J Math Chem (2011) 49:643–665
DOI 10.1007/s10910-010-9760-9

ORIGINAL PAPER

Mathematical criteria to observe mesoscopic emergence of protein biochemical properties

Anirban Banerji · Indira Ghosh

Received: 27 May 2010 / Accepted: 14 October 2010 / Published online: 16 November 2010

Abstract Proteins are regularly described with some general indices (entropy, enthalpy, free energies, hydrophobicity, denaturation temperature etc.), which are inherently statistical in nature. These general indices emerge from innumerable (innately context-dependent and time-dependent) interactions between various atoms of a protein. Many studies have been performed on the nature of these interatomic interactions and the change of profile of atomic fluctuations that they cause. However, we still do not know, under a given context, for a given duration of time, how does a macroscopic property emerge from the cumulative interatomic interactions. An exact answer to that question requires bridging the gap between nano-scale distinguishable atomic description and macroscopic indistinguishable (statistical) measures with mesoscopic description. In this work we propose a mathematical framework that derives expressions to observe emergence of a macroscopic biophysical property from a set of interacting (fluctuating) atoms. Since most of the protein interior interactions are non-linear in nature; observability criteria are derived for both linear and the non-linear descriptions of protein interior. Care has been taken (with extensive literature survey) to ensure that every pertinent bio-physical and bio-chemical facet of protein interior is taken into account, without compromising with the mathematical rigor. Two theorems (concerning mathematical description of protein structure) are proposed here that helps the nascent field of mesoscopic protein studies. Categorical schemes to relate

Electronic supplementary material The online version of this article (doi:10.1007/s10910-010-9760-9) contains supplementary material, which is available to authorized users.

A. Banerji
Bioinformatics Centre, University of Pune, Pune, Maharashtra 411007, India
e-mail: anirbanab@gmail.com

I. Ghosh (✉)
School of Information Technology, Jawaharlal Nehru University, New Delhi 110067, India
e-mail: indirag@mail.jnu.ac.in

mathematical formulations to studies of pKa shift, residual dipolar coupling, origin of hydrophobicity, drug discovery etc.—are provided to ensure their easy applicability. While the present work helps the theoretical discourse by providing a framework to understand the origin of a macroscopic property; ability of it to predict a priory whether the dynamics in a certain set of atoms or the couplings between them, can at all produce a biological property of interest or not, will account for tremendous saving of resource and effort.

Keywords Mesoscopic origin · Biochemical properties · Emergence · Observability · Predictive mathematical framework · Non-linear differential equations

1 Introduction

Recent works have described proteins as "complex systems" [1, 2] and as "deformable polymers" [3]. The mesoscopic nature of protein structures has been reported by crystallographers too [4]. Furthermore, it has been found recently that proteins exist in a state of 'self organized criticality' [5, 6]. Along with all these, recent [7] and previous [8] characterizations of inhomogeneous distributions of mass and hydrophobicity merely serve to complicate an effort to construct a simple and linear scheme for description of protein interior. An approach to study protein interior that describes the inhomogeneous and nonlinear behaviors of protein structural parameters can be constructed by describing them with self similarity prevalent in their distributions. Indeed many previous studies on this topic (a dreadfully undersized representation is references [9–16] had hinted that with an objective quantification of self-similarity, we can decipher the hidden symmetry, which connects global patterns of macroscopic properties in proteins (say hydrophobicity distribution, polarizability distribution etc.) with the local (atomic) interactions that produce them [16]. However the basic question, that, precisely when does the macroscopically measurable quantities emerge from the microscopic interactions between the atoms,—could not be answered from any of these approaches. Such an examination of protein interior is necessary not only for purely theoretical discourse, but also for emerging practical applications that attempt to describe proteins from paradigms of nanotechnology, mesoscopic-science and protein engineering.

Over the years some attempts have been made in the paradigm of protein biophysics, to establish the relationship between coupled microscopic fluctuations and their effect on causing the macroscopic behaviour [17, 18]; the scope of these efforts were limited to certain specialized fields and were not general. The singlemost problem in attaining such generality can be traced back to the shear scale of dimensionality (huge number of atoms, many properties, etc.) associated with the description of a process, without compromising with its scope and depth. In this work, we attempt to overcome this difficulty by detecting the commonalities in patterns of emergence of different biophysical properties before describing them with accurate mathematical constructs.

Bulk properties (properties with statistical nature), as we know them, might not acquire their known features suddenly; instead they may appropriate and gather their special characteristics gradually as the description of the system changes gradually from nano to macroscopic scale. Mesoscopic states are states containing the

intermediate details. It is at this particular scale that we can expect to observe the origin and gradual coming to being of many (if not most) biophysical properties. Hence, it assumes enormous importance to construct objective frameworks compatible to the mesoscopic scale of protein description, so that one would be able to scrutinize the multifaceted characteristics of the origin and development of the biophysical property of his/her interest. The assumption of the present work is that the emergence of macroscopic properties can be studied with accuracy and consistency, by studying the relevant aspects of the fluctuating interaction profiles between atoms in sufficient details over a course of time. Assertions from some recent works [19–21] support this assumption strongly. Interatomic interactions manifest themselves through their fluctuation profiles. In the interior of proteins, atoms are tightly packed and the interactions between them assume complicated nature, but fluctuations prevail there too. These fluctuations have been studied from various perspectives. The thermal, conformational fluctuations of a globular protein were decomposed into collective motions and studied extensively [22–25]. In normal modes analysis, the fluctuations are expressed by a linear combination of normal modes [24, 25]. However, construction of a method to observe the emergence of a macroscopically measured property from these microscopic fluctuations along the $micro \longrightarrow meso \longrightarrow macro$ pathway, had never been tried before. The present algorithm attempts to trace back any macroscopically measured property of statistical origin to time-dependent and context-dependent microscopic fluctuations, to observe at which mesoscopic limit of number of atoms, does the property emerge and how it grows, before attaining its macroscopic statistical nature.

This task is (somewhat) daunting, because the probabilistic and structural features of the entire spectrum of microstates sampled by proteins is not clearly known [26, 27]. Sensitivity (quintessentially nonlinear in nature) of the ensemble of microstates to changes in environmental conditions (i.e., pH, temperature, pressure, ligand binding, and concentrations of osmolytes and denaturants) is not well understood either [27]. Most importantly, the manner in which local fluctuations are coupled to larger, more global structural transitions—is far from being deciphered. Hence a mathematical construct that attempts to model the situation must essentially be top-down in its approach (to circumvent the time dependent couplings between each and every local fluctuations), without compromising with ability to capture the emergence of any biophysical property. Rather than assuming the aforementioned interatomic interactions to be linear, we have resorted to model the situation from a (realistic) nonlinear perspective (reasons behind this assertion and the relevance of it in the present study are numerous and are briefly described in the Supplementary Material-01).

It is easy to note that it would be difficult for the simulation-centric studies to answer the questions like, (a) precisely how many number of atoms are necessary to observe the emergence of hydrophobicity? or (b) given the information regarding residual dipolar coupling, to what extent can we observe the dynamics of protein atoms? or (c) given a particular magnitude of ΔpK_a, for residues either in the surface or in the interior, what should be the minimum number of residues that might produces it?—The difficulty in answering these questions arise not due to the limitations of our computational prowess (which, nevertheless, is indeed a problem,—but only of relatively easy nature); but principally due to the non-availability of the exact

mathematical procedures that connect microscopic properties to macroscopic (bulk) properties of macromolecules. The work presented here suggests such a scheme that is capable of answering aforementioned questions in precise terms. The theorems, namely, 'Connection between independence of protein structural parameter and possibility of them being observed' and 'Criteria for protein structural parameters to be observed as components of a linear system' help in establishing a concrete ground for future studies that attempt to study protein mesoscopic properties and emergence of biophysical features.

