

MODULATING PROTEIN–DNA INTERACTIONS BY POST-TRANSLATIONAL MODIFICATIONS AT DISORDERED REGIONS

DANA VUZMAN

Department of Structural Biology, Weizmann Institute of Science, Rehovot, 76100, Israel

Email: dana.golbin@weizmann.ac.il

YONIT HOFFMAN

Department of Structural Biology, Weizmann Institute of Science, Rehovot, 76100, Israel

Email: yonit.hoffman@weizmann.ac.il

YAAKOV LEVY*

Department of Structural Biology, Weizmann Institute of Science, Rehovot, 76100, Israel

Email: Koby.Levy@weizmann.ac.il

Intrinsically disordered regions, particularly disordered tails, are very common in DNA-binding proteins (DBPs). The ability of disordered tails to modulate specific and nonspecific interactions with DNA is tightly linked to their being rich in positively charged residues that are often non-randomly distributed along the tail. Perturbing the composition and distribution of charged residues in the disordered regions by post-translational modifications, such as phosphorylation and acetylation, may impair the ability of the tail to interact nonspecifically with DNA by reducing its DNA affinity. In this study, we analyzed datasets of 3398 and 8943 human proteins that undergo acetylation or phosphorylation, respectively. Both modifications are common on the disordered tails of DBPs (3.1 ± 0.2 (0.07 ± 0.007) and 2.0 ± 0.2 (0.02 ± 0.003) acetylation and phosphorylation sites per tail (per tail residue), respectively). Phosphorylation sites are abundant in disordered regions and particularly in flexible tails for both DBPs and non-DBPs. While acetylation sites are also frequently occurred in the disordered tails of DBPs, in non-DBPs they are often found in ordered regions. This difference may indicate that acetylation has different function in DBPs and non-DBPs. Post-translational modifications, which often take place at disordered sites of DBPs, can modulate the interactions of proteins with DNA by changing the local and global properties of the tails. The effect of the modulation can be tuned by adjusting the number of modifications and the cross-talks between them.

1. Introduction

Post-translational modifications (PTMs) are widely used to modulate protein function in the cell. PTMs can therefore be viewed as covalent modifications that increase the structural and biophysical diversity of proteins and thus enrich the information stored in the genomes. There are dozens of different PTMs that are incorporated by the cell. To achieve the required effect, a protein may undergo a single PTM or several PTMs that may

engage in cross-talk. In many cases, a single position on the protein can be altered by different modifications so that switching between several effects (or functions) can be regulated by the identity of the PTM at that position. PTMs are often classified according to the mechanisms involved: the addition of functional groups (e.g., phosphorylation and glycosylation); attachment of other polypeptides (e.g., ubiquitination and SUMOylation); changing of the chemical nature of amino acids (e.g., acetylation, deamidation and oxidation); and cleavage of the backbone by proteolysis ^{1, 2}. PTMs can be categorized according to the conformational preference of the modification sites; namely, if the PTM occurs on a structured or disordered region. PTMs at structured domains are crucial, for example, for modifying enzymatic activities or stabilizing protein structure. PTMs at disordered regions are advantageous because of the high-exposure of these sites, which enables them to be accessed easily by the modifying enzymes through high-specificity and low-affinity interactions that also often involve a disorder-to-order transition. Indeed, the abundance of disordered sites as targets for PTMs is evident from many experimental studies ³ and various predictive approaches indicate that sequences surrounding PTMs have features similar to those of disordered regions ⁴.

In this study, we focus on two widespread PTMs, acetylation and phosphorylation, that play important roles in modulating various regulatory processes. We focus on the analysis of these two PTMs in DNA-binding proteins (DBPs), which are a unique but highly central group of proteins that mediate numerous genetic processes, such as transcription, repression, and replication via interaction with DNA. Very often, DBPs include a patch of positively charged residues that serve as the binding site for interacting with the negatively charged DNA. The location of the positively charged residues (Lys and Arg) in the 3D structure supports the participation in hydrogen bonding with the bases of the nucleotides that constitute the cognate site. The positive patches on DBPs not only contribute to specific recognition of the protein residues and the DNA nucleotides, but also play a pivotal role in allowing DBPs to interact non-specifically with DNA. While the affinity of DBPs for non-specific DNA is relatively low and can be strongly affected by salt concentrations that mask the electrostatic complementarity, non-specific interactions are nevertheless central to the achievement of fast recognition.

