THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

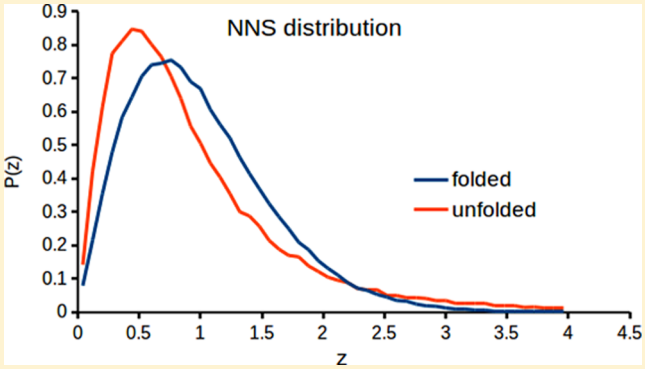
Protein States as Symmetry Transitions in the Correlation Matrices

Luigi L. Palese*

Department of Basic Medical Sciences, Neurosciences and Sense Organs (SMBNOS), University of Bari "Aldo Moro", Piazza G.Cesare − Policlinico, 70124 Bari, Italy

Supporting Information

ABSTRACT: Over the last few years, there has been significant progress in the knowledge on protein folding. However, some aspects of protein folding still need further attention. One of these is the exact relationship between the folded and unfolded states and the differences between them. Whereas the folded state is well known, at least from a structural point of view (just think of the thousands of structures in online databases), the unfolded state is more elusive. Also, these are dynamic states of matter, and this aspect cannot be overlooked. Molecular dynamics-derived correlation matrices are an invaluable source of information on the protein dynamics. Here, bulk eigenvalue spectra of the correlation matrices obtained from the Trp-cage dynamics in the folded and unfolded states have been analyzed. The associated modes represent localized vibrations and are significantly affected by the fine details of the structure and interactions. Therefore, these bulk modes can be used as probes of the protein local dynamics in different states. The results of these analyses show that the correlation matrices describing the folded and unfolded dynamics belong to different symmetry classes. This finding provides new support to the phase-transition models of protein folding.



INTRODUCTION

Proteins are linear polymers of amino acid residues that undergo reversible folding/unfolding transitions. In general, the dynamics of an isolated protein is dictated by the sequence of the protein itself.[1] Also, in protein-folding dynamics, besides the case in which chaperon protein intervention is mandatory, all of the information needed to realize the folded native shape from the unfolded conformations is encoded within the amino acid sequences.[2] Understanding the protein-folding phenomenon is critical not only because, despite considerable advances, it still presents some unresolved issues,[3] but also because a growing number of severe diseases are currently related to impairment of protein folding.[4−6] Given an accurate force field and a sufficiently long simulation time, molecular dynamics (MD) is actually able to predict and describe experimentally observable phenomena, including the atomic-level trajectories of folding and unfolding events.[7−10] State-of-the-art computational protocols and force fields[11] have enabled researchers to obtain repeated folding and unfolding events of a set of small structurally diverse and fast-folding proteins.[12] The obtained structures are strikingly similar to those determined experimentally. So, these fast-folders can provide direct comparisons between experiments and simulations.

In the thermodynamic framework,[2,13−15] protein folding can be described as an equilibrium between the unfolded and native conformations, whose ΔG^0 typically falls in the range between −5 and −15 kcal/mol.[16,17] In this theoretical context, several models of protein folding have been proposed, but all can be grouped into two broad categories: protein folding as a chemical reaction and protein folding as a physical change of state.[15,18] Considering the folding processes as (unimolecular) chemical reactions[19] implies that these are governed by a single rate-limiting step, which corresponds to the transition state. Chemical reaction rate theory recognizes few coordinates, representing the slowly varying degrees of freedom, as the reaction coordinates along which the maximum free energy change occurs.[20,21] Even if suggestive, the chemical analogy of protein folding presents some difficulties. Another proposed analogy that might be even more appropriate is between folding and first-order phase transitions.[18,22] These are highly cooperative processes that typically occur through nucleation and growth events.[15,18,20,23] Although a note of caution is obviously needed while speaking about phase transitions in a single molecule, this model of protein folding is undoubtedly considerably successful and physically sound.

It has been previously demonstrated that the low-energy spectrum of Trp-cage correlation matrices in the folded and unfolded states is different from that of random systems.[24] Here, the focus will be on the high-energy sector of the spectrum. This spectral region describes bulk modes that represent localized vibrations that are significantly affected by fine details of the protein structure.

Received: September 12, 2016
Revised: October 17, 2016
Published: October 18, 2016

DOI: 10.1021/acs.jpcb.6b09216
J. Phys. Chem. B 2016, 120, 11428−11435

METHODS

MD simulations were performed essentially as described previously.[24] The starting structure of Trp-cage was 1L2Y,[25] modeled by VMD.[26] MD simulations were performed with NAMD2,[27] using the CHARMM22 force field with CMAP correction,[28,29] in boxes of TIP3P[30] water molecules (see the Supporting Information (SI) for further details).

The Cartesian coordinates of the Trp-cage backbone atoms were extracted and arranged in a matrix (whose rows contained the different time-dependent conformations) by Tcl (www.tcl.tk) and Vim (www.vim.org) scripting. Numerical and data analyses were carried out using the NumPy,[31] Scipy (Jones E., Oliphant T., and Peterson P., and others; www.scipy.org) and Matplotlib[32] Python (www.python.org) packages, implemented in IPython.[33,34] Also see the SI for numerical details. GIMP 2.8 (www.gimp.org) was used for final editing of the images.

