<table>
<thead>
<tr>
<th>Title</th>
<th>Molecular nonlinear dynamics and protein thermal uncertainty quantification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Xia, Kelin; Wei, Guo-Wei</td>
</tr>
<tr>
<td>Citation</td>
<td>Xia, K., &amp; Wei, G.-W. (2014). Molecular nonlinear dynamics and protein thermal uncertainty quantification. Chaos, 24(1), 013103-.</td>
</tr>
<tr>
<td>Date</td>
<td>2014</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/41112">http://hdl.handle.net/10220/41112</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2014 American Institute of Physics. This paper was published in Chaos and is made available as an electronic reprint (preprint) with permission of American Institute of Physics. The published version is available at: [<a href="http://dx.doi.org/10.1063/1.4861202">http://dx.doi.org/10.1063/1.4861202</a>]. One print or electronic copy may be made for personal use only. Systematic or multiple reproduction, distribution to multiple locations via electronic or other means, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper is prohibited and is subject to penalties under law.</td>
</tr>
</tbody>
</table>
Molecular nonlinear dynamics and protein thermal uncertainty quantification

Kelin Xia¹ and Guo-Wei Wei¹,²,³,a)
¹Department of Mathematics, Michigan State University, Michigan 48824, USA
²Department of Electrical and Computer Engineering, Michigan State University, Michigan 48824, USA
³Department of Biochemistry and Molecular Biology, Michigan State University, Michigan 48824, USA

(Received 16 October 2013; accepted 22 December 2013; published online 8 January 2014)

This work introduces molecular nonlinear dynamics (MND) as a new approach for describing protein folding and aggregation. By using a mode system, we show that the MND of disordered proteins is chaotic while that of folded proteins exhibits intrinsically low dimensional manifolds (ILDMs). The stability of ILDMs is found to strongly correlate with protein energies. We propose a novel method for protein thermal uncertainty quantification based on persistently invariant ILDMs. Extensive comparison with experimental data and the state-of-the-art methods in the field validate the proposed new method for protein B-factor prediction. © 2014 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4861202]

Protein folding produces characteristic and functional three-dimensional structures from unfolded polypeptides or disordered coils.¹⁻⁵ The emergence of extraordinary complexity in the protein folding process poses astonishing challenges to theoretical modeling and computer simulations.⁶,⁷ The present work introduces molecular nonlinear dynamics (MND), or molecular chaotic dynamics, as a theoretical framework for describing and analyzing protein folding. We represent the dynamics of macromolecular particles (i.e., atoms or coarse-grained superatoms) by a set of intrinsically chaotic oscillators. A geometry to topology mapping is employed to create driving and response relations among chaotic oscillators. We unveil the existence of intrinsically low dimensional manifolds (ILDMs) in the chaotic dynamics of folded proteins. Additionally, we reveal that the transition from disordered to ordered conformations in protein folding increases the transverse stability of the ILDM. Stated differently, protein folding reduces the chaoticity of the nonlinear dynamical system, and a folded protein has the best ability to tame chaos. Furthermore, we bring to light the connection between the ILDM stability and the thermodynamic stability, which enables us to quantify the disorderliness and relative energies of folded, misfolded, and unfolded protein states. Finally, we exploit chaos for protein uncertainty quantification and develop a robust chaotic algorithm for the prediction of Debye-Waller factors, or temperature factors, of protein structures.

I. INTRODUCTION

Protein folding process is one of the most important processes in life. How exactly a random coil of polypeptide chain folds into functional structure or configuration after being translated from a sequence of mRNA remains a mystery. Anfinsen’s dogma of sequence-structure-function,⁸ in which a protein’s function depends on its uniquely folded three-dimensional (3D) structure and its structure is determined by the amino acid sequence, is challenged due to the discovery that many partially folded or intrinsically unstructured proteins remain functional despite of the lack of uniquely folded 3D structures.⁹⁻¹⁰ Kinetically and thermodynamically regulated competing pathways, including disordered aggregation, degradation, folding and unfolding, convert linear chains of amino acids translated from sequences of mRNA into degraded fragments, protofibrils, amyloid-like fibrils, amyloids, intrinsically disordered proteins, partially disordered proteins, and folded structures.¹⁻⁵ The formation of disordered proteins is often exploited by living systems to perform novel and diverse biological functions. Unfortunately, aggregated or misfolded proteins are often associated with sporadic neurodegenerative diseases, such as mad cow disease, Alzheimer’s disease, and Parkinson’s disease.¹¹,¹² Currently, there is a lack of efficient means for the characterization of disordered aggregation and the quantification of orderliness, which are crucial to the understanding of the molecular mechanism of degenerative diseases.

The emergence of complexity in self-organizing living systems, including protein folding, poses fabulous challenges to their quantitative description and prediction.⁶,⁷ Cyrus Levinthal suggested that there are near 10⁹⁵ possible conformations for a relatively small polypeptide of 100 residues,¹³ while an average human protein of 480 residues might have an astronomical number of conformations. The complexity is extraordinary since human proteins are coded by over 20 000 genes. The straightforward sampling of the full conformational space becomes unfeasible for large proteins. It takes many months for molecular dynamics (MD) simulations, the main workhorse of computational protein folding,¹⁴ to come up with a very poor copy of what Nature administrates perfectly within a tiny fraction of a second. In fact, disordered aggregation, unfolding and folding often occur at slower time scales and involve larger length scales, which are essentially intractable to full atomic simulations.⁶,⁷ Coarse-grained (CG) representations of polypeptides are employed to reduce the number of degrees of freedom, extend molecular modeling, and bridge with experimental...
observations. An active research topic is how to improve the accuracy of CG models so as to differentiate near degenerate energy landscapes of some conformations generated by the protein folding process. Elastic network models (ENMs), including Gaussian network model (GNM) and anisotropic network model, represent folded proteins as elastic mass-spring networks to investigate its mechanical flexibility and long-time stability beyond the reach of molecular dynamics.\textsuperscript{15–20} In general, ENMs can be viewed as a time-independent molecular mechanics derived from their corresponding time dependent molecular mechanics by using the time-harmonic approximation. An underlying assumption adopted in all of the above-mentioned theoretical models is that protein folding, misfolding, and aggregation are to be modeled with some deterministic dynamical systems, which reinforce Anfinsen’s dogma and exclude any unpredictability and molecular degradation. However, the existence of degraded fragments, protofibrils, amyloid-like fibrils, amyloids, misfolds, and intrinsically disordered proteins highlights the fundamental limitation of current simulation models.

