Disorder profile of nebulin encodes a vernierlike position sensor for the sliding thin and thick filaments of the skeletal muscle sarcomere

Ming-Chya Wu,[1,2,*] Jeffrey G. Forbes,[3] and Kuan Wang[2,4,†]

[1]*Research Center for Adaptive Data Analysis, National Central University, Chungli, Taoyuan 32001, Taiwan*
[2]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*
[3]*TWIC, Inc, College Park, Maryland 20740, USA*
[4]*Nanomedicine Program, and Institute of Biological Chemistry, Academia Sinica, Nankang, Taipei 11529, Taiwan*

Nebulin is an about 1 μm long intrinsically disordered scaffold for the thin filaments of skeletal muscle sarcomere. It is a multifunctional elastic protein that wraps around actin filament, stabilizes thin filaments, and regulates Ca-dependent actomyosin interactions. This study investigates whether the disorder profile of nebulin might encode guidelines for thin and thick filament interactions in the sarcomere of the skeletal muscle. The question was addressed computationally by analyzing the predicted disorder profile of human nebulin (6669 residues, $\sim$200 actin-binding repeats) by PONDR and the periodicity of the A-band stripes (reflecting the locations of myosin-associated proteins) in the electron micrographs of the sarcomere. Using the detrended fluctuation analysis, a scale factor for the A-band stripe image data with respect to the nebulin disorder profile was determined to make the thin and thick filaments aligned to have maximum correlation. The empirical mode decomposition method was then applied to identify hidden periodicities in both the nebulin disorder profile and the rescaled A-band data. The decomposition reveals three characteristic length scales (45 nm, 100 nm, and 200 nm) that are relevant for correlational analysis. The dynamical cross-correlation analyses with moving windows at various sarcomere lengths depict a vernierlike design for both periodicities, thus enabling nebulin to sense position and fine tune sarcomere overlap. This shows that the disorder profile of scaffolding proteins may encode a guideline for cellular architecture.

I. INTRODUCTION

The sarcomere, as a structural and functional basic unit of skeletal muscle tissue, is organized into a regular array of thin and thick filaments of well-defined length. The filament length is relevant to muscle function because a muscle generates force in proportion to thin and thick filament overlap [1–7]. The thin filament is a multimeric protein complex formed from actin, troponin, and tropomyosin, which are compressed longitudinally and radially by nebulin [8,9], similar to the organization of a finger trap toy, while the thick filament is formed from myosin that is grouped in a bundle. As the biggest actin-binding subunit protein (molecular weight 700–800 kDa) localized to the thin filament and serving as a scaffold of about 1 μm long, nebulin is an elastic tether linking as many as 200 actin monomers to myosin. Each monomer is bound by 35 amino acid residues that may also bind calmodulin, tropomyosin, and troponin [10]. Nebulin is an intrinsically disordered protein. The exact structure of nebulin on the thin filament is unknown at present. The bulk of the nebulin sequence is constructed of approximately 150 tandem copies of modules that can be classified into seven-module super-repeats. These super-repeats are further divided into eight segments: with six segments containing adjacent, highly homologous super-repeats, one single repeat segment consisting of eight nebulin modules of the same type, and a nonrepeat segment terminating with a SH3 domain at

the C terminus [7,10]. A single nebulin molecule spans the thin filament with its C terminus anchored at the Z disk and its N-terminal region directed towards the thin filament pointed end [8], associating with tropomodulin [11]. The cross-bridge attachments between actin and myosin are independent force generators for sarcomere force production [1–4]. The schematic organization of a part of the sarcomere is shown in Fig. 1 (not to scale). Because the length of expressed nebulin is associated with the length of the sarcomere [12,13], nebulin is proposed to form a molecular ruler specifying the length of the thin filament during sarcomere assembly [11,14–19]. As a multifunctional protein [20], nebulin also regulates Ca-dependent actomyosin interactions [19,21], maintains intermyofibrillar connectivity [22–24], and sets physiological Z-disk widths [25]. It works cooperatively during muscle contraction and relaxation with titin kinase, which functions as a stretch-dependent digital force sensor [26–28], and a flexible titin PEVK region that serves as a stretch-gated signal hub [29,30]. For discussion of mechanisms of force generation in skeletal muscle, see Ref. [6], and for a review of recent developments on the research of structural roles of nebulin, see, for example, Ref. [7].

Muscle proteins associated with force generation are generally intrinsically disorder or mechanoactivated. The emerging concept of stretch-assisted assembly and stretch-dependent function has attracted much attention for it describes unique features of muscle-associated proteins. Since the human skeletal muscle nebulin sequence encodes a blueprint for thin filament architecture [10] and length specification of thin filament by expressed nebulin is essential for length-dependent force generation of a sarcomere [5–7] according to the the sliding filament [1,2] and the cross-bridge theories [3,4], it

---
*mcwu@ncu.edu.tw
†Present address: College of Biomedical Engineering, Taipei Medical University, Taipei 110, Taiwan.
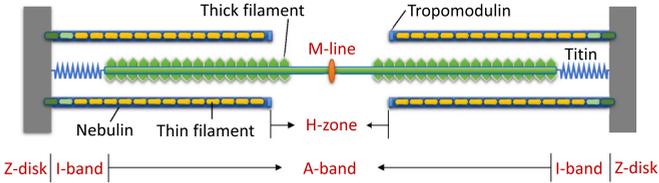
FIG. 1. Schematic representation of the structural organization of a part of the skeletal muscle sarcomere (not to scale).

is interesting to investigate how this intrinsically disordered protein carries out its function as a ruler in sarcomere assembly. In this paper, we focus on whether the disorder profile of nebulin might encode guidelines for thin and thick filament interactions in the sarcomere of the skeletal muscle, with an aim of providing a clear physical picture for a possible scenario of carrying out its role in regulating the length of the thin filament. The question is to be addressed computationally by analyzing the disorder profile of nebulin amino acid sequence [10] and the A-band intensity data in the electron micrographs of the sarcomere [31]. These are raw data. The former is a linear data sequence correlating with nebulin structural features, and the latter is an image, where the grayscale intensity of the image reflects the locations of myosin-related proteins in the thick filament. Both the two data will be formatted as time sequences and will be analyzed using time series analysis techniques. Among others, adaptive data analysis approaches are advantageous choices since there is no *a priori* knowledge for the property of the empirical data. We will first use the detrended fluctuation analysis (DFA) [32] to explore the intrinsic correlation property in each data, from which a scale factor between the data of the nebulin disorder profile and the A-band intensity data is determined for a reasonable alignment of the two data sequences. To study the correlation between the two sequences, we use the empirical mode decomposition (EMD) [33] and the Hilbert spectral analysis to identify hidden periodicities in them. There are several methods developed to characterize possible periodicities in data, including conventional Fourier-based spectral analysis, wavelet [34], and complex network method [35,36]. The EMD method is more suitable for this study because it extracts intrinsic mode functions (IMFs) in time domain, each with a small range of periodicity. It works on the types of data where Fourier analysis fails and is ideally suited for nonlinear and nonstationary data [33] and clearly shows trends that are hidden by other signals. These advantages are useful for extracting relevant components from our data. Finally, the dynamical cross-correlation (DCC) analysis is applied on the relevant components. The DCC results then form a basis to depict a model of the position sensory for the sliding thin and thick filaments of the skeletal muscle sarcomere.