RESULTS AND DISCUSSION

The bulk behavior of the local eigenvalue correlation statistics has been analyzed using the Trp-cage dynamics. Trp-cage (NLYIQWLKDGGPSSGRPPPS) is a 20-residue polymer derived from the C-terminus of the exendin-4 peptide.[25] It folds spontaneously in about 4 μs[35] into a cage structure that contains a short α-helix from residues 2 through 8, a 3_{10}-helix from residues 11 through 14, and a polyproline II helix at the C-terminus[25] (see Figure S1). Trp-cage shows a globular-like fold, where three proline residues (Pro-12, Pro-18, and Pro-19) and a glycine (Gly-11) pack against the aromatic side chains of the central tryptophan (Trp-6) and tyrosine (Tyr-3) residues.[25] This structure is stable below 315 K, which is the melting temperature (T_m) of this miniprotein.[25,36] The small size of Trp-cage, besides its very short folding time, has made this miniprotein an ideal model system for protein-folding studies under different conditions.[24,35,37−44] For this purpose, a series of simulation runs have been conducted, essentially as described,[24] at temperatures above and below the T_m of Trp-cage, as representative of the unfolded and folded conditions, respectively. Unless otherwise stated, simulations that have been described below last for 1 μs, with a sampling time in the picosecond range.

The root-mean-square deviations (RMSDs) of the sampled conformations aligned to a reference structure were calculated for monitoring the large-scale structural fluctuations of the protein. The simulation performed at 300 K confirms the Trp-cage stability at this temperature.[24,25,36] The RMSD (as the mean overall deviation from the reference structure and its relative standard deviation) was 3.09 ± 0.36 Å. The Trp-cage structure remained essentially unaltered on the microsecond time scale, with major fluctuations observed in the extreme N- and C-terminal regions (see Figures S2 and S3).

High-temperature simulations were performed in two different ways: in the first, the starting Trp-cage structure was completely unfolded by a harsh heating protocol, whereas in the second, the starting structure was the folded Trp-cage (see SI). In the following discussion, we will refer to these protocols as the unfolded high-temperature (UHT) simulation and folded high-temperature (FHT) simulation, respectively. In the UHT simulation, the RMSD (and its relative standard deviation) was 7.90 ± 1.92 Å. It should be noted that this simulation ended with the Trp-cage protein in a fully folded structure (see Figure S2). Root-mean-square fluctuation (RMSF) analysis showed that the overall residue mobility was higher than that at 300 K, again, with the N- and C-terminal regions of the cage characterized by a greater mobility compared to that of the rest of the protein (Figure S3). In the FHT protocol, the Trp-cage three-dimensional structure was essentially unchanged during the simulation time course, with only transient unfolding events (Figure S2). The RMSD for this quasi-denatured simulation was 3.25 ± 0.95 Å. Also, the RMSF analysis suggested that the overall mobility was only slightly higher than that observed in the 300 K simulation (see Figure S3).

To further characterize the Trp-cage dynamics under different conditions, we carried out principal component analysis (PCA), as described previously.[45] PCA was based on eigenvector decomposition of the correlation matrix. This matrix, which is square and symmetric by construction, was diagonalized to obtain the so-called principal components (see the SI for a detailed description of the method and its numerical implementation). The eigenmodes of this matrix that are associated with the largest eigenvalues are responsible for collective and large-scale motions and correspond to low-energy excitations. These modes are consistent and robust when compared among different and independent simulations, and it has been demonstrated that they are well distinguishable from a multidimensional random dynamics.[46] In the MD run performed at 300 K, a few adjacent and narrow regions in the conformational landscape were recursively visited (see Figure S4), which correspond to small changes in the relative orientations of the secondary structure motifs.[24] In the PCA of the FHT simulation, reported in Figure S5, the most populated region was represented by a narrow and deep cluster, composed of three potential wells, that corresponds to the conformations near the native structure observed over most of the time course of the simulation. To this cluster, a vast and shallow region of the conformational landscape was added, which was sparsely populated, corresponding to the conformations explored during the transient unfolding events. Conversely, for the simulation at 335 K, performed using the UHT protocol (see Figure S6), PCA showed that a single large region in the conformational landscape was explored, in which no separated clusters could be described. However, a single deep well was observed, corresponding to frequently and recurrently visited conformations having native-like secondary structures (particularly the polyproline II helix and the 3_{10}-helix) that are arranged in a globule-like conformation but in which the native well-packed cage is not yet assembled.

A convenient starting point for the analysis of protein dynamics is the study of the correlation matrices that characterize the system's dynamics. Random matrix theory[47,48] (RMT) offers universal forecasts on the behavior of the eigenvalue distributions of some symmetric random matrix ensembles and, particularly, the expected value for maximum eigenvalue distribution.[46] These properties of random matrix ensembles permit us to estimate the boundaries at the low-energy site of the eigenspectrum for random, noncorrelated systems. The study and application of matrices with random entries have a long tradition, dating back to the Wishart studies on mathematical statistics.[49] Random matrices were introduced later in physical modeling by Wigner and Dyson.[47,48] After a long period, during which RMT was confined to the study of the statistics of eigenvalues and eigenfunctions of complex, many-body quantum systems,[50,51] over the last few years, this theory has undergone rapid development. Originally, it was conceived as an "alternative" statistical mechanics, in which,