Chaos is ubiquitous in nature. The discovery of the sensitivity of initial conditions, one of three signatures of chaos, dated to the 1880s by Henri Poincaré.\textsuperscript{21} However, little attention was paid to chaos until Edward Lorenz’s work on nonlinear dynamics and description of butterfly effect in weather forecasting in the 1960s, which underpin the modern theory of deterministic chaos.\textsuperscript{22} Understanding deterministic chaos is of theoretical and practical importance.\textsuperscript{23–26} Some detailed analysis of chaos dynamics can be found in the literature.\textsuperscript{27,28} Mathematically, a chaotic dynamics also exhibits dense periodic orbits and topologically mixing of its phase space open sets.\textsuperscript{27,28} Chaos has been observed in a vast variety of realistic systems, including Belousov-Zhabotinski reactions, nonlinear optics, Chua-Matsumoto circuit, Rayleigh-Bénard convection, meteorology, population dynamics, psychology, economics, finance solar system, protein dynamics,\textsuperscript{29} and heart and brain of living organisms.\textsuperscript{30} Various chaos control strategies have been proposed.\textsuperscript{23–26} However, the natural ability of protein folding in controlling chaos has not been unveiled yet.

Imagine that a folded protein is a Greek chorus where all particles sing with a synchronized voice on the dramatic action, while an unfolded protein is an anharmonic chaotic orchestra where each particle plays its own rhythm with its own instrument. In this work, we propose a method, called MND, to describe this chaotic behavior of a protein. The fundamental assumption of the MND is that each protein configuration (structure) is the direct manifestation of the internal and external interactions of the protein. Protein functions, such as stability and flexibility, are entirely determined by the protein structure or geometry. Therefore, there is no need to resort to protein interactions to predict thermal fluctuations. Our approach has two basic ingredients: the connectivity matrix extracted from the geometry. In this section, we discuss the details of our methods, algorithms, and their further applications to protein folding.

### A. Geometry to topology mapping

Topological relations or connectivities among molecular particles are basic ingredients in MND model. As discussed above, this topological information can be extracted from the geometric properties of a given molecule of interest. Let us consider a molecule of $N$ particles located at $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$, where $\mathbf{r}_j \in \mathbb{R}^3$, where particles are either atoms, amino acids residues, or other superatoms in the molecule. The distance between the $i$th and $j$th particles is given by $d_{ij}(\mathbf{r}_i, \mathbf{r}_j) = \| \mathbf{r}_i - \mathbf{r}_j \|_2$. The connectivity matrix must satisfy the driven and response relation between two dynamics systems. Additionally, we assume that all particles are mutually connected and their interactions decay as a function of their distance $A_{ij}(d_{ij})$. The simplest form for the connectivity matrix
is the Kirchhoff (or connectivity) matrix generated by cutoff distances $\sigma_{ij}$

\[
A_{ij} = \begin{cases} 
1, & \forall d_{ij} \leq \sigma_{ij}, i \neq j \\
0, & \forall d_{ij} > \sigma_{ij}, i \neq j \\
- \sum_{j \neq i} A_{ij}, & \forall i = j.
\end{cases}
\] (1)

To account for the distance effect in a more realistic manner, it is also convenient to employ smooth and monotonically decreasing radial basis functions or delta sequence kernels of positive type.\(^{31}\) Here, we consider generalized exponential functions

\[
A_{ij} = \begin{cases} 
e^\frac{-d_{ij}/\sigma_{ij}}{k}, & \forall i \neq j, k = 1, 2, \cdots \\
- \sum_{j \neq i} A_{ij}, & \forall i = j,
\end{cases}
\] (2)

and power-law functions

\[
A_{ij} = \begin{cases} \left(\frac{d_{ij}/\sigma_{ij}}{\sigma_{ij}}\right)^{-v}, & \forall i \neq j, v > 1, \\
- \sum_{j \neq i} A_{ij}, & \forall i = j,
\end{cases}
\] (3)

where $\sigma_{ij}$ are characteristic distances between particles. In the present model, $\sigma_{ij}$ can be used as a set of fitting parameters. Note that some of the above matrix expressions have also been used in other flexibility analysis approaches.\(^{19,32–36}\)

However, the present construction of these functional forms was based on the driven and response relation of coupled dynamical systems.\(^{27}\)

Expressions (1)–(3) map a molecular geometry into topological relations or connectivities. The connectivity matrix $A$ is an $N \times N$ symmetric, diagonally dominant matrix. Note that elements in the connectivity matrix are not interaction potentials among particles. For simplicity, we set characteristic distances to $\sigma_{ij} = \sigma$ for all amino acid residues. Figure 1 illustrates the non-diagonal elements of two connectivity matrices, one for the native protein structure of 2mcm and the other for completely unfolded conformation generated from pulling the structure of 2mcm. It is seen that the native protein shows a heterogeneity (or small-world property)\(^{37}\) in its connectivity matrix, which contributes to the protein stability. Whereas the connectivity matrix of the completely unfolded conformation has little nonlocal connection.

### B. Molecular nonlinear dynamics

To introduce the MND, let us consider a folding protein that constitutes $N$ particles and has the spatiotemporal complexity of $\mathbb{R}^{3N} \times \mathbb{R}^+$. Assume that the molecular mechanics of the protein is described by molecular nonlinear dynamics having a set of $N$ nonlinear oscillators of dimension $\mathbb{R}^{nN} \times \mathbb{R}^+$, where $n$ is the dimensionality of a single nonlinear oscillator. Let us consider an $n \times N$-dimensional nonlinear system for $N$ interacting chaotic oscillators

\[
\frac{d u}{d t} = F(u) + E u,
\] (4)

where $u = (u_1, u_2, \cdots, u_N)^T$ is an array of state functions for $N$ nonlinear oscillators, $u_i = (u_{i1}, u_{i2}, \cdots, u_{in})^T$ is an $n$-dimensional nonlinear function for the $i$th oscillator, $F(u) = (f(u_1), f(u_2), \cdots, f(u_N))^T$ is an array of nonlinear functions of $N$ oscillators, and $E = e A \otimes \Gamma$. Here, $e$ is the overall interaction strength, $A$ is the $N \times N$ connectivity matrix defined in Sec. II A, and $\Gamma$ is an $n \times n$ linking matrix. Although it is possible to consider all the physical interactions among protein particles, we feature the importance of the protein distance geometry in this work. As such, we map a protein geometry into a set of topological relations or connectivities.

