PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

A hidden aggregation-prone structure in the heart of hypoxia inducible factor prolyl hydroxylase

Hamid Hadi-Alijanvand,[1,2] Elizabeth A. Proctor,[3] Feng Ding,[4,5] Nikolay V. Dokholyan,[4,6,7] and Ali A. Moosavi-Movahedi[2,8]*

[1] Department of Biological Sciences, Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan, Iran

[2] Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran

[3] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

[4] Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, North Carolina 27599

[5] Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29634

[6] Curriculum in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, North Carolina 27599

[7] Program in Molecular and Cellular Biophysics, University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, North Carolina 27599

[8] Center of Excellence in Biothermodynamics, Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran

ABSTRACT

Prolyl hydroxylase domain-containing protein 2 (PHD2), as one of the most important regulators of angiogenesis and metastasis of cancer cells, is a promising target for cancer therapy drug design. Progressive studies imply that abnormality in PHD2 function may be due to misfolding. Therefore, study of the PHD2 unfolding pathway paves the way for a better understanding of the influence of PHD2 mutations and cancer cell metabolites on the protein folding pathway. We study the unfolding of the PHD2 catalytic domain using differential scanning calorimetry (DSC), fluorescence spectroscopy, and discrete molecular dynamics simulations (DMD). Using computational and experimental techniques, we find that PHD2 undergoes four transitions along the thermal unfolding pathway. To illustrate PHD2 unfolding events in atomic detail, we utilize DMD simulations. Analysis of computational results indicates an intermediate species in the PHD2 unfolding pathway that may enhance aggregation propensity, explaining mutation-independent PHD2 malfunction.

Proteins 2016; 84:611–623.

Key words: HIF-1; aggregation prone intermediate state; protein misfolding; angiogenesis; discrete molecular dynamics; calorimetry; potential of mean force.

INTRODUCTION

Studies of the protein folding process provide insight that can be applied in the development of therapeutics for protein folding diseases.[1] Various cancer cell metabolites may have undesirable effects on critical proteins in the cell signal transduction pathway,[2] not only through kinetic change in function, but also by alteration of the cellular environment (e.g. pH change or accumulation of cell metabolites), which can affect the protein folding pathway. The study and elucidation of the protein folding process, especially the protein folding pathway, is a challenge in the fields of biophysics and molecular biology. The precise details of the folding and unfolding process are a focus point for researchers

Additional Supporting Information may be found in the online version of this article.

Abbreviations: ASA, accessible surface area; ΔCpu, changes of heat capacity in unfolding process; Cv, specific heat; 2D-PMF, two-dimensional potential of mean force; 2D-PMF-g, 2D-PMF for global structure of protein; 2D-PMF-w, 2D-PMF for tryptophan residues and their neighbors; DMD, discrete molecular dynamic; DSC, differential scanning calorimetry; HIF, hypoxia inducible factor; ΔH_u, changes of enthalpy in unfolding process;; PHD2, prolyl hydroxylase domain-containing protein 2; Qsh, fraction of long range native contact; Qsh-g, fraction of long range native contact for global structure of protein; Qsh-w, fraction of long range native contact for tryptophan residues and their neighbors; RMSF, root mean square fluctuation; RSA, relative solvent accessible surface area; STS, spatial TANGO score; T_m, melting temperature; WHAM, weighted histogram analysis method.

Grant sponsor: Institute for Advanced studies in Basic Sciences (IASBS), University of Tehran, Iran National Science Foundation (INSF), Center of Excellence in Biothermodynamics (CEBiotherm), Iran National Elites Foundation (INEF), and UNESCO Chair on Interdisciplinary Research in Diabetes.

*Correspondence to: Ali A. Moosavi-Movahedi, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran. E-mail: moosavi@ut.ac.ir

Received 28 October 2015; Revised 5 February 2016; Accepted 8 February 2016
Published online 12 February 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.25011

specializing in both computational methods and experimental tools.

Protein unfolding studies have the ability to shed light on the answers to fundamental questions such as the nature of the folding pathway and the characteristics of folding intermediates. The qualitative and quantitative characterizations of both stable and meta-stable species on the folding pathway are the main aim of protein folding and unfolding studies.[3]

Prolyl hydroxylase domain-containing protein 2 (PHD2) is a mid-size protein with a jellyroll fold, a member of the non-heme dioxygenases family. PHD2 is characterized by a double-stranded beta helix containing eight strands, with three additional alpha helices supporting its beta helix structure. PHD2 acts as the main regulator of angiogenesis in cancer cells, inducing hypoxia inducible factor-1 (HIF-1) degradation and inhibiting angiogenesis in normoxic conditions.[4,5] This pathway is a key example of a system whose proteins folding can be targeted as a therapeutic strategy.[6–10] Several studies report sense mutations in the PHD2 gene[11] and explore the kinetics of the protein,[12] but no experimental information about the folding pathway of PHD2 is currently available. Mutations to the PHD2 gene have been linked to endometrial cancer,[13] and in some cases wild type PHD2 has also been found to malfunction/over expression.[14] In these sporadic cases, PHD2 somehow fails to induce HIF-1 degradation, and tumor types in normoxic conditions in which no PHD2 mutation is reported still undergo activation of the angiogenesis pathway.[15,16] In some breast cancer xenografts, breast invasive carcinoma, hepatocellular carcinoma, and ovarian serous cyst adenocarcinoma, PHD2 amplification is reported in DNA/mRNA copy.[17,18] The mechanism of PHD2 inhibition or malfunction is still unknown in these conditions, but could be caused by environmental instabilities or protein misfolding. Therefore, knowledge of PHD2 folding intermediates is important because these intermediates are potential players in disease development.

An ideal two-state protein unfolding model assumes that the protein has two stable populations during the unfolding process at the extreme points of the protein unfolding reaction coordinate: a folded population and an unfolded population. This theoretical model assumes that no stable intermediate states are present in the transition from the folded to the unfolded state.[19] Although many proteins have been shown to undergo a two-state unfolding process in experiments and *in silico*,[20] proteins of various topological classes and sizes have been found to undergo a multi-stage protein unfolding process.[21–24] This latter type of protein exhibits at least one stable intermediate between the folded and unfolded states. Various methods exist to detect and describe intermediate states. Experimental methods[25] such as HX-NMR,[26] relaxation dispersion NMR,[27] single molecular FRET,[28] and differential scanning calorimetry (DSC)[29] may be used to study the kinetics of protein folding in detail.

Differential scanning calorimetry is a powerful experimental tool to find and characterize the transition temperature and stable unfolding intermediate states during protein thermal denaturation.[30] DSC results have the form of molar heat capacity as a function of temperature. A peak in the DSC profile indicates the amount of resistance to an increase in temperature of the system, and occurs when the protein undergoes a major thermodynamic transition. In this study, we utilize DSC to provide experimental verification of PHD2 protein unfolding intermediates.

Computational studies of protein unfolding provide atomic details of this complicated process that are inaccessible to experiments. Energetic surface analysis,[31,32] graphical analysis of the folding pathway,[33,34] network signal propagation analysis,[35] and Markov state calculations for large trajectories[36,37] are all common approaches and tools for identifying protein folding intermediates from molecular dynamics (MD) simulations. MD simulation is capable of describing protein dynamics at the atomic level, but the sampling efficiency and time scaling with system size of all-atom MD is in many cases prohibitive for studying the folding of biologically-relevant proteins. Discrete molecular dynamics (DMD) is a form of MD that uses discrete energetic potentials in place of continuous ones to describe interactions between atoms and study protein dynamics.[38–40] DMD has been used to study various protein folding, unfolding,[41–43] and aggregation systems in detail,[44,45] as well as to describe the effects of post-translational modifications on protein dynamics and stability.[46,47] The low computational cost of DMD, while providing high sampling efficiency and a robust all-atom force field, allows us to perform the extensive simulations of PHD2 protein unfolding necessary for identification of unfolding intermediate states.

PHD2 is a main regulator of cell fate in response to oxygen availability, and it may play a critical role in cancer therapies. In this work, we elucidate details of the PHD2 unfolding process and describe intermediates along the PHD2 unfolding pathway using an iterative hybridized computational/experimental workflow. We utilize DSC and fluorescence spectroscopy to identify protein unfolding events, and perform DMD simulations to provide detailed information about these intermediate states. We then test and confirm our computational findings with further analysis of our experimental results. Using this synergistic approach, we provide an illustrative profile of the unfolding pathway of PHD2 using simultaneous analysis of experimental and simulation data, and suggest probable toxic intermediate of PHD2 unfolding.