The rest of this paper is organized as follows. In Sec. II, the data acquisition for the nebulin disorder profile and the A-band intensity data is explained. In Sec. III, the DCC analysis is introduced to determine the scale factor of the raw data of the A-band intensity data with respect to the raw data of the nebulin disorder profile. The DFA analysis and the EMD procedures for the data are briefly introduced and carried out in Sec. IV and Sec. V, respectively. The DCC analysis for the

models of sarcomere normal extension and supercontraction are discussed in Sec. VI, accompanying with an proposal of a ruler model for the thin and thick filament interactions. Finally, we conclude the study in Sec. VII.

II. DATA

Figure 2(a) shows the nebulin disorder profile as a function of residue sequence number. The nebulin disorder profile describes the predicted structural disorder of the nebulin, composed of a sequence of 6669 amino acid residues [10], by PONDR using a VL3H predictor [37]. The value, ranging from 0–1, represents the probability of the region where a residue is located being disordered. The larger the value is, the higher is the probability of disorder. The actual length scale of the sequence is inexplicit but is compatible with half of the length of the A-band stripe in Fig. 1. Here, for simplicity, we assume residues in nebulin have a uniform size.

Figure 2(b) shows the raw data of the A-band intensity data obtained by scanning half of the A-band segment in the electron micrograph presented in Ref. [31] and by taking an average of the intensity value of the image segment in width as a linear sequence in length. The contrast of the A-band stripes in grayscale in the electron micrograph of the sarcomere shown in Ref. [31] reflects the distribution of myosin-associated proteins. The number of data points in the sequence depends on the image quality and data sampling size in image processing. The resultant data shown in Fig. 2(b) is a sequence of 2291 data points in a length of 800 nm, i.e., half of the full length of the A band. In contrast to the nebulin disorder profile, this length is real in space. The relative higher values at the two ends of the A-band intensity data is attributed to the original image that covers the M line and the joints with the titin PEVK region as described in Ref. [31].

III. DYNAMICAL CROSS-CORRELATION ANALYSIS

The raw data of the nebulin disorder profile as a function of the residue sequence number and the A-band intensity data calculated from real-space image are presented in different measuring units. The actual length and the sampling scale of the former are indefinite. To analyze the cross correlation between the two data, we assume that the data represent particular time sequences with cross correlation, and there is a common and basic time unit t defined to measure the two data. Taking into account a possible scale factor γ between the nebulin data f and the A-band intensity data g, the DCC function is defined as

$$c_T(\gamma, \tau) = \frac{1}{T} \int_{T_1}^{T_2} f(\gamma t) g(t + \tau) dt, \qquad (1)$$

where $T = T_1 - T_2$ is the size of the moving window, and τ is a time shift of g with respect to f. The scale factor γ is associated the sampling size of the A-band micrograph intensity. The DCC function $c_T(\gamma, \tau)$ for a specified T reaches maxima at certain τ's for correlated and properly detrended functions f and g. In the following, we first determine the scale factor γ, and then carry out the DCC analysis of the two data with proper alignments, via sliding g with respect to f using various time shifts τ.
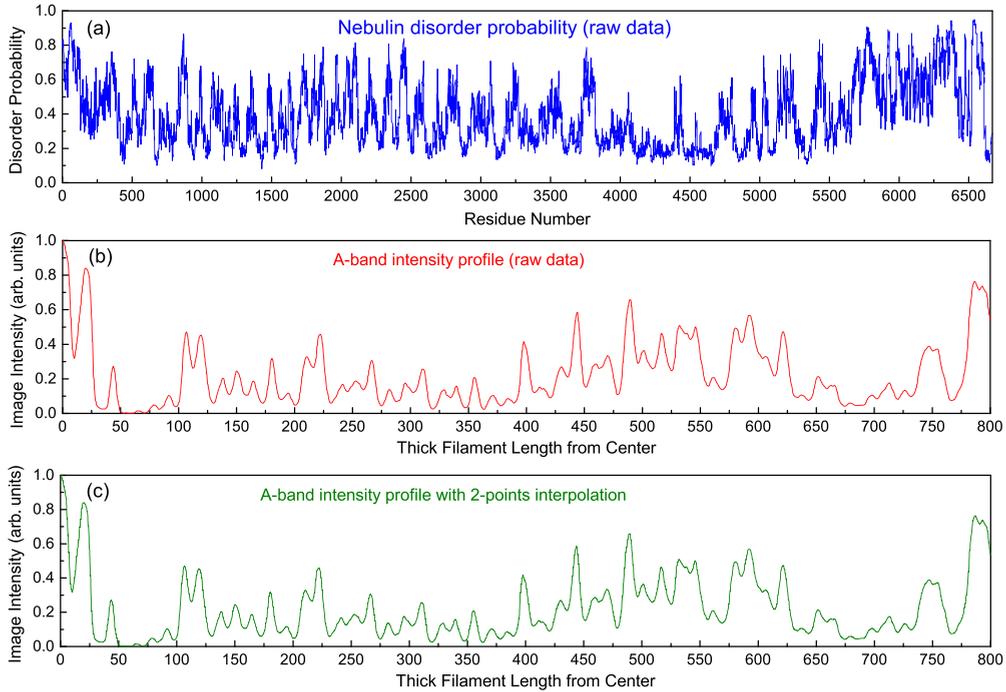
FIG. 2. The nebulin disorder profile and the A-band intensity data. (a) The nebulin disorder profile as a function of residue sequence number (6669 dpts), predicted by PONDR using a VL3H predictor [37]. (b) The A-band intensity raw data (2291 dpts) obtained by scanning the A-band stripes in the electron micrographs of the sarcomere in Ref. [31] and by taking an average of the intensity value of the image in width as a linear sequence in length. (c) The rescaled A-band intensity data with two-points interpolation (6873 dpts) on the raw data of (b).

IV. DETRENDED FLUCTUATION ANALYSIS

We assume that there are intrinsic correlations in both of the data sequences, which are characterized by their own length scales. The scale factor γ will be chosen under the criterion of matching the scale-dependent properties of the two data sequences while keeping their intrinsic correlations unchanged. This is achieved by applying the DFA method [32] to explore the correlations in the data. The DFA measures the fluctuation $F(n)$ of a walk $w(t)$ with respect to a linear fit of the data in a moving window n in time [expressed as $w_n(t)$], and uses an index α defined from

$$F(n) = \sqrt{\frac{1}{T}\sum_t [w(t) - w_n(t)]^2} \sim n^\alpha, \qquad (2)$$

to describe the correlation property of the data: $\alpha = 0.5$ corresponds to white noise, 1.0 refers to $1/f$ noise, 1.5 indicates Brownian noise, and 2.0 represents black noise. Besides, the data sequence with $1.0 < \alpha < 1.5$ displays negative correlation, and $1.5 < \alpha < 2.0$ for positive correlation [32,38]. The raw data of the nebulin disorder profile and the A-band intensity data representing nonaccumulative quantities are here considered as sequences of walks, such that f and g will be regarded as $w(t)$ in Eq. (2).