DOI: 10.1021/acs.jpcb.6b09216
J. Phys. Chem. B 2016, 120, 11428−11435

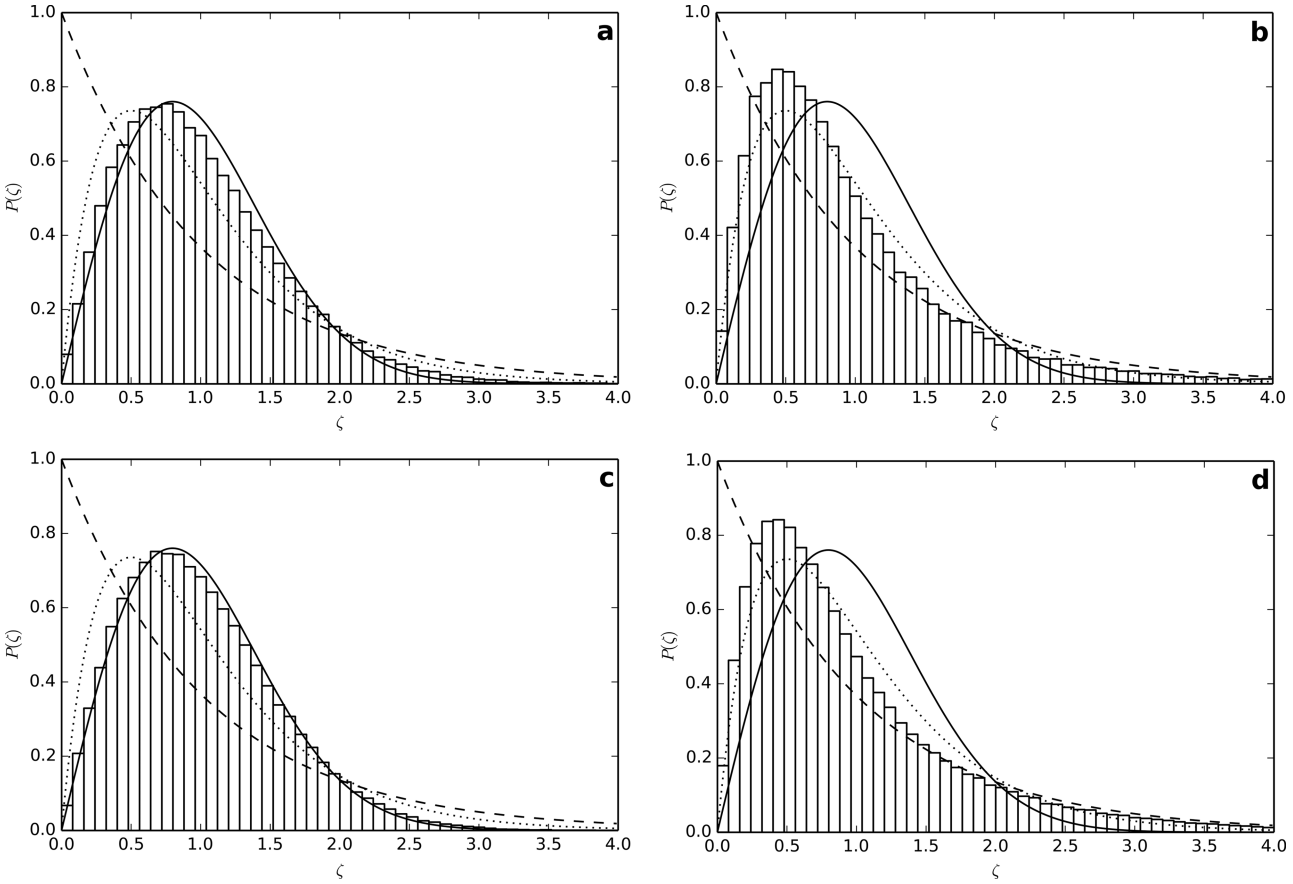


Figure 1. NNS distributions of the MEs obtained from the 300 K (a) and UHT (b) simulations. NNS distributions of the numerically calculated ensembles: the random band matrix (RBM) ensemble with $B = 32$ (c) and the perturbed Poisson (PP) ensemble with $\alpha = 0.005$ (d). The Wigner–Dyson (continuous line), semi-Poisson (dotted line), and Poisson (dashed line) distributions are also reported.

instead of considering an ensemble of identical physical systems governed by the same Hamiltonian, the focus was on an ensemble of Hamiltonians that share the same symmetry property.[47,48,51] Currently its applications extend well beyond the field of many-body quantum systems.[52] Random matrices exhibit a peculiar pattern in the eigenvalue distribution, a phenomenon called "universality". Universality also arises when a system is composed of many parts that interact strongly with each other. The pertinent random matrix ensemble for MD-deriving data sets is the Wishart one.[24,46,49] Here, the upper boundary of this random ensemble has been obtained by the shuffling method,[24,46] and from this, the number of large nonrandom eigenvalues of the MD-deriving correlation matrices has been calculated (see the SI for numerical details). A significant number of eigenvalues of these empirical correlation matrices are above the RMT expected threshold. The exact number of large nonrandom eigenvalues depends on the simulation experimental setup, the sampling frequency, and the time scale of the empirical data sets. The calculated number of nonrandom eigenvalues for the 1 μs simulation at 300 K was 24, whereas for the simulations at 335 K, it was 22 and 20 under the FHT and UHT conditions, respectively.

These simulations were used to generate an ensemble of nonoverlapping short dynamics for each experimental setup. Each ensemble contained 300 matrices of 333 time points; so, each single matrix corresponded to 3.3 ns sampled every 10 ps (we will refer to these matrix ensembles as ME; see the SI for numerical details). The mean nonrandom eigenvalue number for the ME obtained from the 300 K simulation was 7, and for the ensembles from the 335 K simulations, it was 7 and 5 under the FHT and UHT conditions, respectively. Despite the relatively low number of data points in these matrices, at least two eigenvalues were above the Wishart threshold in the worst case. Anyway, and this being an important point concerning this work, these results confirm that each matrix in these ME is not a pure random one, which means that the low-energy side of the spectra of these matrices contains a description of truly correlated (and nonrandom) motions.[24,46]