MATERIALS AND METHODS

Experimental procedures

We transfer the pET 28a+ plasmid carrying the catalytic domain of human wild type prolyl hydroxylase domain-containing protein 2 (PHD2) gene[48,49] to the *Escherichia coli* BL21 DE3 pLysS strain by the CaCl2 transfection method.[50] The transfected bacteria grow to an optical density of 0.8 in a 2TY medium containing 0.1% glucose and 50 μg/mL Kanamycin. The expression of His-tagged PHD2 is induced by addition of IPTG (Promega) to the culture medium to a final concentration 0.5 m*M*. After an additional 3 h of culture at 37°C with 250 rpm shaking, we extract the purified 27 kD protein with the aid of His Buster Amocol resin (based on the manufacturer's protocol). To detach the bound protein from the column, an elution buffer containing 1*M* imidazole according to the Hewitson protocol is utilized.[48] We verify protein purity by SDS-PAGE with coomassie blue staining, and test its activity based on a fluorescence derivative method.[51] Protein purity is >95%. Before experiments, we filter the samples through nitrocellulose article with 0.4 μm mesh.

PHD2 solution contains 40 m*M* Tris-HCl and 0.5*M* NaCl at pH 7.9 in experiments. The concentration of purified protein is determined using the micro-Bradford assay.

We utilize a Nano-DSC machine (Nano-DSC II model 6100, Calorimetry Sciences Corporation) to study the thermal unfolding of PHD2. The cell volume of our DSC machine is approximately 0.9 mL. We preheat and equilibrate the machine by the following protocol: before the start of the scanning process, we preheat the DSC machine to room temperature. DSC scans of deionized water in sample and reference cells are conducted for several times at a heating rate of 1°C per minute, allowing the protein buffer to run through the DSC machine several times (at least three times). The reversibility test indicates that the thermogram is reversible for temperatures <320 K. The protein concentration is 0.8 mg/mL in DSC assay. To find substates of PHD2 unfolding, by using OriginPro software, we perform deconvolution of excess heat capacity of PHD2 unfolding that measured by DSC.

Cary Eclipse spectrofluorimeter equipped with Varian water thermal bath and thermal controller (Cary Eclipse, Varian Co., Australia) is utilized to study protein thermal unfolding and measure intrinsic tryptophan fluorescence. We perform fluorescence studies using a fluorescence cuvette with 1 cm path length. The excitation wavelength is set to 280 nm, and the maximum emission is found at 341 nm. The heating rate is the same used for DSC. A concentration of 0.2 mg/mL PHD2 is used in fluorescence spectroscopy studies. Protein thermal aggregation is measured by recording the absorbance of 0.8 mg/mL protein at 360 nm by using a Cary spectrophotometer with a quartz cuvette with 1 cm path length. The heating rate in the aggregation study is 1°C per minute. We repeat the spectrometric assay a minimum of twice.

Computational procedures

We utilize the all-atom discrete molecular dynamics (DMD) simulation method to study thermal unfolding *in silico*. Briefly, DMD employs discretized energetic potentials in place of continuous potentials. The velocity of each atom is therefore kept constant until a collision occurs, at which time the velocities of the two atoms change instantaneously according to the laws of conservation of energy, momentum, and angular momentum. The Medusa force field is utilized for the implementation of non-bonded potentials.[52] VDW interactions are treated via the Lennard-Jones potential, solvation effects are modeled using the Lazaridis-Karplus implicit solvent model,[53] and distance and angular dependent hydrogen bond interactions are implemented using the reaction algorithm.[54]

To obtain sufficient structural sampling of PHD2 during thermal unfolding, we perform constant temperature DMD simulations for 10^6 time steps (∼50 ns). Temperatures in DMD simulations are maintained by an Anderson's thermostat.[55] We discard ligands from the crystal structure to obtain the initial conformation of PHD2 in its apo-form (PDB ID: 3OUJ, chain A). After reconstruction of omitted residues in 3OUJ by MODELLER V. 9.9,[56] we perform simulations of the finalized structure in DMD. The heat exchange coefficient is set to 10^{-5} for all simulations to avoid temperature-induced jump in system potential energy. We conduct simulations at temperatures: 0.550 (∼276 K), 0.575 (∼289 K), 0.600 (∼302 K), 0.625 (∼314 K), 0.650 (∼327 K), 0.675 (∼340 K), 0.700 (∼352 K), 0.725 (∼365 K), 0.750 (∼377 K), 0.775 (∼390 K), 0.800 (∼402 K), and 0.825 (∼415 K) kcal (mol k_B)$^{-1}$ (simulation temperature units are equivalent to Kelvin multiplied by the Boltzmann factor, k_B[38]) resulting in a total of 600 ns of simulation time. The DMD simulations are repeated with randomized initial velocities to alleviate the bias toward initial state for each ensemble production run.

We utilize the weighted histogram analysis method (WHAM, Grossfield A. Weighted Histogram Analysis Method Version 2.0.6.) for constructing the two-dimensional potential of mean force (2D PMF).[57] The instant potential energy and long range contact order are used as reaction coordinates in constructing the 2D-PMFs. The constant volume heat capacity (C_v) curve of the unfolding process is defined using WHAM method.

Topological description of structures (especially transition states) is possible using a variety of criteria. Deviation from the native state is a valuable descriptor when describing structures along the unfolding pathway, which

can be measured by computing the fraction of native contacts. If residues x and y form a contact in the native structure and are also present in the structure under consideration, we count it as a native contact. We may also take into consideration the separation in sequence, r, between x and y. By considering long r, we detect global topology and long-range native contacts. The parameter Qsh measures the fraction of native C_α contacts with sequence separation of at least seven residues ($r>7$) and spatial separation <10 Å,[3] calculated as:

$$\mathrm{Qsh}=\frac{\sum_{i=1}^{i=N}\sum_{j=1}^{j=N}\left(1/1+e^{10(d_{ij}-(d_{ij}^{\mathrm{native}}+1)}\right)}{N^{native}} \quad (1)$$

where N is the number of residues, d_{ij} denotes the distance between the contacting residues in the query structure, d^{native} denotes the distance between these same residues in the native structure, and N^{native} is the total number of contacts satisfying the Qsh criteria in the native structure (Qsh inferred from "How fast-folding proteins fold" article[3]).

To find the transition temperatures, ΔCp_u, and ΔH_u based on Qsh and fluorescence spectroscopy data, we find the best fit of the observed experimental or computational data with a sigmoid function:[58]

$$\begin{gathered}\mathrm{M1}=\left((F(\mathrm{pre})-F(\mathrm{post})\right)\\ \mathrm{M2}=\left(1+\exp\left((1/T-1/T\mathrm{m})\right.\right.\\ \left.\left.\Delta H\mathrm{u}/R+\Delta\mathrm{Cpu}/R\left(\ln\left(T/T\mathrm{m}\right)+T\mathrm{m}/T-1\right)\right)\right.\\ F(T)=F(\mathrm{post})+{}^{M1}\!/_{M2}\end{gathered} \quad (2)$$

where T_m is the transition temperature of the fitted curve, F(pre) indicates the pre-transition, F(post) the post-transition value for fluorescence and Qsh data. ΔH_u and ΔCp_u are changes of enthalpy and heat capacity upon unfolding. Here, sections of data with large RMSD to the fitted function are discounted and excluded from the fit.

We compute hydrophobic solvation energy using the NAMD-Generalized Born implicit solvent (GBIS) feature.[59] The GBIS model employs solvent accessible surface area (SASA) of the protein to estimate hydrophobic energy. NAMD utilizes the LCPO procedure to compute SASA.[60] We determine the electrostatic energy of a structure in two independent single point MD runs. The GBIS option is activated in the first run, then inactivated for the last run. The difference between the system energies calculated for these runs yields the hydrophobic energy.

We also compute the relative solvent accessible surface area (RSA) of STRIDE-defined strand structures.[61] The RSA of hydrophobic sheets is also computed for ensemble structures.