Electrostatic interactions between DBPs and DNA mediate the search mechanism for the cognate target site, which is surrounded by many alternative sites. Non-specific electrostatic interactions govern the one-dimensional sliding of DBPs along the linear contour of the DNA ⁵. The time a DBP spends in the different search modes (sliding, hopping, inter-segment transfer, and 3D diffusion) is dictated by the electrostatic features of the protein as well as by the salt concentration. It was shown that DBPs may not always slide using the patch that is used for specific interactions ⁶. The non-specific and specific interactions of DBPs with DNA can be supported by disordered regions, often at either or both the C- and N-termini. The disordered tail on DBPs is positively charged and can be viewed as another sub-domain that interacts with DNA. The high specificity and high selectivity of the recognition motif at the structured domain involved in interacting with DNA can be enhanced by the disordered sub-domain. The importance of the N-terminal disordered tail for protein folding and for specific or non-specific interactions with DNA

was studied recently for several homeodomain transcription factors, both experimentally^{7, 8} and computationally^{9, 10}. These disordered tails, which are positively charged and highly disordered in solution, can cause thermodynamic destabilization or stabilization of the protein in the absence or presence of DNA, respectively^{8, 11, 12}. The tails of the homeodomain may fold upon binding to the DNA in the minor groove or remain partly flexible or disordered¹³. A number of studies have demonstrated that N-tails can contribute to selective sequence binding specificity^{8, 13-16}. A recent computational study of three homeodomain proteins with different tail lengths and net charges demonstrated the role of the disordered tail in facilitating DNA search⁹, in agreement with kinetic NMR studies¹⁷. The presence of an N-tail increases the affinity of the protein to the DNA, and therefore enhances its sliding propensity at the expense of hopping and 3D diffusion. However, a higher propensity for sliding has its price: the linear diffusion coefficient of the protein moving along the DNA is lower, which results in a slower search⁹. The disordered tail can also assist intersegment transfer^{18, 19} via the monkey-bar mechanism without accumulation of free protein during the jump from one segment of DNA to another^{9, 10}.

Clearly, PTMs, such as acetylation, which masks the positive charge of Lys, or phosphorylation, which introduces negative charges to Ser, Thr, or Tyr, can reversibly change the electrostatic potential of DBPs and therefore their nonspecific and specific interactions with DNA. Indeed, many DBPs, especially transcription factors, undergo acetylation²⁰ and phosphorylation²¹. Two well known examples of the important role played by a PTM on a disordered tail in modulating interactions with DNA are the modifications on p53 transcription factor C-tail and on the histone tails. The p53 protein is regulated mostly through covalent modifications²² that occur on both the C- and N-tails^{23, 24}. Acetylation events within the C-tail, which is positively charged and strongly interacts with the DNA, significantly enhance site-specific DNA-binding activity and the degree of acetylation is well correlated with p53-mediated activation *in vivo*²⁵. Phosphorylation of the C-tail positively regulates DNA binding and tetramerization of p53²⁶. The histone tails are subject to a variety of PTMs that increase the accessibility of nucleosomal DNA by weakening histone–DNA interactions²⁷⁻³⁰ and consequently influence various cellular processes that depend on the state of the chromatin²⁸.

To quantify the abundance of acetylation and phosphorylation in DBPs and in particular at ordered or disordered sites, we analyzed human DBPs with experimentally characterized PTM sites. We then refined our analysis by investigating the occurrence of acetylation and phosphorylation on disordered tails and illustrated the effects of gradual modifications on DNA search dynamics by coarse-grained molecular dynamics simulations of the Ets-1 transcription factor.

2. Methods

2.1. Statistical analysis of acetylation and phosphorylation in DNA-binding proteins

Datasets of acetylated and phosphorylated human proteins were constructed using PhosphoSite (<http://www.phosphosite.org>). The Uniprot site (<http://www.uniprot.org>)

was used to download the sequences of annotated human proteins and to obtain a subset of DNA-binding proteins (DBPs). Next, the sequences of the proteins from the PhosphoSite datasets were retrieved from the Uniprot sequences. The charges on the protein were calculated by assigning a single positive charge (+1) to each Lys or Arg and a single negative charge (-1) to each Glu or Asp. Acetylated Lys was considered a neutral residue and phosphorylated Ser, Thr, or Tyr were analyzed as negative residues (with a charge unit of -2). Determination of disordered regions was performed using IUPred in the long-range³¹. A protein was considered to have a disordered tail if an unstructured segment of at least five consecutive amino acids was predicted at either its N-terminus or its C-terminus (if the protein was predicted to have unstructured segments at both ends, two separate tails were counted). The ordered region of a protein was defined by excluding the disordered segments from the whole protein sequence. The effects of acetylation and phosphorylation on the composition and distribution of charges on the disordered segments or on the ordered parts of DBPs and non-DBPs can be characterized by various measures that aim to capture the complexity introduced to the protein sequence by the PTMs. In the current study, we focused on two measures: the number of PTMs, calculated as the sum of all PTMs in the tail or globular part and the density of PTMs, calculated as the number of the PTMs normalized by tail length. We emphasize that the influence of the number of each type of PTMs, their location on the protein sequence, and, in particular, the pattern of the charged residues can be further characterized by other parameters.

2.2. Simulation model for searching the DNA of modified proteins

We studied the dynamic nature of DNA search by phosphorylated variants of the Ets-1 protein using a reduced model^{5,9} that allows sampling of long timescale processes, such as sliding, hopping, 3D diffusion, and intersegment transfer. We modeled the DNA as having three beads per nucleotide, representing phosphate, sugar, and base. Each bead was located at the geometric center of the group it represents and a negative point charge was assigned to beads representing the DNA phosphate groups. In the simulations, a 100 bp B-DNA molecule was used to study protein diffusion and the DNA remained in-place and rigid throughout the simulations.