### C. Stability analysis of the ILDM

We use the protein connectivity matrix to define the driving and response relation of nonlinear chaotic oscillators. Amazingly, an $N$-time reduction in the spatiotemporal complexity can be achieved, leading to an ILDM of dimension $\mathbb{R}^n \times \mathbb{R}^+$. Formally, the $n$-dimensional ILDM is defined as

\[
u(t) = \nu(t) \cdots \nu(N(t) = s(t)),
\] (5)

where $s(t)$ is a synchronous state or reference state.

To understand the stability of the ILDM of protein chaotic dynamics, we define a transverse state function as $w(t) = u(t) - S(t)$, where $S(t)$ is a vector of $N$ identical components $(s(t), s(t), \cdots, s(t))^T$. Obviously, the invariant ILDM is given by $w(t) = u(t) - S(t) = 0$. Therefore, the stability of the ILDM can be analyzed by $\frac{d w(t)}{d t} = (D F(s) + E) w$, which can be studied by the following linearized equation:

\[
\frac{d w(t)}{d t} = (D F(s) + E) w,
\] (6)

where $DF(s)$ is the Jacobian of $F$.

To further analyze the stability of Eq. (6), we diagonalize connectivity matrix $A$

\[
A \phi_j = \lambda_j \phi_j, \quad j = 1, 2, \cdots, N,
\] (7)

where $\{\phi_j\}^N_{j=1}$ are eigenvectors and $\{\lambda_j\}^N_{j=1}$ are the associated eigenvalues. These eigenvectors span a vector space in which a transverse state vector has the expansion

---

FIG. 1. The connectivity matrices for two conformations of protein 2mcm generated by using Eq. (2) with $k = 2$ and $\sigma_2 = 6$. Diagonal elements have been excluded to emphasize non diagonal interactions. (a) The connectivity matrix of the native structure of 2mcm shown in Fig. 3(a) indicates much nonlocal interactions. (b) The connectivity matrix of a completely unfolded conformation shown in Fig. 3(g) demonstrates little nonlocal interactions.
\[ w(t) = \sum_j v_j(t) \phi_j(t). \] (8)

Therefore, the stability problem of the ILDM is equivalent to the following stability problem:
\[ \frac{d\mathbf{v}_j(t)}{dt} = (Df(s) + \varepsilon \lambda_j \Gamma) \mathbf{v}_j(t), \quad j = 1, 2, \ldots, N, \] (9)
where \( Df(s) \) is the diagonal component of \( \mathbf{DF}(s) \). The stability of Eq. (9) is determined by the largest Lyapunov exponent \( \lambda_{\text{max}} \), namely, \( \lambda_{\text{max}} < 0 \), which can be decomposed into two contributions
\[ \lambda_{\text{max}} = \lambda_f + \lambda_c, \]
where \( \lambda_f \) is the largest Lyapunov exponent of the original \( n \)-dimensional chaotic system \( \frac{d\mathbf{x}}{dt} = f(\mathbf{x}) \), which can be easily computed for most chaotic systems. Here, \( \lambda_c \) depends on \( \lambda_i \) and \( \Gamma \). The largest eigenvalue \( \lambda_1 \) equals 0, and its corresponding eigenvector represents the homogeneous motion of the ILDM, and all of other eigenvalues \( \lambda_j, j = 2, 3, \ldots, N \) govern the transverse stability of the ILDM. Let us consider a simple case in which the linking matrix is the unit matrix \( (\Gamma = I) \). Then stability of the ILDM is determined by the second largest eigenvalue \( \lambda_2 \), which enables us to estimate the critical interaction strength \( \varepsilon_c \) in terms of \( \lambda_2 \) and \( \lambda_f \).
\[ \varepsilon_c = \frac{\lambda_f}{-\lambda_2}. \] (10)

The dynamical system reaches the ILDM when \( \varepsilon > \varepsilon_c \) and is unstable when \( \varepsilon \leq \varepsilon_c \). The eigenvalues of protein connectivity matrices are obtained with a standard matrix diagonalization algorithm.

### III. NUMERICAL RESULTS

To demonstrate the full strength of the molecular nonlinear dynamics, we present three major results in this work. First, we show that the transition from chaos to periodicity can be induced by protein interactions. Additionally, we demonstrate the correlation between protein randomness and the dimensionality of chaos in the protein MND. Finally, we utilize the persistence of the invariant ILDM for the uncertainty quantification of protein structures due to thermal fluctuations.

In our MND model Eq. (4), the exact expressions of the nonlinear functions \( \mathbf{F}(\mathbf{u}) \) is not specified. Mathematically, it is well known that a double well or triple well function can lead to multiple local minima, which support bistable or multiple stable states. The system becomes unstable or chaotic under certain perturbation. In this paper, for simplicity, we just choose a set of Lorenz attractors\(^{22}\) to illustrate our ideas. The Lorenz equation is three-dimensional \( \mathbf{u}_i = (x_i, y_i, z_i)^T \),
\[ \begin{align*}
\frac{dx_i}{dt} &= \alpha(y_i - x_i) \\
\frac{dy_i}{dt} &= \gamma x_i - y_i - x_i z_i \\
\frac{dz_i}{dt} &= x_i y_i - \beta z_i, \quad i = 1, 2, \ldots, N,
\end{align*} \] (11)
where parameters \( \alpha > 0, \beta > 0 \) and \( \gamma > 0 \) are to be specified for each given system.

The Lorenz Eq. (11) can be analyzed with the Poincaré section and the first return map.\(^{38}\) Parameter \( \gamma \) has been used to classify certain behavior of the Lorenz Eq. (11). The origin is a fixed point when \( \gamma < 1 \). When \( \gamma = 1 \), the system is at the saddle-node bifurcation point. For \( \gamma > 1 \), there is a pair of fixed points given by
\[ u^\pm = (\pm \sqrt{\beta(\gamma - 1)}, \pm \sqrt{\beta(\gamma - 1)}, \gamma - 1)^T. \]

For \( \alpha > \beta + 1 \), the above twin fixed points are stable if
\[ \frac{\gamma}{\alpha} < \frac{\alpha + \beta + 3}{\alpha - (\beta + 1)}. \]

At \( \alpha = \beta + 1 \), the twin fixed points are no longer stable due to the Hopf bifurcation.

By choosing the classical parameter values \( \alpha = 10 \), \( \beta = 8/3 \) and \( \gamma = 28 \), almost all points in the phase space go to a strange attractor—the Lorenz attractor. The Lorenz equations are solved by using the forward Euler scheme and/or the fourth order Runge-Kutta scheme in this work.

Note that the Lorenz oscillators used in the present work do not directly govern the realistic dynamics of protein residues or atoms. Instead, they just provide a simple representation of the protein dynamics. It turns out that such a simple representation is able to capture certain fundamental features of the protein as described below.

