared to that of the native state at the same temperature (Table 1). Therefore, ILMs of the soliton type might play a functional role in misfolded proteins and in unfolded proteins.

Methods

MD simulations. All-atom MD simulations in explicit water (TIP3P force field[54]) with Trp-cage (model 1 in PDB ID: 1L2Y)[41], the chicken villin headpiece subdomain HP-36 (model 1 in PDB ID: 1VII)[44] and the protein VA3 (model 1 in PDB: 1ED0)[55] were conducted using the GROMACS software package[56] and the AMBER99sb-ILDN force field[57]. In addition, one all-atom MD run of the Trp-cage at $T = 300$ K and one at 380 K using CHARMM27[58,59] and AMBER99SB*-ILDN-q[60] force fields were conducted. The results found using these force fields are reported in the Supplementary Table 4 and were similar to those obtained with AMBER99sb-ILDN force field. The time step used in all simulations was 1 fs, and the list of neighbors was updated every 5 fs with the grid method and a cutoff radius of 1.0 nm. The coordinates of all the atoms in the simulation box were saved every 1 ps. The initial velocities were chosen randomly. We used the NPT ensemble with a cubic box of 4.55 nm for Trp-cage, 5.21 nm for HP-36 and 5.27 nm for VA3. The temperature and pressure were kept to the desired value by using the Nosé-Hoover thermostat[61,62] and the Parrinello-Rahman[63] barostat, respectively. The electrostatic term was computed by using the particle mesh Ewald algorithm[64] (with a radius of 1 nm) using the Fast Fourier Transform optimization (with an order equal to four for the interpolation). The cutoff algorithm was applied for the non-coulomb potentials with a

radius of 1.0 nm. The system was warmed up for 50 ps and equilibrated for 1 ns with lower restraints, finishing with no restraints at the desired temperature. We performed three MD runs of Trp-cage and two MD runs of HP-36 at $T = 380$ K using different initial conditions and one MD run at $T = 300$ K for each protein (named run 1 in Table 1). In addition, we performed one MD run for each mutant of the Trp-cage ([G15A] and [G10A-G11A-G15A]) and for VA3 at $T = 300$ K and one at $T = 380$ K. Each MD run was of 500 ns duration. In total, the thirteen long MD runs corresponded to 6.5 μs of simulations.

In addition, we performed eight extra short MD runs of Trp-cage of 1 ns duration at $T = 380$ K using the same procedure as described above, except that the coordinates of all the atoms in the simulation box were saved every 1 fs. The initial structures of those MD runs were extracted at different times from the MD run 1 of Trp-cage at $T = 380$ K, corresponding to $t = 100, 161, 209, 283, 300, 441$ and 459 ns. Data from the MD run using the frame at $t = 441$ ns are presented in the present paper (Figs 5 and 6). The results for the other seven runs were similar. Finally, we performed an MD run of Trp-cage of 20 ns duration at $T = 300$ K (named run 2 in Table 1) using a completely unfolded structure of the protein obtained in a previous work[39] at $T = 450$ K using the same procedure described above.

Free-energy map and free-energy profiles. An effective free-energy map $V((\gamma, \theta)_i)$ and two free-energy profiles $V(\gamma_i)$ and $V(\theta_i)$ were computed for each pair of CGAs $(\gamma, \theta)_i$ by using

$$V(\gamma, \theta)_i = -k_B T \ln [P((\gamma, \theta)_i)]$$
$$V(\gamma_i) = -k_B T \ln [P(\gamma_i)]$$
$$V(\theta_i) = -k_B T \ln [P(\theta_i)] \tag{1}$$

where k_B is the Boltzmann constant, T is the temperature and $P((\gamma, \theta)_i)$, $P(\gamma_i)$, $P(\theta_i)$ are the probability density functions (PDF) of the pair $(\gamma, \theta)_i$, of γ_i and of θ_i, respectively. The PDFs were computed from the MD trajectories on a time-scale of 500 ns.