The protein was represented by a single bead for each residue located at the C α of that residue. Beads representing charged amino acids (Lys, Arg, Asp, and Glu) were charged in the model. Non-specific protein-DNA interactions were modeled by electrostatic interactions between all charged residues of the protein and the phosphate bead of the DNA using the Debye-Huckel potential, which accounts for the ionic strength of a solute immersed in aqueous solution⁵. The dynamics of Ets-1 was studied at salt concentrations in the range of 0.01–0.12 M using a dielectric constant of 80 and a temperature at which the protein is completely folded. More details of the simulations can be found in refs. ^{5, 9, 10, 32}. Using this model, we studied the interaction of several phosphorylated variants of Ets-1 in which a phosphorylated residue was modeled as a point charge of -2. The non-specific interactions between the protein and DNA were quantified by classifying each snapshot as performing 1D diffusion (either sliding or hopping) or 3D diffusion.

3. Results

3.1. The occurrence of acetylation and phosphorylation at disordered regions

Intrinsically disordered regions are often found as primary sites for PTMs³³⁻³⁵. Some PTMs are more frequently found at disordered sites than others³⁶ (e.g., acetylation and phosphorylation are abundant in the disordered regions of DBPs^{37, 38}). In this study, we focus on quantifying the existence of acetylation and phosphorylation modifications in DBPs as these PTMs involve changes to the electrostatic charges of the modified sites in a way that may affect interactions with the negatively charged DNA molecules. We found that among 3398 human proteins that undergo acetylation, 65% of the acetylated Lys are located in intrinsically disordered regions. Among the 8943 human proteins that undergo phosphorylation, 79% of the modified Ser, Thr, or Tyr are located in intrinsically disordered regions. The occurrence of acetylation sites in disordered regions is much smaller in non-DBPs (34%), while phosphorylation sites are still widespread in disordered sites (69%) (Fig. 1).

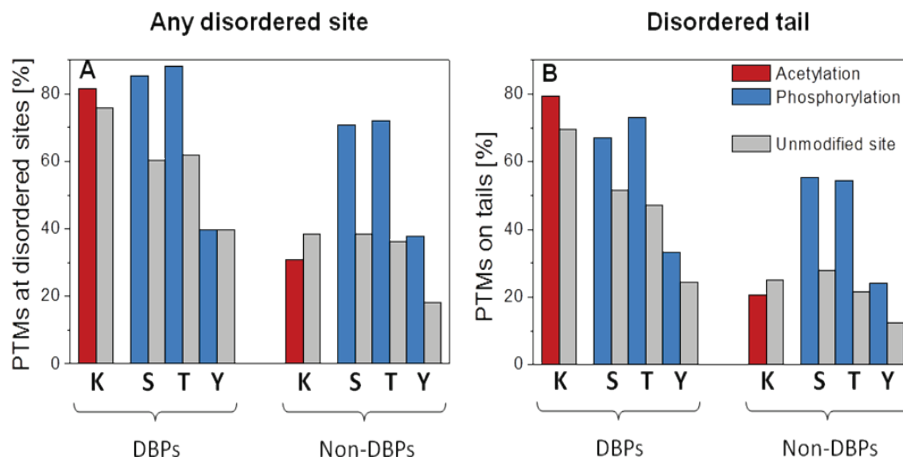


Fig. 1. The occurrence of acetylation and phosphorylation modification on proteins. The occurrence of acetylation (on Lys) and phosphorylation (on Ser, Thr, or Tyr) (shown in red and blue, respectively) post-translational modifications (PTMs) at the disordered sites of proteins (A) and disordered tails (B) is analyzed for two datasets of human acetylated and phosphorylated proteins (including 671 and 1673 proteins, respectively). Each dataset of modified proteins was divided into groups of DNA-binding proteins (DBPs; 81 and 132 acetylated and phosphorylated, respectively) and non DNA-binding proteins (non-DBPs; 590 and 1541 acetylated and phosphorylated, respectively). The values for the occurrence of the corresponding unmodified residues at disordered or tail regions are shown in gray.

To simplify the analysis, we focused on smaller proteins that are composed of <300 amino acids. This dataset includes 671 and 1673 human proteins that undergo acetylation and phosphorylation, respectively. Among the acetylated proteins, 8% are DBPs, which account for 30% of the acetylated Lys in the dataset (517 modified Lys). Among the phosphorylated proteins, 8% are DBPs, which account for 13% of the modified phosphor-sites (298 Ser, 127 Thr, and 196 Tyr) in the dataset. The involvement of acetylation and phosphorylation in proteins that are not DBPs provides further illustration of the diversity of their cellular function in addition to their modulatory interactions with DNA.

Importantly, the majority of the modification sites in the DBPs (about 81% and 77% of the acetylation and phosphorylation sites, respectively) are in disordered regions. In non-DBPs, the opposite trend is observed for acetylation, as only 30% of the acetylation sites in these proteins are in disordered sites. This is in agreement with analysis of the mouse acetylome, which showed the high preference of acetylation sites to locate in structured regions³⁹. This trend was also reported recently in an analysis of acetylation using a dataset of ubiquitinated proteins³⁶. For phosphorylation, whose sites generally show a high preference for disordered regions³, there is a smaller difference between DBPs and non-DBPs. However, a larger fraction of phosphorylation sites is found in disordered sites on DBPs than non-DBPs (63% on DBPs vs 46% on non-DBPs) (Fig. 1).

