CRYPTIC DISORDER: AN ORDER-DISORDER TRANSFORMATION REGULATES THE FUNCTION OF NUCLEOPHOSMIN

DIANA M. MITREA

St. Jude Children's Research Hospital, Department of Structural Biology

262 Danny Thomas Place, Memphis TN, 38103, USA

E-mail: Diana.Mitrea@StJude.org

RICHARD W. KRIWACKI

St. Jude Children's Research Hospital, Department of Structural Biology

262 Danny Thomas Place, Memphis TN, 38103, USA

E-mail: Richard.Kriwacki@StJude.org

Abstract

It is now well appreciated that disordered proteins and domains are prevalent in eukaryotic proteomes and that disorder is critically linked with their regulation and functionality. However, our recent observations with the multi-domain protein, nucleophosmin (Npm), suggest that the biological palette of disorder is more diverse than currently understood. The N-terminal oligomerization domain of Npm (Npm-N) can be transformed from a folded, pentameric structure to a monomeric, disordered state through changes in solution ionic strength and, importantly, through physiologically relevant post-translational modifications. Thus, it appears that Npm has been evolutionarily tuned to exist in equilibrium between disordered and ordered states. Results from us and others [1,2] suggest that the function of Npm is regulated through shifts in this equilibrium via post-translational modifications. Interestingly, this polymorphic behavior is not detected using standard secondary structure and disorder prediction algorithms, which show Npm-N to be folded into β-strands, consistent with the structure of the pentameric form. We have used a combination of computational tools, including structure-based analysis, sequence analysis algorithms (NetPhos 1.0, SCRATCH, KinasePhos, GPS2.1, PONDR) and molecular mechanics energy calculations [3,4], to test the hypothesis that the polymorphic behavior of Npm-N can be understood on structural and energetic grounds. This computational strategy has resulted in the identification of unfavorable energetic "hot-spots" within the Npm-N structure which coincide with experimentally observed sites of post-translational modification. Based on these observations, we propose that Npm-N has evolved energetic switches within its structure to enable transformation to a disordered state through phosphorylation. We further propose that the transformation process is triggered by sequential phosphorylation of solvent exposed hot-spots followed by exposure and modification of additional but initially buried sites to completely shift the equilibrium to the disordered state. This regulated, shifting equilibrium is associated with control of Npm localization within the nucleolus, nucleoplasm and cytoplasm, and with its role in regulation of centrosome duplication through interactions with Crm1-Ran [5]. More broadly, we present a general computational strategy to identify transformational hot-spots within proteins and to test the hypothesis that other proteins currently understood to be folded participate in functionally-relevant order-disorder equilibria as we have observed for Npm. The identification of such polymorphic proteins would broaden the palette of protein disorder utilized in biological systems.

1. Biological Function and Structural Features of Npm

The non-ribosomal nucleolar protein nucleophosmin (Npm here; also termed B23 and numatrin) has recently received attention due to its involvement in cell cycle regulation, chromatin remodeling, ribosome biogenesis and most importantly pro- and anti-oncogenic functions, which make it a desirable anticancer drug target [6-8]. Npm primarily resides in the nucleolous, where it regulates, among other systems, the Arf-Hdm2-p53 tumor suppression pathway by stabilizing Arf through nucleolar sequestration [9-11]. It can also shuttle between the nucleus and cytoplasm to fulfill its function in transporting pre-ribosomal particles from their site of synthesis to their site of assembly [7]. Npm exhibits three functional domains (Figure 1A). The 130 N-terminal residues encode the oligomerization domain (Npm-N) [12]; the middle segment encodes a RNA binding and ribonuclease domain; and the C-terminal 54 residues encode a DNA-binding domain [12,13]. Computational analysis of the primary sequence of Npm using the disorder prediction algorithm PONDR [14] indicated that the N- and C-terminal domains are folded and that the middle domain of ~120- residues is disordered (Figure 1B). In agreement with this prediction, the N-terminal domain has been crystallized as a β-sheet-rich pentamer [2,15]. The solution structure for the C-terminal domain has also been elucidated by NMR as being a right-handed 3-helix bundle [13]. To date, there is no structural information on the central region.
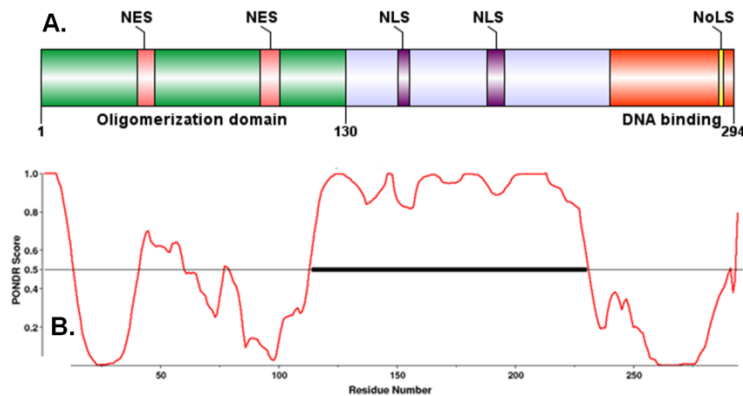