The results of the DFA analysis on both of the raw data are shown in Figs. 3(a) and 3(b). Remarkably, this scaling analysis reveals crossover phenomenon [38], manifesting that each data has different correlation properties at its short and long length scales are characterized by two indices α_1 and α_2, respectively. This reflects that the properties of the time series in short and long time scales are governed by different

dynamics. For the raw data of the nebulin disorder profile, we have $\alpha_1 = 1.30$ and $\alpha_2 = 0.78$, and two length scales are demarcated at $n = 10^2 = 100$. Thus, the nebulin disorder profile displays negative correlation in the length scale less than 100 dpts, and its fluctuation in longer length scale behaves in a way between white noise and $1/f$ noise, displaying weak long-range correlation. This property is associated with the specific modular super-repeats in nebulin sequence [8,10]. Negative correlation in shorter length scale is a consequence of the modular structure in the order of hundreds of residues, while the long-range correlation in intermodular structure is due to the specificity of nebulin structural features with
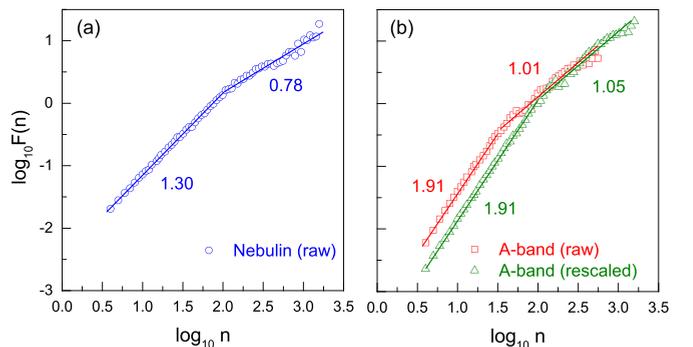

FIG. 3. Detrended fluctuation analysis (DFA) of the nebulin disorder profile and the A-band intensity data. The DFA α index value is an indicative of the slope of the linear regime of the plot. (a) DFA of the nebulin disorder profile raw data. (b) DFA of the A-band raw data and the rescaled version with two-points interpolation.

disorder probability bounded between 0 and 1. In this regard, the length scale demarcation at $n = 100$ is interesting.

For the A-band intensity raw data, we have $\alpha_1 = 1.91$ and $\alpha_2 = 1.01$, and two length scales are demarcated at $n = 10^{1.5} \approx 32$. The A-band intensity data displays black noise property in the shorter length scale, and $1/f$-like behavior in the longer length scale. Black noise property is not interesting here because it simply reflects the data is quite smooth below the length scale. In contrast, the long-range correlation in longer length scale is a result of the specific arrangement of the myosin-associated proteins in thick filament. This suggests that we should pay more attention to the longer time scale. According to the above analysis, apparently a reasonable choice of the scale factor γ is $100/32 \approx 3$. Such rescaling operation makes the longer length scales of the two data match with each other. We then interpolate the A-band intensity data by inserting two points between two sequential raw data points, using the algorithm of cubic spline fitting. The uniform interpolation is applicable from the consideration of regular arrangement of myosin-associated protein in the A band. The resultant A-band intensity data is a sequence of 6873 data points, as shown in Fig. 2(c). Since all the raw data of the A-band intensity data are preserved in the interpolated version, this process does not introduce errors and the intrinsic correlation property is substantially kept unchanged. The DFA analysis of the interpolated version is shown in Fig. 3(b), where we have $\alpha_1 = 1.91$ and $\alpha_2 = 1.05$, compatible with those of the raw data. Meanwhile, short and long length scales now border at $n = 10^2$ as expected.

To verify the robustness of the above results, we have performed the same calculations on three types of surrogates obtained from (i) shuffling the raw data of nebulin disorder profile and A-band intensity data, (ii) the disorder profile predicted using the shuffling of the full nebulin sequence, and (iii) the disorder profile predicted using arbitrary amino acid sequence. The result of case (i) shows that there is no correlation in surrogates (i.e., single DFA index with average value $\langle \alpha \rangle \approx 0.5$), which shows the existence of scale-dependent intrinsic correlations in the raw data. In contrast, both of case (ii) and case (iii) have two DFA indices with an alternative demarcation varying in their scaling analysis for different realizations and the index values are non-self-averaging. Different demarcation for α_1 and α_2 from that of raw data may imply $n = 100$, which for our data is relevant. We have also found that their $\alpha_1 (\gtrsim 1.8$, indicating positive correlation) is much larger than 1.3 of the raw data, but $\alpha_2 \approx 0.78$ is similar to that of the raw data. The long-range correlation indicated by α_2 is a result of the disorder probability bounded between 0 and 1. Hence, this reveals a physical picture that the disorder profile corresponding to the nebulin sequence fluctuates within $n < 100$ in a way with negative correlation, while it is bounded between 0 and 1 and shows a property between white noise and $1/f$ noise. For surrogates (ii) and (iii), the disorder profile fluctuates within some alternative and indefinite range in a way with positive correlation, while it is also bounded between 0 and 1. Disorder probability sets the boundaries leading to long-range correlation α_2 value, while (positive or negative) correlation indicated by α_1 and the demarcation for short (α_1) and long (α_2) length scales are relevant to sequence property and make the raw data

distinct from surrogates. As a result, above observations confirm the intrinsic correlation in nebulin disorder profile is a consequence of sequence specificity.

For simplicity, in the following, if not specifically stated, the rescaled A-band intensity data will be termed briefly as the A-band intensity data.

V. EMPIRICAL MODE DECOMPOSITION

Next, we apply the EMD [33] method to separate oscillatory components with different intrinsic length scales in the data. The existence of characteristic oscillatory components in the nebulin disorder profile is attributed to the structural properties encoded in the residue sequence for actin-binding repeats, while the same feature for the A-band intensity data is a result of spacial arrangement of myosin-associated proteins. These form the basis suitable for the EMD analysis. The EMD method assumes that a time sequence is composed of a number of IMFs, each with a distinct frequency range. An IMF is a well-defined function for Hilbert transform, which is symmetric with respect to local mean and has the same numbers of zero crossings and extremes. The decomposition is adaptive without prior bases, and is achieved by an iterative sifting algorithm [33]. The resultant IMFs are a set of components, which are complete and orthogonal among one another. For the details of the algorithm, see Refs. [33,39]. Briefly, the decomposition can be expressed as, for data $x(t)$,

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t), \tag{3}$$

where $c_i(t)$, $i = 1, \ldots, n$, are IMFs, and $r_n(t)$ is a residue, which is a constant or a monotonic trend. Figures 4(a) and 4(b) show the decompositions of the nebulin disorder profile and the A-band intensity data, respectively. The data is decomposed into 11 components. The orthogonality index [33],

$$\kappa = \frac{\sum_{i,t} c_i(t) c_{i+1}(t)}{\sqrt{\sum_{i,t} c_i^2(t) \sum_{j,t} c_{j+1}^2(t)}}, \tag{4}$$

has been used to choose a reasonable decomposition by minimizing the value of κ [40]. The mode mixing problem was handled using the ensemble empirical mode decomposition (EEMD) [41], which is an extended version of EMD. Different sifting criteria have been used in EEMD to confirm the robustness of the decomposition. As mentioned, IMFs c_i's are characterized by their own oscillatory frequencies, and the averaged frequency of a lower-index IMF is larger than that of a higher-index IMF. The instantaneous frequency of the data is a function of t, the sequence number of data. We perform the Hilbert transform on each IMF $c_k(t)$, according to

$$c_k(t) + i\bar{c}_k(t) = A_k(t) e^{i\phi_k(t)}, \tag{5}$$

where $\bar{c}_k(t)$ is the conjugate of the IMF $c_k(t)$, derived from

$$\bar{c}_k(t) = \frac{1}{\pi} \mathrm{P} \int_{-\infty}^{\infty} \frac{c_k(t')}{t - t'} dt', \tag{6}$$

where P indicates the Cauchy principle value; the amplitude $A_k(t)$ is defined as