3.2. The occurrence of acetylation and phosphorylation at disordered tails

In addition to analyzing the occurrence of PTMs at disordered regions, it is of interest to focus specifically on disordered tails (at either the C- or N-termini) because they are quite common in proteins, especially DBPs (Fig. 2). We have previously shown that ~70% of DBPs and only ~50% of non-DBPs in humans possess disordered tails^{10, 32}. The tails of human DBPs are longer than those of other human proteins (~70% of the tails of DBPs are longer than 5 residues while only 40% of all human proteins have tails longer than 5 residues, Fig. 2B). The disordered tails of DBPs have a larger net positive charge than the tails of all human proteins and their charges tend to be much more clustered (Fig. 2C). We note that the composition of positively charged residues and their degree of clustering in the structured parts of DBPs and all human proteins set are very similar³⁷. We have previously reported that the lengths and the net charges of the disordered tail are correlated with the efficiency of the DNA search (e.g., the ability of the tail to promote sliding or jumping between different stretches of DNA).

The analysis of PTM sites on disordered tails (Figs. 1A and 1B) shows that the tails of DBPs are richer in acetylation and phosphorylation sites than the corresponding tails of non-DBPs. The occurrence of acetylation and phosphorylation sites on DBP tails is similar to that in any disordered protein region, indicating that most of the modification sites are located at the tails (Fig. 1). In the non-DBPs, most acetylation and phosphorylation sites are located at structured sites, and when they are located at disordered regions these may be tails or other disordered regions (e.g., internal loops).

Since the interactions of the disordered tails with DNA are governed by the number of positive and negative charges on the tail and their organization, one must also characterize how many acetylation and phosphorylation sites there are in the tails of DBPs and non-DBPs. The average number of acetylation sites is 3.1 ± 0.2 on the tails of DBPs and 0.6 ± 0.1 on the tails of non-DBPs. On the globular parts of DBPs and non-DBPs, the average numbers of acetylation sites are 1.8 ± 0.2 and 1.6 ± 0.2 , respectively (Fig. 3). The trend remains when the number of modification sites is normalized by the length of the domain (or by the disordered tail of the globular domain). Thus, the acetylation density is much higher in the tails of DBPs than in the tails of non-DBPs and the density is lower and very similar in the globular domains of DBPs and non-DBPs (Fig. 3a and 3c (inserts)).

Similarly to acetylation, phosphorylation is ubiquitous on disordered tails and there are DBPs with even 10 phosphorylation sites in the tail. However, the occurrence of phosphorylation in the disordered tails of DBPs is very similar to that in non-DBPs (the average numbers of phosphorylation sites in the tails of DBPs and non-DBPs are 2.0 ± 0.2 and 1.7 ± 0.1 , respectively). The occurrence of phosphorylation sites in the globular part of DBPs and non-DBPs is similar and resembles that found in the tail sub-domains.

Clearly, the number of acetylation and phosphorylation sites in the disordered tail provides only partial information about the potency of the PTMs to modulate function (e.g., to affect the interaction with DNA). It is important to quantify how the modifications communicate with the charged residues of the disordered region. The location of acetylation disrupts charge clustering patterns in a way that not only reduces the net charge but also reduces the local charge density in some regions of the tail more than in other regions.

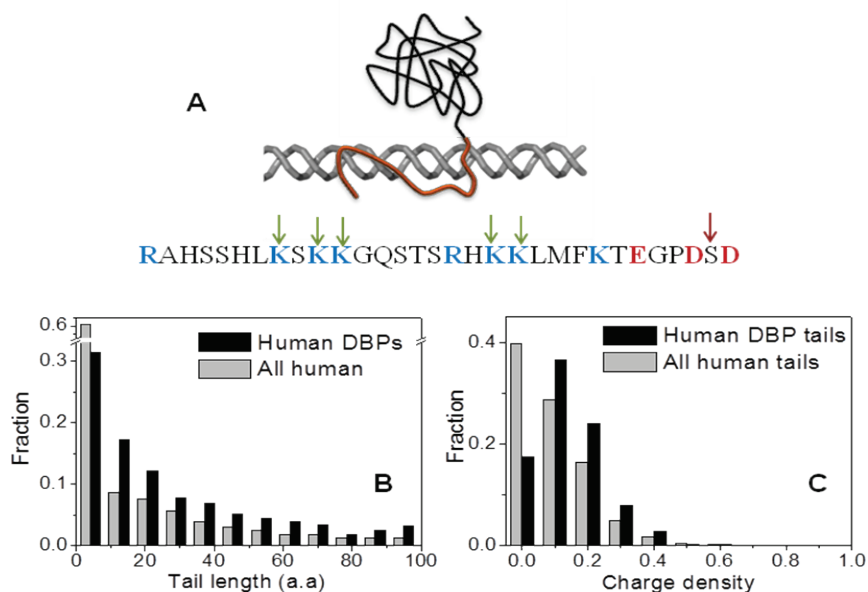


Fig. 2. Analysis of the properties of the disordered tails of DNA-binding proteins in the human genome. A). A schematic illustration of the disordered tail (in orange) of a DNA-binding protein. The tail can interact either non-specifically or specifically with DNA and affect various biophysical properties of the interaction. The sequence of the C-tail of p53 is shown as an example with the positively and negatively charged residues shown in blue and red, respectively. Acetylation sites (green arrows) and phosphorylation sites (red arrow) are shown as well. B). Histograms of tail lengths in human DNA-binding proteins (DBPs; a total of 864 proteins) and in all human proteins (which total 20,334 proteins). A tail is defined as a disordered region predicted by IUPred. C). Distribution of the positive charge density in the tails of DBPs.