### A. Folding induced transition from chaos to periodicity

As a proof of principle, we first demonstrate that folded proteins are able to control chaos. To this end, we consider the MND generated by the coarse-grained representation of bacteriocin AS-48 (protein data bank (PDB) ID: 1e68) by using its 70 amino acid residues. As a comparison, we create a reference dynamical system with 70 weakly coupled Lorenz attractors. In this case, we set \( \mathbf{E} \) in Eq. (4) as \( \mathbf{E} = \eta \mathbf{B} \otimes \Gamma \), where \( \mathbf{B} \) is given by
\[ B_{ij} = \begin{cases} 1, & \forall d_{ij} \leq \sigma_{ij}, j > i \\ -1, & \forall d_{ij} \leq \sigma_{ij}, j < i \\ 0, & \text{otherwise.} \end{cases} \] (12)

The linking matrix is given as
\[ \Gamma = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \] (13)

In this setting, there is no realistic protein structure information in the connectivity matrix. Consequently, the dynamics of original seventy weakly coupled Lorenz attractors is chaotic as shown in Figs. 2(a) and 2(b). In fact, each chaotic attractor resembles the well-known wings of butterfly as plotted in Fig. 2(a). Here, \( \eta = 7 \) and \( \sigma_{ij} = 4 \Lambda \) are used. The Lorenz parameters are set to \( \alpha = 1 \), \( \gamma = 60 \), and \( \beta = \frac{8}{3} \).
To appropriately consider protein interactions, we set \( E = (\varepsilon A + \eta B) \otimes \Gamma \), where the connectivity matrix \( A \) takes the form given in Eq. (2) and \( \varepsilon = 10 \). Surprisingly, the MND of bacteriocin AS-48 exhibits a shocking transition from chaos to periodicity as depicted in Figs. 2(c) and 2(d). The chaotic dynamics of each oscillator undergoes a Hopf bifurcation as illustrated in Fig. 2(c). It is interesting to note that there is a constant delay in the dynamics of any two adjacent oscillators and there is a lag synchronization in the protein dynamics. As such, the protein dynamics can be described by an \( \mathbb{R}^3 \times \mathbb{R}^+ \) dimensional ILDM, which achieves a stunning 70-fold reduction in complexity and dimensionality.

### B. High dimensional chaos associated with unfolding proteins

Having demonstrated the ability of transforming high dimensional chaos to a three-dimensional periodic orbit by a folded protein, we further analyze the dynamics of a set of forty one conformations generated from pulling the structure of macromomycin (PDB ID: 2mcm) with a constant force. It is quite standard to create partially folded and unfolded protein conformations by pulling the structure of macromomycin. Forty partially folded or totally unfolded conformations are generated by pulling the folded structure (PBD ID: 2mcm). (a) The \( C_x \) atoms of the native folded structure. (b) The synchronous dynamics of a 3D ILDM for the native folded structure. (c) The \( C_x \) atoms of partially ordered Conformation 14. (d) The partial ordered chaotic dynamics of Conformation 14. The first 30 residues are unfolded, which leads to the chaotic dynamics in their nonlinear oscillators. The dynamics of the last two residues are also chaotic for the same reason. Interesting synchronized domains occur in the middle for partially folded residues. (e) The \( C_x \) atoms of more disordered Conformation 28. (f) The more chaotic dynamics of Conformation 28 showing a higher degree of randomness. Unfolded residues are in their chaotic motions. Compared to the dynamical behavior of Conformation 14, there are fewer synchronized domains and an average synchronized domain involves fewer oscillators. (g) The \( C_x \) atoms of completely unfolded Conformation 41. (h) The completely chaotic dynamics of Conformation 41 showing 336-dimensional chaotic motions.

FIG. 2. Transition from chaos to periodicity in the chaotic dynamics model (CDM) of bacteriocin AS-48 (PDB ID 1e68). (a) The butterfly wing pattern for one of 70 chaotic oscillators. (b) The solution of original 70 chaotic oscillators. (c) The periodic orbit of the ILDM for bacteriocin AS-48. (d) Bacteriocin AS-48 induced Hopf bifurcation from chaos. All of 70 nonlinear oscillators are in one lag synchronized periodic orbit.

FIG. 3. The spectacular correlation between folded residues and controlled chaos in the dynamics of macromomycin. Forty partially folded or totally unfolded conformations are generated by pulling the folded structure (PBD ID: 2mcm). (a) The \( C_x \) atoms of the native folded structure. (b) The synchronous dynamics of a 3D ILDM for the native folded structure. (c) The \( C_x \) atoms of partially ordered Conformation 14. (d) The partial ordered chaotic dynamics of Conformation 14. The first 30 residues are unfolded, which leads to the chaotic dynamics in their nonlinear oscillators. The dynamics of the last two residues are also chaotic for the same reason. Interesting synchronized domains occur in the middle for partially folded residues. (e) The \( C_x \) atoms of more disordered Conformation 28. (f) The more chaotic dynamics of Conformation 28 showing a higher degree of randomness. Unfolded residues are in their chaotic motions. Compared to the dynamical behavior of Conformation 14, there are fewer synchronized domains and an average synchronized domain involves fewer oscillators. (g) The \( C_x \) atoms of completely unfolded Conformation 41. (h) The completely chaotic dynamics of Conformation 41 showing 336-dimensional chaotic motions.
of 80 000 simulation steps is employed for each protein and new conformations are extracted at the frequency of every 2000 steps.

We make use of the generalized exponential functions given in Eq. (2) to construct the connectivity matrix. Related parameters in the MND are chosen as follows: the characteristic distance \( s_{ij} = 10\text{Å}, \) \( k = 2, \) and the interaction strength \( \varepsilon = 0.12. \) For parameters in the Lorenz dynamic system, we use \( \alpha = 10, \) \( \gamma = 28, \) and \( \beta = \frac{8}{3}. \) Random numbers of range \([0,1]\) are used for as initial conditions for all oscillators. The unit matrix is used for linking matrix (\( F = I \)) in this case and in the rest of this paper. The forward Euler scheme with the time increment of \( h = 10^{-2} \) is used for the time integration.