2 Methodology

While the present work owes its philosophical basis to many [28–33] control theoretic and differential equation based studies (discussed in Supplementary Material-02 with references), it differs from all of the above; because to our knowledge, this is the first attempt to propose a theoretical framework that attempts to observe how the measurable macroscopic biophysical properties of proteins come to being, from (microscopic) time-dependent and context-dependent interatomic interactions. The mathematical structures employed in the present work are not uncommon, but it is their unique and first application in the realm of emergence studies of protein biophysical properties, which makes the present work important.

2.1 Section 1: Definition of the system (a single protein)

2.1.1 Scheme 1: General representation of protein interior parameters with linear differential equation

We approach to objectively describe these time-dependent and context-dependent correlations amongst protein structural parameters by representing any arbitrarily chosen protein with a linear differential equation with sufficient capacity to describe the (time-dependent) dynamic-dependencies of protein structural parameters on one another.

$$\dot{x}\,(t) = A(t)\,x(t) \;+\; f(t) \tag{1}$$

where, x is a n-vector, $A\,(t)$ is an $n \times n$ continuous matrix on an open interval I in R, and $f\,(t)$ is locally square integrable on some arbitrarily chosen interval $(a,\ b)$, viz. $f\,(t) \in L_n^2\ ([a, b]\,)$. In other words, the space of all measurable n-vector functions $f(t)$ defined for $t \in [a, b] = J$ with values $f\,(t) \in R^n$, $t \in J$ such that $\int_a^b \mid f\,(t)\mid^2 dt \;<\; \infty$. In physical interpretation, $\mathbf{A}$ describes the contact-map information for a sub-set of atoms. While the contact-map itself does not change, coordinates of the aforementioned set of atoms might undergo small changes due to time-dependent fluctuations in them. Thus the parameter describing positions of atoms within a protein must necessarily be time-dependent; $x\,(t)$ stands for that. However, thinking from a general perspective, positional coordinate of an atom within a protein is nothing but a property of it, amongst many other properties that characterize that

particular atom. Thus instead of merely position, $x(t)$ can describe any other property of that atom too. In that case, since the contact-map $\mathbf{A}$ remains static, $A(t)\,x(t)$ would be a reliable parameter to model any potential interatomic interaction within a protein. But not all microscopic interactomic interactions account for the emergence macroscopic parameter of interest (studying the cumulative effect due to dispersion interactions might be of interest in order to to observe the emergence of hydrophobicity, however the same might be of little or no use while attempting to observe the emergence of specific heat or mass fractal dimension or denaturation temperature of the protein). The term $f(t)$ captures the trend in the time-dependent variation of the relevant parameter in the microscopic realm, cumulative change of which (supposedly) accounts for the emergence of the macroscopic property under consideration.

We can write Eq. 1, in the following equivalent form as :

$$x(t) = x_0 + \int_{t_0}^{t} [A(s)\,x(s) + f(s)]\,ds \tag{2}$$

Subsequently we can define the successive approximations by the relations:

$$x_0(t) = x_0$$

and

$$x_{n+1}(t) = x_0 + \int_{t_0}^{t} [A(s)\,x_n(s) + f(s)]\,ds, \quad t \in J, \quad n = 0, 1, 2, \ldots$$

The solution of Eq. 2 with $x(t_0) = x_0$ is given by:

$$x(t) = X(t, t_0)\,x_0 + \int_{t_0}^{t} X(t, s)\,f(s)\,ds, \tag{3}$$

where $X(t, t_0)$ is the fundamental matrix solution of homogeneous equation, $\dot{x}(t) = A(t)\,x(t)$, which has the following properties:

(1) $X(t_0, t_0) = I$ (the identity matrix).
(2) $X(t, t_0) = X(t, s)\,X(s, t_0), \qquad t_0 \le s \le t$
(3) $X(t, s) = X^{-1}(s, t)$

It is easy to note that, from a physical perspective, the matrix $X(t, t_0)$ describes a framework to capture correlation profiles of an ongoing interaction over the time domain under consideration. Hence, with suitable context-dependent modifications, it can be used elsewhere for the auto-correlation and cross-correlation studies.

2.1.2 Case 2: General representation of protein interior parameters with non-linear differential equation

Of course, a representation scheme similar to Eq. 3 can be obtained for a non-linear differential equation of the form:

$$\dot{x}\,(t) = A(t)\,x(t) \;+\; f(t,\,x) \tag{4}$$

where $f\,(t,\,x)$ is continuous on $J \times R^n$. For Eq. 4, the solution of $x\,(t)$ with $x\,(t_0) = x_0$ can be written as:

$$x\,(t) \;=\; X\,(t,\,t_0)\,x_0 \;+\; \int_{t_0}^{t} X\,(t,\,s)\,f\,(s,\,x\,(s))\,ds, \tag{5}$$

where $X\,(t,\,t_0)$ is the fundamental matrix solution of homogeneous equation.

2.2 Section 2: Criteria to observe emergence from protein interior dynamics

Based on the definition of the system as provided above (that is, an arbitrarily chosen protein), we now proceed to derive the conditions under which one can observe emergence of any biophysical property from interaction profile between any arbitrarily chosen (sub)set of atoms belonging to the protein. Here we note that exact (context-dependent) causality for the emergence of a macroscopic property from a set of microscopic interactions, might not be known always. However, regardless the physical nature of these properties, mathematical scheme that attempts to describe them do not vary by much. Hence a general scheme to describe emergence, from mathematical standpoint, is attempted here.

2.2.1 Case 1: Criteria to observe emergence of a biophysical property, assuming that proteins are linear systems

We continue with our description of interactions of various structural parameters in protein interior and the emerging property with the differential equation, $\dot{x}\,(t) = A(t)\,x \;+\; f(t)$ (all the symbols retain their meaning from Eq. 1)

Here, in Sect. 2, we attempt to approach the description of the process when Eq. 1 is subjected to a linear observation process, described with simple form, viz:

$$y \;=\; O\,(t)\,x \;+\; \hat{O}\,(t)\,f, \quad y \;\in\; R^n \tag{6}$$

Assuming that the interaction of various structural parameters was taking place in a time interval $[t_0,\,t_1] \subset (a, b)$ and $x\,(t_0) = x_0 \in R^n$, we had arrived at Eq. 3. To lay down the criteria to observe emergence of any biophysical property (with statistical nature) from the set of microscopic interactions between pertinent properties at the atomic scale, we start by admitting:

[1] observation of the relevant phenomenon under question, is itself a time-dependent process; and
[2] if f is known function, for example $f(t) = B(t)\,u(t)$ with $u(t)$ being a control then, in principle the term $\hat{O}(t)\,f$ in Eq. 6 and $O(t)$ times the integral of Eq. 3 can be subtracted from:

$$y(t) = O(t)\;X(t,\,t_0)\,x_0 \;+\; O(t)\int_{t_0}^{t} X(t,\,s)\,f(s)\,ds + \;\hat{O}(t)\,f(t)$$

to yield the modified closed form expression for observation given by:

$$\hat{y}(t) = O(t)\;X(t,\,t_0)\,x_0 \tag{7}$$

The term $X(t, t_0)\,x_0$ in Eq. 7 satisfies the homogeneous equation

$$\dot{x} = A(t)\,x \tag{8}$$

Therefore the expression to represent linear observation on a protein whose interior structural dependencies can be described by a linear differential equation, is obtained as:

$$y(t) = O(t)\,x(t) \tag{9}$$

Hence, the original question about obtaining information about a system described by Eq. 1 with the help of an observation scheme described by Eq. 6, reduces to the same question for the corresponding homogeneous system, described by Eq. 8 and the homogeneous observation described by Eq. 9. This transformation between paradigm of questions marks the change in modes of studies; because the present platform (comprised of Eq. 8 and Eq. 9) offers us a homogeneity in the treatment of the problem in general; something that wasn't ensured in the platform comprised of Eq. 1 and Eq. 6.