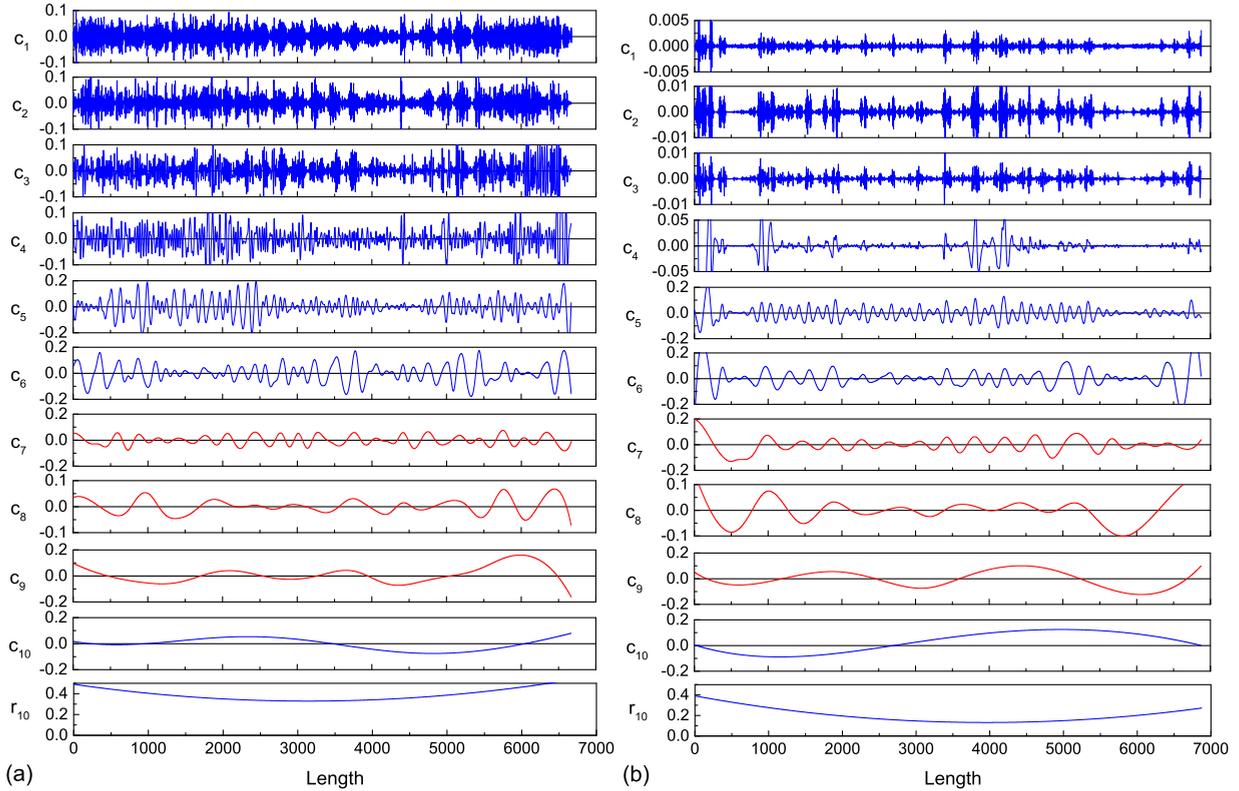$$A_k(t) = \sqrt{c_k^2(t) + \bar{c}_k^2(t)}, \tag{7}$$

FIG. 4. Empirical mode decomposition of (a) the nebulin disorder profile and (b) the A-band intensity data. Each data is decomposed into 10 IMFs and 1 residue. IMFs c_7, c_8, and c_9 with average periods of 40 nm, 100 nm, and 200 nm, highlighted by red color are relevant for dynamical cross-correlation analysis. See main text for details.

and the instantaneous phase $\phi_k(t)$ is calculated through

$$\phi_k(t) = \tanh^{-1} \frac{\bar{c}_k(t)}{c_k(t)}. \tag{8}$$

The instantaneous frequency is calculated by taking time derivative of $\phi_k(t)$ in Eq. (8). The power spectral density (PSD) analysis of the two data calculated accordingly is shown in Fig. 5(a), where significant frequency contents locate at the frequency less than 0.005, corresponding to a length scale of about 200 data points (dpts). Remarkably, the IMFs c_7, c_8,

FIG. 5. Power spectral density (PSD) analysis of the A-band intensity data and the nebulin disorder profile. (a) PSD of all IMFs. (b) PSD of IMFs c_7, c_8, and c_9.

and c_9 contribute predominantly to the main peaks in the PSD, as shown in Fig. 5(b). The periodicities of the IMFs c_7, c_8, and c_9 roughly correspond to average module lengthes of 430, 860, and 1720 residues, locating in the regime of longer-length scale with long-range correlation in our DFA analysis. It is thus reasonable to assume that these three IMFs are key components involved in the correlation properties associated with position sensory. Also note that in Fig. 5(b), the PSDs of the paired IMFs c_7, c_8, and c_9 of the nebulin disorder profile and the A-band intensity data do not exactly match with each other but only partially overlap. This is clear from the different wave forms of the paired IMFs of the same indices shown in Fig. 4. This property has significant effects in the DCC analysis discussed in the next section.

Before discussing the results, we shall remark that the EMD method and the Hilbert spectral analysis are used here to provide good resolution on frequency content in PSD by means of superiority of processing nonlinear and nonstationary time series by EMD than other methods [33]. The Fourier-based spectral analysis unavoidably introduces harmonics in the power spectrum, which obscures intensity of actual frequency components in the PSD analysis. We have skipped comparison of results using different methods and focused on the subject of the present paper. For comparison of decompositions of typical data using EMD and other methods see Ref. [33], and for implementations on real data from biomedical recordings, financial, and biological systems see, for example, Refs. [28,39,42–47]. Meanwhile, we have also used the three types of surrogates discussed in Sec. IV as
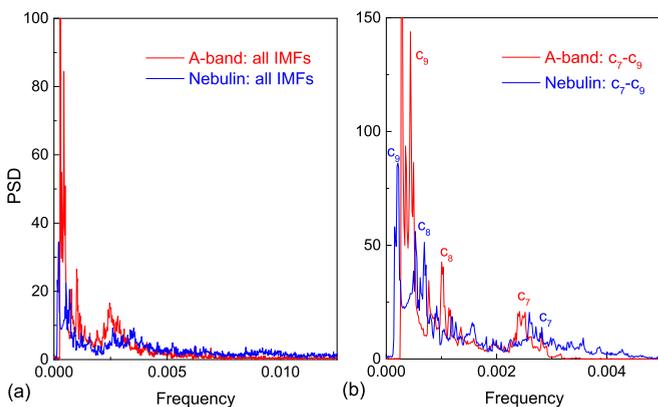
a control group and performed the same EMD and Hilbert spectral analysis on them. However, without any regularity and periodicity, such surrogates do not show significance in further DCC analysis.

VI. RESULTS AND DISCUSSIONS

A. Normal extension

To mimic the relative movements of the thin and thick filaments during muscle extension and contraction, we consider a sliding filament model of the spacial relationship between A band and nebulin for normal resting, normal extension and supercontraction, as shown in Figs. 6(a), 6(b), and 6(c), respectively. In this model, the A band is sandwiched between two subsequent nebulins. The titin elastic region enables the reversibility of the movements, and the muscle contraction and extension is realized through the sliding of the A band on nebulin. The degree of the compression and stretch of the sarcomere length is associated with their relative positions. According to Ref. [5], the tension of the sarcomere depends on its striated spacing (stretch length), and the optimal stretch length for tension generation was found to be from 70–130 % of the normal resting length, depending on species and muscle type [48]. For the case of the sarcomere half-resting length of 800 nm considered in this study, this corresponds to an amount of 240 nm (or 2000 dpts) reduction for contraction or increment for stretch. In principle, an existence of some kind of periodicities in the site correlation between paired IMFs at certain length scales allows stepwise sliding of A-band on nebulin. Moreover, an existence of aperiodicities in the site correlation between paired IMFs at certain length scales can function as a marker of the position
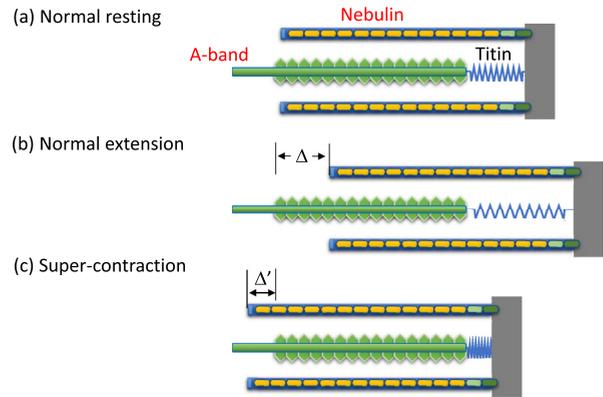