Figure 1. Domain organization of human Npm. (A) Cartoon representation of the three structural domains: oligomerization (green), RNA binding and ribonuclease activity (grey) and DNA binding (orange); (B) Disorder prediction in the secondary structure of full length Npm, using the PONDR algorithm [14].

Mammalian Npm orthologs exhibit a high degree of amino acid sequence conservation with over 69% sequence identity and over 90% sequence similarity observed over the entire protein. The sequence of the oligomerization domain is even more highly conserved, with greater than 90% sequence identity observed across all analyzed mammalian species (Figure 2), suggesting that this domain plays an important, evolutionarily conserved biological role in Npm function. Furthermore, the three dimensional structure of the N-terminal domain is conserved amongst metazoan species [2,15-18]. Initial insights into the importance of Npm oligomerization in biological function emerged from studies in which small molecules or RNA aptamers were used to disrupt its pentameric structure. In these studies, fluorescence microscopy was utilized to show that disruption of oligomers was associated with re-localization of Npm from the nucleolus to the

nucleoplasm [19,20]. These results suggest that the pentameric state of Npm is important in the biological function of Npm, with sub-cellular localization being one property that is affected by the presence or absence of folded oligomers.
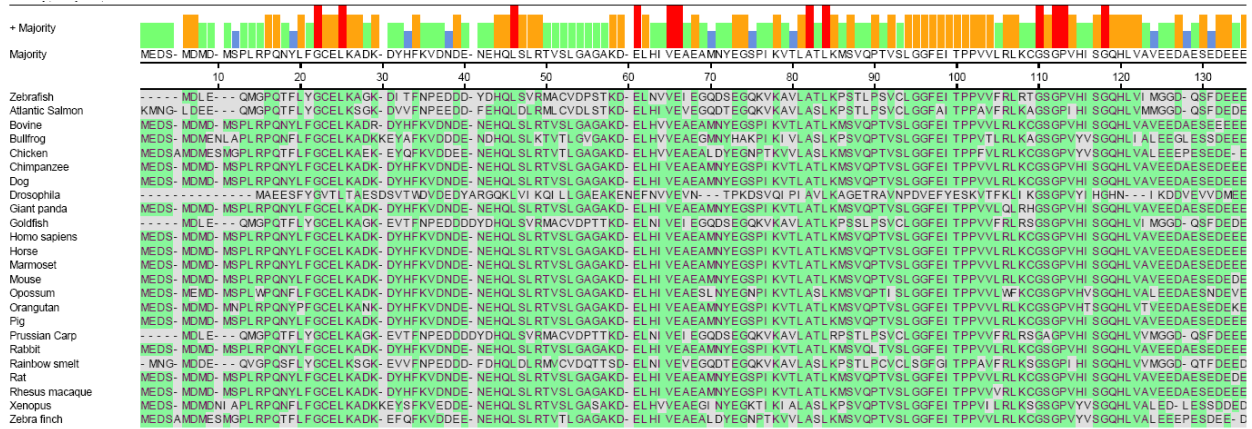

Figure 2. Sequence alignment of Npm-N orthologs using MegAlign (DNASTAR Lasergene 8)

Intriguingly, Olson and co-workers showed in the late 1990s using analytical ultracentrifugation that the oligomerization state of Npm was dependent upon ionic strength *in vitro*. At low and high cation concentrations, Npm sedimented as a monomer and oligomer, respectively [1]. Subsequently, we have shown that the low salt monomeric state of Npm-N is largely disordered, while the high salt state is folded as a pentamer, similar to the crystallographically observed structures [2,15]. Based upon this past report and our recent findings, we initiated investigations into the molecular basis of this apparent ionic strength-dependent, order-disorder transformation.

To gain insight into the intrinsic structural features of the Npm-N amino acid sequence, we performed several different types of sequence analysis. First, secondary structure prediction was performed using the SCRATCH and PredictProtein servers [21,22]. The results of these analyses (Figure 3) were in good agreement with the crystallographic results, with all of the observed β-strands within the pentameric structure accurately predicted by the two servers (only the results from PredictProtein are illustrated in Figure 3). This type of analysis provides little insight, however, into the molecular basis of the ionic strength-dependent structural transformation exhibited by Npm-N. For example, these algorithms do not take into account the influence of ionic strength on secondary and/or higher order structure.