However, to make meaningful predictive studies, the present framework needs to be modified further, regarding a suitable description scheme to model temporal frame of reference. Thus, without any loss of generality, we perform the translation of the the origin, so that $\tau = t - t_0$. This accounts for the limits $t_0 \longrightarrow 0$, whereby $t_1 \longrightarrow (t_1 - t_0) = T$.

To formalize the problem, we define the criteria to observe emergence in the manner that, the system represented by Eq. 8 and Eq. 9 is observable (that is, the pair is observable) on the time interval $[0, T]$

$$\textbf{iff}\; y(t) = O(t)\,x(t) = 0, \qquad t \in [0, T] \quad \text{implies}\; x(t) = 0, \quad t \in [0, T]\,.$$

(which is equivalent to the assertion $x(0) = x_0 = 0$).

Thus, the re-defined version of the problem is to identify (and/or develop) the criteria to observe emergence on the matrices $A(t)$ and $O(t)$.

We approach the problem by denoting the space of square integrable r–vector functions on $[0,\ T]$ by $L_r^2\ [0,\ T]$.

At this point we propose the theorem of 'connection between independence of protein structural parameter and possibility of them being observed'.

Theorem (Connection between independence of protein structural parameter and possibility of them being observed) *If the structural parameters corresponding to any protein can be represented by vectors* $x_1, x_2, \ldots, x_k$ *in finite dimensional Euclidean space* R^n*, and if* $x_1(t), x_2(t), \ldots, x_k(t)$ *be the corresponding solutions of Eq.* 8 *for them in* $[0, T]$ *with* $x(0) = x_i,\ i = 1, 2, \ldots, k$*; further if the corresponding observations* y_i *on* $[0, T]$ *can be defined by* $y_i(t) = O(t)\, x_i(t)\ ,\ t \in [0, T]$*; then the observed linear system described by Eq.* 8 *and Eq.* 9 *is observable on* $[0, T]$*; if and only if,* y_i *are linearly independent in* $L_r^2\ [0, T]$ *whenever the* x_i *are linearly independent in the same finite dimensional Euclidean space* R^n.

Proof The solutions $x_i(t)$ are linearly independent in $L_r^2\ [0, T]$ only in the case when x_i are linearly independent in R^n. If Eqs. 8 and 9 is observable and

$$y(t) = \sum_{i=1}^{k} c_i\, y_i(t) = 0 \tag{10}$$

then the corresponding solution also vanishes. In other words, it implies:

$$x(t) = \sum_{i=1}^{k} c_i x_i(t) = 0 \tag{11}$$

and in particular

$$\sum_{i=1}^{k} c_i x_i(0) = \sum_{i=1}^{k} c_i x_i = 0 \tag{12}$$

In the case where x_i's are linearly independent, we have $c_1 = c_2 = \cdots = c_k = 0$. Hence, from Eq. 10, we can conclude that in such a case y_i's will be linearly independent too.

On the other hand (evidently, in the more *general case*) if there exists linearly independent $x_1, x_2, \ldots, x_k$ such that the associated observations of them, namely $y_1(t), y_2(t), \ldots, y_k(t)$ are *not* independent (that is, are dependent on $L_r^2\ [0, T]$), then letting $c_1 = c_2 = \cdots = c_k$ are *not* all zero, such that

$$y(t) = \sum_{i=1}^{k} c_i\, y_i(t) = 0,$$

we notice that $y(t)$ becomes an identically vanishing observation on the solution $x(t) = \sum_{i=1}^{k} c_i x_i(t)$, which is *not* the zero solution of Eq. 8, because in such a case, $x_1 = x_1(0), x_2 = x_2(0), \ldots, x_k = x_k(0)$ *are* linearly independent.

Hence, in such a case, we can conclude that Eqs. 8 and 9 will not be observable.

□

Proof of this theorem ('Connection between independence of protein structural parameter and possibility of them being observed') paves the way for the more general theorem, 'Criteria for protein structural parameters to be observed as components of a linear system'.

Theorem (Criteria for protein structural parameters to be observed as components of a linear system) *The system(protein), described by Eqs.* 8 *and* 9*, is observable on time interval* $[0, T]$ *iff the observability Grammian matrix, given by:*

$$\Phi(0, T) = \int_0^T X^*(t, 0)\, O^*(t)\, O(t)\, X(t, 0)\, dt$$

(where $X^*(t, 0)$ *and* $O^*(t)$ *are the transposes of* $X(t, 0)$ *and* $O(t)$*) is positive definite.*

Proof The solution of $x(t)$ of Eq. 8 corresponding to the initial condition $x(0) = x_0$ is given by:

$$x(t) = X(t, 0)\, x_0$$

and we obtain $y(t) = O(t)\, x(t) = O(t)\, X(t, 0)\, x_0$

$$\begin{aligned} \| y \|^2 &= \int_0^T y^*(t)\, y(t)\, dt \\ &= x_0^* \int_0^T X(t, 0)\, O^*(t)\, O(t)\, X(t, 0)\, dt\, x_0 \\ &= x_0^*\, \Phi(0, T)\, x_0 \end{aligned}$$

- a quadratic form in x_0.

Clearly $\Phi(0, T)$ is a symmetric $n \times n$ matrix.

If $\Phi(0, T)$ is positive definite then:

$$y = 0 \quad \Rightarrow x_0^*\, \Phi(0, T)\, x_0 = 0 \quad \Rightarrow x_0 = 0$$

and then the system described by Eqs. 8 and 9 is observable on $[0, T]$. If, $\Phi(0, T)$ is *not* positive definite then it implies that there exists some $x_0 \neq 0$ such that $x_0^*\, \Phi(0, T)\, x_0 = 0$.

In such a case, $x(t) = X(t, 0)\ x_0 \neq 0$ for $t \in [0, T]$ but since $\|y\|^2 = 0$, it implies $y = 0$.

And therefore, we can conclude that system described by Eqs. 8 and 9 is not observable on $[0, T]$. □

Corollary *If the system described by Eqs. 8 and 9 is observable on* $[s, t]$ *then it is also observable on any interval* $[0,\ T]$ *such that* $0 \leq s \leq t \leq T$.

Proof of corollary We have:

$$\begin{aligned}\Phi(0, T) &= X^*(s, 0) \int_0^T X^*(\tau, s)\, O^*(\tau)\, O(\tau)\, X(\tau, s)\, d\tau\, X(s, 0) \\ &\geq X^*(s, 0)\, \Phi(s, t)\, X(s, 0) \\ &> 0\end{aligned}$$

Hence the proof. □

2.2.2 *Case 2: Criteria to observe emergence of a biophysical property, assuming that proteins are non-linear systems*

In certain cases the dependencies amongst structural parameters within any protein might not be governed by equations with simple linear dependencies. This is probable too, considering that interactions amongst protein structural determinants are time-dependent and context-dependent. Hence, in such a case, without resorting to the linear (simplistic and approximated) case, we will have to describe proteins as non-linear systems. Here, instead of referring to Eq. 1, we start our descriptions with Eq. 4, viz: $\dot{x}(t) = A(t)\, x(t) + f(t,\ x)$,
where x and f are n-vectors, $t \in I$, the real time interval with linear observation $y = O(t)\, x(t)$, where y is an m-vector $(m < n)$ and $A(t)$, $f(t,\ x)$, $O(t)$—are continuous with respect to their arguments.