FIG. 6. Schematic illustration of the sliding filament model, proposed for the relationship between A-band and nebulin in the process of sliding thin and thick filaments during resting, muscle stretch and contraction. (a) Normal resting, (b) normal extension, and (c) supercontraction.

sensor signifying the sliding of the A band with respect to nebulin. This conjecture is to be verified by examining the DCCs of individual paired IMFs c_7's, c_8's, and c_9's, each of which has significant contribution to the PSDs shown in Fig. 5.

First we consider the situation of the paired IMFs c_8's. These components are detrended and have an average period of 860 dpts (or 100 nm), estimated through zero-crossing calculation of c_8 in Fig. 4. Figure 7 shows the IMFs c_8's of the two data with various relative position shiftings of nebulin with respect to the A band, DCC of the two IMFs, and a comprehensive schematic drawing of the DCC maximum
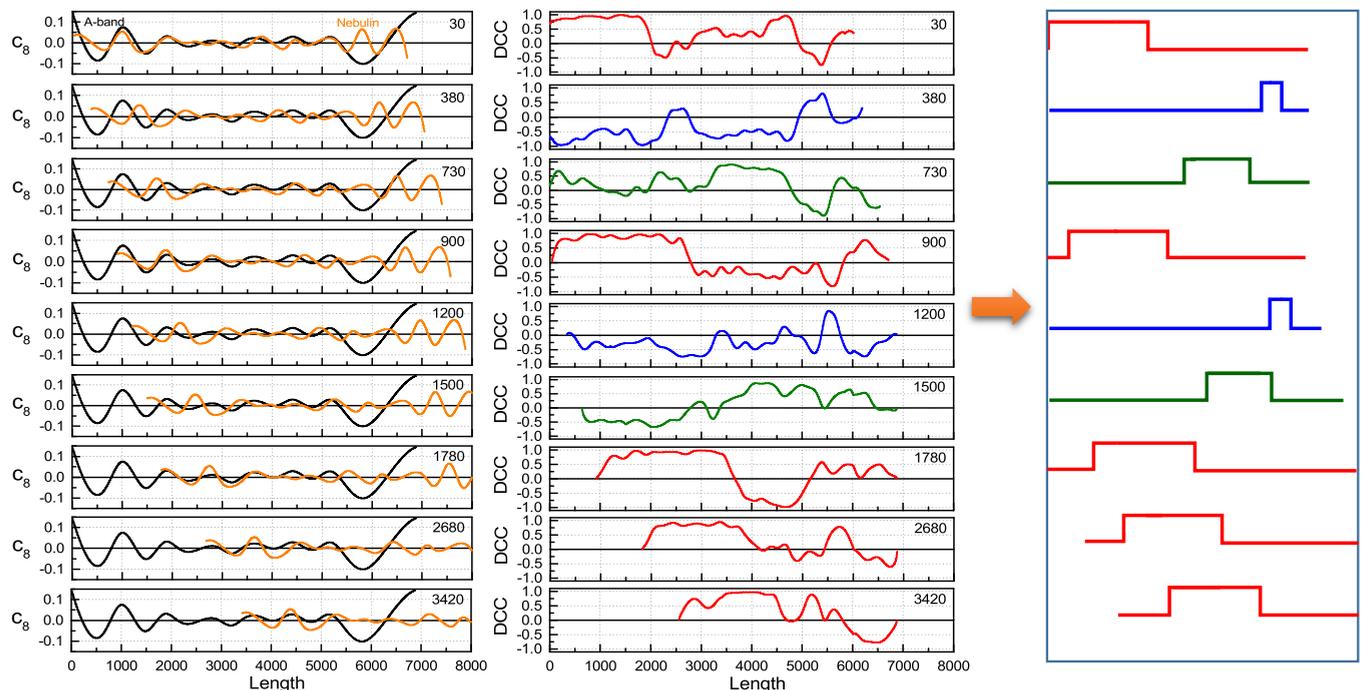

FIG. 7. Dynamical cross-correlation (DCC) analysis of the IMF c_8 with $T = 860$ dpts and a comprehensive schematic drawing of the DCC maximum profile at various data shiftings, plotted in the A-band coordinate.
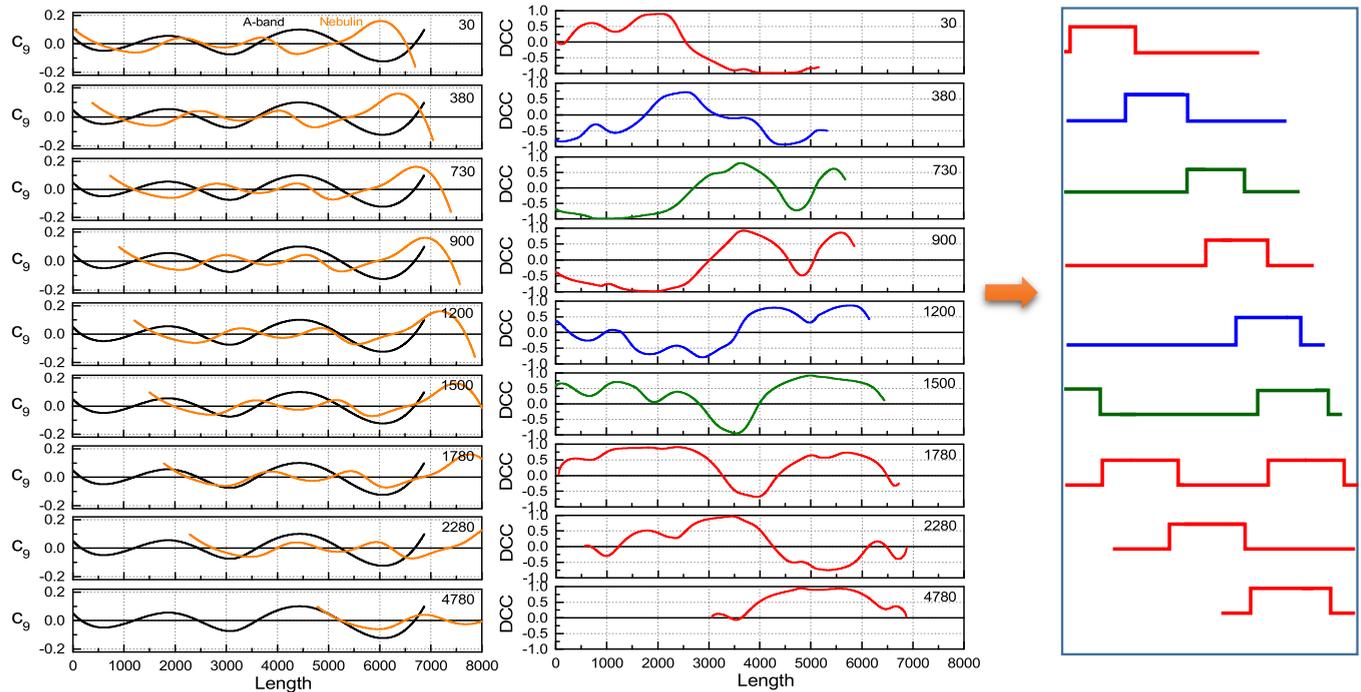
FIG. 8. Dynamical cross-correlation (DCC) analysis of the IMF c_9 with $T = 1720$ dpts and a comprehensive schematic drawing of the DCC maximum profile at various data shiftings, plotted in the A-band coordinate.