Second, disorder predictions were performed using the PredictProtein and PONDR servers [14,22]. The results were again generally in good agreement with the crystallographic results, with loops that connect β-strands predicted to be disordered (Figures 1A and 3). Again, however, these algorithms have not been developed to allow the influence of ionic strength on disorder or, alternatively, on higher order structure (tertiary and quaternary structure) to be assessed. Consequently, these analyses did not provide insight into the molecular basis of the ionic strength-dependent, order-disorder transformation of Npm-N. A need thus arises for tools that enable us to predict a protein's potential polymorphic behavior.
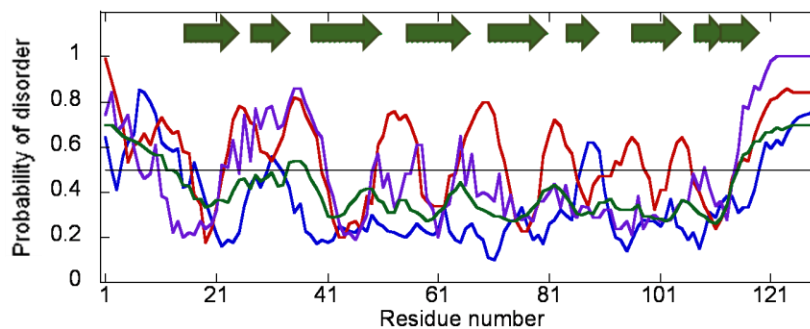
Figure 3. Results of secondary structure and disorder predictions for Npm-N by PredictProtein [22]. Solid traces represent probability for disorder, determined by Blue – NORSnet, Red – PROFbval, Purple - Ucon, Green – MD algorithms; Arrows represent locations of the predicted β-strands.

Next, we discuss our computational strategy to gain insights into the mechanisms responsible for inducing the switch from order to disorder in Npm-N.

2. Alteration of the electrostatic features of Npm-N through phosphorylation

We reasoned that, while variation of ion strength is not a physiologically relevant regulatory mechanism, the order-disorder transformation of Npm-N might be modulated by post-translational modifications of residues within this domain. In particular, we focused attention on phosphorylation as a possible mechanism for perturbation of the Npm-N order-disorder equilibrium. An analysis of the literature, together with results from prediction algorithms, revealed numerous known or putative phosphorylation sites that were evenly distributed through the amino acid sequence (Table 1).

Table 1. Analysis of experimentally characterized and putative phosphorylation sites within Npm-N in the context of the pentameric structure, including the extent of conservation of the modified residue (with the human sequence as reference), solvent accessibility, and the kinase(s) responsible for the modification.

Phosphorylation site	Residue conservation	Solvent accessibility	Modifying kinase
S4	S	Y	PLK2 [23,24]
S10	S/G/N[*]	Y	CDK [24,25]
Y17	Y/F	Y	TK, PDGFR [24,26]
S43	S	N	PKA, CaM-II, CKI [27]
T46	T	N	CKI [28], CAMK [26 24]
S48	S/T	N	PKA [24,29]
Y67	Y	Y	TK, Abl [24,26]
S70	S/N	Y	CDK [24,25]
T75	T/V[*]	N	PKA [29,30]
T78	T/S	N	PKC [29]
S82	S	Y	CKI [28]

S88	S	N	PKA, PKG, ATM [27,28,31]
T95	T	Y	PDK [5], cdc2, p38MAPK [29]
S106	S	Y	[32]
S112	S	N	PKA [29]
S125	S	Y	CKII [27,29], CDK2 [30]

[*] Another aliphatic, hydroxylated phosphorylatable residue (T/S) can be found in close proximity (~ 2-3 residues away from the conserved site).