Normalized Shannon entropy S and localized excitations. The quantity p_i measures the fluctuations of the pair of CGA i $(i = 2$ to $N - 2)$ relative to the $N - 3$ pairs of CGAs along the sequence and is defined by

$$p_i(t) = \frac{\Delta u_i(t)^2}{\sum_{j=2}^{N-2} \Delta u_j(t)^2}, \tag{2}$$

where Δu_i is the displacement of the vector $\mathbf{u}_i$ between two consecutive snapshots $\Delta \mathbf{u}_i(t) = \mathbf{u}_i(t) - \mathbf{u}_i(t-1)$ and $\mathbf{u}_i = (\cos(\gamma_i)\sin(\theta_i), \sin(\gamma_i)\sin(\theta_i), \cos(\theta_i))$. By definition, $0 \leq p_i(t) \leq 1$. The localization of the structural fluctuations of the protein main chain can be quantified by

$$S(t) = -\frac{\sum_{i=2}^{N-2} p_i(t) \ln(p_i(t))}{\ln(N-3)} \tag{3}$$

where $S(t)$ can be interpreted as a normalized Shannon entropy $0 \leq S(t) \leq 1$. The maximum localization of the fluctuations occured for $S = 0$ ($p_i = 0$ for $i \neq k$ and $p_k = 1$), and the maximum delocalization occured for $S = 1$ ($p_i = 1/(N-3)$ for all i).

An ILM is defined by a value of $S \leq S_{cutoff}$, where S_{cutoff} is chosen such that the correlation coefficient between the sequence of the $\Delta u_i(t)^2$ of all ILMs localized at the same site i is larger than 0.9. For example, in the MD run 1 ($T = 380$ K), the most strongly localized excitation at $i = 10$ had a value of $S = 0.40$ with a maximum amplitude $\Delta u_{10}(t)^2 = 2.42$. All excitations located at $i = 10$ are highly correlated ($\rho > 0.9$) if $S \leq 0.5$, i.e., they all represent the same single peak of localized excitation (Supplementary Figure 7). It is worth noting that the number of ILMs at $i = 10$ in this MD run increases as a function of the S_{cutoff} value (see Supplementary Table 5) as expected. However, the probability to find an ILM at $i = 10$ in the non-native state is larger than in the native state for different values of S_{cutoff}. The ratio P_{nn}/P_n is infinite for $S_{cutoff} = 0.45$ and decreases to 4.6 at $S_{cutoff} = 0.6$ (Supplementary Table 5) because at $S_{cutoff} = 0.6$ we include excitations which are not centered on a single site as the one shown in red in the Supplementary Figure 7. A similar conclusion was drawn for all of the MD runs studied here, which sets the value of S_{cutoff} to 0.5.

References

1. Flach, S. & Willis, C. R. Discrete breathers. *Phys. Rep.* **295,** 181–264 (1998).
2. Dauxois, T. & Peyrard, M. *Physics of Solitons* (Cambridge University Press, 2006).
3. Sievers, A. J. & Takeno, S. Intrinsic localized modes in anharmonic crystals. *Phys. Rev. Lett.* **61,** 970–973 (1988).
4. Schwarz, U. T., English, L. Q. & Sievers, A. J. Experimental Generation and Observation of Intrinsic Localized Spin Wave Modes in an Antiferromagnet. *Phys. Rev. Lett.* **83,** 223–226 (1999).
5. McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267,** 585–590 (1977).
6. Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* **280,** 558–563 (1979).
7. Parak, F. & Knapp, E. W. A consistent picture of protein dynamics. *Proc. Natl. Acad. Sci. USA* **81,** 7088–7092 (1984).
8. Doster, W., Cusack, S. & Petry, W. Dynamical transition of myoglobin revealed by inelastic neutron scattering. *Nature* **337,** 754–756 (1989).
9. Hayward, S., Kitao, A. & Gō, N. Harmonicity and anharmonicity in protein dynamics: A normal mode analysis and principal component analysis. *Proteins* **23,** 177–186 (1995).
10. García, A. E. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* **68,** 2696–2699 (1992).
11. Doster, W. The dynamical transition of proteins, concepts and misconceptions. *Eur. Biophys. J.* **37,** 591–602 (2008).

12. Gerstein, M., Lesk, A. M. & Chothia, C. Structural Mechanisms for Domain Movements in Proteins. *Biochemistry* **33,** 6739–6749 (1994).

13. Benkovic, S. J. & Hammes-Schiffer, S. Enzyme Motions Inside and Out. *Science* **312,** 208–209 (2006).

14. Rashin, A. A., Rashin, A. H. & Jernigan, R. L. Diversity of function-related conformational changes in proteins: coordinate uncertainty, fragment rigidity and stability. *Biochemistry* **49,** 5683–5704 (2010).