Our basic results are demonstrated in Fig. 3. Conformation 1, shown in Fig. 3(a), is the folded structure obtained by a short period of relaxation of the crystal structure from the PBD, while Conformations 2 to 41 are increasingly less folded due to the increase in the pulling force during their generation, see three typical ones in Figs. 3(c), 3(e), and 3(g). The same set of random initial data is assigned to all the nonlinear oscillators in all conformations. Figure 3(b) shows the occurrence of a stable ILDM, a synchronous chaos, in the dynamics of the folded structure. What spotlights the uniqueness and importance of the native structure is that none of any other conformations that are partially ordered and/or essentially disordered is able to tame chaos in their nonlinear dynamics under the same condition. Indeed, the dynamical systems of Conformations 2 to 41 are unstable, chaotic, and 336-dimensionally complex due to 112 residues. Consequently, they are very sensitive to initial values, simulation algorithms, and time increments. However, it is still possible to extract some useful physical information from their unstable chaotic dynamics with well designed numerical experiments. Figures 3(d) and 3(f) indicate synchronized domains in overall non-synchronized chaotic dynamics. Amazingly, there is a miraculous correspondence between synchronized domains and the locations of partially folded amino acid residues, which clearly indicates that protein folding leads to the control of chaos. It is interesting to note that synchronized domains are strikingly persistent over time. They appear to be locked in certain ranges of solution values and show much less fluctuation and volatility than fully chaotic oscillators do. Finally, as shown in Fig. 3(h), the dynamics of the completely unfolded conformation is fully chaotic, which reinforces our observation that protein folding tames chaos and induces ILDM.

### 3. Critical interaction strength

It remains to understand why the critical interaction strength in Eq. (10) is able to determine the transverse stability of the chaotic dynamics of a given molecular conformation. It is seen that the critical interaction strength is an inverse function of the second largest nonzero eigenvalue of the molecular connectivity matrix. The latter is a manifestation of the molecular structure. As shown in Fig. 1, the structure of a folded protein gives rise to a “small-world” connectivity network, while the structure of completely unfolded conformation has little nonlocal interactions. This implies that the transverse stability of the ILDM of a molecule is ultimately determined by its structure. Since the thermodynamical stability of a molecular structure is characterized by its total energy, there must be a one-to-one correspondence between the critical interaction strength and the total energy. As such, we can estimate the relative energies of protein folding conformations based on their second largest nonzero eigenvalue of their connectivity matrix. To verify this hypothesis, we analyze the nonlinear dynamics of protein structures 1ubq, 7rsa, and 2,mcm. For each structure, we create a set of unfolded conformations by molecular dynamics. Figures 4(b), 4(d), and 4(f) show good agreements between energies estimated by using the largest eigenvalues and those obtained from molecular dynamics simulations.
Various procedures of conformation generation are also considered, including using different initial structure preparations, number of integration steps, and pulling velocities, which lead to different sets of conformations. However, our findings presented in Figs. 3 and 4 are not affected by these variations.

C. ILDM based B-factor prediction

In this section, we introduce a new approach for protein flexibility analysis. The proposed method is based on transverse stability of the ILDM for the prediction of atomistic B-factors, or temperature factors of a given molecular structure. B-factor is a measure of the mean-squared atomic displacement due to thermal motion and possible experimental uncertainties. In general, an atom with a larger B-factor implies it is more flexible and atoms with smaller B-factors are relatively rigid. The analysis of B-factors provides insights on the large-scale and long-time functional behaviors of native state macromolecules. This information is complementary to that obtained from atomic detail simulation techniques.

There are many other interesting approaches for the flexibility analysis in the literature. For example, normal mode analysis (NMA) has been proposed to uncover the intrinsic structural flexibility of the protein and study biomolecular systems like lysozyme. NMA approach has a large number of variations. Tirion proposed ENM by simplifying the interaction potential in the NMA. By introducing the idea from polymer science, Bahar et al. use GNM to describe protein flexibility. They employed the C2 atom representation of proteins based on local packing density and contact topology. Anisotropic fluctuations are considered in anisotropic network model (ANM). Parameters in these models are calibrated with Debye-Waller factor or B-factors. Crystal structures have been also taken into considerations. Due to the simplified potential and reduced representation, these coarse-grained based ENM and GNM approaches gain popularity and have been applied to the study of macromolecules or protein complexes, such as, hemoglobin, F1 ATPase, chaperonin GroEL, viral capsids, and ribosome. More applications can be found in a few good review papers. Our ILDM based method should be potentially useful for solving these problems as well.

It is interesting to note that the intriguing dynamics of a protein ILDM is exponentially stable and persistently invariant. It will be wonderful if one can take advantage of such properties in practical biophysical studies. To this end, we propose an ILDM based new method for protein rigidity analysis. Our essential idea is to perturb the dynamics of each particle in a macromolecule in the transverse direction. Because of the stability of the ILDM, the nonlinear system must return to its original orbit, just like the free induction decay of the spin dynamics in nuclear magnetic resonance (NMR) experiments. Similar to the T2 and T1 relaxations in an NMR experiment, the total relaxation time after the transverse perturbation, defined as the time used to recover the original state within a factor of 1/e, is a measure of the strength of its particle-particle and particle-environment interactions. For a given particle, stronger interactions with neighboring particles and environment lead to a shorter relaxation time, which translates into higher rigidity and lower B-factor. Therefore, the direct connection between the thermodynamical stability and the ILDM stability enables us to quantitatively estimate atomic temperature factors in a molecule.

Figure 5(a) illustrates the relaxation process of a perturbed nonlinear dynamics of protein 2nuh. The instantaneous perturbation is propagated from the nearest neighboring amino acid residues to a wider region over a time period before gradually fades off. By recording the relaxation time, one is able to predict the B-factor of an amino acid residue and compare it to the experimental data given by X-ray crystallography. Figures 5(b)–5(d) provide such comparisons for
three protein structures, namely 2nuh, 1aru, and 4dr8. It is seen that our results obtained from the perturbation of the ILDM are in a very good consistency with those of X-ray data.

The parameter details used in our B-factor prediction are as follows. The chaotic dynamical system parameters are chosen as $\alpha = 1, \gamma = 12, \beta = \frac{8}{3}$. Both forward Euler scheme and the fourth order Runge-Kutta scheme are used in the time integration to validate each other. Appropriate time increment that satisfies the stability requirement under given interaction strength is used. Initially, all oscillators rest in their steady states. A perturbation is then employed on the $j$th particle: $z_j \rightarrow 2z_j$. The perturbation procedure is repeatedly carried out for all particles in the protein to compute their relaxation time values, which are converted to B-factors by linear regressions. For the predictions of B-factors in Figs. 5(b)–5(d), $\sigma_{ij} = 20\AA$ is employed for 2nuh, $\sigma_{ij} = 20\AA$ is used for 1aru, and $\sigma_{ij} = 5\AA$ is chosen for 4dr8.