Results from several large scale proteomic studies demonstrate that numerous residues within Npm-N are phosphorylated in living cells [24,30-32]. Phopshorylation levels are increased in cells undergoing mitosis, when nucleolar localization of Npm is not required [24,30]. Interestingly, while the location and number of Npm's phosphorylatable residues are highly conserved among mammalian species, not all sites shown to be modified are solvent accessible (Figure 4). Based on this observation, we hypothesize that an evolutionarily conserved regulatory mechanism mediates unfolding, or some other type of structural reorganization of pentameric Npm-N to provide kinases access to the buried phosphorylation sites. Our rational was that if phosphorylation at one or more accessible sites destabilized the pentameric protein fold, local or global unfolding may occur, thereby exposing additional sites for modification.
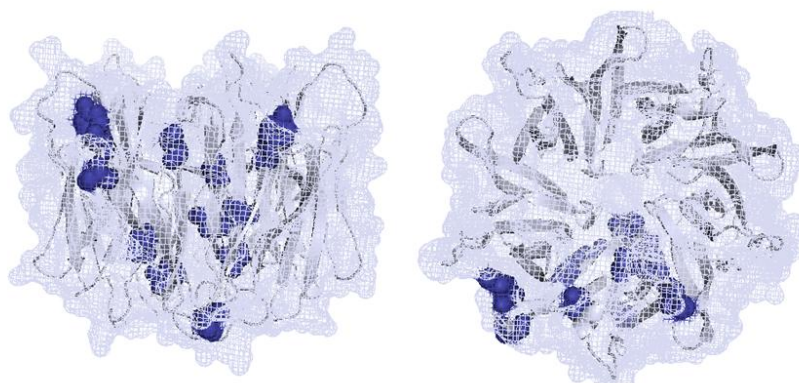

Figure 4. Location of phosphorylation sites (indicated with blue spheres) within the pentameric, β-strand-rich structure of Npm-N. These residues exhibit varying degrees of solvent accessibility (left, side view with modified residues illustrated for one of the five symmetric subunits of the pentamer; right, top view in the same format).

3. *In Silico* site-directed mutagenesis

In order to test the hypothesis discussed above, we employed the empirical force field algorithm FoldX to calculate the energetic effects of phosphorylation at each conserved, potential phosphorylation site of Npm-N. The energy function used by FoldX utilizes the full atomic details of the protein and empirically-derived, weighted physical and statistical energy terms, and is computationally efficient [3,4]. In a recent comparative study, the FoldX algorithm was judged to be among the most accurate protein stability predictors available. Notably, FoldX was one of only

two predictors that were able to correctly identify stabilizing mutations [33]. This time and resource efficient approach relies on the availability of a high resolution structural data for the protein under investigation. A reliable force field which accurately estimates the effects of both stabilizing and destabilizing mutations is also critical. For best results, structural models with a resolution of minimum 2.2 Å should be used. Due to the empirical nature of the energy calculations, poorly defined side-chain or backbone orientations in the input model can lead to significant errors [34].

As a starting point for the energy calculations, we utilized the Npm-N pentamer structure refined at 1.8 Å resolution [2]. Due to disorder in the N- and C-termini, residues 1-13 and 119-130 are not visible in the crystal structure, and consequently, we were unable to assess the effect of phosphorylation at Ser4, Ser10 and Ser125 on the stability of Npm-N pentamer. We speculate that, due to the dynamic nature of these regions, their post-translational modification will have only limited impact on the overall fold of the Npm-N oligomer. Modification of these residues, rather, may serve to regulate recognition by various binding partners.

We performed all calculations at the physiological ionic strength of 0.15 M, where experimental evidence places the Npm-N order-disorder equilibrium in favor of the folded pentameric species (ΔG = -149.5 kcal/mol). Reducing the ionic strength to null during FoldX simulations recapitulates the effect observed *in vitro*, namely a considerable reduction in the thermodynamic stability of the pentameric structure ($\Delta \Delta G$ ~ +75 kcal/mol). FoldX simulations revealed that phosphorylation of each individual site (within each of the five polypeptide chains present in the pentamer) destabilized the Npm-N structure, with the effects ranging from mild to severe, depending upon the extent of burial on the modified site within the pentameric fold. Due to the empirical nature of the force field, and the rather limited amount of experimental data describing the effect of phosphorylation on protein stability compared to mutagenic studies, we performed calculations of phopsho-mimetic mutations (Ser to Glu and Thr to Asp). The results were in agreement with the phosphorylation results, namely a decrease in the Npm-N stability was observed for all mutated residues. A collection of energetic calculations for the phosphorylated species is given in Table 2. The energy terms within the FoldX force field that contributed (increased in magnitude, corresponding to thermodynamic unfavorability) most significantly to destabilization of the Npm-N pentamer were those for Van de Waals interactions, hydrogen bonds and solvation energy. Counterintuitively, the term for electrostatic interactions was not a major contributor to the change in total stability of the protein, nor did the value of this term for the different phosphorylation mutants correlate with experimental findings on the extent of perturbation of the order-disorder equilibrium destabilization associated with site-specific phosphorylation [2]. Modifications at positions Thr46 or Ser48, shown to be phosphorylated in mitotic cells [24], reduced the stability of Npm-N to extents that were similar to or greater than those associated with reducing the ionic strength parameter to null for the unmodified amino acid sequence. Similarly, modifications at Thr78 and Ser112 resulted in severe energetic penalties (between 60-123 kcal/mol). From a structural and functional perspective, these observations suggest that phosphorylation of these residues causes a shift of the order-disorder equilibrium