Here, although admitting that case of non-linear representation scheme is more general (and perhaps, more accurate), we resort mathematically to describe the case as a linear system (Eq. 4) with perturbation $f(t,\ x)$. We assume further that the it is feasible to describe the situation as one where (Eq. 4) is being observed by a quantity y. In such a framework of description, the problem of identifying the criteria-set to observe any emergent property can be formulated as one where: it is required to find an unknown state at the present time t from the quantity y, over the interval $[\theta,\ t]$ where θ denotes some past time; because, since $(m < n)$, the equation $y = O(t)\, x(t)$ does not allow immediate finding of x and y.

At this point, having loosely describing the framework to describe the situation, we proceed to formally define the system as:

NL-Defn: 1 The system can be defined as suitable for observation of emergence of a biophysical property at time t, if there exists $\theta < t$ such that the state of the system at time t, can be identified from the knowledge of the system output over the interval $[\theta,\ t]$.

NL-Defn: 2 If the system is suitable for observation of emergence of a biophysical property at every $t \in I$, it can be called 'completely observable'.

NL-Defn: 3 If the interval of output observation mentioned in NL-Defn-1, can be made arbitrarily small, we speak of differential observability over those intervals.

We start our analysis of non-linear description of protein interior by assuming that (Eq. 4) has a unique solution for any initial condition. If we denote τ as $\theta < \tau < t$, the solution for (Eq. 4) can be asserted to be uniquely defined for $x = x(\tau)$ as the initial condition and (drawing from aforementioned non-linear description of proteins in Case 2 of Section 1) given by:

$$x(t) = X(t, \tau)\, x(\tau) + \int_{\tau}^{t} X(t, s)\, f(s, x(s))\, ds$$

However, since the fundamental matrix is invertible in nature, we have:

$$x(t) = X(\tau, t)\, x(t) - \int_{\tau}^{t} X(\tau, s)\, f(s, x(s))\, ds \tag{13}$$

Correspondingly the $y(\tau)$ will be given by:

$$y(\tau) = O(\tau)\, X(\tau, t)\, x(t) - O(\tau) \int_{\tau}^{t} X(\tau, s)\, f(s, x(s))\, ds \tag{14}$$

Describing the transpose of any matrix, with a star symbol, with a little bit of rearrangement (by multiplying Eq. 14 with $X^*(\tau, t)\, O^*(\tau)$ from the left and integrating within the interval θ to t), we obtain:

$$\begin{aligned}
&\int_{\theta}^{t} X^*(\tau, t)\, O^*(\tau)\, y(\tau)\, d\tau \\
&= \left[\int_{\theta}^{t} X(\tau, t)\, O^*(\tau)\, O(\tau)\, X(\tau, t)\, dr\, x(t)\right] \\
&\quad - \left[\int_{\theta}^{t} X(\tau, t)\, O^*(\tau)\, O(\tau) \int_{\tau}^{t} X(\tau, s)\, f(s, x(s))\, ds\, dr\right] \\
&= \Phi(\theta, t)\, x(t) \\
&\quad - \left[\int_{\theta}^{t} X^*(s, t) \int_{\theta}^{s} X^*(\tau, s)\, O^*(\tau)\, O(\tau)\, X(\tau, s)\, d\tau f(s, x(s))\, ds\right]
\end{aligned}$$

$$= \Phi(\theta, t)\, x(t) - \left[\int_{\theta}^{t} X^{*}(s, t)\, \Phi(\theta, s)\, f(s, x(s))\, ds \right]$$

If the matrix $\Phi(\theta, t)$ is invertible, that is, for a truncated linear system,

$$\dot{x} = A(t)\, x, \qquad y(t) = O(t)\, x \tag{15}$$

is observable; then from the last equation of the array of equations,

$$\left(\int_{\theta}^{t} X^{*}(\tau, t)\, O^{*}(\tau)\, y(\tau)\, d\tau \right.$$
$$\left. = \Phi(\theta, t)\, x(t) - \left[\int_{\theta}^{t} X^{*}(s, t)\, \Phi(\theta, s)\, f(s, x(s))\, ds \right] \right)$$

we obtain:

$$x(t) = \Phi^{-1}(\theta, t) \int_{\theta}^{t} X^{*}(s, t)\, O^{*}(s)\, y(s)\, ds \; + \; \Phi^{-1}(\theta, t)$$
$$\times \int_{\theta}^{t} X^{*}(s, t)\, \Phi(\theta, s)\, f(s, x(s))\, ds$$

If we assign:

$$U_1(t, \theta, s) = \Phi^{-1}(\theta, t) \int_{\theta}^{t} X^{*}(s, t)\, O^{*}(s)$$

and

$$U_2(t, \theta, s) = \Phi^{-1}(\theta, t) \int_{\theta}^{t} X^{*}(s, t)\, \Phi(\theta, s)$$

then the following compact relation can be obtained:

$$x(t) = \int_{\theta}^{t} U_1(t, \theta, s)\, y(s)\, ds \; + \; \int_{\theta}^{t} U_2(t, \theta, s)\, f(s, x(s))\, ds \tag{16}$$

Equation 16 represents the relation of the unknown state x with the observed output y over he interval $[\theta, t]$.

In Eq. 16 the time θ may not be necessarily fixed, and therefore θ can be replaced by τ. Upon carrying out this change, Eq. 16 can be substituted into Eq. 13, and we obtain:

$$x(\tau) = X(\tau, t) \int_{\tau}^{t} U_1(t, \tau, s)\, y(s)\, ds + X(\tau, t) \int_{\tau}^{t} U_2(t, \tau, s)\, f(s, x(s))\, ds - \int_{\tau}^{t} X(\tau, s)\, f(s.x(s))\, ds \quad (17)$$

In compact form,

$$x(\tau) = \int_{\tau}^{t} U_3(t, \tau, s)\, y(s)\, ds + \int_{\tau}^{t} U_4(t, \tau, s)\, f(s, x(s))\, ds \quad \text{for } (\tau < t) \quad (18)$$

where $U_3(t, \tau, s) = X(\tau, t)\, U_1(t, \tau, s)$, and $U_4(t, \tau, s) = X(\tau, t)\, U_2(t, \tau, s) - X(\tau, s)$

On the basis of the derivations above we can put forward a proposition as

NL-Proposition: 1 *Under the condition that the system (protein interior) is describable by the differential equations* $\dot{x}(t) = A(t)\, x(t) + f(t, x)$ *and* $y(t) = O(t)\, x$, *it is globally*

(a) observable at any time instance t,
(b) completely observable, or
(c) differentially observable, if the following conditions hold:
(1) *there exists a constant* $c > 0$, *such that* $det\, \Phi(t, \theta) \geq c$.
(2) *Equation* 16 *has a unique solution for any y, which is continuous on* $[\theta, t]$ *(a) for some* $\theta < t$, *in case of an observable system at time t, (b) for all t and for some* $\theta < t$, *in the case of a completely observable system, or (c) for all t and for all* $\theta < t$, *in the case of a differentially observable system*

A careful of reading of the description of the situation suggests that Eq. 16 in 'NL-proposition-1' can be replaced by Eq. 17. In the case of such replacement the same results will be valid, but with some simple change of variables. However the question that whether Eq. 16 or Eq. 17 has a unique solution is difficult to evaluate. The difficulty arises because we notice that if the interval of integration in Eq. 16 or Eq. 17 is suitably changed, Eq. 16 or Eq. 17 may then be considered as a nonlinear operator equation on a continuous function space. Thus, we will have to resort to Banach's contraction mapping theorem to these nonlinear equations.