### 1. Comparison with GNM

To demonstrate the robustness of the present chaotic dynamics model, a set of 60 protein structures downloaded from the PDB is considered. All the structures are obtained by the X-ray diffraction with resolution about 2.0 Å, and they are free from multiple conformations, which mean that for each protein, all the atomic occupancies equal 1.0. The correlation coefficient $C_r$ is used to measure the consistency between theoretical predictions and experimental data.

![Graphs showing protein flexibility analysis](image1)

**FIG. 5.** Protein flexibility analysis by the present ILDM method. (a) The impact and relaxation of the transverse perturbation of the ILDM at a given amino acid residue (the 52th C$_\alpha$). The residues in horizontal axis are listed in the descending order according to their distances with respect to the perturbed residue. (b)–(d) The experimental B-factors and ILDM predictions for protein 1aru, 2nuh, and 4dr8. The correlation coefficients are respectively 0.913, 0.866, and 0.751 for three predictions.

![Graph showing optimal parameter search](image2)

**FIG. 6.** The optimal parameter search for exponential type of probability density estimators in Eq. (2). A set of 60 proteins is used and their correlation coefficients are averaged over the set. The averaged correlation coefficients are plotted with respect to the change of parameter $\sigma_{ij}$ and $k$. It is seen that except for $k = 0.5$, all other kernels achieve their maximal correlation coefficient in the range $\sigma_{ij} \in [4,10]$.

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>$N$</th>
<th>MND</th>
<th>GNM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1AIE</td>
<td>31</td>
<td>0.405</td>
<td>0.155</td>
</tr>
<tr>
<td>1AKG</td>
<td>16</td>
<td>0.223</td>
<td>0.185</td>
</tr>
<tr>
<td>1BX7</td>
<td>51</td>
<td>0.679</td>
<td>0.706</td>
</tr>
<tr>
<td>1ETL</td>
<td>12</td>
<td>0.833</td>
<td>0.628</td>
</tr>
<tr>
<td>1ETM</td>
<td>12</td>
<td>0.683</td>
<td>0.432</td>
</tr>
<tr>
<td>1ETN</td>
<td>12</td>
<td>0.007</td>
<td>$-0.274$</td>
</tr>
<tr>
<td>1FF4</td>
<td>65</td>
<td>0.676</td>
<td>0.674</td>
</tr>
<tr>
<td>1GK7</td>
<td>39</td>
<td>0.718</td>
<td>0.821</td>
</tr>
<tr>
<td>1GVD</td>
<td>52</td>
<td>0.618</td>
<td>0.591</td>
</tr>
<tr>
<td>1HJE</td>
<td>13</td>
<td>0.681</td>
<td>0.616</td>
</tr>
<tr>
<td>1KYC</td>
<td>15</td>
<td>0.860</td>
<td>0.754</td>
</tr>
<tr>
<td>1NOT</td>
<td>13</td>
<td>0.801</td>
<td>0.523</td>
</tr>
<tr>
<td>1O06</td>
<td>20</td>
<td>0.886</td>
<td>0.844</td>
</tr>
<tr>
<td>1OB4</td>
<td>16</td>
<td>0.764</td>
<td>0.750*</td>
</tr>
<tr>
<td>1OB7</td>
<td>16</td>
<td>0.637</td>
<td>0.652*</td>
</tr>
<tr>
<td>1PJ1</td>
<td>29</td>
<td>0.569</td>
<td>0.625</td>
</tr>
<tr>
<td>1PEF</td>
<td>18</td>
<td>0.915</td>
<td>0.808</td>
</tr>
<tr>
<td>1PEN</td>
<td>16</td>
<td>0.267</td>
<td>0.270</td>
</tr>
<tr>
<td>1Q9B</td>
<td>43</td>
<td>0.762</td>
<td>0.656</td>
</tr>
<tr>
<td>1RUJ</td>
<td>36</td>
<td>0.391</td>
<td>0.431</td>
</tr>
<tr>
<td>1U06</td>
<td>55</td>
<td>0.350</td>
<td>0.434</td>
</tr>
<tr>
<td>1UOY</td>
<td>64</td>
<td>0.690</td>
<td>0.671</td>
</tr>
<tr>
<td>1USE</td>
<td>40</td>
<td>0.097</td>
<td>$-0.142$</td>
</tr>
<tr>
<td>1VRZ</td>
<td>21</td>
<td>0.633</td>
<td>0.677*</td>
</tr>
<tr>
<td>1XY2</td>
<td>8</td>
<td>0.181</td>
<td>0.562</td>
</tr>
<tr>
<td>1YJO</td>
<td>6</td>
<td>0.344</td>
<td>0.434</td>
</tr>
<tr>
<td>1YZM</td>
<td>46</td>
<td>0.815</td>
<td>0.901</td>
</tr>
<tr>
<td>2DSX</td>
<td>52</td>
<td>0.298</td>
<td>0.127</td>
</tr>
<tr>
<td>2JKU</td>
<td>35</td>
<td>0.854</td>
<td>0.656</td>
</tr>
<tr>
<td>2NLS</td>
<td>36</td>
<td>0.580</td>
<td>0.530</td>
</tr>
<tr>
<td>2OL9</td>
<td>6</td>
<td>0.565</td>
<td>0.689</td>
</tr>
<tr>
<td>2OLX</td>
<td>4</td>
<td>0.794</td>
<td>0.885</td>
</tr>
<tr>
<td>6RXN</td>
<td>45</td>
<td>0.559</td>
<td>0.594</td>
</tr>
</tbody>
</table>

### TABLE I. Comparison of B factor prediction by parameter free MND and Gaussian normal mode (GNM) for small-sized data set.$^*$ The asterisk sign indicates improved prediction with modified protein data.
where \( \{ B'_i, i = 1, 2, \cdots, N \} \) are theoretical prediction of B-factors and \( \{ B''_i, i = 1, 2, \cdots, N \} \) are a set of experimental B-factors obtained directly from the PDB. Here, \( B' \) and \( B'' \) are the statistical averages of theoretical and experimental B-factors, respectively.

Through the calculation of correlation coefficients averaged over 60 proteins, we search the optimal value of parameter \( \sigma_i \) for a set of \( k \) values (0.5, 1.0, 1.5, 2.0, 2.5). It is seen from Figure 6 that the best average correlation coefficient is achieved around \( \sigma_{ij} \in [4, 10] \), except when \( k = 0.5 \).

Based on this analysis, we create a parameter free version of our MND model by setting \( \sigma_{ij} = 4 \AA \) and \( k = 1 \). This version is employed in all of our comparison with GNM.

TABLE II. Comparison of B factor prediction in terms of correlation coefficients by parameter free MND and Gaussian normal mode (GNM) for medium-sized data set.\(^6\) The asterisk sign indicates improved prediction with modified protein data.