15. Henzler-Wildman, K. A. *et al.* A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450,** 913–916 (2007).

16. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450,** 964–972 (2007).

17. Scott, A. Davydov's soliton. *Phys. Rep.* **217,** 1–67 (1992).

18. Davydov, A. S. Solitons and energy transfer along protein molecules. *J. Theoret. Biol.* **66,** 379–387 (1977).

19. Xie, A. H., Meer, L., Hoff, W. & Austin, R. H. Long-lived amide I vibrational modes in myoglobin. *Phys. Rev. Lett.* **84,** 5435–5438 (2000).

20. Edler, J., Pfister, R., Pouthier, V., Falvo, C. & Hamm, P. Direct Observation of Self-Trapped Vibrational States in α-Helices. *Phys. Rev. Lett.* **93,** 106405 (2004).

21. Austin, R. H. *et al.* Tilting after Dutch windmills: probably no long-lived Davydov solitons in proteins. *J. Biol. Phys.* **35,** 91–101 (2009).

22. Kobus, M., Nguyen, P. H. & Stock, G. Coherent vibrational energy transfer along a peptide helix. *J. Chem. Phys.* **134,** 124518 (2011).

23. Juanico, B., Sanejouand, Y.-H., Piazza, F. & De Los Rios, P. Discrete Breathers in Nonlinear Network Models of Proteins. *Phys. Rev. Lett.* **99,** 238104 (2007).

24. Piazza, F. Nonlinear excitations match correlated motions unveiled by NMR in proteins: a new perspective on allosteric cross-talk. *Phys. Biol.* **11,** 036003 (2014).

25. Krokhotin, A., Liwo, A., Maisuradze, G. G., Niemi, A. J. & Scheraga, H. A. Kinks, loops, and protein folding, with protein A as an example. *J. Chem. Phys.* **140,** 025101 (2014).

26. Feddersen, H. Localization of vibrational energy in globular protein. *Phys. Lett. A* **154,** 391–395 (1991).

27. Karplus, M. & Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **347,** 631–639 (1990).

28. Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M. & Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **433,** 128–132 (2005).

29. Shaw, D. E. *et al.* Atomic-level characterization of the structural dynamics of proteins. *Science* **330,** 341–346 (2010).

30. Peregrine, D. H. Water waves, nonlinear Schrödinger equations and their solutions. *Anziam J.* **25,** 16–43 (1983).

31. Kibler, B. *et al.* The Peregrine soliton in nonlinear fibre optics. *Nat. Phys.* **6,** 790–795 (2010).

32. Gō, N., Noguti, T. & Nishikawa, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA* **80,** 3696–3700 (1983).

33. Bahar, I., Atilgan, A. R., Demirel, M. C. & Erman, B. Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability. *Phys. Rev. Lett.* **80,** 2733–2736 (1998).

34. Keskin, O., Jernigan, R. L. & Bahar, I. Proteins with Similar Architecture Exhibit Similar Large-Scale Dynamic Behavior. *Biophys. J.* **78,** 2093–2106 (2000).

35. Nishikawa, K., Momany, F. A. & Scheraga, H. A. Low-Energy Structures of Two Dipeptides and Their Relationship to Bend Conformations. *Macromolecules* **7,** 797–806 (1974).

36. Senet, P., Maisuradze, G. G., Foulie, C., Delarue, P. & Scheraga, H. A. How main-chains of proteins explore the free-energy landscape in native states. *Proc. Natl. Acad. Sci. USA* **105,** 19708–19713 (2008).

37. Korkut, A. & Hendrickson, W. A. A force field for virtual atom molecular mechanics of proteins. *Proc. Natl. Acad. Sci. USA* **106,** 15667–15672 (2009).

38. Maisuradze, G. G., Senet, P., Czaplewski, C., Liwo, A. & Scheraga, H. A. Investigation of Protein Folding by Coarse-Grained Molecular Dynamics with the UNRES Force Field. *J. Phys. Chem. A* **114,** 4471–4485 (2010).

39. Cote, Y., Maisuradze, G. G., Delarue, P., Scheraga, H. A. & Senet, P. New Insights into Protein (Un)Folding Dynamics. *J. Phys. Chem. Lett.* **6,** 1082–1086 (2015).