TANGO is an empirical function to predict aggregation prone regions of a protein sequence at desired temperature.[62] Using the following equation, we translate the one-dimensional TANGO score to a three-dimensional metric, denoted as the spatial TANGO score (STS):

$$\mathrm{STS_s}=\frac{\sum_{i=1}^{i=N}\left((\mathrm{TANGO}_i)\times\frac{\mathrm{ASA}_i}{ASA_{\mathrm{E}}}\right)}{\mathrm{TANGO_s}} \quad (3)$$

Here the "s" subscript denotes the segment score, "i" indicates the amino acid of a segment, and "E" denotes the extended structure. "TANGO" is the TANGO score assigned to that segment's amino acids, or the sum of segment scores.

VMD 1.9[63] is used to perform accessible surface area (ASA) analysis on DMD-based trajectory of PHD2 unfolding and, the electrostatic potential maps of unfolding intermediates are defined using APBS 0.5.0.[64]

RESULTS

To study the unfolding steps of PHD2 protein, we perform experimental and computational studies. To ease following the results, the experimental findings are presented at first then computational results are explained thereafter.

Multiple unfolding transitions detected by DSC

We determine thermal profiles of PHD2 unfolding using nanodifferential scanning calorimetry (nano-DSC). Four transitions are observed in the unfolding pathway of PHD2 from the nano-DSC profile (Fig. 1). PHD2 undergoes transitions upon thermal unfolding at 306 K, 320 K, 335 K, and 345 K. Therefore, PHD2 is not a simple two-state folder, but instead features four transition points, equating to five states: a native state, an unfolded state, and three stable intermediates along the unfolding pathway. To capture additional details of these transitions, we deconvolute excess heat capacity peaks in the DSC profile of PHD2 unfolding to obtain four transitions (I1–I4), each with specific thermodynamic properties (Fig. 1 inset, Table I). In deconvolution, we assume sequential unfolding transitions, counting each as a simple two-state transition. While DSC profile shows transitions in stability of PHD2 along unfolding pathway, we are also curious to study the possible changes in structure of PHD2 during thermal unfolding.

Unfolding transitions monitored by local fluorescence of TRP

We use fluorescence spectroscopy to study changes in tertiary structure of PHD2 upon denaturation. The fluorescence emission of tryptophan residues in the protein

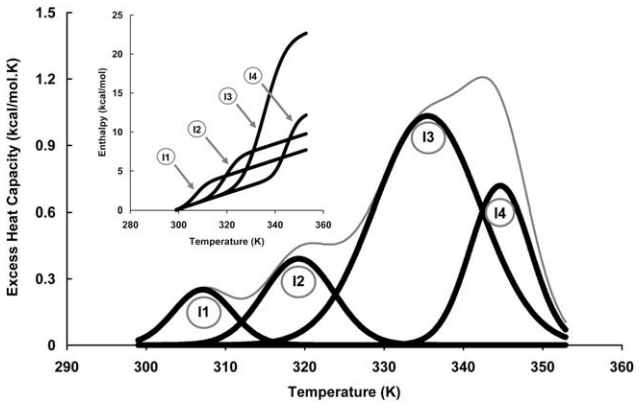

Figure 1

Multistate PHD2 unfolding pathway defined using DSC. We utilize DSC to measure the molar heat capacity and identify protein unfolding intermediate states and transitions. Four distinct transitions (I1-4) are detected in DSC profile of PHD2. The deconvoluted peaks are shown with a bold black line, and the sum of the deconvolution peaks is shown as a gray line. Inset shows the enthalpy change of each transition represent as a function of temperature.

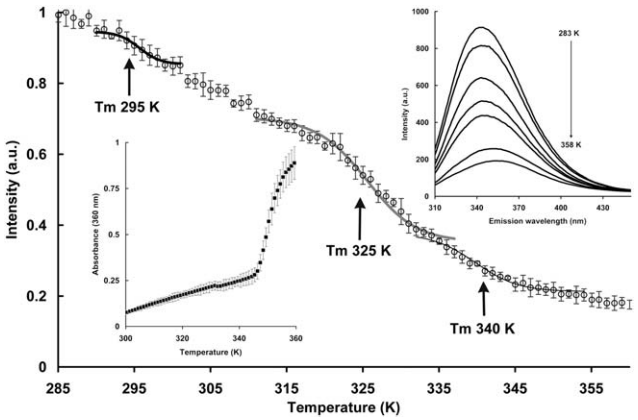

Figure 2

Multistate PHD2 unfolding and aggregation as a function of temperature defined by spectroscopic approaches. Intrinsic fluorescence of tryptophan residues as a function of temperature is depicted in main figure. The T_m values of three transitions are denoted with arrows in the main panel. Tryptophan emission decreases in intensity with increase in temperature (upper right inset). The peaks show measured tryptophan emission at 283 K, 293 K, 305 K, 319 K, 333 K, 344 K, and 358 K. Lower Inset shows thermal aggregation of PHD2 measured by UV absorbance at 360 nm. Error bars are the standard error of the mean for duplicated experiments.

decreases incrementally with tryptophan accessibility. This method is sensitive to changes in the local environment of tryptophan. PHD2 is unfolded upon heating, and therefore its tryptophan residues feature increased accessibility, decreasing their fluorescence emission at high temperature (Fig. 2, right inset). The continuous regions of the fluorescence thermal unfolding profile of PHD2 are fitted to sigmoid functions [Methods, Eq. (2)], and three transitions detected in current fluorescence spectroscopy output (Fig. 2). The innate multi states of PHD2 unfolding are suggested by fluorescence spectroscopy, also. Thermal aggregation of PHD2 unfolding process is assessed by measuring absorbance at 360 nm (Fig. 2, left inset). This measurement notes a jump in aggregation at 345 K where the signal of UV absorbance at 280 nm, as an indication of protein unfolding, also suddenly increases.

Although these results give clues to large structural rearrangements, fluorescence spectroscopy and DSC are unable to describe atomic level structural changes in the protein during the unfolding process. Therefore computational studies seem inevitable to seek unfolding process in atomic level.

Table I

Comparison of PHD2 Thermal Unfolding Thermodynamic Data Obtained Using Different Approaches

	Transition	T_m (K)	ΔH (kcal/mol)	ΔCp (kcal/mol K)
DSC[a] (excess parameters)	1	306	2.3	0.3
	2	320	4.4	0.4
	3#	335	17.3	1.0
	4	345	6.8	0.7
Qsh-g[b]	1	284	148.1	−4.0
	2	294	340.0	−43
	3#	320	74.6	4.5
	4	345	56.8	0.9
Flouresence	1	295	125.3	0.9
	2	325	64.3	0.8
	3	340	80.3	1.0
Qsh-w[c]	1	285	103.7	0.0
	2	321	98.8	6.6
	3	346	61.0	0.4

[a]DSC data are the result of general model deconvolution approaches.
[b]Qsh-g is the states determined by analyzing global-Qsh as a function of temperature
[c]Qsh-w is from Qsh analysis of tryptophan and its neighbors. Non-DSC parameters derived from fitting of data to a sigmoidal function. The main transition is indicated by "#" superscript.

Investigation of unfolding intermediates *in silico*

To investigate the experimental observations, we perform molecular dynamics simulations of protein thermal unfolding. To detect the unfolding states of PHD2 *in silico*, we use the discrete molecular dynamics (DMD) simulation engine[40] to generate thermal unfolding trajectories (Methods). Total simulation time for PHD2 thermal unfolding is approximately 0.6 μs. We calculate specific heat using the weighted histogram analysis method (WHAM), a common approach to obtain the melting transition temperature from MD data (Fig. 3). The sequence of computed changes in specific heat of PHD2 unfolding is similar to its experimental counterpart. These results show a multi states profile with two minor transitions precedes the main transition, a sharp

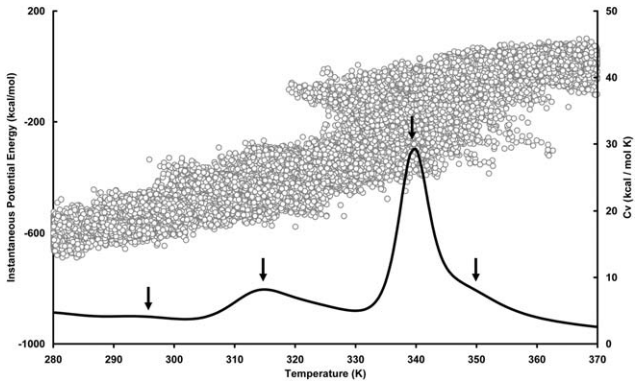

Figure 3

Multistate PHD2 thermal denaturation defined using computational approaches. We determine the changes in instantaneous potential energy of the system upon heating. Melting transitions are apparent at approximately 295 K, 315 K, 340 K (main transition) and 350 K based on the computed change in heat capacity with temperature (indicated by vertical arrows). The instantaneous potential energies of structures as calculated in DMD simulations are indicated as gray squares (left vertical axis), and the constant volume heat capacity is shown as a solid curve (right vertical axis).

main transition and a small transition following the main transition.