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>N</th>
<th>MND</th>
<th>GNM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1ABA</td>
<td>87</td>
<td>0.757</td>
<td>0.613</td>
</tr>
<tr>
<td>1CYO</td>
<td>88</td>
<td>0.729</td>
<td>0.741</td>
</tr>
<tr>
<td>1FK5</td>
<td>93</td>
<td>0.590</td>
<td>0.485</td>
</tr>
<tr>
<td>1GXU</td>
<td>88</td>
<td>0.702</td>
<td>0.421</td>
</tr>
<tr>
<td>1H71</td>
<td>83</td>
<td>0.332</td>
<td>0.549</td>
</tr>
<tr>
<td>1LR7</td>
<td>73</td>
<td>0.684</td>
<td>0.620</td>
</tr>
<tr>
<td>1N7E</td>
<td>95</td>
<td>0.451</td>
<td>0.497</td>
</tr>
<tr>
<td>1NNX</td>
<td>93</td>
<td>0.761</td>
<td>0.631</td>
</tr>
<tr>
<td>1NOA</td>
<td>113</td>
<td>0.601</td>
<td>0.615</td>
</tr>
<tr>
<td>1OPD</td>
<td>85</td>
<td>0.315</td>
<td>0.398</td>
</tr>
<tr>
<td>1QAU</td>
<td>112</td>
<td>0.665</td>
<td>0.620</td>
</tr>
<tr>
<td>1R7J</td>
<td>90</td>
<td>0.365</td>
<td>0.368</td>
</tr>
<tr>
<td>1UHA</td>
<td>83</td>
<td>0.709</td>
<td>0.638*</td>
</tr>
<tr>
<td>1ULR</td>
<td>87</td>
<td>0.612</td>
<td>0.495</td>
</tr>
<tr>
<td>1USM</td>
<td>77</td>
<td>0.794</td>
<td>0.798</td>
</tr>
<tr>
<td>1V05</td>
<td>96</td>
<td>0.582</td>
<td>0.632</td>
</tr>
<tr>
<td>1W2L</td>
<td>97</td>
<td>0.638</td>
<td>0.397</td>
</tr>
<tr>
<td>1X3O</td>
<td>80</td>
<td>0.536</td>
<td>0.654</td>
</tr>
<tr>
<td>1Z21</td>
<td>96</td>
<td>0.545</td>
<td>0.433</td>
</tr>
<tr>
<td>1ZVA</td>
<td>75</td>
<td>0.604</td>
<td>0.690</td>
</tr>
<tr>
<td>2BP9</td>
<td>36</td>
<td>0.541</td>
<td>0.680*</td>
</tr>
<tr>
<td>2BRF</td>
<td>100</td>
<td>0.768</td>
<td>0.710</td>
</tr>
<tr>
<td>2CE0</td>
<td>99</td>
<td>0.619</td>
<td>0.529</td>
</tr>
<tr>
<td>2E3H</td>
<td>81</td>
<td>0.663</td>
<td>0.605</td>
</tr>
<tr>
<td>2EAO</td>
<td>89</td>
<td>0.750</td>
<td>0.695</td>
</tr>
<tr>
<td>2EHS</td>
<td>75</td>
<td>0.722</td>
<td>0.747</td>
</tr>
<tr>
<td>2FO3</td>
<td>85</td>
<td>0.720</td>
<td>0.348</td>
</tr>
<tr>
<td>2IP6</td>
<td>87</td>
<td>0.564</td>
<td>0.572</td>
</tr>
<tr>
<td>2MCM</td>
<td>112</td>
<td>0.803</td>
<td>0.820</td>
</tr>
<tr>
<td>2NUH</td>
<td>104</td>
<td>0.794</td>
<td>0.771</td>
</tr>
<tr>
<td>2PKT</td>
<td>93</td>
<td>0.220</td>
<td>−0.193*</td>
</tr>
<tr>
<td>2PLT</td>
<td>99</td>
<td>0.461</td>
<td>0.509*</td>
</tr>
<tr>
<td>2QIL</td>
<td>99</td>
<td>0.548</td>
<td>0.594</td>
</tr>
<tr>
<td>2RB8</td>
<td>93</td>
<td>0.650</td>
<td>0.517</td>
</tr>
<tr>
<td>3BZQ</td>
<td>99</td>
<td>0.484</td>
<td>0.466</td>
</tr>
<tr>
<td>5CYT</td>
<td>103</td>
<td>0.440</td>
<td>0.331</td>
</tr>
</tbody>
</table>

TABLE III. Comparison of B factor prediction in terms of correlation coefficients by parameter free MND and Gaussian normal mode (GNM) for large-sized data set.\(^6\) The asterisk sign indicates improved prediction with modified protein data.