profiles, plotted in the A-band coordinate. In the calculation of the DCC, we have chosen a moving window of $T = 860$ dpts, which is about the average period of the IMF c_8. This enables a reasonable resolution for the detection of site correlation between the paired IMFs c_8. The schematic drawing of the DCC maximum profile is a binary representation of the DCC, in which 1 is used to replace $c_T \geqslant 0.5$ and 0 for $c_T < 0.5$. Further, short and transient (unstable) segments with $c_T \geqslant 0.5$ in DCC have been neglected in the schematic version to make the DCC maximum profiles comprehensive. Thus, the DCC maximum profile is a simplified schematic representation of the DCC, composed of a sequence of step functions or square waves. For clarity, different colors have been used to highlight certain DCC maximum profiles in Fig. 7. At particular shifting lengths ranging from 30–1500 dpts, the structure of the DCC maximum profile changes, from red curve to blue curve, then to green curve, then to red curve and repeats (see Fig. 7). Two cycles can be observed within this shifting range. In the first cycle, three different step functions indicated by red, blue, and green, respectively, are produced and show up in sequent DCC maximum profiles as the increasing of position shifting. In the second cycle, a slight position shift of the three step functions (showing up in the same ordering) to the shifting direction of nebulin is observed. This implies that the structure property specified by the IMF c_8 of the disorder profile of nebulin is essential for the DCC observed. There is indeed a mixture of periodicity and aperiodicity in the IMFs, resulting in the cyclic property of the DCC maximum profiles with a period of about 870 dpts (101 nm). The cyclic profile of the DCC maximum repeats for a shifting larger than this length. A maximum shifting of 3420 dpts is considered here to have the plateau of the DCC maximum profile with a width of about 2000 dpts (i.e., the plateau width of curves in red in the

schematic drawing of the DCC in Fig. 7). The shifting length of 3420 dpts is about a half of the whole length of A band or nebulin. For the shiftings larger than 2000 dpts, the DCC maximum profiles plotted in blue and green color in Fig. 7 no longer exist. Interestingly, this is roughly the upper bound of the optimal stretch length.

Next, we consider the situation of the paired IMFs c_9's. The paired IMFs c_9's have an average period of 1720 dpts (or 200 nm), which is about two times the characteristic length scale of the IMF c_8. Figure 8 shows the IMFs c_9's of the two data with various shiftings of nebulin with respect to the A band, DCC of the two IMFs, and a comprehensive schematic drawing of the DCC maximum profiles, plotted in the A-band coordinate. In this case, the moving window size T has been chosen as the average period of the IMF c_9, which is about 1720 dpts. Similarly, there is also cyclic property of the profile of the DCC maximum with a period of about 1750 dpts (or $\sim$200 nm). In contrast to the case of the paired IMFs c_8's, the DCC maximum profiles of the paired IMFs c_9's have more but shorter-width step functions as a result of a relatively shorter consistent oscillatory periods between the two IMFs. Further, within the first cycle, the DCC maximum profile shows a moving step function in the shifting direction of nebulin. A second step function appears when the shifting length is larger than 1500 dpts, and its profile completely shows up at a shifting length of 1750 dpts. This is about the upper bound of the optimal stretch length. The maximum shifting considered here is 4780 dpts, which is about 70% of the whole length of the A-band image data.

Finally, we consider the case of the paired IMFs c_7's. The average period of the IMF c_7 is 430 dpts (or 47 nm), which is about a half of the characteristic length scale of the IMF c_8. Figure 9 shows the IMFs c_7's of the two data with various
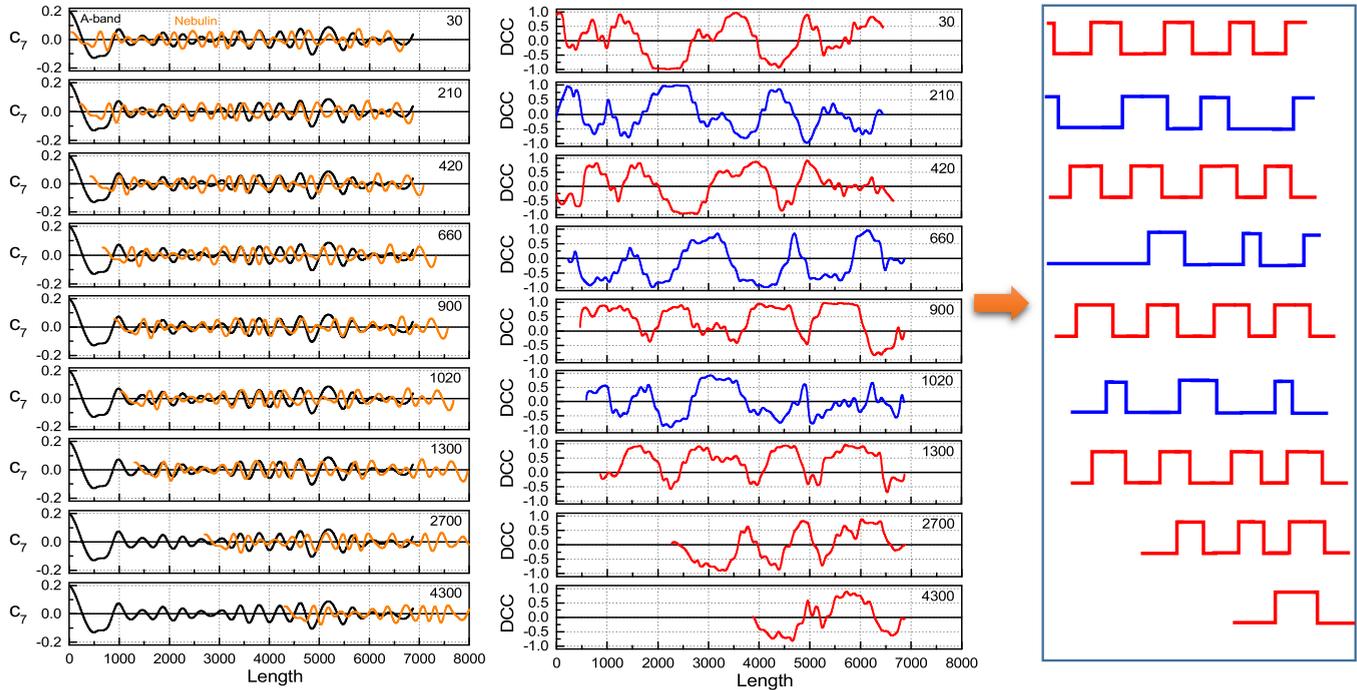
FIG. 9. Dynamical cross-correlation (DCC) analysis of the IMF c_7 with $T = 430$ dpts and a comprehensive schematic drawing of the DCC maximum profile at various data shiftings, plotted in the A-band coordinate.

shiftings of nebulin with respect to the A band, DCC of the two IMFs, and a comprehensive schematic drawing of the DCC maximum profiles, plotted in the A-band coordinate. Here we have chosen the moving window size of $T = 430$ dpts, which is about the average period of the IMF c_7. Again, there is also cyclic property of the profile of the DCC maximum, while there are more step functions coexisting in one DCC curve due to the fact of shorter oscillatory periods. The step functions move as the amount of the position shifting is increased. The cyclic property in this case has a period of about 400 dpts (45 nm), and the maximum shifting considered here is 4300 dpts. Interestingly, main structures of the DCC maximum profiles (curves in red color in Fig. 9) remain unchanged before the shifting reach the upper bound of the optimal stretch length (i.e., about 2000 dpts).