To illustrate the importance of the location of the PTMs, we focus on the disordered C-terminus of p53, which is positively charged and thus strongly interacts with DNA and can affect sliding features very significantly. Recent single-molecule experiments⁴⁰ and a coarse-grained simulation study⁴¹ showed that the C-tails of p53 increase its sliding speed along DNA. The C-tail of p53 protein undergoes several covalent modifications^{22, 42} (Fig. 2)^{23, 24}. Dynamic p53 acetylation and deacetylation events were observed in response to DNA

damage²⁵. The acetylation events within the C-tail significantly enhance site-specific DNA-binding activity^{25, 43}. Moreover, C-tail acetylation levels are well correlated with p53-mediated activation *in vivo*²⁵. Phosphorylation of the C-tail positively regulates DNA binding and tetramerization of p53²⁶. The C-tail of p53 can undergo acetylation at 5 different sites and this can affect the number of clusters of consecutive charges and the local effective charge along the tail. The single phosphorylation site on the C-tail is isolated from the acetylation sites and apparently does not interfere with them and its effect is to increase the negative charges at the end of the C-tail.

We propose that acetylation may directly modulate DNA-binding while phosphorylation may indirectly influence attraction to DNA for example, via intra- or inter-molecular protein-protein interactions. The different biological functions of acetylation and phosphorylation may explain the different occurrence of these PTMs in DBPs and non-DBPs. While phosphorylation, which serves in protein function and interactions, is abundant in the tails of all proteins, acetylation, which modulates DNA-binding, is highly pronounced in the tails of DBPs.

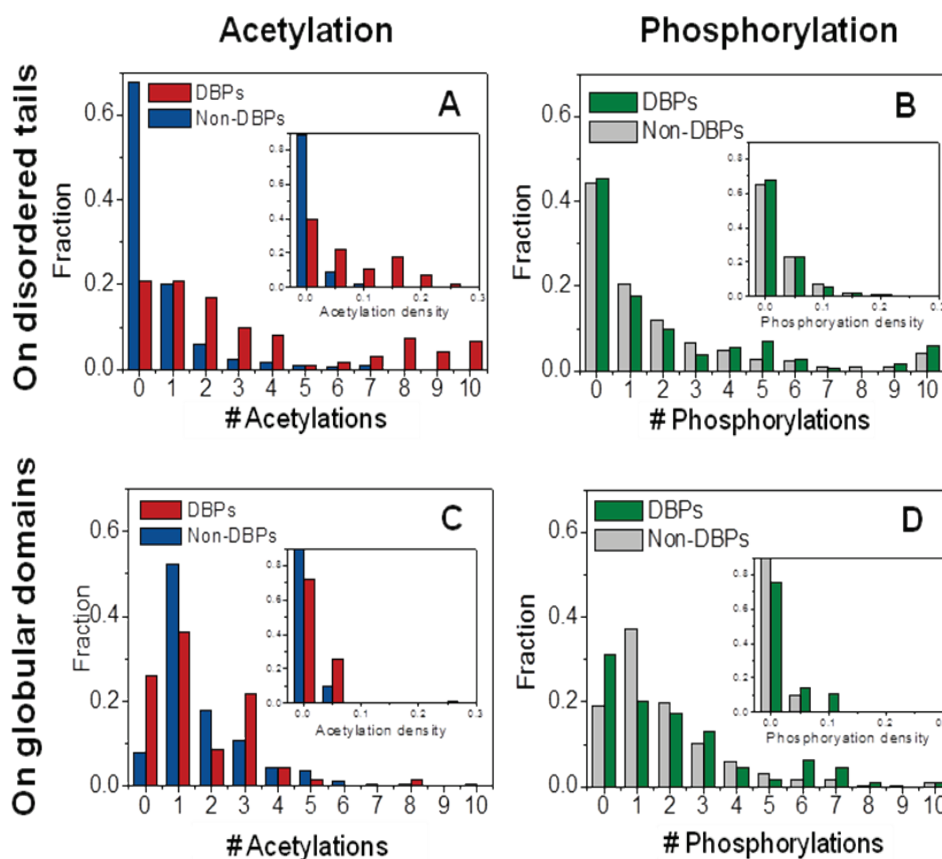


Fig. 3. Bioinformatic analysis of acetylation and phosphorylation on DNA-binding proteins (DBPs). A and B) The number of acetylation (A) and phosphorylation (B) sites on the disordered tails of DBPs (red and green, respectively) and non-DBPs (blue and gray, respectively). The insets indicate the density of acetylation and phosphorylation sites (i.e., the number of sites normalized by the length of the tail sub-domain). C and D). Similar to A and D but for the globular domains of DBPs and non-DBPs, respectively.