In order to characterize the unfolding states in more detail, we study the fraction of long-range native contacts (Qsh) [Methods, Eq. (1)] as a function of the unfolding progression. The fraction of native contacts is a common reaction coordinate to illustrate protein folding phenomena.[65] We calculate Qsh for the global protein structure (Qsh-g) by considering all amino acids as well as by considering just tryptophan residues and their neighbors in the native structure (Qsh-w) (Fig. 4). We observe that the fraction of long-range native contacts undergoes two clear transitions upon PHD2 heating, producing two stable populations during protein unfolding. Discrepancy in Figure 4 indicates that there are two additional small transitions preceding main transition in Qsh-g graph and one small transition in Qsh-w.

To illustrate the flanking intermediate states of each transition which observed in Figure 4, we compute the two-dimensional potential of mean force (2D-PMF) using the WHAM approach. We choose Qsh as the first coordinate, and instantaneous potential energy calculated from DMD simulations as the second coordinate. Here, we calculate two different types of 2D-PMF for the PHD2 unfolding pathway: "2D-PMF-g," using Qsh of the global protein structure (Fig. 5), and "2D-PMF-w," using Qsh of tryptophan and its neighbors (Fig. 6).

Characterization of unfolding intermediate's states requires elucidation of the intermediates' structural features. Here, we suppose the structure of PHD2 consists of three significant regions in regard to the focus of current study:

- The active site lumen (residues 256, 299, 301, 303, 310, 325, 327, 329, 343, 366, 376, 385, 387, 389), which surrounds the active site residues (ACS lumen)
- The docking site (residues 197, 207, 209, 251, 241, 311, 338, 342, 346, 353, 354, 356, 378, 380), the proposed region in which PHD2 interacts with its partners (Docking site)[66]

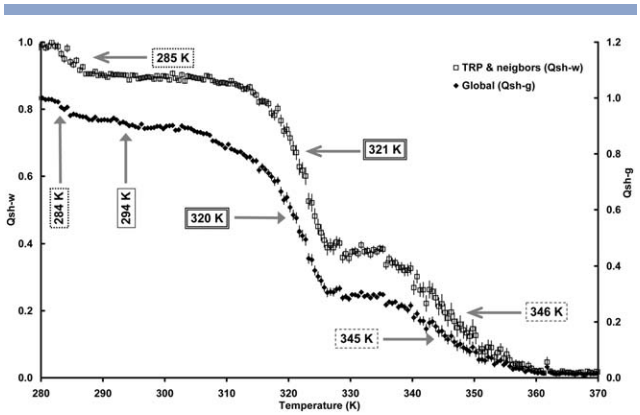

Figure 4

Major structural transitions determined using analysis of DMD's results. Qsh (long-range native contacts) of global protein structure (Qsh-g) and Qsh of tryptophan residues and their neighbors' changes (Qsh-w) are defined as a function of temperature to determine major structural transitions. The right vertical axis depicts Qsh-g changes and left vertical axis depicts Qsh-w. Gray arrows indicate the transitions temperatures. Error bars indicate the standard error of mean for the specific population of structures in the corresponding temperature.

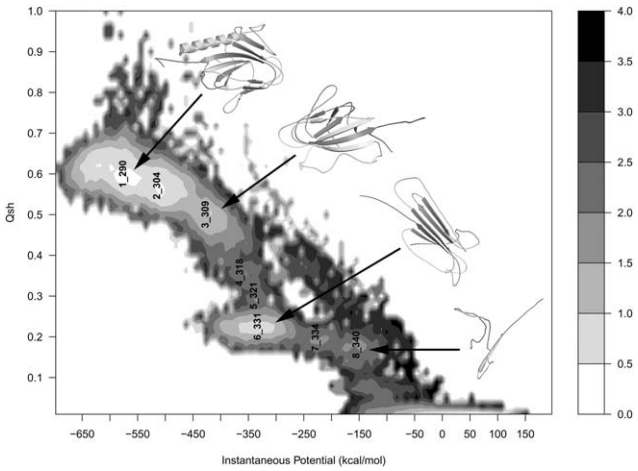

Figure 5

DMD derived PHD2 thermal unfolding intermediates characterized by considering deviation of all residue interactions in native structure. 2D-PMF-g is depicted as the result of WHAM analysis of Qsh-g versus the instantaneous potential of structures from DMD. Numbers in each well indicate intermediate number and corresponding temperature (Kelvin). Ribbon structure is the average structure of the population of each energetic well. The scale bar at right indicates the contour for potential of mean force energy in kcal/mol at 310 K.

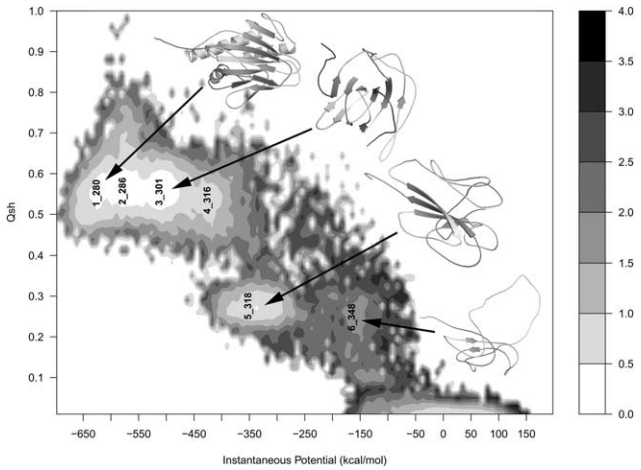

Figure 6

DMD-based PHD2 thermal unfolding intermediates characterized by deviation of tryptophan residue interactions in starting structure. 2D-PMF-w is represented as the result of WHAM analysis of Qsh-w versus the instantaneous potential of structures from DMD. Numbers in each well indicate intermediate number and corresponding temperature (Kelvin). Ribbon structure is the average structure of the population of each energetic well. The scale bar at right indicates the contour for potential of mean force energy in kcal/mol at 310 K.

- The home of lip helix (residues 241–2, 258, 260–2, 274, 277, 281, 293–7, 390–1,393, 396)

In PHD2 crystals (PDB ID: 3OUJ) that lack the substrate peptide (part of the HIF-1α protein), residues 400 to 416, which we name the lip, cover the region of the PHD2 protein that we name the lip home. In PHD2 crystal structures (PDB ID: 3HQR) that contain the substrate peptide, the lip leaves the lip home and is replaced by part of the substrate peptide. The Catalogue of Somatic Mutations in Cancer (COSMIC) reports mutations to residue 396 in some breast tumor cells (Mutation ID: 160659). This residue resides in the lip home, motivating study of lip home changes along the PHD2 unfolding pathway. To study the species that belong to intermediate's ensemble detected in "2D-PMF-g" the average accessible surface area (ASA) of mentioned critical regions of PHD2 is computed for each state's structural ensemble (Fig. 7).

The appearance of hydrophobic strand structures in an intermediate is a hallmark for high aggregation propensity of the structure.[67,68] We follow the fraction of ASA of strands and hydrophobic sheets, and compute the hydrophobic energy for protein structures along the unfolding pathway of PHD2 (Fig. 8). A jump in the value of the hydrophobic energy occurs in the PHD2 unfolding process before 340 K. Strand structures, notably hydrophobic strands, persist until 340 K.

TANGO detects three aggregation prone regions in the PHD2 sequence (Fig. 9): S1 = residues 138–145, S2 = 177–182, S3 = 198–205. TANGO assumes an extended structure for hotspots, and identifies these same regions as hotspots of PHD2 regardless of assumed temperature (300–350 K), although the assigned score does vary with temperature. The TANGO score of S1 is the most temperature-sensitive, while the score of S3 is the most temperature-resistant, indicating that S1 may play a role in aggregation at low temperature, while S3 acts over a wide range of temperature.