toward disorder and that this shift may down-regulate Npm function(s) associated specifically with the ordered, pentameric form. These sites, however, are not solvent accessible in the pentameric form of the protein (Table 1 and Figure 4); therefore, it is unlikely that the responsible kinases can directly phosphorylate these sites due to steric inaccessibility. Moderate energetic effects (+10-35 kcal/mol) were observed for *in silico* phosphorylation of Tyr17, Ser43, Thr75, Ser88, a mix of buried and surface exposed residues. The results of our calculations suggest that modifications of surface-exposed sites also induce, albeit to a more limited extent, structural destabilization (e.g., residues Tyr67, Ser82 and Thr95; Table 2). Perhaps not sufficient for inducing complete unfolding of the pentamer, these surface phosphorylation events may shift the population distribution of the ordered pentamer-disordered monomer Npm-N equilibrium, thus providing a mechanism for a buried site to be transiently exposed for phosphorylation.

Table 2. Effects of phosphorylation on the thermodynamic stability of Npm-N pentamer (energies in units of kcal/mol pentamer; $\Delta\Delta G_{total}$ – total free energy; $\Delta\Delta G_{vdw}$ – Van der Waals energy; $\Delta\Delta G_{el}$ – electrostatic energy; $\Delta\Delta G_{solv}$ – solvation energy). Phosphorylation and Alanine substitutions were introduced simultaneously in all five chains of Npm-N; for the AlaScan, the sites were replaced in one polypeptide chain at a time.

Site	Effect of Phosphorylation				Effect of substitution with Alanine				# of chains stabilized by AlaScan
	$\Delta\Delta G_{total}$	$\Delta\Delta G_{vdw}$	$\Delta\Delta G_{el}$	$\Delta\Delta G_{solv}$	$\Delta\Delta G_{total}$	$\Delta\Delta G_{vdw}$	$\Delta\Delta G_{el}$	$\Delta\Delta G_{solv}$	
Y17	14.80	10.69	-0.11	6.38	13.47	10.46	-0.04	5.62	0
S43	18.11	0.62	-3.65	16.49	-5.47	0.81	0.03	-6.99	5
T46	122.80	86.15	6.19	16.72	11.83	3.68	0.00	-2.55	0
S48	66.71	31.02	32.48	-0.70	-2.49	0.74	0.00	-7.78	5
Y67	6.56	7.18	0.35	4.49	4.21	6.81	0.08	4.43	0
S70	-0.30	-1.58	0.67	-1.24	0.15	0.38	0.0	-1.97	2
T75	25.93	8.92	0.85	-0.87	9.12	3.93	-0.89	0.62	0
T78	78.37	44.20	-4.80	19.58	11.58	4.14	0.00	-1.02	0
S82	4.85	-0.85	2.83	-0.31	1.23	0.15	-0.14	-2.44	2
S88	34.93	6.88	2.85	17.99	-1.39	1.33	0.0	-5.24	5
T95	5.33	-1.02	0.83	-3.12	1.66	2.61	0.0	-1.67	2
S112	64.67	35.94	-6.4	30.01	-1.71	1.44	0.0	-8.37	5

4. Probing for structural strain in Npm-N

While the thermodynamic analyses discussed above provide insights into how phosphorylation affects the thermodynamic stability of Npm-N, some questions remain. For example, what are the evolutionary features that enable these types of proteins to switch between an ordered and a disordered species? More importantly, can we identify and understand the energetic hallmarks of this type of structural regulatory mechanism? Can we use the insights gained in our studies of Npm-N to develop general rules to identify other proteins that exhibit similar polymorphic behavior?