Here we remark that in this study the maximum position shifting (sliding) of nebulin with respect to the A band in the DCC analysis is defined through the existence of a specified DCC maximum profile. Both the window size T and individual DCC maximum profiles are relevant factors, whereby the DCC analysis is applicable only when there is a significant overlap between the nebulin and the A-band data sequences under position shiftings. Consequently, the situation of overstretch (and the completely extended case discussed in Ref. [49]), corresponding to an insufficient overlap (and no overlap in the case of completely extension) of the nebulin and the A-band data sequences, cannot be considered in the framework of our DCC analysis.

B. Supercontraction

According to Ref. [5], a moderate contraction of the sarcomere is applicable in practice for tension generation, as the normal extension case discussed above. This belongs

to the situation of supercontraction [49], which is an inverse of the normal extension. The full length of the sarcomere system under supercontraction is shortened, and the primary length of the A band sets a natural upper limit for the contraction if it is incompressible. To consider such a situation, we simulate the sarcomere contraction via negative position shiftings (i.e., sliding in opposite direction) of nebulin with respect to the A band, and the results of the DCC analysis and the DCC maximum profiles of the paired IMFs c_8's plotted in the nebulin coordinate for some typical shiftings ranging from -860 dpts to -4700 dpts, are shown in Fig. 10. The feature of the DCC maximum profile for supercontraction is different from the cases of normal extension, both in shape and in distribution of step functions. It can be seen from the schematic drawing of the DCC maximum profile that there are two step functions when the paired IMFs c_8's have a relative position shifting of -860 dpts. The width of the first step function narrows down as the shifting length is increased, and the first step function disappears for the shifting larger than -2640 dpts. In contrast, the second step function resides at the same position for a shifting length ranging from -860 dpts to -2640 dpts. It disappears after a larger position shifting than -2640 dpts, and a relative narrower step function is further produced for a position shifting of -4700 dpts. Thus, the appearance of the markerlike second step function is relevant to the position sensing before breakdown.

Note that our DCC analysis does not take into account the situation of complete supercontraction where the shifting length is larger than the whole length of the A band and nebulin, which involves more than one nebulin and one A-band segment. In short, the above analysis suggests that the DCC concept is generally applicable to both normal extension and supercontraction.
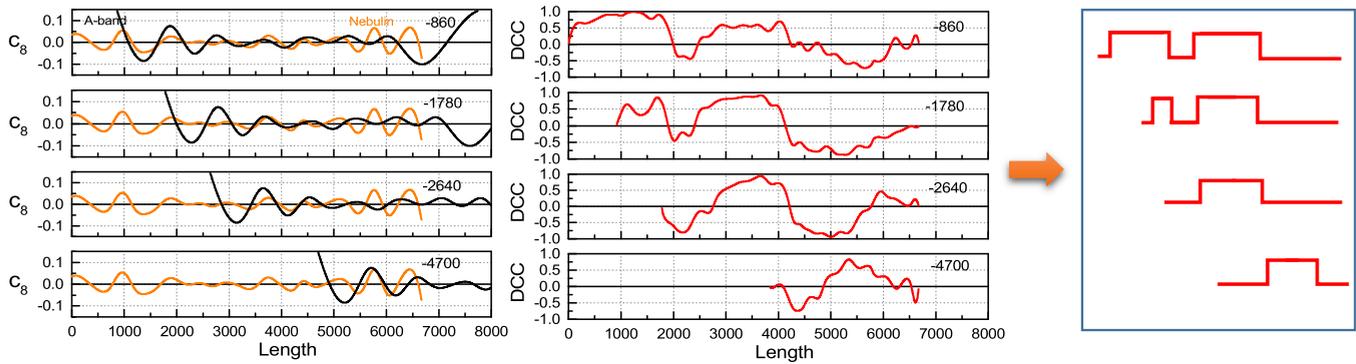
FIG. 10. Dynamical cross-correlation (DCC) analysis of the IMF c_8 with $T = 860$ dpts and a comprehensive schematic drawing of the DCC maximum profile for some typical data shiftings for supercontraction, plotted in the nebulin coordinate.

C. Vernierlike design for position sensory

The DCC maximum profile reflects some kind of structure match and consistency between the A-band intensity data and the nebulin disorder profile. The cycle representing reentrant patterns in DCC of one IMF forms a periodic marker for the sliding machinery functioning as a linear position sensor. Combination of the scale-dependent correlations allows finer position sensing for sarcomere overlap. The distinct cyclic properties of the DCC maximum profiles of the paired IMFs c_7's, c_8's, and c_9's with characteristic length scales of 45 nm, 100 nm, and 200 nm respectively then depict a vernierlike design for position sensing of sliding A band on nebulin, as shown in Fig. 11. For simplicity, here we have used a number of geometrical objects in distinct shapes, sizes, and colors to represent least distinct structural specificities in the paired IMFs at different length scales. When the structural properties encoded in the two sequences at the same scales match as a result of proper shiftings (stretches), the resultant DCC maximum profile shows a sequence of moving step functions, which are similar to the schematic drawings of the DCC maximum profiles in Figs. 7, 8, 9, and 10. The purple block for c_7 in Fig. 11 is applicable for both red and blue curves in Fig. 9. Remarkably, as shown in Fig. 11, to reproduce the DCC maximum profiles at the characteristic length scale of the paired IMFs c_8 with various position shiftings (A band is pulled leftward), the structure information encoded in the first cyclic period measured from the N terminus of the sequence (the left end of the disorder profile) of nebulin is sufficient
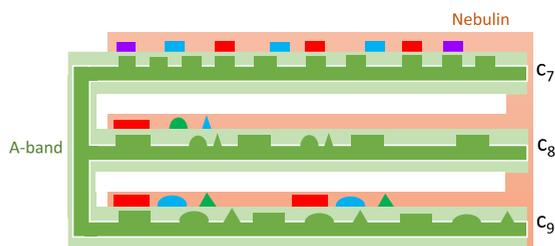
and adequate. More structural properties can be added, but these are redundant. This sufficient and adequate information is the least structural specificities. Meanwhile, the structural property of the A band emerging from the myosin-binding repeats shows relatively higher regularity over the characteristic length scales. Therefore, this inference from the vernierlike design suggests the least specificities of the nebulin amino acid sequence at different scale levels in combination with the A-band structural regularity for position sensing. A still unclear point is the biochemical basis for the recognition of the DCC maximum profile at various length scales. Experimental tests and supports to the proposed model require further investigations.

VII. CONCLUSIONS

In conclusion, we have investigated whether the disorder profile of nebulin may encode guidelines for the thin and thick filament interactions in the sarcomere of the skeletal muscle, by analyzing the DCC of the predicted disordered probability of human nebulin and the A-band image intensity data using adaptive time series analysis approaches. The data sequences of the nebulin disorder profile predicted by PONDR using a VL3H predictor and the A-band intensity data obtained from image data were preprocessed to align using the concept of consistent scale dependence in their intrinsic correlations assessed by DFA, and were then respectively decomposed by EMD into finite components each with a distinct oscillatory frequency. The two data sequences possess significant periodicities predominantly in the components with the characteristic length scales of 45 nm, 100 nm, and 200 nm belonging to the long-range correlation regime defined in DFA analysis, and the results of the DCC analysis of the three paired components with various relative position shiftings depict a vernierlike design for position sensing. Such a design enables nebulin to sense position and fine tune sarcomere overlap. Consequently, this study shows that the disorder profile of scaffolding proteins may encode a guideline for cellular architecture.