We defined the spatial TANGO score (STS) as a translation of TANGO into three dimensions [Methods, Eq. (3)]. The STS allows us to follow the aggregation propensity of structures along the unfolding pathway. We

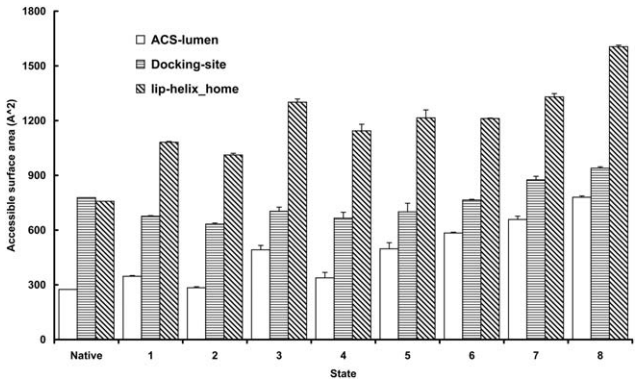

Figure 7

Different regions of PHD2 become accessible to solvent as unfolding progresses. Intermediate state number is indicated on the *x* axis, with native state information derived from the crystal structure. Error bars represent the standard error of the mean.

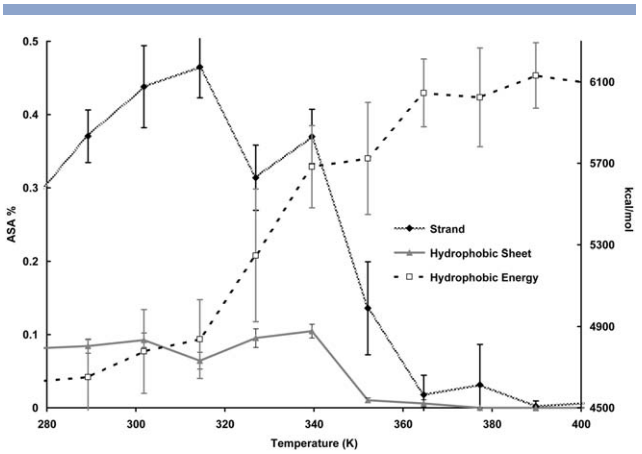

Figure 8

Variation in accessibility of strand of structure and hydrophobicity of states along the unfolding pathway of PHD2. The hydrophobic solvation energy (right axis) is computed for PHD2 structures obtained from simulation using the GBIS model of implicit solvation from NAMD. The relative ASA of β-strand structure (derived from STRIDE) and the ASA of hydrophobic strand residues are provided for reference (left axis).

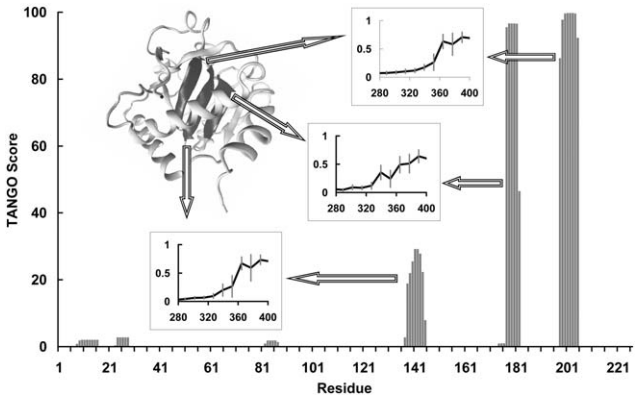

Figure 9

Spatial TANGO score (STS) tracks innate aggregation propensity in three dimensions. Main graph indicates the position of TANGO-predicted hotspots along the PHD2 sequence. The STS value of each hotspot is defined as a function of temperature. The position of each hotspot in the 3D structure of PHD2 is depicted by black ribbons.

compute STS as a function of temperature for each TANGO-defined hotspot (panels in Fig. 9), using temperature as an implicit reaction coordinate of protein unfolding. The values of temperature that we obtain indicate an energetic barrier to disruption of the protein structure.

In order to determine the stability of the TANGO-detected segments as a sheet during unfolding, we calculate the fraction of long-range native contacts for the β-barrel of PHD2. The TANGO-detected hotspots comprise half of the barrel (Fig. 10), and persist structurally intact up to 340 K.

DISCUSSION

PHD2 is the main regulator of HIF-1α stability in normoxia. HIF is a transcription factor that orchestrates the expression of a set of genes that accelerate cell survival under hypoxic conditions. In the hypoxic state of cells, PHD2 is unable to trigger degradation of HIF-1α. PHD2 expression increases in some cancer types, and high concentrations of PHD2 may trigger protein aggregation by bypassing the pathways of checking protein folding quality. On the other hand, evidence also exists that in some cancers the level of HIF-1α is high even in high level of PHD2,[16] and therefore members of the PHD2 pool might be potentially misfolded or inactive. Further, Rantanen *et al.* have shown that upon ectopic expression of PHD3, a PHD2 homolog, detectable aggregates of PHD3 appear in normoxic cells.[69] The authors demonstrate that a member of HIF-regulatory network may become inactive as a consequence of this protein aggregation in the cell. These findings lead us to question the possibility of PHD2 aggregation, as well as which specific steps along the PHD2 folding pathway lead to aggregation. Answering such questions may lead to a more complete understanding of PHD2/PHD3-related disease pathogenesis and aid in the discovery or design of possible therapies.

Differential scanning calorimetry (DSC) is a standard and sensitive method to study protein thermal transitions.[70] A protein's structure resists increases in its potential energy with increasing temperature, but at a certain temperature (the melting temperature, T_m) the protein can no longer increase its kinetic energy without breaking bonds, and therefore undergoes a transition in potential energy. This transition appears as a peak in the DSC profile. Most small proteins feature a single peak visible in a DSC profile upon heating, meaning that the protein has a single thermal transition. The appearance of multiple peaks in a DSC profile is a sign of the existence of several transitions, and consequently several populations of intermediate states, during protein thermal denaturation. The PHD2 active domain is represented structurally as a single-domain protein in Pfam (Pfam ID: PF13640) and the NCBI conserved domain database (CDD ID: cl15773); our DSC results indicate that PHD2 is a single structural domain protein with at least three stable thermodynamical intermediates, implying that at least three species with different stabilities exist in the thermal unfolding process of PHD2. PHD2 undergoes critical structural changes at I3 (Fig. 1), where the maximum changes in heat capacity and unfolding enthalpy occur, potentially representing the unfolding of the main hydrophobic region of PHD2.

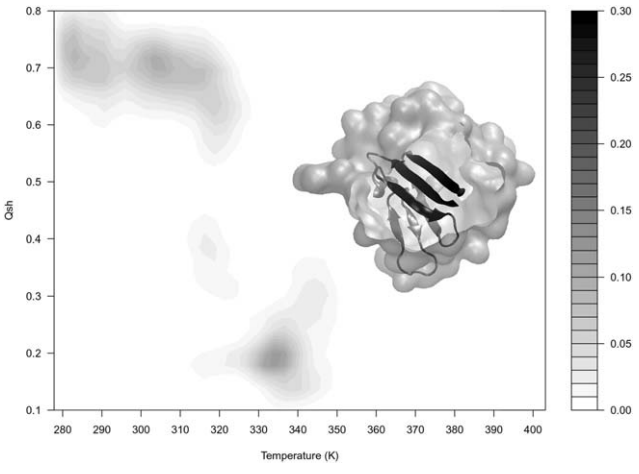

Figure 10

Stability of aggregation-prone regions of PHD2 measured by Qsh. In order to track structural changes in the TANGO-detected regions (black strands) along simulation trajectories, we compute the fraction of long-range native contacts (Qsh) for the β-barrel regions of PHD2 as a function of simulation temperature. Color bar indicates the contact density; β-barrel is highlighted as ribbon within the surface represented structure of PHD2.

Intrinsic tryptophan emission provides valuable information about tryptophan and its neighboring residues, while DSC mainly detects changes in exposure of hydrophobic groups, which exert their effects by changing the heat capacity. We observe the multi-state unfolding process using an intrinsic fluorescence assay (Fig. 2). The absorbance scans at 360 nm reveal PHD2 aggregation upon unfolding, in which we find that the aggregation step of PHD2 unfolding increases suddenly at 345 K (Fig. 2, left inset), corresponding to a sharp decrease in DSC signal after the I4 transition (Fig. 1). The study of PHD2 aggregation indicates that aggregation-prone states potentially exist even at low temperatures, where the protein mostly resides in a native-like state.