3.3. The interplay between post-translational modifications of disordered tails

In many cases, several modifications are needed to achieve the required effect. For example, the 31- and 40-residue long C-tails of p53 and H3 histone tail involve 6 modifications (5 acetylations and a single phosphorylation) and 10 modifications (5 acetylations, 6 methylations, and 3 phosphorylations)⁴⁴, respectively. Various types of cross-talks may come into play between different PTMs. Some modifications require the existence of other modifications. In some cases, two PTMs may compete for the same modification site so that the timing of the effect of one PTM is determined by the removal of the other PTM. In other cases, the effects of multiple PTMs accumulate when the level of modification is incrementally changed. The latter scenario suggests that each PTM may not be sufficient to achieve the needed outcome alone but that several PTMs of the same type gradually achieve the required outcome.

A gradual change by successive PTMs was reported for the Ets-1 transcription factor, whereby gradual phosphorylation of its disordered tail results in gradual attenuation of its binding affinity to DNA. This ability to gradually tune the binding affinity was viewed as an "incremental rheostat" rather than as a switch on/off by the modifications⁴⁵. It was shown for Ets-1 that phosphorylation of its disorder tail interferes with formation of intramolecular interactions, which results in a lower binding affinity to DNA^{38, 45}. The five phosphorylation events on the disordered tail of Ets-1 were reported to attenuate affinity to DNA not by directly affecting its interactions with DNA but by modulating its internal flexibility. Nevertheless, it is likely that the phosphorylated tail interacts directly with non-specific DNA. Here, we followed this scenario and studied the non-specific interactions of the Ets-1 transcription factor with a disordered tail in which different numbers of phosphorylation sites were phosphorylated.

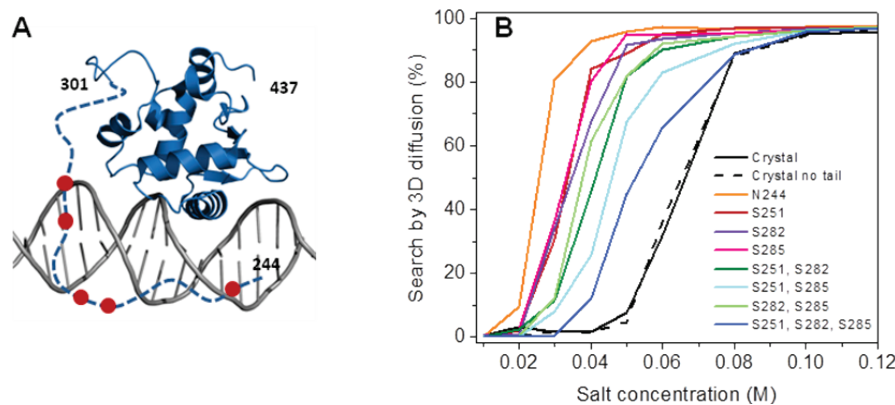


Fig. 4. The effect of incremental phosphorylation on non-specific protein-DNA interactions. A). The Ets-1 transcription factor includes a long disordered region with five phosphorylation sites that were found to attenuate specific binding to DNA indirectly by affecting internal protein dynamics so as to weaken specific affinity. The disordered tail can also interact directly with DNA and the strength of the interaction might be dependent on the degree of phosphorylation. B). The populations of 3D diffusion for 10 variants of Ets-1 as a function of salt concentration indicate a clear linkage between the number of phosphorylations and the salt concentration needed to detach the protein from the DNA to the bulk. As phosphorylation sites are mutated, a higher salt concentration is needed to populate the 3D search (at the expense of linear diffusion).

Phosphorylation of the disordered tail of the Ets-1 transcription factor results in weaker non-specific binding to DNA. This weaker binding is reflected in the lower salt concentration needed to detach the protein from the DNA and to switch from DNA search via linear diffusion to 3D diffusion (Fig. 4). The attraction of the folded part of Ets-1 to the DNA is gradually attenuated as the level of phosphorylation increases (namely, gradually mutating the phosphorylation sites results in search characteristics that are more similar to that of Ets-1 with a truncated disordered tail). Interestingly, similarly to the gradual and indirect effect of phosphorylation on specific binding that arises from modulating the internal protein dynamics, our results suggest that phosphorylating the disordered tail may directly perturb nonspecific binding to DNA.