Whereas large eigenmodes that contain low-energy excitations play a relevant role in protein functionality and account for a large fraction of the overall fluctuations, the bulk of eigenvalues describes high-energy, localized vibrations. These bulk, noise-dressed eigenvalues can be studied only from a statistical point of view. RMT predicts well-defined distributions for the nearest neighboring spacings (NNSs).[53,54] The individual eigenvalue spacings are calculated as

$$\zeta_i = \frac{(\lambda_{i+1} - \lambda_i)}{\langle \lambda_{i+1} - \lambda_i \rangle}$$

where $\langle \cdots \rangle$ is the average over all matrices in the empirical ensemble. The purpose of this procedure, which is generally indicated as "unfolding", is to separate the average spectral density, which is not universal, from the spectral fluctuations, which are instead universal. This procedure is essentially a local rescaling of the eigenvalues, such that in the unfolded spectrum the mean level spacing equals unity. Because the analysis of the

DOI: 10.1021/acs.jpcb.6b09216
J. Phys. Chem. B 2016, 120, 11428–11435

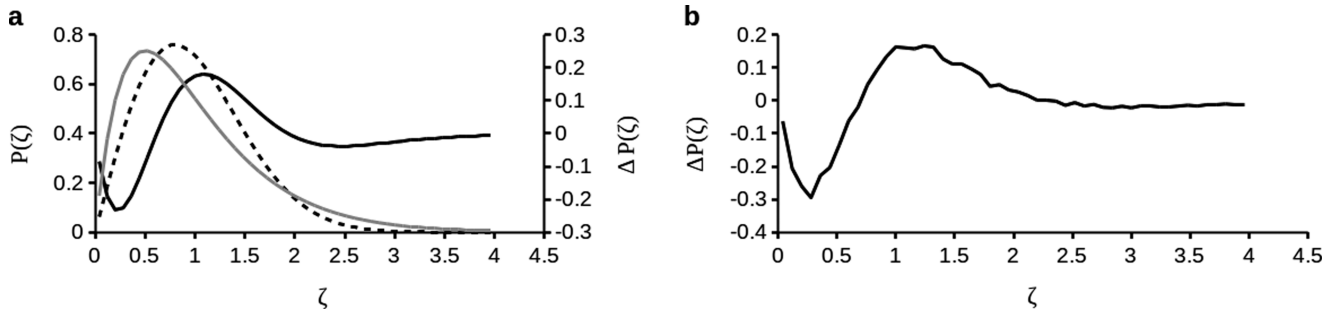


Figure 2. (a) Calculated NNSs of the Wigner–Dyson (black dashed line) and semi-Poisson (gray continuous line) distributions and their difference (Wigner–Dyson minus semi-Poisson; black continuous line). (b) Difference in the NNS distributions obtained from the 300 K (folded dynamics) and UHT (unfolded dynamics) simulations.

NNS distributions requires a suitable number of matrices, we use the above-mentioned ME for this purpose. The distances between different NNS distributions were calculated by the Kolmogorov–Smirnov (KS) statistic on two samples (see the SI for numerical details). The KS distance between the 300 K and FHT simulations was 0.0269 ($p = 2 \times 10^{-23}$), whereas the distance between the folded and unfolded simulations (UHT) was 0.1112 ($p = 0$). The KS distance between the two high-temperature simulations (FHT and UHT) was 0.0982 ($p \approx 0$). Therefore, this analysis confirms that the above-described NNS distributions are really different.

For the Gaussian orthogonal ensemble (GOE) matrices,[55] the NNS distribution is given by

$$P(\zeta) = \frac{\pi}{2}\zeta\ e^{-\pi\zeta^2/4}$$

also known as the "Wigner surmise" or Wigner–Dyson distribution (see Figures 1 and 2).

The unfolded NNS distributions of the ME described above reveal a series of interesting differences under the different simulation conditions. In the case of the ensemble obtained from the 300 K simulation, the Wigner–Dyson distribution fits reasonably well to the dataset (although not completely satisfactory; see below), as reported in Figure 1a. However, it is when we analyze the ensembles obtained from the high-temperature simulations that the fitting by the Wigner–Dyson distribution becomes increasingly unsatisfactory. These NNS distributions are reported in Figure S7 for the ensemble obtained under the FHT conditions and in Figure 1b for the ensemble derived from the simulation under the UHT condition. Even if the Wigner–Dyson distribution is a good description for the observed NNS distributions in several real-world examples, in general, even in the RMT field, the one-parameter Brody distribution is considered for the fitting of the NNS statistics.[54,56,57] This distribution can be represented as

$$P(\zeta) = B(1+\beta)\zeta^{\beta}\ e^{-B\zeta^{1+\beta}}$$

where

$$B \equiv \left[\Gamma\left(\frac{\beta+2}{\beta+1}\right)\right]^{1+\beta}$$

and Γ is the gamma function. The case $\beta = 1$ corresponds to the GOE, whereas $\beta = 0$ corresponds to uncorrelated eigenvalues and hence to Poisson-distributed spacings. The β-values of the best-fitting Brody distributions were 0.82, 0.76, and 0.73 for the ME obtained from the simulations performed at 300 K, 335 K under FHT conditions, and 335 K under UHT conditions, respectively. These values are very similar to those reported before using a different model.[54] However, the resulting fitting is very poor, as demonstrated by Figures S8–S10, particularly for the ME obtained from the UHT simulation. A distribution whose appearance is significantly better for the description of this NNS statistic is apparently the so-called semi-Poisson distribution[58]

$$P(\zeta) = 4\zeta\ e^{-2\zeta}$$

at least with respect to the peak position (see Figure 1b).