While experimental approaches aid in the description of larger-scale, molecular level aspects of the unfolding process, computational methods are capable of shedding light on the unfolding mechanism in atomic detail. In earlier computational studies, we found evidence that PHD2 unfolds by a multistate process.[71] In the current study, increased computational power allows for richer sampling of the unfolding trajectory, permitting determination of the changes in PHD2 unfolding heat capacity during unfolding. We find in our computations that the major melting transition of PHD2 upon heating occurs around 340 K, with smaller transitions near 295 K, 315 K, and 350 K (Fig. 3). *In silico* heat capacity peaks are broad, as in their experimental counterparts, suggesting an ensemble of structural species with complex multistate unfolding behavior. We find a sharp enthalpic transition at 340 K in the DMD-derived unfolding potential energy. If we focus on the appearance sequence of transitions in changes of specific heat, it reveals that two small transitions precede the main transition and one minor transition detected following the main one in experiment and computation. This consistency supports comparing the DSC results with DMD-based data. The number of native contacts present in the structure at a given time is commonly used as a reaction coordinate for the trajectory of the folding/unfolding process to illustrate the behavior of a protein in response to environmental changes. We utilize two types of measures of native contacts (Fig. 4); the overall similarity between values of Qsh-g and Qsh-w of PHD2 unfolding indicates that global protein structural changes are correlated with changes in the microenvironment of the tryptophan residues, so methods like intrinsic fluorescence are therefore appropriate to trace PHD2 behaviors in thermal unfolding.

Calculated thermodynamic quantities appear to correspond to experimental unfolding transitions, as estimated from Qsh changes (Table I), demonstrating the consistency between the experimentally-obtained and computationally-derived values of T_m (with 10% deviation), number of transitions and the sequence of their appearance. Two minor transitions exist in the fraction of long-range native contacts in the global PHD2 structure. These transitions (transitions 1 and 2 for Qsh-g) have negative heat capacity changes, which we do not observe in our experiments. This discrepancy may be a result of protein packing in early steps of simulations. The number and the appearance order of transitions after/before main transition are similar between DSC and Qsh-g analysis or between fluorescence spectroscopy and Qsh-w data (Table I).

Inspection of DSC-derived T_m values reveals that the first intermediate population is the least stable among the intermediate states (Table I), implying transiency even at room or body temperature. This low-stability state of PHD2 may play functional roles contributing to activity. Instability may originate from dynamic regions of the protein, where fluctuations are potentially necessary for regulation of substrate binding.

Differences between calorimetric changes of enthalpy and changes of enthalpy measured with fitting to spectroscopy data reveals information about the ionization enthalpy. Calorimetric enthalpy measures the ionization enthalpy of protein residues and buffer components in addition to changes in protein surface area as a function of temperature, while fitting a derived enthalpy masks these ionization terms. Spectroscopy-derived enthalpy mainly considers changes in enthalpy due to conformational changes, implying that PHD2 thermal unfolding proceeds by expansion of hydrophobic regions, where the change in both heat capacity and enthalpy increase simultaneously. The disruption of a hydrogen bond-rich region is likely the dominant event of the last transition, where the change in heat capacity and enthalpy decreases, as observed by DSC and they increase as detected in analyzing fluorescence spectroscopy data.

Constructing a potential of mean force (PMF) curve allows for description of the complex thermodynamic processes in a simple representation. The set of structures residing in each potential well of the 2D-PMF is an ensemble representing that respective intermediate state during PHD2 unfolding. Since we consider all regions of the protein in "2D-PMF-g" (Fig. 5), the properties of these states may agree more closely with DSC data. Because fluorescence data originates from the emission of tryptophan residues and their microenvironment, and to a lesser extent from other aromatic residues, the fluorescence-detected intermediate states exhibit higher correspondence with the states determined from "2D-PMF-w" (Fig. 6). The last intermediate state (number 6) in 2D-PMF-w features a partially packed structure, in agreement with the remained tryptophan fluorescence signal at high temperature (Fig. 2).

We infer from the "2D-PMF-g" that the unfolding process likely proceeds through two parallel pathways (Fig. 5), one of which is more highly populated. In the more populated route, PHD2 reaches its unfolded pool passing through intermediate state 1 to 8.

A description of unfolding intermediates must necessarily consider PHD2 structural changes. We calculate the average accessible surface area (ASA) of the ACS lumen, Docking site and the home of lip helix (see results) of the PHD2 protein for species that belong to intermediate's ensemble in "2D-PMF-g" (Fig. 7). The lip home ASA increases suddenly in the first intermediate, denoting exposure of this region. We observe from the representative structure of early intermediate (the average structure of species in intermediate's ensemble) that the lip leaves the lip home region (Fig. 5). In addition, this intermediate exhibits loss of structure in the native helical regions of PHD2. Solvent exposure of the active site lumen region does not change critically until intermediate 3 (Fig. 5). However, from intermediate 4, its ASA increases gradually. This phenomenon is indicative of the exposure and consequent unfolding of the active site region. The representative structures of intermediates 3 to 5 illustrate the corruption of active site region.

Because the docking site resides in an exposed region of the protein, its exposure does not change until the sixth intermediate (Fig. 7). The representative structure of intermediate 6 contains a β-strand-rich structure (Fig. 5). The existence of this prominent structure, a β-sheet reminiscent of the active site lumen roof, with antiparallel strands, in its folded state increases the probability of formation of aggregated structures under conditions similar to those in well 6 (Fig. 5).[72] This aggregation-prone structure lacks the inhibitory mechanisms present in natural β-sheets to prevent aggregation.[73] We find that the representative structures of ensembles that reside in well 7 and 8 do not contain this β-sheet structure (Fig. 5). Changes in accessibility of tryptophan residues further support these findings, the consequences of changes in accessibility of tryptophan residues are represented in supplementary result by analyzing the detected intermediate states of tryptophan-based 2D-PMF (Supporting Information Fig. S1). We find in computations that some species with buried tryptophan residues persist even at high temperatures.

Analyzing the accessibility of PHD2 secondary structures reveals that the most hydrophobic strands become exposed at the main thermodynamic transition (340 K), as determined by computed heat capacity (Fig. 8). This unfolded structure provides the prerequisite for beta aggregation. A TANGO-predicted aggregation-prone segment is available to participate in aggregation if it becomes accessible to solvent. Using the spatial TANGO score (STS), we determine the time of maximal aggregation propensity along the DMD simulation trajectories (Fig. 9). Among the TANGO-suggested PHD2 hotspots, S2 reaches the aggregation-prone state most quickly. Solvent accessibility alone is not sufficient to trigger aggregation; the strand-forming propensity of the segment is also important. We use TANGO to determine the propensity of each segment to form a β-strand structure, and find that the long-range native contacts of the TANGO-suggested segments maintain β-sheet structure up to a temperature of 340 K (Fig. 10), consistent with our findings in DMD simulations. The strand-rich structure of intermediate 6 (Fig. 5) features sufficient stability and aggregation propensity to act as an aggregation-prone structure along PHD2 unfolding pathway.

Because simulation system temperature is not identical to the *in vitro* temperature, we relate the sequence of transitions observed from DSC to the sequence of structural changes that occur in the three parts of the PHD2 protein. Before the first transition in DSC, the lip structure leaves the lip home, exposing the lip home region, causing the α-helices to unfold. In the second DSC transition, the active site lumen unfolds. Before the fourth transition, the most populated species features one stable β-sheet structure, which melts after the fourth transition. We infer from these findings that the PHD2 structure between third and the fourth transition would be a hydrogen bond-saturated species with its hydrogen bond network melting in the fourth transition.

The existence of a β-rich structure at high temperature (Fig. 10) points to possible simultaneous secondary structure formation and nucleation around tryptophan during the "forward" process of folding (If we suppose the unfolding states are similar to folding states but in reverse order, this conclusion seems possible). The existence of such an aggregation-prone intermediate may indicate the high sensitivity of PHD2 to changes in cellular milieu, including increases in metal ion concentrations in tumor cells[74] that promote oxidative damage, a precursor to protein aggregation. The presence of this beta-rich intermediate in the wild type PHD2 unfolding pathway (intermediate 6 in Fig. 5, intermediate 5 in Fig. 6) may offer insight to aberrations in wild type PHD2 in some cancer types.