To begin to address these questions, we performed systematic *in silico* screens of the effects of structural alteration of each residue in Npm-N on the thermodynamic stability of the pentameric structure. Our rationale was that this strategy would identify residues that experience unfavorable steric or electrostatic clashes within the native pentameric fold. The experimental observations that high solution ionic strength results in stabilization of the ordered, pentameric form of Npm-N [1,2] suggested that unfavorable electrostatic interactions are screened by salt ions and consequently allow the hydrophobic core to more readily collapse and fold. This would not be surprising, considering the acidic loops composed of poly-Glu and -Asp, which connect the β-strands in Npm-N pentamer. As a methodological approach to identify unfavorable side-chain interaction, we utilized the AlaScan function in FoldX. In this procedure, the program replaces every single residue with Ala, on each one of the five polypeptide chains, one at a time, and calculates the resulting change in free energy. The observation of increased stability upon Alanine substitution, which essentially corresponds to removal of the side-chain, suggests that the substituted residue experienced unfavorable interactions with one or more of its neighbors. As depicted in Table 2, replacement of either of the four highly conserved, buried Ser residues (43, 48, 88 and 112) with Alanine, in individual subunits of the Npm-N pentamer, resulted in enhanced stability of the protein. This suggests that removal of the hydroxyl group of the Ser side-chain releases some form of unfavorable "conformational strain". We further simulated the effect of side-chain removal by uniformly mutating each conserved Ser residue within all five protomers within the pentamer. Again, the four Ser residues are notable as introducing unfavorable conformational strain within the pentameric structure, as shown in the selected energetic terms highlighted in Table 2. In accord with the hydrophobic environment of these amino acids, the electrostatic contribution to the increase in stability is negligible. Van der Waals clashes are reduced, while the solvation energy becomes more favorable. These results support the hypothesis that conformational strain within the wild-type Npm-N pentamer provides a mechanism for thermodynamic destabilization of this ordered state. The energy landscape created enables structural polymorphism through an equilibrium with an unfolded, disordered state.

5. Mechanistic insights on Npm's order-disorder polymorphism

The concept that the normal function of proteins involves conformational fluctuations between multiple states is well established. There are several examples wherein the structural and dynamic features associated with transitions amongst alternative conformations have been characterized in

detail (*e.g.* ATP synthase[35], calmodulin [36]). However, the function of most proteins is understood in terms of static, folded structures. Our results for Npm-N illustrate that, in contrast to the examples noted above, that conformational fluctuations can occur between two dramatically different, functionally relevant conformations: a folded, pentameric state and a highly disordered, monomeric state. Due to the bias in secondary structure prediction algorithms which are based on the structural data present in the Protein Data Bank, and that of the disorder prediction algorithms, which are based on the current understanding of intrinsically disordered proteins, only the folded state of Npm-N was evidenced in these types of analyses. By using a computational approach to analyze the thermodynamics of conformational strain within the wild-type Npm-N pentamer, we identified residues we describe as structural hot-spots which appear to have dual functions: *(1)* as sites of thermodynamic destabilization and *(2)* as sites for post-translational modification (phosphorylation) to augment thermodynamic destabilization. Furthermore, our results suggest a regulatory mechanism, in which sequential phosphorylations regulate Npm-N structure and function (Figure 5). Modification of solvent exposed residues weakens the pentameric structure of Npm-N, shifting the order-disorder equilibrium toward the disordered state. This enhances the accessibility of buried sites for phosphorylation, leading to further destabilization of the ordered state and consolidation of the disordered state. The first step of this proposed mechanism is supported by a report from Bañuelos and co-workers, who have shown that phosphomimetic mutations at the surface exposed sites in the nucleoplasmin core results in structural destabilization, with no evident changes in the overall packing in the crystal structure. Interestingly, the enhanced dynamics in the protein correlated well with an increase in functional activity in decondensing sperm chromatin [37]. Phosphorylation events have been previously linked to increase in structural flexibility in proteins, often times in order to facilitate binding of a particular partner. In several instances, local unfolding occurs as part of the function regulatory process.   The order-to-disorder transitions are restricted, however, to small structural elements (*e.g.* unfolding of an α-helix) or occur between two energetically proximal states [38-41].

It is well understood that post-translational modifications occur on a significant fraction of eukaryotic proteins and that they play a crucial role in modulating their function. Due to technical difficulties, however, more often than not, the structural characterization of mammalian proteins is performed in the absence of these modifications. We propose a perspective from which to view newly characterized proteins and a re-visiting of the previously solved structures of, particularly, multi-functional proteins, in search of alternative, potentially biologically relevant conformations. An energetic assessment of the protein, in the presence of simulated post-translational modifications, can provide insights into the existence of destabilizing hot-spot positions and/or the energy penalty or benefit associated with any particular modification. Taking this strategy a step further, one could compare the location of the hot-spot residues, not only to the positions for post-translational modifications, but also those corresponding to ligand binding or partner-protein binding, which could become accessible for interaction only after a certain set of already exposed sites have been modified. This type of mechanism seems to be quite efficient for regulation of

multi-domain, multi-functional proteins, where one segment of the polypeptide chain is unfolded in response to phosphorylation, for example, without the need for protein degradation, a process which would have inactivated the other domains as well.
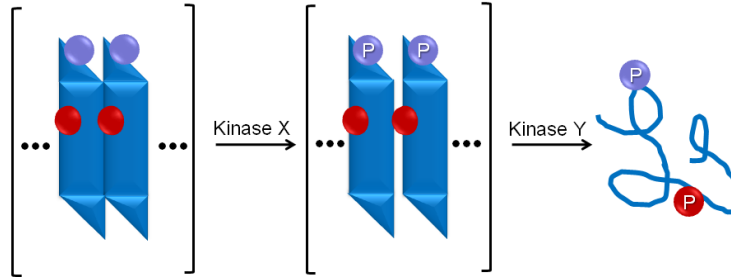