4. Conclusions

PTMs, which are widely used by the cell to increase the structural and biochemical diversity of proteins, often take place at intrinsically disordered sites. These disordered regions, particularly at the protein termini (namely, tails), are more abundant in DNA-binding proteins than in other proteins and serve as potential sites for PTMs. In this study, we analyzed the occurrence of acetylation and phosphorylation in human DBPs in comparison to non-DBPs at disordered versus ordered sites. There is a strong preference for acetylation to occur at either disordered sites or in the tail regions of DBPs while in non-DBPs an opposite trend is seen in which acetylation sites are mostly located in structured regions. With respect to phosphorylation, there is general tendency to find phosphorylated Ser and Thr in disordered regions, irrespective of whether the protein binds DNA or not. However, the occurrence of phosphorylation sites in disordered sites is still higher for DBPs than for non-DBPs. While both PTMs have similar effect of reducing the net charge, phosphorylation is known to be important in protein activation and protein-protein interaction and acetylation is thought to modulate DNA binding such as serving as primary device of basic chromatin modification leading to transcription activation. The difference between the two PTMs may explain the widespread occurrence of phosphorylation in the disordered regions of all human proteins, and the high occurrence of acetylation sites mainly in tails of DBPs, where the acetylation primarily affect the protein-DNA binding.

Both acetylation and phosphorylation decrease the number of positive charges (by masking the positive charge of Lys in acetylation and introducing a negative charge on either Ser, Thr, or Tyr in phosphorylation). Clearly, tuning the positive potential of the protein will affect association with DNA. This tuning can be achieved in principle on either the ordered or disordered subdomains of DBPs. The observation that most acetylation and phosphorylation sites in DBPs are located in the tails provides additional evidence for the role of the tail as an affinity tuner for interactions with DNA that can be modulated by PTMs³⁷.

PTMs may exert a complex effect when they occur at multiple sites. In some cases, the effects of PTMs accumulate additively to regulate an outcome in a graded manner (e.g., the regulation of DNA-binding affinity in a progressive manner by multiple

phosphorylations of the Ets-1 transcription protein⁴⁵). One can easily envision cases in which the effect of multiple PTMs at disordered region will not be simply additive but will depend on the local environment of each PTM.

Acetylation and phosphorylation affect the composition of the charges of the disordered tail. The global and local changes in the net charge of the disordered tail may directly affect not only its attraction to nonspecific DNA, but also the internal properties of the tail itself (such as its conformational preferences). Recent findings call for a clear linkage between the degree of structure of intrinsically disordered proteins and the charge content^{10, 46-48}. The effect of the PTMs on the conformational ensemble of the disordered region may depend both on the net charges and on how they communicate with the original charges in the sequence and thus the local effective charges. Our future goals are to characterize the effect of the PTMs on the charge pattern of the disordered tails and to analyze the evolutionary conservation of the PTMs sites on tails.

In summary, the disordered regions and particularly disordered tails of DBPs have important features that allow them to modulate interactions with DNA. Some of these features can be manipulated in an incremental manner by PTMs that gradually shift the composition and distribution of charges and consequently fine tune the strength of protein–DNA interactions but in others the cross-talks between the different PTMs sites might be more complex.

5. Acknowledgments

This work was supported by the Kimmelman Center for Macromolecular Assemblies and the Minerva Foundation with funding from the Federal German Ministry for Education and Research. Y.L. is the incumbent of the Lillian and George Lyttle Career Development Chair. We thank Tzachi Hagai for the inspiration of studying acetylation pattern in DBPs.

- 1 H. Xie, S. Vucetic, L. M. Iakoucheva, C. J. Oldfield, A. K. Dunker, Z. Obradovic and V. N. Uversky, *Journal of proteome research*, 2007, **6**, 1917-1932.
- 2 V. N. Uversky, C. J. Oldfield, U. Midic, H. Xie, B. Xue, S. Vucetic, L. M. Iakoucheva, Z. Obradovic and A. K. Dunker, *BMC genomics*, 2009, **10 Suppl 1**, S7.
- 3 C. R. Landry, E. D. Levy and S. W. Michnick, *Trends Genet*, 2009, **25**, 193-197.
- 4 L. M. Iakoucheva, P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic and A. K. Dunker, *Nucleic acids research*, 2004, **32**, 1037-1049.
- 5 O. Givaty and Y. Levy, *Journal of the molecular biology*, 2009, **385**, 1087-1097.
- 6 A. Marcovitz and Y. Levy, *Proc Natl Acad Sci USA*, *In press*
- 7 A. I. Dragan, Z. L. Li, E. N. Makeyeva, E. I. Milgotina, Y. Y. Liu, C. Crane-Robinson and P. L. Privalov, *Biochemistry*, 2006, **45**, 141-151.
- 8 C. Crane-Robinson, A. I. Dragan and P. L. Privalov, *Trends in Bioch Sci*, 2006, **31**, 547-552.
- 9 D. Vuzman, A. Azia and Y. Levy, *J Mol Biol*, 2010, **396**, 674-684.
- 10 D. Vuzman and Y. Levy, *Proc Natl Acad Sci USA*, 2010, **107**, 21004-21009.
- 11 A. Tóth-Petróczy, M. Fuxreiter and Y. Levy, *J Amer Chem Soc*, 2009, **131**, 15084-15085
- 12 P. L. Privalov, A. I. Dragan, C. Crane-Robinson, K. J. Breslauer, D. P. Remeta and C. A. S. A. Minetti, *J Mol Biol*, 2007, **365**, 1-9.