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>N</th>
<th>MND</th>
<th>GNM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1AH0</td>
<td>64</td>
<td>0.617</td>
<td>0.562</td>
</tr>
<tr>
<td>1ATG</td>
<td>231</td>
<td>0.606</td>
<td>0.497</td>
</tr>
<tr>
<td>1BY1</td>
<td>224</td>
<td>0.464</td>
<td>0.552</td>
</tr>
<tr>
<td>1CCR</td>
<td>111</td>
<td>0.562</td>
<td>0.351</td>
</tr>
<tr>
<td>1E5K</td>
<td>188</td>
<td>0.751</td>
<td>0.859</td>
</tr>
<tr>
<td>1EW4</td>
<td>106</td>
<td>0.596</td>
<td>0.547</td>
</tr>
<tr>
<td>1IFR</td>
<td>113</td>
<td>0.684</td>
<td>0.637</td>
</tr>
<tr>
<td>1INKO</td>
<td>122</td>
<td>0.561</td>
<td>0.368</td>
</tr>
<tr>
<td>1NLS</td>
<td>238</td>
<td>0.634</td>
<td>0.523*</td>
</tr>
<tr>
<td>1IO8</td>
<td>221</td>
<td>0.410</td>
<td>0.309</td>
</tr>
<tr>
<td>1IPM</td>
<td>123</td>
<td>0.646</td>
<td>0.685</td>
</tr>
<tr>
<td>1PZ4</td>
<td>113</td>
<td>0.859</td>
<td>0.843</td>
</tr>
<tr>
<td>1QTO</td>
<td>122</td>
<td>0.378</td>
<td>0.334</td>
</tr>
<tr>
<td>1RO</td>
<td>108</td>
<td>0.334</td>
<td>0.529</td>
</tr>
<tr>
<td>1UKU</td>
<td>102</td>
<td>0.650</td>
<td>0.742</td>
</tr>
<tr>
<td>1V70</td>
<td>105</td>
<td>0.409</td>
<td>0.162</td>
</tr>
<tr>
<td>1WBE</td>
<td>204</td>
<td>0.531</td>
<td>0.549</td>
</tr>
<tr>
<td>1WHI</td>
<td>122</td>
<td>0.407</td>
<td>0.270</td>
</tr>
<tr>
<td>1WPA</td>
<td>107</td>
<td>0.542</td>
<td>0.417</td>
</tr>
<tr>
<td>2AGK</td>
<td>233</td>
<td>0.687</td>
<td>0.512</td>
</tr>
<tr>
<td>2C71</td>
<td>205</td>
<td>0.662</td>
<td>0.560</td>
</tr>
<tr>
<td>2CG7</td>
<td>90</td>
<td>0.477</td>
<td>0.379</td>
</tr>
<tr>
<td>2CW</td>
<td>227</td>
<td>0.659</td>
<td>0.696</td>
</tr>
<tr>
<td>2HQK</td>
<td>213</td>
<td>0.793</td>
<td>0.365</td>
</tr>
<tr>
<td>2HYK</td>
<td>237</td>
<td>0.583</td>
<td>0.515</td>
</tr>
<tr>
<td>2I24</td>
<td>113</td>
<td>0.380</td>
<td>0.494</td>
</tr>
<tr>
<td>2IMF</td>
<td>203</td>
<td>0.615</td>
<td>0.514</td>
</tr>
<tr>
<td>2PPN</td>
<td>107</td>
<td>0.637</td>
<td>0.668</td>
</tr>
<tr>
<td>2R16</td>
<td>176</td>
<td>0.461</td>
<td>0.618*</td>
</tr>
<tr>
<td>2V9V</td>
<td>135</td>
<td>0.580</td>
<td>0.528</td>
</tr>
<tr>
<td>2VIM</td>
<td>104</td>
<td>0.375</td>
<td>0.282</td>
</tr>
<tr>
<td>2VPA</td>
<td>204</td>
<td>0.771</td>
<td>0.576</td>
</tr>
<tr>
<td>2VYO</td>
<td>206</td>
<td>0.705</td>
<td>0.761</td>
</tr>
<tr>
<td>3SEB</td>
<td>238</td>
<td>0.777</td>
<td>0.826</td>
</tr>
<tr>
<td>3VUB</td>
<td>101</td>
<td>0.638</td>
<td>0.607</td>
</tr>
</tbody>
</table>

We validate our method by the comparison with the GNM. The code for the GNM is adopted from Jernigan Laboratory (http://ribosome.bb.iastate.edu/software.html). The cutoff distance used is 7 Å and some minor modification is done to enhance its performance on the coarse-grained PDB data. Three sets of protein data utilized by Park, Jernigan, and Wu\(^6\) are considered in our work. We carefully go through these data to delete the multiple conformations and add in the missing residues. The modified data lead to improved GNM predictions compared to those in the literature.\(^6\) Tables I–III list our results for three data sets, respectively. Figure 7 plots the results of the comparison over three data sets. Apparently, the proposed MND method outperforms GNM for most proteins. The average correlation coefficients are listed in Table IV. It is seen that the proposed method achieves about 8% improvement over the GNM.
2. Improvement of the flexibility prediction by considering co-factors

Protein structures downloaded from the PDB typically contain many cofactors, i.e., coenzymes and prosthetic groups, which are important for proteins’ biological functions. Cofactors also contribute to protein structural rigidity. Therefore, the inclusion of cofactors in the present chaotic dynamics model will improve the prediction of protein B-factor. In the present work, we consider a simple treatment of cofactors in which the nonlinear oscillators of cofactors are used as part of driven sources while those of residues are treated as a response system. This treatment reduces computational cost if the MND is on a stable fixed point, because one does not need to actually compute cofactors’ dynamics.

Figure 8 depicts the improvement in the theoretical prediction of protein 1fac B-factors due to the consideration of two metal clusters, i.e., Fe4S4 clusters. We plot the interaction strength of cofactors with respect to the correlation coefficient. Clearly, the consideration of cofactors leads to a five percent improvement in our prediction.

IV. CONCLUDING REMARKS

Protein folding and the related structure function relation are of fundamental interest and importance to scientific community. The present work proposes MND to represent the protein folding process and illustrate the structure function relation. We unveil the ability of folded protein structures for controlling the chaos in the underlying MND and show the existence of an ILDM for each folded protein. We demonstrate that protein unfolding leads to high dimensional chaotic states. The orderliness of unfolded protein configurations generated by using steered molecular dynamics simulations directly correlates with the dimensionality of chaotic states. We further utilize the transverse stability of the ILDM for protein uncertainty quantification, i.e., the prediction of protein temperature factors due to thermal fluctuations. By a comparison with cutting edge method in the field, namely, the GNM, it is found that the proposed ILDM based approach works well for protein flexibility analysis.

A discussion of several aspects is in order. First, although the Lorenz system is employed in our study, our findings about the protein control of chaos and the existence of ILDM associated with the folded protein conformation are general and can be regenerated by using other nonlinear dynamical systems. The ability of the ILDM for protein flexibility analysis can also be realized by using other nonlinear dynamical systems because it is the connectivity matrix, rather than the specific dynamical system, that captures the fundamental physics of protein dynamics. Secondly, a simple method for steered molecular dynamics simulations is utilized in the present work to generate partially folded and unfolded protein conformations. However, our findings about the order of unfolded proteins and eigenvalue analysis of protein orderliness are independent of the method used in generating unfolded conformations and protein unfolding pathways. Similar results are obtained by analyzing unfolded protein conformations generated by a different steered molecular dynamics approach. Finally, the prediction of the B-factor can be further improved by the consideration of different types of the residues and the crystal structure of a given protein.

ACKNOWLEDGMENTS

This work was supported in part by NSF Grants DMS-1160352 and IIS-1302285, and NIH Grant R01GM-090208. The authors acknowledge the Mathematical Biosciences Institute for hosting valuable workshops.


59 L. W. Yang and C. P. Chng, “Coarse-grained models reveal functional
dynamics–I. elastic network models–theories, comparisons and
60 L. Skjaerven, S. M. Hollup, and N. Reuter, “Normal mode analysis for
61 Q. Cui and I. Bahar, Normal Mode Analysis: Theory and Applications to
Biological and Chemical Systems (Chapman and Hall/CRC, 2010).
62 J. K. Park, R. Jernigan, and Z. Wu, “Coarse grained normal mode analysis
vs. refined Gaussian network model for protein residue-level structural