By computing the RMSF of PHD2 residues for structures belong to intermediates of the 2D-PMF-g, we find that the order of melting of the various regions of the structure as calculated using RMSF is the same as proposed by DSC (animated Supporting Information Fig. S2). Intermediate 1 loses the lip helix; this region has the highest RMSF of all PHD2 regions in this state (Supporting Information Fig. S2, yellow regions). In intermediate 2, 3, and 4, the helices melt (Supporting Information Fig. S2, regions colored by blue, red, and orange, respectively). The floor of the active site lumen is disrupted in intermediate 5 (Supporting Information Fig. S2, green regions). Some critical events occur in well 6: small regions of the active site's lumen roof melt (Supporting Information Fig. S2, pink strands), but the majority of the active site roof remains stable (Supporting Information Fig. S2, black strands). This stable sheet may serve as a nucleus for the formation of a sandwich of sheets that remains stable in this well as an aggregate form. The average structure of well 6 (Fig. 5) features one side of

the sheet with purely positive electrostatic potential and the other side a net negative electrostatic potential (Supporting Information Fig. S3). We propose a stack of β-sheets is made by the aid of electrostatic interactions between β-sheets with opposite electrostatic potential. This finding may provide additional proof of an aggregation-prone intermediate in the PHD2 unfolding pathway.

The high levels of PHD2 mRNA in many cancer cells may induce increased PHD2 translation. PHD2 overexpression, as in its homolog PHD3, may potentially induce protein aggregation and hence inactivation in normoxic conditions. Overexpression of PHD2 protein may result in the bypass of cell quality control checks for protein folding, promoting the presence of misfolded protein, especially given our observation of an aggregation-prone state of PHD2. Aggregation-dependent PHD2 inactivation in normoxia, as a possible justification, might be allow HIF protein to remain active, so that HIF dependent pathways such as angiogenesis and metastasis hence become promoted in normoxic tumors, permitting the development of malignancy.

CONCLUSION

In summary, with simultaneous analysis of experimental and MD data, we provide an illustrative portrait for the unfolding pathway of the PHD2 protein. We find that PHD2 unfolding occurs sequentially in the lip region, helices, active site, and finally in the beta-sheet structure at the roof of active site lumen. The β-sheet-rich structure of this intermediate may increases the possibility of protein aggregation in early stage of PHD2 folding process. Characterization of the unfolding pathway aids in developing a better understanding of PHD2 mutation-independent abnormalities, as well as the potential influence of specific conditions such as PHD2 genetic amplification on the PHD2 protein folding process.

ACKNOWLEDGMENTS

The plasmid carrying the catalytic domain of human wild-type prolyl hydroxylase domain-containing protein 2 (PHD2) gene is a generous gift from Professor Schofield and Dr. Flashman at the University of Oxford, the authors gratefully acknowledge such kind gift. The authors gratefully acknowledge the help and guidance of Dr. Emily Flashman and Dr. KarKheng Yeoh from the Department of Chemistry at the University of Oxford during the process of expression and purification of PHD2 protein.

REFERENCES

1. Govindarajan S, Goldstein RA. On the thermodynamic hypothesis of protein folding. Proc Natl Acad Sci USA 1998; 95:5545–5549.
2. Tennant DA, Duran RV, Boulahbel H, Gottlieb E. Metabolic transformation in cancer. Carcinogenesis 2009; 30:1269–1280.
3. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. Science 2011; 334:517–520.
4. Berra E, Benizri E, Ginouves A, Volmat V, Roux D, Pouyssegur J. HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1alpha in normoxia. EMBO J 2003; 22: 4082–4090.
5. Ozer A, Bruick RK. Non-heme dioxygenases: cellular sensors and regulators jelly rolled into one? Nat Chem Biol 2007; 3:144–153.
6. Scott MD, Frydman J. Aberrant protein folding as the molecular basis of cancer. Methods Mol Biol 2003; 232:67–76.
7. Symposium: biochemistry of cancer–irreversible transformation in the metabolic regulatory mechanism and cancer. Seikagaku 1967;39: 525–529.
8. Chan DA, Giaccia AJ. PHD2 in tumour angiogenesis. Br J Cancer 2010; 103:1–5.
9. Rankin EB, Giaccia AJ. The role of hypoxia-inducible factors in tumorigenesis. Cell Death Differ 2008; 15:678–685.
10. Ginouves A, Ilc K, Macias N, Pouyssegur J, Berra E. PHDs overactivation during chronic hypoxia "desensitizes" HIFalpha and protects cells from necrosis. Proc Natl Acad Sci USA 2008; 105:4745–4750.
11. Astuti D, Ricketts CJ, Chowdhury R, McDonough MA, Gentle D, Kirby G, Schlisio S, Kenchappa RS, Carter BD, Kaelin WG,J, Ratcliffe PJ, Schofield CJ, Latif F, Maher ER. Mutation analysis of HIF prolyl hydroxylases (PHD/EGLN) in individuals with features of phaeochromocytoma and renal cell carcinoma susceptibility. Endocr Relat Cancer 2011; 18:73–83.
12. Dao JH, Kurzeja RJ, Morachis JM, Veith H, Lewis J, Yu V, Tegley CM, Tagari P. Kinetic characterization and identification of a novel inhibitor of hypoxia-inducible factor prolyl hydroxylase 2 using a time-resolved fluorescence resonance energy transfer-based assay technology. Anal Biochem 2009; 384:213–223.
13. Kato H, Inoue T, Asanoma K, Nishimura C, Matsuda T, Wake N. Induction of human endometrial cancer cell senescence through modulation of HIF-1alpha activity by EGLN1. Int J Cancer 2006; 118:1144–1153.
14. Couvelard A, Deschamps L, Rebours V, Sauvanet A, Gatter K, Pezzella F, Ruszniewski P, Bedossa P. Overexpression of the oxygen sensors PHD-1, PHD-2, PHD-3, and FIH is associated with tumor aggressiveness in pancreatic endocrine tumors. Clin Cancer Res 2008; 14:6634–6639.
15. Zhong H, De Marzo AM, Laughner E, Lim M, Hilton DA, Zagzag D, Buechler P, Isaacs WB, Semenza GL, Simons JW. Overexpression of hypoxia-inducible factor 1alpha in common human cancers and their metastases. Cancer Res 1999; 59:5830–5835.
16. Jokilehto T, Rantanen K, Luukkaa M, Heikkinen P, Grenman R, Minn H, Kronqvist P, Jaakkola PM. Overexpression and nuclear translocation of hypoxia-inducible factor prolyl hydroxylase PHD2 in head and neck squamous cell carcinoma is associated with tumor aggressiveness. Clin Cancer Res 2006; 12:1080–1087.
17. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013; 6:pl1.
18. Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, Gelmon K, Chia S, Mar C, Wan A, Laks E, Biele J, Shumansky K, Rosner J, McPherson A, Nielsen C, Roth AJ, Lefebvre C, Bashashati A, de Souza C, Siu C, Aniba R, Brimhall J, Oloumi A, Osako T, Bruna A, Sandoval JL, Algara T, Greenwood W, Leung K, Cheng H, Xue H, Wang Y, Lin D, Mungall AJ, Moore R, Zhao Y, Lorette J, Nguyen L, Huntsman D, Eaves CJ, Hansen C, Marra MA, Caldas C, Shah SP,