These data collectively suggest that the distributions that directly originate from the GOE (the Wigner–Dyson distribution and its surrogate, i.e., the Brody distribution) can be used only as approximate descriptions of what is obtained from the MD data. An intriguing possibility is that different random matrix ensembles could be better descriptors of the eigenvalue statistics of MD-derived correlation matrices and that a kind of transition between different classes of random matrices occurs when the dynamics of Trp-cage switches between the folded and unfolded states.

As mentioned above, RMT is extremely powerful in describing the statistical behavior of quantities arising in complex systems. Its wide range of successfully applications is rooted on the fact that RMT describes the quantities of a system that do not depend on the detailed dynamical properties of the system itself. Instead, the quantities that are predicted by RMT are the consequence of global symmetries that are common to all systems that belong to a given symmetry class. Some physical systems consist of parts with different symmetries or, equivalently, classically integrable and chaotic parts. Billiards[50] are paradigmatic examples of these systems: in fact, it has been shown that certain billiards can undergo Poisson to GOE transitions.[59,60] In systems in which the time-reversal symmetry is broken down, a GOE to Gaussian unitary ensemble (GUE) transition has been described.[60,61] Transitions from Poisson to GOE and from GOE to GUE have been observed in the spectrum of hydrogen atoms in a magnetic field.[60,62] Transitions from Poisson to GOE or GUE have been reported in the metal–insulator (or Anderson) transition.[63] Intriguingly, transition from Poisson to GOE statistics was reported for the interpoint distance as the embedding dimension changes.[64] In this context, the Brody parameter, which does not have a precise physical meaning in the Hamiltonian dynamics framework, can be represented as a measure of a fractal dimension.

What is needed here is a model for the Wigner to Poisson transition that could account for the peculiar distributions of NNSs observed in the reported ME. One useful starting point

DOI: 10.1021/acs.jpcb.6b09216
J. Phys. Chem. B 2016, 120, 11428–11435

is the observation that whereas the GOE matrices lead to the Wigner−Dyson distribution the Poisson distribution of the NNS statistics can be easily obtained from a random diagonal matrix ensemble. This, together with fact that the symmetries of matrices belonging to a particular class are most important in determining the statistical properties of random matrices, suggests immediate exploration of the consequences of matrix symmetry alterations on the NNS of the ensemble. Of the examined random symmetric matrix ensembles, two appeared to be more interesting in the description of the ME-deriving NNS distributions. The first was the random band matrix (RBM) ensemble with uncorrelated entries. A band matrix is a particular form of sparse matrix whose nonzero entries are confined to one or more diagonals on either side. The number of diagonals with nonvanishing elements is indicated as band width B (upper and lower). It is interesting to note that for RBM of dimensions N by N, above a certain threshold (of the order of $B \approx \sqrt{N}$), the eigenvalue NNS statistics is essentially of the Wigner−Dyson type (Figure 1c).[65] Below this threshold, semi-Poisson statistics is observed (Figure S7), which decays progressively into Poisson statistics with the decrease in B. However, even if perfect semi-Poisson statistics can be obtained with a band width of less than 15 (consider the dimensions of the correlation matrices that we are interested in), the obtained distributions are not satisfactory as descriptors of the unfolded dynamics under UHT conditions (not shown). We also analyzed ensembles obtained by adding to a random diagonal matrix a symmetric random matrix. These ensembles are composed of matrices having the form

$$H = H_P + \alpha H_{WD}$$

where H_P is the random diagonal matrix characterized by Poisson-distributed NNS, H_{WD} is a classical random self-adjoint matrix whose NNS follows Wigner−Dyson statistics, and $0 \leq \alpha \leq 1$ is a real number representing the strength of the symmetric perturbation. We call this random matrix ensemble as perturbed Poisson (PP). At the extreme values of the perturbation, the PP ensemble recovers perfect Poisson and Wigner−Dyson statistics. But for intermediate values, it can be observed a striking similarity between the PP ensemble and the observed NNS statistics of the MD-derived correlation matrices (Figure 1d).

The distance of each experimental NNS distribution with respect to the numerically obtained ones has been evaluated. For the RBM ensemble, the number of nonzero diagonals was the tuned parameter, whereas in the case of the PP ensemble, this was represented by the α parameter. In the case of the ensemble derived from the 300 K folded simulation, the KS distance of its NNS distribution compared to that of the GOE ensemble (an ensemble containing 1000 matrices was used; all the numerically obtained ensembles described here contained the same number of matrices) was 0.0258, with $p = 6.6 \times 10^{-11}$. The best-performing RBM ensemble was characterized by a value of $B = 21$, with a KS distance of 0.0047 ($p = 0.8255$), whereas for the PP ensemble, the best results were obtained using $\alpha = 0.03$, with a distance of 0.0051 and $p = 0.1129$. It can be concluded that in the case of this simulation the best model is the RBM ensemble with a large number of nonzero diagonal elements (very similar to the Wigner−Dyson distribution). The FHT simulation leads to an NNS distribution whose best model is the RBM ensemble ($B = 16$, KS distance 0.0082, $p = 0.1695$). The best-fitting PP ensemble ($\alpha = 0.02$) was characterized by a KS distance of 0.0202 but with probability $p = 7 \times 10^{-7}$; the GOE performed poorly, with a KS distance of 0.0453 and $p = 9 \times 10^{-33}$. So, similar to that in the case of the 300 K ensemble, the best symmetry class for describing the NNS distribution obtained from the FHT simulation was the RBM ensemble. The NNS distribution obtained from the simulation under UHT conditions was very poorly described by both the GOE and RBM ensembles. In the case of the GOE ensemble, the calculated KS distance was 0.1323 and the probability that the two distributions are the same was absolutely negligible ($p \approx 0$). Also, the results obtained from the best-performing RBM ensemble led to the conclusion that the two NNS distributions cannot be the same (KS distance 0.0518, $p = 8 \times 10^{-43}$, obtained at $B = 8$). Of the tested ensembles, the best results were obtained with the PP ensemble characterized by an α factor equal to 0.005. The KS distance was 0.0152, with calculated probability $p = 0.0004$. It should be noted that on dividing the 1 μs simulations in time windows up to 33 ns the symmetry classes that are found to be the best models for the NNS distributions obtained under different simulation conditions are those reported above (not shown).