Without venturing into the general case, we consider a simple non-linear system (particular case) described by:

$$\dot{x} = A(t)\,x + \epsilon f(t, x) \tag{19}$$

$$y = O(t)\,x \tag{20}$$

where ϵ is a scaler positive constant. We assume that the following condition is satisfied:

$$\| f(t, x_1) - f(t, x_2) \| \leq K \| x_1 - x_2 \|, \quad (K \geq 0) \tag{21}$$

A general solution $x(t)$ for Eq. 18 with $x = x(\tau)$ as a formal initial condition is:

$$x(\tau) = X(\tau, t)\,x(t) - \epsilon \int_{\tau}^{t} X(\tau, t)\, f(s, x(s))\, ds \tag{22}$$

We create an analogue of Eq. 16 derivation from Eq. 13, by starting with Eq. 22. Therefore:

$$x(t) = \Phi^{-1}(\theta, t) \int_{\theta}^{t} X^*(s, t)\, O^*(s)\, y(s)\, ds + \epsilon \Phi^{-1}(\theta, t) \int_{\theta}^{t} X^*(s, t)\, \Phi(s, \theta)\, f(s, x(s))\, ds \tag{23}$$

Substituting Eq. 23 on Eq. 22, we obtain:

$$x(\tau) = X(\tau, t)\,\Phi^{-1}(t, \theta) \int_{\theta}^{t} X^*(s, t)\, O^*(s)\, y(s)\, ds + \epsilon X(\tau, t)\,\Phi^{-1}(\theta, t) \int_{\theta}^{t} X^*(s, t)\,\Phi(s, \theta)\, f(s, x(s))\, ds - \epsilon \int_{\tau}^{t} X^*(\tau, s)\, f(s, x(s))\, ds \quad (\theta \leq \tau \leq t)$$

We denote this condition as 'NL-cond-1'.

Consequently, to ensure that the system described by Eqs. 19 and 20, is observable, it is sufficient that the inverse of $\Phi(t,\ \theta)$ exists, the solution of Eq. 23 exists and it is unique.

At this point, if we assume that there exists solution of $x_1, x_2\ (x_1 \neq x_2)$ of Eq. 23 for a given y, then, using Eq. 21, we obtain:

$$\begin{aligned}
&(|x_1\ (\tau) - x_2\ (\tau))\\
&\leq \epsilon \int_{\tau}^{t} |\ X\ (\tau,\ s)\ |K|x_1\ (s) - x_2\ (s)\ |ds\\
&\quad + \epsilon |X\ (\tau,\ t)\ \Phi^{-1}\ (t, \theta)\ | \int_{\theta}^{t} |X^*\ (s, t)\ \Phi\ (s,\ \theta)\ |K|x_1\ (s) - x_2\ (s)\ |ds\\
&\leq \epsilon k_1\ (t,\ \theta)\ (t - \tau)\ \|\ x_1 - x_2\ \| + \epsilon k_2\ (t,\ \theta)\ \|\ x_1 - x_2\ \|\ (t - \theta)
\end{aligned}$$

where

$$k_1\ (t - \theta) = \max_{\theta < \tau < s < t} |X\ (\tau,\ t)\ \Phi^{-1}\ (t,\ \theta)\ ||X^*\ (s,\ t)\ \Phi\ (s,\ \theta)\ K$$

From this, there exists a $k\ (t,\ \theta)$ such that:

$$\|\ x_1 - x_2\ \| \leq \epsilon\ k\ (t,\ \theta)\ (t - \theta)\ \|\ x_1 - x_2\ \| \tag{24}$$

where $k\ (t, \theta) = k_1\ (t,\ \theta) + k_2\ (t,\ \theta)$

Hence, most importantly, if ϵ satisfies the inequality:

$$\epsilon k\ (t,\ \theta)\ (t - \theta) < 1 \tag{25}$$

it follows that $x_1 = x_2$ on $[\theta,\ t]$.

This contradiction leads to the next proposition for a sufficient condition to ensure observation of emergence of a biophysical property, belonging to the the system described by Eqs. 19 and 20; since the condition, 'NL-cond-1' necessarily guarantees the existence of solutions of Eq. 23.

Hence, we can finally state:

NL-Proposition: 2 *The system described by Eqs.* 19 *and* 20*, is globally (a) observable at the instance t, (b) completely observable or (c) differentially observable, if the following conditions hold:*

(1) *there exists a constant* $c > 0$*, such that* $det\ \Phi\ (t,\ \theta)\ \geq\ c$.
(2) *a positive constant,* ϵ *, satisfies*
$\epsilon < 1/k\ (t,\ \theta)\ (t - \theta)$
(a) for some $\theta < t$*, in case of an observable system at time t,*
(b) for all t and for some $\theta < t$*, in the case of a completely observable system, and*
(c) for all t and for all $\theta < t$*, in the case of a differentially observable system.*

3 Results and discussion

3.1 Applicability of the algorithm on three different spheres in protein biophysics

The present algorithm proposes a rigorous and reliable template for a series of algorithms that can be constructed to study the emergence of various biophysical factors within biological macromolecules. However, the actual implementation of these ideas might require superlative computational facilities that are not easily available in contemporary scenario; though the possibility of using such facilities in near future seems genuine. Due to prohibitive computational cost, actual implementation of these algorithms could not be achieved in the present work. In the absence of obtained data, in this section we present exact schemes to study the emergence of measurable biophysical properties from the mathematical discourse presented above. Out of innumerable possibilities of application of this algorithm, we talk about three paradigms; on which, the present study can (tangibly) be enormously impacting. In each of these (extremely well-studied) spheres of protein biophysics, the applicability of the present algorithm is clearly mentioned, alongside the utilitarian benefits that application of this algorithm can provide it with.

3.2 Applicability 1: Case of hydrophobicity

3.2.1 Scope of applicability of the present algorithm (in studying origin of hydrophobicity)

The origin of hydrophobicity can be traced back to some kind of inter-atomic interactions, was never questioned by any of the proposed theories on the origin of hydrophobicity. These inter-atomic interactions are bound to cause certain fluctuations in the spatial coordinate of the atoms. The present work has derived the conditions where macroscopically measured hydrophobicity can be traced back to the precise number of atoms (described by the fluctuations in their spatial coordinate due to inter-atomic interactions), responsible for the emergence of the macroscopic property named hydrophobicity. Categorically, the user needs to assign the measured (computationally or experimentally) change of (cumulative) hydrophobicity content (as provided by the constructs proposed in [16] or [34]) due to any set of atoms, to the variable $\dot{x}$; the contact-map information for these atoms in the matrix A; and the partial charge of atoms or any other (microscopic atom-specific) property that the user considers responsible to cause (macroscopic) hydrophobicity, to the (context-dependent) function $f(x, t)$, of the Eq. 4. Or alternatively, he can input the (time-dependent) coordinate information of the relevant set of atoms in the variable x and keeping the rest of the enlisted parameters invariant, attempt to measure the change of hydrophobicity content by scrutinizing the magnitude of $\dot{x}$. Owing to the high degree of flexibility inherent in the algorithm, the effectiveness of the proposed scheme lies primarily with the discretion of the user; especially in the appropriate choice of parameters to be assigned to the function $f(x, t)$. For example, a researcher might assign parameters related

to dispersion forces to $f(x, t)$ of the Eq. 4, or else he can assign the atomic hydrophobicity magnitudes [35] for every atom, to the same. Efficiency with which the change in the content of hydrophobicity for the system of atoms is measured would be different.