40. Cote, Y., Senet, P., Delarue, P., Maisuradze, G. G. & Scheraga, H. A. Anomalous diffusion and dynamical correlation between the side chains and the main chain of proteins in their native state. *Proc. Natl. Acad. Sci. USA* **109,** 10346–10351 (2012).

41. Neidigh, J. W., Fesinmeyer, R. M. & Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **9,** 425–430 (2002).

42. Qiu, L., Pabit, S. A., Roitberg, A. E. & Hagen, S. J. Smaller and faster: the 20-residue trp-cage protein folds in 4 μs *J. Am. Chem. Soc.* **124,** 12952–12953 (2002).

43. McKnight, J. C., Doering, D. S., Matsudaira, P. T. & Kim, P. S. A Thermostable 35-Residue Subdomain within Villin Headpiece. *J. Mol. Biol.* **260,** 126–134 (1996).

44. McKnight, C. J., Matsudaira, P. T. & Kim, P. S. NMR structure of the 35-residue villin headpiece subdomain. *Nat. Struct. Mol. Biol.* **4,** 180–184 (1997).

45. Zhou, R. Trp-cage: Folding free energy landscape in explicit water. *Proc. Natl. Acad. Sci. USA* **100,** 13280–13285 (2003).

46. Paschek, D., Nymeyer, H. & García, A. E. Replica exchange simulation of reversible folding/unfolding of the Trp-cage miniprotein in explicit solvent: On the structure and possible role of internal water. *J. Struct. Biol.* **157,** 524–533 (2007).

47. Paschek, D., Hempel, S. & García, A. E. Computing the stability diagram of the Trp-cage miniprotein. *Proc. Natl. Acad. Sci. USA* **105,** 17754–17759 (2008).

48. Bandyopadhyay, S., Chakraborty, S., Balasubramanian, S., Pal, S. & Bagchi, B. Atomistic Simulation Study of the Coupled Motion of Amino Acid Residues and Water Molecules around Protein HP-36: Fluctuations at and around the Active Sites. *J. Phys. Chem. B* **108,** 12608–12616 (2004).

49. Lucas, A. A. Phonon Modes of an Ionic Crystal Slab. *J. Chem. Phys.* **48,** 3156–3168 (1968).

50. Maisuradze, G. G., Liwo, A., Senet, P. & Scheraga, H. A. Local vs Global Motions in Protein Folding. *J. Chem. Theory Comput.* **9,** 2907–2921 (2013).

51. Nicolaï, A., Delarue, P. & Senet, P. Decipher the Mechanisms of Protein Conformational Changes Induced by Nucleotide Binding through Free-Energy Landscape Analysis: ATP Binding to Hsp70. *PLoS Comput. Biol.* **9,** e1003379 (2013).

52. Woutersen, S., Mu, Y., Stock, G. & Hamm, P. Subpicosecond conformational dynamics of small peptides probed by two-dimensional vibrational spectroscopy. *Proc. Natl. Acad. Sci. USA* **98,** 11254–11258 (2001).

53. Zhou, H.-X., Wlodek, S. T. & McCammon, J. A. Conformation gating as a mechanism for enzyme specificity. *Proc. Natl. Acad. Sci. USA* **95,** 9280–9283 (1998).

54. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).

55. Romagnoli, S. *et al.* NMR structural determination of viscotoxin A3 from Viscum album L. *Biochem. J.* **350,** 569–577 (2000).

56. van der Spoel, D. *et al.* Gromacs User Manual version 4.6.6 (2011). URL http://manual.gromacs.org/documentation/ Date of access: 15 October 2015.

57. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99sb protein force field. *Proteins* **78,** 1950–1958 (2010).

58. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102,** 3586–3616 (1998).

59. Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25,** 1400–1415 (2004).
60. Best, R. B., De Sancho, D. & Mittal, J. Residue-specific α-helix propensities from molecular simulation. *Biophys. J.* **102,** 1462–1467 (2012).
61. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81,** 511–519 (1984).
62. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31,** 1695–1697 (1985).
63. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).
64. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An Nlog(N) method for ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).

Acknowledgements

Author Contributions

P.S. designed the research. A.N. and P.D. performed the simulations. A.N., P.D. and P.S. analyzed the data. A.N. and P.S. wrote the paper.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Nicolaï, A. *et al.* Intrinsic Localized Modes in Proteins. *Sci. Rep.* **5**, 18128; doi: 10.1038/srep18128 (2015).