Aparicio S. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. Nature 2015; 518:422–426.
19. Prinz JH, Keller B, Noe F. Probing molecular kinetics with Markov models: metastable states, transition pathways and spectroscopic observables. Phys Chem Chem Phys 2011; 13:16912–16927.
20. Mirny L, Shakhnovich E. Protein folding theory: from lattice to all-atom models. Annu Rev Biophys Biomol Struct 2001; 30:361–396.
21. Mayne L, Englander SW. Two-state vs. multistate protein unfolding studied by optical melting and hydrogen exchange. Protein Sci 2000; 9:1873–1877.
22. Ionescu RM, Smith VF, O'Neill JC,J, Matthews CR. Multistate equilibrium unfolding of Escherichia coli dihydrofolate reductase: thermodynamic and spectroscopic description of the native, intermediate, and unfolded ensembles. Biochemistry 2000; 39:9540–9550.
23. Bhuyan AK, Udgaonkar JB. Observation of multistate kinetics during the slow folding and unfolding of barstar. Biochemistry 1999; 38:9158–9168.
24. Bismuto E, Colonna G, Irace G. Unfolding pathway of myoglobin. Evidence for a multistate process. Biochemistry 1983; 22:4165–4170.
25. Sanchez IE, Kiefhaber T. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. J Mol Biol 2003; 325:367–376.
26. Feng H, Zhou Z, Bai Y. A protein folding pathway with multiple folding intermediates at atomic resolution. Proc Natl Acad Sci USA 2005; 102:5026–5031.
27. Korzhnev DM, Religa TL, Banachewicz W, Fersht AR, Kay LE. A transient and low-populated protein-folding intermediate at atomic resolution. Science 2010; 329:1312–1316.
28. Laptenok SP, Visser NV, Engel R, Westphal AH, van Hoek A, van Mierlo CP, van Stokkum IH, van Amerongen H, Visser AJ. A general approach for detecting folding intermediates from steady-state and time-resolved fluorescence of single-tryptophan-containing proteins. Biochemistry 2011; 50:3441–3450.
29. Moosavi-Movahedi AA, Gharanfoli M, Nazari K, Shamsipur M, Chamani J, Hemmateenejad B, Alavi M, Shokrollahi A, Habibi-Rezaei M, Sorenson C, Sheibani N. A distinct intermediate of RNase A is induced by sodium dodecyl sulfate at its pK(a). Colloids Surf B 2005; 43:150–157.
30. Naganathan AN, Perez-Jimenez R, Munoz V, Sanchez-Ruiz JM. Estimation of protein folding free energy barriers from calorimetric data by multi-model Bayesian analysis. Phys Chem Chem Phys 2011; 13:17064–17076.
31. Onuchic JN, Wolynes PG. Theory of protein folding. Curr Opin Struct Biol 2004; 14:70–75.
32. Garcia AE, Onuchic JN. Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. Proc Natl Acad Sci USA 2003; 100:13898–13903.
33. Vendruscolo M, Dokholyan NV, Paci E, Karplus M. Small-world view of the amino acids that play a key role in protein folding. Phys Rev Evol Stat Nonlin Soft Matter Phys 2002; 65:061910.
34. Dokholyan NV, Li L, Ding F, Shakhnovich EI. Topological determinants of protein folding. Proc Natl Acad Sci USA 2002; 99:8637–8641.
35. de Graff AM, Shannon G, Farrell DW, Williams PM, Thorpe MF. Protein unfolding under force: crack propagation in a network. Biophys J 2011; 101:736–744.
36. Pande VS, Beauchamp K, Bowman GR. Everything you wanted to know about Markov State Models but were afraid to ask. Methods 2010; 52:99–105.
37. Bowman GR, Huang X, Pande VS. Using generalized ensemble simulations and Markov state models to identify conformational states. Methods 2009; 49:197–201.
38. Ding F, Tsao D, Nie H, Dokholyan NV. Ab initio folding of proteins with all-atom discrete molecular dynamics. Structure 2008; 16: 1010–1018.
39. Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Discrete molecular dynamics studies of the folding of a protein-like model. Fold Des 1998; 3:577–587.
40. Proctor EA, Ding F, Dokholyan NV. Discrete molecular dynamics. Wiley Interdisciplinary Rev Comput Mol Sci 2011; 1:80–92.
41. Ding F, Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI. Direct molecular dynamics observation of protein folding transition state ensemble. Biophys J 2002; 83:3525–3532.
42. Ding F, Borreguero JM, Buldyrey SV, Stanley HE, Dokholyan NV. Mechanism for the alpha-helix to beta-hairpin transition. Proteins 2003; 53:220–228.
43. Proctor EA, Kota P, Aleksandrov AA, He L, Riordan JR, Dokholyan NV. Rational coupled dynamics network manipulation rescues disease-relevant mutant cystic fibrosis transmembrane conductance regulator. Chem Sci 2015; 6:1237–1246.
44. Ding F, Furukawa Y, Nukina N, Dokholyan NV. Local unfolding of Cu, Zn superoxide dismutase monomer determines the morphology of fibrillar aggregates. J Mol Biol 2012; 421:548–560.
45. Nedd S, Redler RL, Proctor EA, Dokholyan NV, Alexandrova AN. Cu,Zn-superoxide dismutase without Zn is folded but catalytically inactive. J Mol Biol 2014; 426:4112–4124.
46. Proctor EA, Ding F, Dokholyan NV. Structural and thermodynamic effects of post-translational modifications in mutant and wild type Cu, Zn superoxide dismutase. J Mol Biol 2011; 408:555–567.
47. Redler RL, Wilcox KC, Proctor EA, Fee L, Caplow M, Dokholyan NV. Glutathionylation at Cys-111 induces dissociation of wild type and FALS mutant SOD1 dimers. Biochemistry 2011; 50:7057–7066.
48. Hewitson KS, Schofield CJ, Ratcliffe PJ. Hypoxia-inducible factor prolyl-hydroxylase: purification and assays of PHD2. Methods Enzymol 2007; 435:25–42.
49. Chowdhury R, McDonough MA, Mecinovic J, Loenarz C, Flashman E, Hewitson KS, Domene C, Schofield CJ. Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. Structure 2009; 17:981–989.
50. Sambrook J, Russell DW. Molecular cloning: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2001.
51. McNeill LA, Bethge L, Hewitson KS, Schofield CJ. A fluorescence-based assay for 2-oxoglutarate-dependent oxygenases. Anal Biochem 2005; 336:125–131.
52. Yin S, Biedermannova L, Vondrasek J, Dokholyan NV. MedusaScore: an accurate force field-based scoring function for virtual drug screening. J Chem Inf Model 2008; 48:1656–1662.
53. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999; 35:133–152.
54. Ding F, Buldyrev SV, Dokholyan NV. Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model. Biophys J 2005; 88:147–155.
55. Andersen HC. Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys 1980; 72:2384–2393.
56. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993; 234:779–815.
57. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 1992; 13: 1011–1021.
58. Senisterra GA, Finerty PJJ. High throughput methods of assessing protein stability and aggregation. Mol Biosyst 2009; 5:217–223.
59. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins 2004; 55:383–394.
60. Weiser J, Shenkin PS, Still WC. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). J Comput Chem 1999; 20:217–230.
61. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. Proteins 1995; 23:566–579.

62. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 2004; 22:1302–1306.
63. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996; 14:33–38. 27-38.
64. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci USA 2001; 98:10037–10041.
65. Cho SS, Levy Y, Wolynes PG. P versus Q: structural reaction coordinates capture protein folding on smooth landscapes. Proc Natl Acad Sci USA 2006; 103:586–591.
66. McDonough MA, Li V, Flashman E, Chowdhury R, Mohr C, Lienard BM, Zondlo J, Oldham NJ, Clifton IJ, Lewis J, McNeill LA, Kurzeja RJ, Hewitson KS, Yang E, Jordan S, Syed RS, Schofield CJ. Cellular oxygen sensing: crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). Proc Natl Acad Sci USA 2006; 103: 9814–9819.
67. Khare SD, Wilcox KC, Gong P, Dokholyan NV. Sequence and structural determinants of Cu, Zn superoxide dismutase aggregation. Proteins 2005; 61:617–632.
68. Khare SD, Ding F, Gwanmesia KN, Dokholyan NV. Molecular origin of polyglutamine aggregation in neurodegenerative diseases. PLoS Comput Biol 2005; 1:230–235.
69. Rantanen K, Pursiheimo J, Hogel H, Himanen V, Metzen E, Jaakkola PM. Prolyl hydroxylase PHD3 activates oxygen-dependent protein aggregation. Mol Biol Cell 2008; 19:2231–2240.
70. Privalov PL, Dragan AI. Microcalorimetry of biological macromolecules. Biophys Chem 2007; 126:16–24.
71. Hadi-Alijanvand H, Proctor EA, Goliaei B, Dokholyan NV, Moosavi-Movahedi AA. Thermal unfolding pathway of PHD2 catalytic domain in three different PHD2 species: computational approaches. PLoS One 2012; 7:e47061.
72. Bellesia G, Shea JE. Effect of beta-sheet propensity on peptide aggregation. J Chem Phys 2009; 130:145103.
73. Richardson JS, Richardson DC. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc Natl Acad Sci USA 2002; 99:2754–2759.
74. Pasha Q, Malik SA, Iqbal J, Shaheen N, Shah MH. Comparative evaluation of trace metal distribution and correlation in human malignant and benign breast tissues. Biol Trace Elem Res 2008; 125: 30–40.