Such categorical information about the number and character of atom-cluster that produces hydrophobicity due to their inter-atomic interactions, becomes indispensable in order to probe recent questions regarding the nature of hydrophobicity. For example, a steady flow of opinions could be recorded over the last decade, which argued that hydrophobic effect is not necessarily an entropic phenomenon; it can be enthalpic or entropic depending on the temperature and the geometric characteristics of the solute [36–39]. The difficulty with attempting this problem stems primarily from the inherent contradiction, namely, geometrical descriptions (distinguishable object based) and thermodynamic (statistical) descriptions work at two different levels of systemic descriptions. While the former is primarily bottom-up (nano scaled) in its nature the later is top-down (macroscopic). The present theory provides a quantitative tool-set to examine the emerging properties in their nascent form in mesoscopic scale. Utility of the present scheme therefore lies in the fact that it can study unambiguously how exactly from the nano-scale (small number of distinguishable atoms), the macroscopic property of thermodynamic nature (hydrophobicity) is emerging.

3.2.2 Scope of applicability of the present algorithm (understanding hydrophobicity and PMF)

It has been found experimentally that the relationship between "bulk hydrophobic interaction", exposure of hydrophobic residues from its pure phase to water and "pair hydrophobic interaction" potential of mean force (PMF) in water is nonlinear [39, 40]. Although various experimental studies from varying perspectives (studies related to virial coefficients, Kirkwood-Buff integrals, and on related spatially integrated quantities [39–42], have studied the nature of hydrophobicity and many have attempted to focus purely on the multiple facets of PMF [43–45], the spatial dependence of PMF with direct experimental mechanism is difficult to obtain. In this context, to know the precise nature of spatial extent of PMF, rather than resorting to the simulation-centric studies, a researcher can resort to the rigorous mathematical treatise presented here. By describing the fluctuating magnitude of temperature dependent spatial extent of PMF with the variable $\dot{x}$, the (time-dependent) contact map with matrix A, and the expression for dependency of exposure of hydrophobic residues from its pure phase to water on the (time-dependent and context-dependent) "pair hydrophobic interaction" PMF of residues in Eq. 4, a consistent scheme can be constructed to solve for the spatial extent of PMF for the concerned set of residues. As has been mentioned before, the scheme constructed in this work is a general one and the effectiveness of its depends crucially on the judicious choice of parameters that are assigned to various terms in Eq. 4 (or in Eq. 1, if the description is linear).

3.3 Applicability 2: Case of polarizability

3.3.1 Scope of applicability of the present algorithm (studying the nature of pK_a shift)

A low polarizability in the interior of the protein implies a low magnitude of dielectric constant, which in turn implies a conducive environment for electrostatic interactions. All pH-dependent properties of proteins are (predictably) governed by the (long range) electrostatic interactions between ionizable side chains. Owing to this coupling to chemical protonation equilibria, protein electrostatics can be probed directly through measurements of pKa values [46–50]. The effect of electrostatic interactions is usually quantified in terms of the shift ΔpK_a, of the pKa value of an ionizable group in a protein relative to the pKa values of the same group in a small reference molecule in dilute aqueous solution.

Many aspects of protein pKa shifts are known to us. However, solution to the basic inverse problem **[P-1]**, viz., given a particular magnitude of ΔpK_a, for residues either in the surface or in the interior, what should be the minimum number of residues, which might produces it ?—is not easily obtainable in a general sense. Furthermore, despite immense efforts, computational and/or theoretical approaches that can reliably predict the large pKa shifts observed for buried residues in a general sense—remains difficult to find. While one aspect of these problems lies in the computational problems while considering ionization-induced water penetration and conformational changes in pKa calculations, the other aspect points to the lack of a theoretically sound schemes that can describe the emergence of a macroscopic property from its inception to the subsequent phases, as the number of residues that contribute to produce the property increases over time. Substituting the pKa values for every amino acid (in a sub-set of amino acids under consideration) in variable x, incorporating the contact-map information in the matrix A, and describing desolvation or conformational fluctuations (or any other parameter that the researcher thinks necessary) in the context-dependent (non-linear) function f in the Eq. 4, one can attempt to obtain a quantitative magnitude for ΔpK_a as the output $\dot{x}$.

The need for such a thorough scheme becomes even acute when one attempts to delve into the uniform dielectric continuum model of protein ineterior electrostatics. In such a model, the entire effect due to polarizability is described through a single dielectric constant (DC). (As a result, electrostatically highly heterogeneous [51] and anisotropic [52] protein interior is represented through an overtly simplified construct. The shortcomings of such model have been commented upon by many [48, 51, 53]). In such a case, the magnitude of DC becomes a complex function of the extents in which formal charges, partial charges, and dipoles [48] are considered. The effective DC is calculated from the response of the entire protein to an externally applied electric field; which in its turn, is calculated from the total dipole moment fluctuation of the protein, through computational methods (however, since the magnitude of dipole moment fluctuation is significantly affected by charged surface residues [54–56], to what accuracy will such a construct be taking into account the self-energy of deeply buried ionizable residues,—remains unclear). Although rigorous computational examinations of the pKa shifts have been undertaken from the framework of

macroscopic dielectric continuum models (semi-macroscopic partial charge [56, 57], lattice dipole [48] models, all-atom simulations [50]) answer to another basic and general inverse problem **(P-2)**, namely, electrostatic effects due to how many buried residues are being reflected in a measured magnitude of dipole moment fluctuation of the protein—could not be found from the purview of the aforementioned studies.

The commonality in **P-1** and **P-2** is striking. Both of them are asking extremely basic questions. Computational constructs, regardless of how sophisticated they are, cannot provide the answers to them. Instead, an elaborate mathematical model, which treats protein interior as a fluctuating, nonlinear (DC behaves in a nonlinear manner [58]) and time-dependent system, might help us in finding the answers to these basic (inverse problems). The mathematical model presented in the present work attempts to achieve precisely the same.

3.3.2 Scope of applicability of the present algorithm (studies with residual dipolar couplings)

Accurate measurement of residual dipolar couplings (RDCs) in weakly aligned proteins can in principle provide incisive information about the structure and dynamics of them in the solution state [59, 60]. But the problem in this operation stems from the nature of measured data, which usually embodies a convolution of the structural and dynamic properties. Amongst many other aspects, the sensitivity of RDCs to internal motions has been recognized by many as an enormously interesting question [61–64]. This is so because, unlike the conventional spin relaxation and chemical exchange-based studies, RDCs are sensitive to motions that span a wide range of time scales; and henceforth, they might be considered as potent probes to monitor biologically relevant motions [61]. But, many structure refinement protocols for analyzing dipolar data implicitly assume that internal motions are either absent, negligibly small, or uniform and axially symmetric in nature [65, 66]. Although some sporadic attempts have been made to study dynamics in protein interior, a rigorous theoretical framework that solves the inverse problem **(P-3)**, namely, given the information regarding residual dipolar coupling, to what extent can we observe the dynamics of protein atoms,—has not been proposed. Study with particular case [61] tends to suggest that some internal motions can well be into the range of observability. Here describing the change in the magnitude of the (emergent) dipolar coupling as the dependent variable of Eq. 4, the contact-map information for neighboring atoms in the matrix A, and relevant information (either the residual pKa values, or some cumulative measure of atomic partial charges at residual level or any other parameter set that the researcher thinks pertinent) in $f(x,t)$,—one can attempt to observe at which threshold level of a number of atoms does the measured magnitude of the residual dipolar coupling emerge.