Whatever is the exact symmetry class that best fits the correlation matrices behind a given simulation condition, it should be stressed that the maximum of the eigenvalue NNS distributions is shifted toward higher values in the case of ensembles derived under folded conditions compared to that obtained under unfolded conditions (this is also valid for all tested ME, obtained also from different simulation setups; not reported). The difference between the distributions arising from the ME under folded and UHT conditions, which is shown in Figure 2, clearly shows this shift. All these distribution differences are similar to the difference of the Wigner−Dyson distribution respect to the semi-Poisson distribution.

The folded and unfolded states of proteins can be thought of as different phases of the protein state of matter, and in fact, the folding/unfolding process has been described as a first-order phase transition (see above). It is interesting to note that most phases differ in their symmetries, even if not all phases differ for some symmetries.[66] It can be stated (and in this case in general) that if two states of the same material have different symmetries then they are different phases. The results reported in this work clearly show that the Trp-cage in the unfolded state and the same molecule in the folded state differ with respect to the symmetry class of the correlation matrices that characterize their dynamics. The symmetry class appropriate for the folded dynamics is the RBM ensemble (but with a large number of nonzero diagonals), whereas the symmetry class for the unfolded dynamics could be the PP ensemble. As it has been demonstrated above, these differences cannot be ascribed to the temperature at which the simulation is performed. In fact, of the two simulations performed at a temperature above the T_m of this protein, that in which the folded conformations dominate the dynamics appears to be in the same symmetry class of the simulation performed below the T_m, in which the protein explores only the folded conformations. We must also consider the possibility that water,[67−69] in which the protein is immersed, may be the cause (at least in part) of the observed differences in the symmetry of the correlation matrices. The considerations made above on the simulation temperature rule out an effect of bulk water, whose dynamics is primarily controlled by this parameter. Furthermore, we can exclude the role of the first hydration layer in the observed effects. In fact, although it is true that the instantaneous water configuration has a structural role, the Trp-cage folding path is independent

DOI: 10.1021/acs.jpcb.6b09216
J. Phys. Chem. B 2016, 120, 11428−11435

of the structure and dynamics of nearby water molecules,[38] as has been previously shown in a different system.[70] Also, considering the symmetries of the ensembles that best describe the folded and unfolded dynamics, it is conceivable that the main difference between these two conditions is that in the unfolded state the protein is expanded above a threshold so as to allow the isomerization of groups, whereas in the folded state the protein is tightly packed and all atom motions are strictly related to those of the neighbors. On the basis of these data, and extrapolating well above Trp-cage, one might suggest that phase transition between the unfolded and folded states of a protein occurs when the protein dynamics shifts from the weakly correlated regime represented by the PP ensembles (where each degree of freedom interacts mostly with itself) to the strongly correlated regime represented by the RBM ensembles (where each degree of freedom interacts with a number of other degrees of freedom). In more general terms, it can be stated that the NNS distributions obtained under the folded conditions are of the Wigner−Dyson type, whereas those obtained under the unfolded conditions are of the semi-Poisson distribution type.

CONCLUSIONS

Understanding the differences between the folded and unfolded dynamics of proteins is extremely important not only for theoretical but also for practical reasons. The growing number of pathological conditions that have their root in inappropriate protein folding is sufficient to be convinced of this, and this is related to the large amount of data and theoretical analysis that is available in the literature concerning the protein-folding phenomenon.[4−6] Moreover, a number of recent works have suggested that liquid protein phases are crucial in the emerging concept of membraneless organelles.[71−73]

Using Trp-cage as a model system, this work suggests an interesting connection between the folding process and phase transitions. It has been demonstrated that the folded and unfolded dynamics differ in an abstract type of symmetry, namely, the one that involves a displacement correlation matrix. For the unfolded dynamics, this matrix is best described by the PP ensemble, whereas the RBM ensemble is the more appropriate class to describe the folded dynamics of a protein. More generally, a switch from the semi-Poisson to the Wigner−Dyson type of distribution can be associated with the transition from an unfolded to a folded type of dynamics. It is suggested that the phase-transition picture of the folding/unfolding process is much more than a metaphor because the folded and unfolded states are related to different symmetries. It is worth noting that RMT-based distributions have already been shown to be involved in phase-transition phenomena.[74]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.6b09216.

MD; PCA; Trp-cage structure; RMSD analysis; RMSF analysis; NNS distribution (PDF)

Numerical calculations details (ZIP)

AUTHOR INFORMATION

Corresponding Author

*E-mail: luigileonardo.palese@uniba.it. Phone: +39 080 5448524.

Notes

The author declares no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by a grant from the University of Bari (CUP H93G13000170005).