It is emphasized that the vernierlike design ruler model is heuristic and may inspire further inferences for the origin of the structural specificities in the nebulin disorder profile in different length scales. However, though the disorder profile of nebulin was predicted through its amino acid sequence

FIG. 11. Model for the vernierlike position sensor based on the different scale properties in the IMF components c_7, c_8, and c_9. Least structural specificities represented by geometric objects have been used to reproduce the DCC maximum profiles in Figs. 7, 8, 9, and 10, when the A band is pulled leftward with respect to nebulin kept fixed.

using PONDR [37], the connection between structural disorder and amino acid sequence is not straightforward. A direct inference from nebulin sequence is unrealistic. Besides, it is also unclear that what is the biochemical basis for recognizing the DCC maximum to implement the vernierlike design ruler. The current work cannot answer these questions. All of these facts are topics for further study in future works.

[1] A. F. Huxley and R. Niedergerke, Nature (London) **173**, 971 (1954).

[2] H. Huxley and J. Hanson, Nature (London) **173**, 973 (1954).

[3] A. F. Huxley, Prog. Biophys. Biophys. Chem. **7**, 255 (1957).

[4] A. F. Huxley and R. M. Simmons, Nature (London) **233**, 533 (1971).

[5] A. M. Gordon, A. F. Huxley, and F. J. Julian, J. Physiol. **184**, 170 (1966).

[6] D. E. Rassier, B. R. MacIntosh, and W. Herzog, J. Appl. Physiol. **86**, 1445 (1999), and references therein.

[7] C. A. C. Ottenheijm and H. Granzier, J. Biomed. Biotech. **2010**, 968139 (2010).

[8] K. Wang and J. Wright, J. Cell Biol. **107**, 2199 (1988).

[9] V. K. Yadavalli, J. G. Forbes, and K. Wang, Langmuir **25**, 7496 (2009)

[10] K. Wang, M. Kniper, Q. Q. Huang, A. van Heerden, L. Hsu, and G. Gutierrez, J. Biol. Chem. **271**, 4304 (1996).

[11] A. S. McElhinny, B. Kolmerer, V. M. Fowler, S. Labeit, and C. C. Gregorio, J. Biol. Chem. **276**, 583 (2001).

[12] M. Kruger, J. Wright, and K. Wang, J. Cell Biol. **115**, 97 (1991).

[13] S. Labiet, T. Gibson, A. Lakey, K. Leonard, M. Zeviani, P. Knight, J. Wardale, and J. Trinick, FEBS Lett. **282**, 313 (1991).

[14] M.-J. G. Chen, C.-L. Shih, and K. Wang, J. Biol. Chem. **268**, 20327 (1993).

[15] O. Ogut, M. M. Hossain, and J.-P. Jin, J. Biol. Chem. **278**, 3089 (2003).

[16] A. S. McElhinny, C. Schwach, M. Valichnac, S. Mount-Patrick, and C. C. Gregorio, J. Cell Biol. **170**, 947 (2005).

[17] J.-P. Jin and K. Wang, J. Biol. Chem. **266**, 21215 (1991).

[18] C. G. dos Remedios, D. Chhabra, M. Kekic, I. V. Dedova, M. Tsubakihara, D. A. Berry, and N. J. Nosworthy, Physiol. Rev. **83**, 433 (2003).

[19] C. C. Witt, C. Burkart, D. Labeit, M. McNabb, Y. Wu, H. Granzier, and S. Labeit, EMBO J. **25**, 3843 (2006).

[20] A. S. McElhinny, S. T. Kazmierski, S. Labeit, and C. C. Gregorio, Trends Cardiovascular Med. **13**, 195 (2003).

[21] D. D. Root and K. Wang, Biochemistry **33**, 12581 (1994).

[22] M.-L. Bang, C. Gregorio, and S. Labeit, J. Struct. Biol. **137**, 119 (2002).

[23] G. M. Conover, S. N. Henderson, and C. C. Gregorio, Mol. Biol. Cell **20**, 834 (2009).

[24] P. Tonino, C. T. Pappas, B. D. Hudson, S. Labeit, C. C. Gregorio, and H. Granzier, J. Cell Sci. **123**, 384 (2010).

[25] C. T. Pappas, N. Bhattacharya, J. A. Cooper, and C. C. Gregorio, Mol. Biol. Cell **19**, 1837 (2008).

[26] O. Mayans, P. F. van der Ven, M. Wilm, A. Mues, P. Young, D. O. Fürst, M. Wilmanns, and M. Gautel, Nature (London) **395**, 863 (1998).

[27] E. M. Puchner, A. Alexandrovich, A. L. Kho, U. Hensen, L. V. Schäfer, B. Brandmeier, F. Gräter, H. Grubmüller, H. E. Gaub, and M. Gautel, Proc. Natl. Acad. Sci. USA **105**, 13385 (2008).

[28] M.-C. Wu, J. G. Forbes, and K. Wang, Soft Matter **9**, 9897 (2013).

[29] W. A. Linke, M. Ivemeyer, P. Mundel, M. R. Stockmeier, and B. Kolmerer, Proc. Natl. Acad. Sci. USA **95**, 8052 (1998).

[30] M. L. Greaser, S. M. Wang, M. Berri, P. Mozdziak, and Y. Kumazawa, Adv. Exp. Med. Biol. **481**, 53 (2000).

[31] M. Sjöström and J. M. Squire, J. Microscopy **111**, 239 (1977).

[32] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Phys. Rev. E **49**, 1685 (1994).

[33] H. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Z. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, Proc. R. Soc. London A **454**, 903 (1998).

[34] C. Torrence and G. P. Compo, Bull. Am. Meteorol. Soc. **79**, 61 (1998).

[35] J. Zhang and M. Small, Phys. Rev. Lett. **96**, 238701 (2006).

[36] J. Zhang, J. Sun, X. Luo, K. Zhang, T. Nakamura, and M. A. Small, Physica D **237**, 2856 (2008).

[37] P. Radivojac, Z. Obradovic, C. J. Brown, and A. K. Dunker, Pac. Symp. Biocomput. **8**, 216 (2003).

[38] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Golderger, Chaos **5**, 82 (1995).

[39] M.-C. Wu and C.-K. Hu, Phys. Rev. E **73**, 051917 (2006).

[40] Alternative definitions of the orthogonality index are possible, but the result will not be substantially affected under the minimization criterion of index values. See Ref. [33] for details.

[41] Z. Wu and N. E. Huang, Adv. Adapt. Data Anal. **01**, 1 (2009).

[42] M.-C. Wu, M.-C. Huang, H.-C. Yu, and T. C. Chiang, Phys. Rev. E **73**, 016118 (2006).

[43] M.-C. Wu, Physica A **375**, 633 (2007).

[44] M.-C. Wu, J. Korean Phys. Soc. **50**, 304 (2007).

[45] M.-C. Wu, E. Watanabe, Z. R. Struzik, C.-K. Hu, and Y. Yamamoto, Phys. Rev. E **80**, 051917 (2009).

[46] M.-C. Wu, Europhys. Lett. **97**, 48009 (2012).

[47] M.-C. Wu and T. Y. Tsong, J. Phys. Soc. Jpn. **82**, 114801 (2013).

[48] R. S. Littlefield and V. M. Fowler, Seminars Cell Dev. Bio. **19**, 511 (2008).

[49] A. Herrel, J. J. Meyers, J.-P. Timmermans, and K. C. Nishikawa, J. Exp. Biol. **205**, 2167 (2002).