Two other cases, where the present algorithm could be utilized to extract (hitherto unexplored) information about electrostatics of protein–protein interactions and details of mean-field nature of Poisson–Boltzmann theory, are explained in Supplementary Material-03.

3.4 Applicability 3: Case of drug-discovery and computational-chemistry

In the paradigm of drug discovery (and computational chemistry, in general) an outstanding problem can be stated as, given the information regarding the structure of a protein active site and a list of potential small molecule ligands, predict the binding mode and estimate the binding affinity for each ligand [67]. This problem has multiple aspects associated with it and has been a field of intense computational and experimental analysis over the last 15 years. However, certain basic questions in this paradigm still remain unanswered and in the absence of theoretically deduced unambiguous criteria set, approaches to these questions often provide inconsistent results [68]. The entire operation of docking can be summarized into two operations; first, the operation of "posing"; viz., the appropriate positioning of the correct conformer of a ligand in the active site (combination of conformation and orientation being known as a "pose"). Second, the operation of "scoring", where poses are selected and ranked with respect to some scoring function [67, 68]. Although apparently straightforward, this two step process involves many a complex (non-linear) physico-chemical interactions from geometric perspective and in their bid to simultaneously address these issues through this two-step process, many approximations and inadequate constructs are resorted to. These have been identified in many recent works [68–71]. (For example, simplistic treatment of electrostatics, electronic polarization, aqueous desolvation, and ionic influences; lack of accounting for entropy changes in the protein and the ligand on binding; inadequate weighting of proton positions (tautomers, rotamers) and charge states (ionization) of both protein and ligand; assumptions in many (but not all) of the cases that active site is rigid (possibly including tightly bound water molecules) and that only the small molecule can move; etc.) A close scrutiny amongst these drawbacks points immediately to an underlying connecting factor. Many of these shortcomings exist because the precise mode of emergence of these features from a certain set of number of atoms, is not known; furthermore, the (non-linear) dependencies that these properties might be having on one another is difficult to decipher too, because of the same reason. A solution to these problems can be found from examining the situation from a coherent perspective where the distinguishability of a non-statistical (non-macroscopic, non-thermodynamic) system of atoms can be ensured; but at the same time, conditions for observability of the emergence of macroscopic (statistical) properties are appropriately identified. The control theoretic approach presented in this work, might help in quantifying many of these features, from the perspective of inverse problem; where the atomic origin of these features will be addressed from a bottom-up mathematical standpoint without delving into the depths of physical and/or chemical dependencies. For example, a computational framework can be set-up where the user assigns measured (computational or experimental) change of entropy of either protein or ligand to the variable $\dot{x}$, the neighboring atom information of it in the form of a contact-map to the matrix A, the conformational fluctuation based information for each amino acids involved, to x; and finally, some relevant parameter that maps residual conformational fluctuations with (macroscopic) entropy, to the nonlinear context-dependent function $f(x,t)$. (Otherwise, the cumulative effect of conformational fluctuations, as the internal energy of the set of atoms under consideration, can be mapped to entropy and

assigned to $f(x, t)$.) With such a scheme, for a known set of $\dot{x}$ values, the corresponding x; or the more direct; known x to unknown $\dot{x}$ studies—can be attempted.—Obviously, an expert in this sphere of knowledge can ascertain the feasibility of attempting certain problems, the general mathematical framework stays valid for every set of parameter.

On the other hand, the scoring functions serve as objective function to classify diverse poses of a single ligand in the receptor binding site before estimating the binding affinities of different receptor–ligand complexes (and ranking them) upon the docking of a compound database is performed [72]. Interestingly, the drawbacks of scoring functions, as elaborated in a recent work [73] (failure to accommodate subtle physical effects affecting the experimental binding energy; viz., the treatment of polar groups in the ligand or the protein being desolvated upon binding but failing to find a matching polar interaction in the complex, treatment of hydrophobic patches of the ligand exposed to the solvent upon binding, a more comprehensive treatment of loss of rotational and translational entropy; etc.)—also suffer from the same nature of problems as the ones explained in that last paragraph. Here also, a consistent scheme that does not involve itself with the mind-boggling complexity of the physico-chemical interactions, but circumvents it by assuming that all the aforementioned properties come to being due to some or the other form of interatomic interactions between a set of atoms involving electromagnetic forces, before attempting to identify the number of atoms necessary to produce the property under consideration—can be of extreme utility. Since the algorithm proposed here targets the transition zone between nano-scale individualistic properties to mesoscopic and subsequently macroscopic properties, by targeting the number of interacting atoms rather than the property itself;—it can overlook the complex physico-chemical details. Yet, it can monitor, from which threshold of atoms, the emergence of a particular property is observed. This helps him to identify the possible dependencies one property can have on the others and predict which ones are more fundamental than the others.

4 Conclusion

An algorithm to study the lower threshold of emergence for various biophysical properties, is presented in this work. Categorical linkages between rigorous mathematical backbone with protein biophysical properties are established. An exact knowledge of these limits hold paramount utiliterian importance in the paradigm of the nascent field Nano-Bioscience. They, on the other hand, provide the contemporary state of protein interior knowledge with constructs to investigate the fundamental questions of protein biophysics. In near future, when the computational facilities become less prohibitive, these algorithms can be implemented to answer the questions like "precisely how many atoms are necessary for us to observe hydrophobicity in a protein under a specified biological context"?

Acknowledgments This study was supported by COE scheme, Department of Biotechnology, Government of India. One of the authors, Anirban, wants to thank Om Prakash Pandey, for all his efforts to make the author understand the key concepts of drug-discovery, as have been discussed in this work.