REFERENCES

(1) Palese, L. L. Protein Dynamics: Complex by Itself. *Complexity* **2013**, *18*, 48−56.
(2) Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **1973**, *181*, 223−230.
(3) Sosnick, T. R.; Barrick, D. The Folding of Single Domain Proteins–Have We Reached a Consensus? *Curr. Opin. Struct. Biol.* **2011**, *21*, 12−24.
(4) Knowles, T. P.; Vendruscolo, M.; Dobson, C. M. The Amyloid State and its Association with Protein Misfolding Diseases. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 384−396.
(5) Luheshi, L. M.; Dobson, C. M. Bridging the Gap: From Protein Misfolding to Protein Misfolding Diseases. *FEBS Lett.* **2009**, *583*, 2581−2586.
(6) Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006**, *75*, 333−366.
(7) Bossis, F.; Palese, L. L. Molecular Dynamics in Cytochrome c Oxidase Mössbauer Spectra Deconvolution. *Biochem. Biophys. Res. Commun.* **2011**, *404*, 438−442.
(8) Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **2012**, *41*, 429−452.
(9) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To Milliseconds and Beyond: Challenges in the Simulation of Protein Folding. *Curr. Opin. Struct. Biol.* **2013**, *23*, 58−65.
(10) Sborgi, L.; Verma, A.; Piana, S.; Lindorff-Larsen, K.; Cerminara, M.; Santiveri, C. M.; Shaw, D. E.; de Alba, E.; Muñoz, V. Interaction Networks in Protein Folding via Atomic-Resolution Experiments and Long-Time-Scale Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2015**, *137*, 6506−6516.
(11) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the Accuracy of Physical Models Used in Protein-Folding Simulations: Quantitative Evidence from Long Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98−105.
(12) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517−520.
(13) Wu, H. Studies on Denaturation of Proteins. XIII. A Theory of Denaturation. *Chin. J. Physiol.* **1931**, *5*, 321−344.
(14) Mirsky, A. E.; Pauling, L. On the Structure of Native, Denatured, and Coagulated Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1936**, *22*, 439−447.
(15) Shakhnovich, E. Protein Folding Thermodynamics and Dynamics: Where Physics, Chemistry, and Biology Meet. *Chem. Rev.* **2006**, *106*, 1559−1588.
(16) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis. *Proteins* **1995**, *21*, 167−195.
(17) Rose, G. D.; Fleming, P. J.; Banavar, J. R.; Maritan, A. A Backbone-Based Theory of Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16623−16633.
(18) Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Rokhsar, D. S. Pathways for Protein Folding: Is a New View Needed? *Curr. Opin. Struct. Biol.* **1998**, *8*, 68−79.
(19) Matthews, C. R. Pathways of Protein Folding. *Annu. Rev. Biochem.* **1993**, *62*, 653−683.
(20) Onuchic, J. N.; Socci, N. D.; Luthey-Schulten, Z.; Wolynes, P. G. Protein Folding Funnels: The Nature of the Transition State Ensemble. *Folding Des.* **1996**, *1*, 441−450.
(21) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. Principal Component Analysis for Protein Folding Dynamics. *J. Mol. Biol.* **2009**, *385*, 312−329.