- 13 J. M. Gruschus, D. H. Tsao, L. H. Wang, M. Nirenberg and J. A. Ferretti, *Biochemistry*, 1997, **36**, 5372-5380.
- 14 R. Joshi, J. M. Passner, R. Rohs, R. Jain, A. Sosinsky, M. A. Crickmore, V. Jacob, A. K. Aggarwal, B. Honig and R. S. Mann, *Cell*, 2007, **131**, 530-543.
- 15 P. L. Privalov, A. I. Dragan, C. Crane-Robinson, K. J. Breslauer, D. P. Remeta and C. A. Minetti, *J Mol Biol*, 2007, **365**, 1-9.
- 16 A. I. Dragan, Z. Li, E. N. Makeyeva, E. I. Milgotina, Y. Liu, C. Crane-Robinson and P. L. Privalov, *Biochemistry*, 2006, **45**, 141-151.
- 17 J. Iwahara and G. M. Clore, *J Am Chem Soc*, 2006, **128**, 404-405.
- 18 T. Hu and B. I. Shklovskii, *Physical Review E*, 2007, **76**, 051909.
- 19 M. Doucleff and G. M. Clore, *Proc Natl Acad Sci USA*, 2008, **105**, 13871-13876.
- 20 A. J. Bannister and E. A. Miska, *Cell Mol Life Sci*, 2000, **57**, 1184-1192.
- 21 A. J. Whitmarsh and R. J. Davis, *Cell Mol Life Sci*, 2000, **57**, 1172-1183.
- 22 A. M. Bode and Z. G. Dong, *Nature Reviews Cancer*, 2004, **4**, 793-805.
- 23 E. Appella and C. W. Anderson, *European Journal of Biochemistry*, 2001, **268**, 2764-2772.
- 24 X. Yang, *Cell Death and Differentiation*, 2003, **10**, 400-403.
- 25 J. Y. Luo, M. Y. Li, Y. Tang, M. Laszkowska, R. G. Roeder and W. Gu, *Proc Natl Acad Sci USA*, 2004, **101**, 2259-2264.
- 26 K. Sakaguchi, H. Sakamoto, D. Xie, J. W. Erickson, M. S. Lewis, C. W. Anderson and E. Appella, *J Prot Chem* 1997, **16**, 553-556.
- 27 A. Eberhardter and P. B. Becker, *Embo Reports*, 2002, **3**, 224-229.
- 28 S. Y. Roth, J. M. Denu and C. D. Allis, *Annual Review of Biochemistry*, 2001, **70**, 81-120.
- 29 Z. Y. Yang, C. Y. Zheng and J. J. Hayes, *Journal of Biological Chemistry*, 2007, **282**, 7930-7938.
- 30 B. M. Turner, *Bioessays*, 2000, **22**, 836-845.
- 31 Z. Dosztanyi, V. Csizmok, P. Tompa and I. Simon, *Bioinformatics*, 2005, **21**, 3433-3434.
- 32 D. Vuzman, M. Polonsky and Y. Levy, *Biophys J*, 2010, **99**, 1202-1211.
- 33 P. Tompa, *Structure and function of intrinsically disordered proteins*, Boca Raton, 2010.
- 34 A. K. Dunker, I. Silman, V. N. Uversky and J. L. Sussman, *Curr Opin Struct Biol*, 2008, **18**, 756-764.
- 35 V. N. Uversky, C. J. Oldfield and A. K. Dunker, *Annual Rev of Biophysics*, 2008, **37**, 215-246.
- 36 T. Hagai, A. Azia, A. Tóth-Petróczy and Y. Levy, *J Mol Biol*, 2011, 412, 319-324, (2011)
- 37 D. Vuzman and Y. Levy, *Submitted*, Mol BioSystems, In press.
- 38 M. Fuxreiter, I. Simon and S. Bondos, *Trends in Bioch Sci*, 36, 415-423, (2011)
- 39 C. Choudhary, C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen and M. Mann, *Science*, 2009, **325**, 834-840.
- 40 A. R. Fersht, A. Tafvizi, F. Huang, L. A. Mirny and A. M. van Oijen, *Proc Natl Acad Sci USA*, 2011, **108**, 563-568.
- 41 N. Khazanov and Y. Levy, *J Mol Biol*, 2011, **408**, 335-355.
- 42 L. Jayaraman and C. Prives, *Cellular and Molecular Life Sciences*, 1999, **55**, 76-87.
- 43 W. Gu and R. G. Roeder, *Cell*, 1997, **90**, 595-606.
- 44 S. J. Nowak and V. G. Corces, *Trends Genet*, 2004, **20**, 214-220.
- 45 M. A. Pufall, G. M. Lee, M. L. Nelson, H. S. Kang, A. Velyvis, L. E. Kay, L. P. McIntosh and B. J. Graves, *Science*, 2005, **309**, 142-145.
- 46 A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine and R. V. Pappu, *Proc Natl Acad Sci USA*, 2010, **107**, 8183-8188.
- 47 M. Borg, T. Mittag, T. Pawson, M. Tyers, J. D. Forman-Kay and H. S. Chan, *Proc Natl Acad Sci USA*, 2007, **104**, 9650-9655.
- 48 D. Potoyan and G. Papoian, *J Amer Chem Soc*, 2011, **133**, 7405-7415.