References

1. M. Vendruscolo, N.V. Dokholyan, E. Paci, M. Karplus, Phys. Rev. E. **65**, 061910 (2002)
2. H. Kaya, H.S. Chan, Phys. Rev. Lett. **85**, 4823–4826 (2000)
3. N.C. Fitzkee, P.J. Fleming, P.J. Gong, N. Panasik Jr., T.O. Street, G.D. Rose, Trends Biochem. Sci. **30**, 73–80 (2005)
4. J. Tissen, J. Fraaije, J. Drenth, H. Berendsen, Acta Cryst. D **50**, 569–571 (1994)
5. J.C. Phillips, Proc. Natl. Acad. Sci. USA **106**, 3107–3112 (2009)
6. J.C. Phillips, Proc. Natl. Acad. Sci. USA **106**, 3113–3118 (2009)
7. A. Banerji, I. Ghosh, Eur. Biophys. J. **38**, 577–587 (2009)
8. K. Rother, R. Preissner, A. Goede, C. Frömmel, Bioinformatics **19**, 2112–2121 (2003)
9. S. Reuveni, R. Granek, J. Klafter, Phys. Rev. Lett. **100**, 208101 (2008)
10. M. MacDonald, N. Jan, Can. J. Phys. **64**, 1353–1355 (1986)
11. H. Li, S. Chen, H. Zhao, Biophys. J. **58**, 1313–1320 (1998)
12. J.S. Helman, A. Coniglio, C. Tsallis, Phys. Rev. Lett. **53**, 1195–1197 (1984)
13. G.L. Millhauser, E.E. Salpeter, R.E. Oswald, Proc. Natl. Acad. Sci. USA **85**, 1503–1507 (1988)
14. T.G. Dewey, Phys. Rev. E **60**, 4652–4658 (1999)
15. M.B. Enright, D.M. Leitner, Phys. Rev. E **71**, 011912 (2005)
16. A. Banerji, I. Ghosh, PLoS ONE **4**(10), e7361 (2009). doi:10.1371/journal.pone.0007361
17. K.A. Peterson, M.B. Zimmt, S. Linse, R.P. Domingue, M.D. Fayer, Macromolecules **20**, 168–175 (1987)
18. O. Annunziata, D. Buzatu, J.G. Albright, . Langmuir **21**, 12085–12089 (2005)
19. T. Lazaridis, M. Karplus, Biophys. Chem. **100**, 367–395 (2003)
20. W. Bialek, R. Ranganathan, arXiv:0712.4397v1 (2007)
21. M. Socolich, S.W. Lockless, W.P. Russ, H. Lee, K.H. Gardner, R. Ranganathan, Nature **437**, 512–518 (2005)
22. B. Brooks, M. Karplus, Proc. Natl. Acad. Sci. USA **80**, 6571–6575 (1983)
23. M.M. Teeter, D.A. Case, J. Phys. Chem. **94**, 8091–8097 (1990)
24. N. Go, T. Noguti, T. Nishikawa, Proc. Natl. Acad. Sci. USA **80**, 3696–3700 (1983)
25. M. Levitt, C. Sander, P.S. Stern, J. Mol. Biol. **181**, 423–447 (1985)
26. H. Yang, G. Luo, P. Karnchanaphanurach, T.M. Louie, I. Rech, S. Cova, L. Xun, X.S. Xie, Science **302**, 262–266 (2003)
27. S.T. Whitten, B. Garcıa-Moreno, V.J. Hilser, Proc. Natl. Acad. Sci. USA **102**, 4282–4287 (2005)
28. K.E. Starkov, J. Math. Sci. **78**, 433–496 (1996)
29. I. Miroshnik, V. Nikiforov, A. Fradkov, *Nonlinear and Adaptive Control of Complex Systems* (Kluwer, Dordrecht, 1999)
30. A.G. Kartsatos, *Advanced Ordinary Differential Equations* (Mariner publishing company, Tampa, FL, 1980)
31. J.P. Gauthier, I.A.K. Kupka, SIAM J. Control Optim. **32**, 975–994 (1994)
32. Y. Yamamoto, I. Sugiura, J. Optim. Theor. Appl. **13**, 660–669 (1974)
33. E.D. Sontag, Eur. J. Control **11**, 396–435 (2005)
34. D. Eisenberg, R.M. Weiss, T.C. Terwilliger, Proc. Natl. Acad. Sci. USA **81**, 140–144 (1984)
35. L.A. Kuhn, C.A. Swanson, M.E. Pique, J.A. Tainer, E.D. Getzoff, Prot. Str. Fun. Gen. **23**, 536–547 (1995)
36. N.T. Southall, K.A. Dill, J. Phys. Chem. B **104**, 1326–1331 (2000)
37. Y.K. Cheng, P.J. Rossky, Nature **392**, 696–699 (1998)
38. T. Lazaridis, Acc. Chem. Res. **34**, 931–937 (2001)
39. S. Shimizu, H.S. Chan, J. Chem. Phys. **113**, 4683–4700 (2000)
40. R.H. Wood, P.T. Thompson, Proc. Natl. Acad. Sci. USA **87**, 946–949 (1990)
41. A. Sacco, M. Holz, J. Chem. Soc. Faraday Trans. **93**, 1101–1104 (1997)
42. W. Blokzijl, J.B.F.N. Engberts, Angewandte Chemie (International Edn in English) **32**, 1545–1579 (1993)
43. S. Ludemann, R. Abseher, H. Schreiber, O. Steinhauser, J. Am. Chem. Soc. **119**, 4206–4213 (1997)
44. D. van Belle, S.J. Wodak, J. Am. Chem. Soc. **115**, 647–652 (1993)
45. G. Hummer, S. Garde, A.E. Garcia, M.E. Paulatis, L.R. Pratt, Proc. Natl. Acad. Sci. USA **95**, 1552–1555 (1998)
46. A. Warshel, Biochemistry **20**, 3167–3177 (1981)

47. J. Antosiewicz, J.A. McCammon, M.K. Gilson, Biochemistry **35**, 7819–7833 (1996)
48. C.N. Schutz, A. Warshel, Proteins **44**, 400–417 (2001)
49. W.R. Forsyth, J.M. Antosiewicz, A.D. Robertson, Proteins **48**, 388–403 (2002)
50. T. Simonson, J. Carlsson, D.A. Case, J. Am. Chem. Soc. **126**, 4167–4180 (2004)
51. V.P. Denisov, J.L. Schlessman, B. Garcıa-Moreno, B. Halle, Biophys. J. **87**, 3982–3994 (2004)
52. X.J. Song, J. Chem. Phys. **116**, 2 (2002)
53. K. Talley, C. Ng, M. Shoppell, P. Kundrotas, E. Alexov, PMC Biophys. **1**, 2 (2008)
54. T. Simonson, C.L. Brooks, J. Am. Chem. Soc. **118**, 8452–8458 (1996)
55. J.W. Pitera, M. Falta, W.F. Gunsteren, Biophys. J. **80**, 2546–2555 (2001)
56. B. Honig, A. Nicholls, Science **268**, 1144–1149 (1995)
57. D. Bashford, D.A. Case, Ann. Rev. Phys. Chem. **51**, 129–152 (2000)
58. M.B. Partenskii, P.C. Jordan, J. Phys. Chem. **96**, 3906–3910 (1992)
59. J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, Proc. Natl. Acad. Sci. USA **92**, 9279–9283 (1995)
60. J.R. Tolman, H.M. Al-Hashimi, L.E. Kay, J.H. Prestegard, J. Am. Chem. Soc. **123**, 1416–1424 (2001)
61. J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, Nat. Struct. Biol. **4**, 292–297 (1997)
62. L.E. Kay, Nat. Struct. Biol. **5**, 513–517 (1998)
63. M.P. Foster, C.A. McElroy, C.D. Amero, Biochemistry **46**, 331–340 (2007)
64. B. Vogeli, L. Yao, J. Am. Chem. Soc. **131**, 3668–3678 (2009)
65. G.M. Clore, A.M. Gronenborn, Proc. Natl. Acad. Sci. USA **95**, 5891–5898 (1998)
66. G.M. Clore, D.S. Garrett, J. Am. Chem. Soc. **121**, 9008–9012 (1999)
67. G.P.A. Vigers, J.P. Rizzi, J. Med. Chem. **47**, 80–89 (2004)
68. P.C.D. Hawkins, A.G. Skillman, A. Nicholls, J. Med. Chem. **50**, 74–82 (2007)
69. G.L. Warren et al., J. Med. Chem. **49**, 5912–5931 (2006)
70. V. Maiorov, R.P. Sheridan, J. Chem. Inf. Model. **45**, 1017–1024 (2005)
71. P.M. Marsden, D. Puvenendrampillai, J.B.O. Mitchell, Org. Biomol. Chem. **2**, 231–237 (2004)
72. T. Schulz-Gasch, M. Stahl, J. Mol. Mod. **9**, 47–57 (2003)
73. M. Jacobsson, A. Karlen, J. Chem. Inf. Mod. **46**, 1334–1343 (2006)