DOI: 10.1021/acs.jpcb.6b09216
J. Phys. Chem. B 2016, 120, 11428−11435

(22) Shakhnovich, E. I.; Finkelstein, A. V. Theory of Cooperative Transitions in Protein Molecules. I. Why Denaturation of Globular Protein Is a First-Order Phase Transition. *Biopolymers* **1989**, *28*, 1667−1680.
(23) Rollins, G. C.; Dill, K. A. General Mechanism of Two-State Protein Folding Kinetics. *J. Am. Chem. Soc.* **2014**, *136*, 11420−11427.
(24) Palese, L. L. Correlation Analysis of Trp-Cage Dynamics in Folded and Unfolded States. *J. Phys. Chem. B* **2015**, *119*, 15568−15573.
(25) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-Residue Protein. *Nat. Struct. Biol.* **2002**, *9*, 425−430.
(26) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(27) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(28) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(29) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(31) van der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13*, 22−30.
(32) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90−95.
(33) Pérez, F.; Granger, B. E. IPython: A System for Interactive Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 21−29.
(34) Oliphant, T. E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 10−20.
(35) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 Micros. *J. Am. Chem. Soc.* **2002**, *124*, 12952−12953.
(36) Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Andersen, N. H. The Trp-Cage: Optimizing the Stability of a Globular Miniprotein. *Protein Eng., Des. Sel.* **2008**, *21*, 171−185.
(37) Snow, C. D.; Zagrovic, B.; Pande, V. S. The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2002**, *124*, 14548−14549.
(38) Juraszek, J.; Bolhuis, P. G. Sampling the Multiple Folding Mechanism of Trp-Cage in Explicit Solvent. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15859−15864.
(39) Juraszek, J.; Saladino, G.; van Erp, T. S.; Gervasio, F. L. Efficient Numerical Reconstruction of Protein Folding Kinetics with Partial Path Sampling and Pathlike Variables. *Phys. Rev. Lett.* **2013**, *110*, No. 108106.
(40) Hałabis, A.; Żmudzińska, W.; Liwo, A.; Ołdziej, S. Conformational Dynamics of the Trp-Cage Miniprotein at its Folding Temperature. *J. Phys. Chem. B* **2012**, *116*, 6898−6907.
(41) Byrne, A.; Williams, D. V.; Barua, B.; Hagen, S. J.; Kier, B. L.; Andersen, N. H. Folding Dynamics and Pathways of the Trp-Cage Miniproteins. *Biochemistry* **2014**, *53*, 6011−6021.
(42) Cote, Y.; Maisuradze, G. G.; Delarue, P.; Scheraga, H. A.; Senet, P. New Insights into Protein (Un)Folding Dynamics. *J. Phys. Chem. Lett.* **2015**, *6*, 1082−1086.
(43) Bille, A.; Linse, B.; Mohanty, S.; Irbäck, A. Equilibrium Simulation of Trp-Cage in the Presence of Protein Crowders. *J. Chem. Phys.* **2015**, *143*, No. 175102.
(44) Levine, Z. A.; Fischer, S. A.; Shea, J.-E.; Pfaendtner, J. Trp-Cage Folding on Organic Surfaces. *J. Phys. Chem. B* **2015**, *119*, 10417−10425.
(45) Bossis, F.; Palese, L. L. Amyloid Beta (1−42) in Aqueous Environments: Effects of Ionic Strength and E22Q (Dutch) Mutation. *Biochim. Biophys. Acta* **2013**, *1834*, 2486−2493.
(46) Palese, L. L. Random Matrix Theory in Molecular Dynamics Analysis. *Biophys. Chem.* **2015**, *196*, 1−9.
(47) Wigner, E. P. Random Matrices in Physics. *SIAM Rev.* **1967**, *9*, 1−23.
(48) Dyson, F. J. Statistical Theory of the Energy Levels of Complex Systems. I. *J. Math. Phys.* **1962**, *3*, 140−156.
(49) Wishart, J. The Generalised Product Moment Distribution in Samples From a Normal Multivariate Population. *Biometrika* **1928**, *20A*, 32−52.
(50) Bohigas, O.; Giannoni, M. J.; Schmit, C. Characterization of Chaotic Quantum Spectra and Universality of Level Fluctuation Laws. *Phys. Rev. Lett.* **1984**, *52*, 1.
(51) Guhr, T.; Müller-Groeling, A.; Weidenmüller, H. A. Random-Matrix Theories in Quantum Physics: Common Concepts. *Phys. Rep.* **1998**, *299*, 189−425.
(52) Edelman, A.; Wang, Y. Random Matrix Theory and Its Innovative Applications. In *Advances in Applied Mathematics, Modeling, and Computational Science*; Melnik, R., Kotsireas, I. S., Eds.; Fields Institute Communications 66; Springer: New York, NY, 2013; pp 91−116.
(53) Müller, M.; Jiménez, Y. L.; Rummel, C.; Baier, G.; Galka, A.; Stephani, U.; Muhle, H. Localized Short-Range Correlations in the Spectrum of the Equal-Time Correlation Matrix. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2006**, *74*, No. 041119.
(54) Potestio, R.; Caccioli, F.; Vivo, P. Random Matrix Approach to Collective Behavior and Bulk Universality in Protein Dynamics. *Phys. Rev. Lett.* **2009**, *103*, No. 268101.
(55) Edelman, A.; Rao, N. R. Random Matrix Theory. *Acta Numer.* **2005**, *14*, 233−297.
(56) Gruebele, M. Bose Statistics Triangle Rule Model for Intramolecular Vibrational Energy Redistribution. *J. Phys. Chem.* **1996**, *100*, 12183−12192.
(57) Brody, T. A. A Statistical Measure for the Repulsion of Energy Levels. *Lett. Nuovo Cimento* **1973**, *7*, 482−484.
(58) Bogomolny, E. B.; Gerland, U.; Schmit, C. Models of Intermediate Spectral Statistics. *Phys. Rev. E* **1999**, *59*, R1315.
(59) Shigehara, T.; Yoshinaga, N.; Cheon, T.; Mizusaki, T. Level-Spacing Distribution of a Singular Billiard. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1993**, *47*, R3822.
(60) Schierenberg, S.; Bruckmann, F.; Wettig, T. Wigner Surmise for Mixed Symmetry Classes in Random Matrix Theory. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2012**, *85*, No. 061130.
(61) Lenz, G.; Haake, F. Reliability of Small Matrices for Large Spectra with Nonuniversal Fluctuations. *Phys. Rev. Lett.* **1991**, *67*, 1.
(62) Wintgen, D.; Friedrich, H. Classical and Quantum-Mechanical Transition Between Regularity and Irregularity in a Hamiltonian System. *Phys. Rev. A* **1987**, *35*, 1464.
(63) Shklovskii, B. I.; Shapiro, B.; Sears, B. R.; Lambrianides, P.; Shore, H. B. Statistics of Spectra of Disordered Systems Near the Metal-Insulator Transition. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 11487.
(64) Sakhr, J.; Nieminen, J. M. Spacing Distributions for Point Processes on a Regular Fractal. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2006**, *73*, No. 036201.
(65) Janssen, M.; Pracz, K. Correlated Random Band Matrices: Localization-Delocalization Transitions. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2000**, *61*, 6278.
(66) Sethna, J. P. *Statistical Mechanics: Entropy, Order Parameters and Complexity*; Oxford University Press: Oxford, U.K., 2006.
(67) Levy, Y.; Onuchic, J. N. Water Mediation in Protein Folding and Molecular Recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389−415.
(68) Ball, P. Water as an Active Constituent in Cell Biology. *Chem. Rev.* **2008**, *108*, 74−108.
(69) Bellissent-Funel, M.-C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E.

DOI: 10.1021/acs.jpcb.6b09216
J. Phys. Chem. B 2016, 120, 11428−11435

Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016**, *116*, 7673−7697.
(70) Rhee, Y. M.; Sorin, E. J.; Jayachandran, G.; Lindahl, E.; Pande, V. S. Simulations of the Role of Water in the Protein-Folding Mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6456−6461.
(71) Shorter, J. Membraneless Organelles: Phasing In and Out. *Nat. Chem.* **2016**, *8*, 528−530.
(72) Brangwynne, C. P.; Tompa, P.; Pappu, R. V. Polymer Physics of Intracellular Phase Transitions. *Nat. Phys.* **2015**, *11*, 899−904.
(73) Nott, T. J.; Craggs, T. D.; Baldwin, A. J. Membraneless Organelles Can Melt Nucleic Acid Duplexes and Act as Biomolecular Filters. *Nat. Chem.* **2016**, *8*, 569−575.
(74) Majumdar, S. N.; Schehr, G. Top Eigenvalue of a Random Matrix: Large Deviations and Third Order Phase Transition. *J. Stat. Mech.* **2014**, *2014*, No. P01012.

DOI: 10.1021/acs.jpcb.6b09216
J. Phys. Chem. B 2016, 120, 11428−11435