Figure 5. A sequential phosphorylation mechanism mediates a structural switch from order to disorder in Npm-N. Ordered, pentameric Npm-N is phosphorylated at surface exposed sites, which results in destabilization and increased dynamics, transiently exposing buried segments of the polypeptide (Step 1). This enables phosphorylation at the buried sites which introduces severe strain within the folded structure, leading to global unfolding (Step 2).

What is the biological relevance of our findings for Npm-N? Npm is a highly abundant, multi-domain protein and the N-terminal oligomerization plays important roles in modulating its biological function, as evidenced by alteration of normal function through mutagenesis within the N domain [2,42] or treatment with agents which disrupt oligomerization [19,20]. Npm function is modulated in a cell cycle-dependent manner. For example, several different cyclin-dependent kinases phosphorylate Npm, with some of these sites identified within the N domain, and modulate its nucleolar localization and interactions with Crm1-Ran at centrosomes [5,43]. This latter interaction controls centrosome duplication, allowing only one such event per division. Based upon the experimentally characterized polymorphic behavior of Npm-N and our growing understanding of the thermodynamic underpinnings for this phenomenon, we hypothesize that Npm, and possibly oligomeric proteins, have evolved to exhibit two dramatically different structural states—order and disorder— to diversify their biological functionality. Here we propose a phosphorylation mechanism which causes a switch from order to disorder. Phosphatases are likely to reverse the switching action of cell cycle kinases, thus resetting the equilibrium in favor of the ordered state. Folded, ordered Npm-N may participate in inter-molecular interactions (e.g., with Crm1/Ran, histone chaperones) that are disrupted when switched to the monomeric, disordered state. Furthermore, phosphorylation may not only cause the order-disorder switch, but also create phosphorylation-dependent recognition motifs that become exposed through the phenomenon of order-disorder switching. By including a disordered state in its structural repertoire, Npm has evolved to possess complex and diverse functionality. However, the biological implications of its disordered features, which have been appreciated for many years, have remained cryptic.

The approach described herein uses existing energy predictors to analyze the effect of phosphorylation at experimentally validated sites. The kinases responsible for postranslational modifications within the structured part of the Npm-N are unknown, thereby making the

experimental measurements quite problematic. Even in the absence of this information, our method allows us to assess the potential repercussions of phosphorylation events on the protein's stability and gain insight into the phosphorylation-dependent regulatory mechanism. We propose that by combining structural information with motif prediction algorithms and energy calculations, the cryptic polymorphic behavior of other biologically important proteins may be revealed. This would add a new level of complexity in protein regulatory mechanisms and further expand our knowledge of the roles of disordered protein/domains in biology.

6. Materials and Methods

FoldX simulations. All thermodynamic calculations were performed using the pentameric structure of human Npm-N determined using X-ray crystallography [2]. The total energy of the structure was first optimized using the RepairPDB command. All simulations, unless otherwise noted, were performed at 298 K, 0.15 M ionic strength and pH 7.0. Each calculation was performed in triplicate, with identical outputs.

References:

1.   Herrera JE., Jones AE., and Olson MO., *Biochemistry* **35**, 2668 (1996).
2.   Kriwacki RK et al. Mitrea DM, (manuscript in preparation).
3.   Nielsen JE, Guerois R, Serrano L., *J. Mol. Biol.* **320** (2), 369 (2002).
4.   Rousseau F., Schymkowitz JW., Martins IC., Ferkinghoff-Borg J., Stricher F., Serrano L., *PNAS* **102**, 10147 (2005).
5.   Budhu A., Wang W., Forgues M. and Wang XW, *Nature Cell Biology* **7** (8), 823 (2005).
6.   Mecucci C., Grisendi S, Falini B., Pandolfi PP. , *Nature Reviews* **6**, 493 (2006).
7.   Lindstrom M., *Biochemistry Research International* **2011** (Article ID 195209) (2010).
8.   Alcalay M., Colombo E., Pelicci PG., *Oncogene* **30** (23), 2595 (2011).
9.   den Besten W., Kuo ML., Bertwistle D., Roussel MF., Sherr CJ., *Genes Dev.* **18** (15), 1862 (2004).
10.   Kuo ML., den Besten W., Williams RT., Sherr CJ., *Cell Cycle* **4** (11), 1593 (2005).
11.   Shan J., Chen D., Zhu WG., Qin J., Gu W., *Nature* **464** (7288), 624 (2010).
12.   Szebeni A., Hingorani K., Olson MO., *J. Biol. Chem.* **275** (32), 24451 (2000).
13.   Townsley FM., Grummitt CG., Johnson CM., Warren AJ., Bycroft M., *J. Biol. Chem.* **283** (34), 23326 (2008).
14.   Dunbrack RL., Xue B., Williams RW., Dunker AK., Uversky VN., *Biochim. Biophys. Acta* **1804** (4), 996 (2010).
15.   Kim HS., Lee HH., Kang JY., Lee BI., Ha JY., Yoon HJ., Lim SO., Jung G., Suh SW., *Proteins* **69** (3), 672 (2007).
16.   Akey IV., Dutta S., Dingwall C., Hartman KL., Laue T., Nolte RT., Head JF., Akey CW., *Mol. Cell* **8** (4), 841 (2001).
17.   Akey IV., Namboodiri VM., Schmidt-Zachmann MS., Head JF., Akey CW., *Structure* **12** (12), 2149 (2004).
18.   Dutta S., Namboodiri VM., Akey IV., Head JF., Akey CW., *Structure* **11** (2), 175 (2003).
19.   Shakalya K., Qi W., Stejskal A., Goldman A., Beeck S., Cooke L., Mahadevan D., *Oncogene* **27** (30), 4210 (2008).

20. Gao Z., Jian Y., Sun J., Shen Q., Feng F., Jing Y., Yang C., *Oncogene* **28** (47), 4201 (2009).
21. http://scratch.proteomics.ics.uci.edu/.
22. http://www.predictprotein.org/.
23. Hoffmann I., Krause A., *PLoS One* **5** (3) (2010).
24. Vermeulen M., Olsen JV., Santamaria A., Kumar C., Miller ML., Jensen LJ., Gnad F., Cox J., Jensen TS., Nigg EA., Brunak S., Mann M., *Science Signaling* **3** (104) (2010).
25. Zhou Y., Du W., Pike S., Pang Q., *Carcinogenesis* **31** (2), 302 (2010).
26. Ren J., Xue Y., Gao X., Jin C., Wen L., and Yao X., *Mol. Cell. Proteomics* **7**, 1598 (2008).
27. http://kinasephos.mbc.nctu.edu.tw/.
28. http://www.phosida.de/.
29. http://www.cbs.dtu.dk/services/NetPhosK/.
30. Zhou C., Dephoure N., Villen J., Beausoleil S.A., Bakalarski C.E., Elledge S.J., Gygi S.P., *PNAS* **105**, 10762 (2008).
31. Yi W., Cantin G.T., Lu B., Park S.K., Xu T., Lee J.-D., Yates J.R. III, *J. Proteome Res.* **7**, 1346 (2008).
32. Gnad F., Oppermann F.S., Olsen J.V., Hornberger R., Greff Z., Keri G., Mann M., Daub H., *Mol. Cell. Proteomics* **8**, 1751 (2009).
33. Vihinen M., Khan S., *Human Mutation* **31** (6), 675 (2010).
34. Serrano L., Nadra AD., Alibes A., *Methods in Enzymology* **498**, 3 (2011).
35. Iino R., Okuno D., Noji H., *J. Biochem.* **49** (6), 655 (2011).
36. Grabarek Z., *Biochim. Biophys. Acta* **1813** (5), 913 (2011).
37. Muñoz IG., Taneva SG., Franco G., Falces J., Arregi I., Muga A., Montoya G., Urbaneja MA., Bañuelos S., *Biochemistry* **47** (52), 13897 (2008).
38. Hollingworth D., Díaz-Moreno I., Frenkiel TA., Kelly G., Martin S., Howell S., García-Mayoral M., Gherzi R., Briata P., Ramos A., *Nat Struct Mol Biol* **16** (3), 238 (2009).
39. Wenta N., Nardozzi J., Yasuhara N., Vinkemeier U., Cingolani G., *JMB* **402** (1), 83 (2010).
40. Ding F., Proctor EA., Dokholyan NV., *JMB* **408** (3), 555 (2011).
41. Traaseth NJ., Gustavsson M., Karim CB., Lockamy EL., Thomas DD., Veglia G., *JMB* **408** (4), 755 (2011).
42. Lindstrom MS., Enomoto T., Jin A., Ke H., Zhang Y. , *J. Biol. Chem.* **281** (27), 18463 (2006).
43. Brancaccio M., Hanashiro K., Fukasawa K., *Oncogene* **30** (19), 2188 (2011).