Understanding evolutionary dynamics of phosphorylation with fungal species as model system

Wenfa Ng

Novena, Singapore, Email: ngwenfa771@hotmail.com

Abstract

Movement in proteins requires energy. To this end, natural selection has selected phosphate as the principal energy currency in cells. On the other hand, depicting the state of transcription could come in the form of epigenetic markers, which are modifications on nucleotide residues. But, what are the deeper evolutionary forces that underpin the selection of phosphorylation as a key process for translating molecular information of cellular state into specific phenotype in movement and metabolism at the cellular level? From another perspective, what are the factors that guide the selection of particular phosphosites as principal phosphorylation sites? Seeking answers to the latter question, Villen and coworkers ("Evolution of protein phosphorylation across 18 fungal species", *Science*, Link) used mass spectrometry to profile the phosphoproteome of 18 fungal species and employed discovery science approaches to elucidate specific phosphosite highly conserved for particular functions such as transcription and translation. Using histone protein (H2) and transcription initiator factor (eIF4E) as model proteins for gaining a deeper understanding of the evolutionary forces that shape the annotation of specific phosphosite (from a large library of possible phosphosites) as key molecular effectors of cellular processes such as conformational changes in enzymes and ion channels. Results obtained suggests possible selection forces that define particular phosphosite for function, which are corroborated through assessing kinase motif usage and biochemical assays for the binding affinity between peptide libraries and cell lysates. Looking at a broader landscape of protein phosphorylation, the paper, however, does not yield sufficient insights to answer questions such as how protein phosphorylation first emerged as a defining mechanism for translating stored cellular energy in phosphate groups into movement necessary for protein function and, by extension, that of the cell. Understanding the evolutionary processes that first potentiated protein phosphorylation as well as the specific natural selection factors that resulted in the definition of specific phosphosite for particular function are fundamental questions important to understanding how biology utilizes chemical and physical principles for powering life.

Keywords: epigenetics, phosphorylation, phosphosite, protein, DNA, genetic marker,

Subject areas: molecular biology, biochemistry, cell biology, genetics, computational biology,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2966v1 | CC BY 4.0 Open Access | rec: 6 May 2017, publ: 6 May 2017

Perspective

Phosphorylation plays important roles in many aspects of cell biology and intra- and intercellular signaling; for example, serving as markers for epigenetic regulation of gene transcription, and more commonly, as a repository of energy for enzymatic and protein function. But, what determines the selection of specific protein sequence motif as sites for inserting a phosphorus atom, and what does this selection suggests of the evolutionary processes that underpin it? More deeply, what comes first? Specifically, was a library of nucleotide and amino acid sites suitable for inserting phosphate groups generated by nature, and later, served as molecular backbone for carrying stored energy or information? Another angle for analyzing the same issue posits the definition of a set of environmental conditions that allow the progressive selection, through successive generations of cells, of a specific nucleotide or amino acid residue that, upon phosphorylation, endow an important function to the cell.

Phosphate groups can be placed either at the DNA or protein level, each with differing physiological implications for the cell. For example, phosphate groups serve as epigenetic markers on DNA and histone proteins, with important roles in regulating the permissiveness of specific segments of the DNA molecule for transcription. Specifically, phosphorylation changes the structural backbone of the DNA molecule, on a dynamic timescale suited for periodic controlled movement, necessary for allowing transient attachment of transcription factors needed for initiation, elongation and termination phase of transcription. On the other hand, protein phosphorylation endows the protein (whether a structural protein or an enzyme) an important energy source or marker for structural modification necessary for carrying out a function. Specifically, phosphorylation and dephosphorylation helps bring about movement important for ligand binding or enzymatic catalysis.

In considering the evolutionary processes that brought about the development of phosphorylation at the protein level in cells, two perspectives are dichotomously different in analyzing this question. Firstly, do the relative conservation of the phosphosite suggests possible use for phylogenetic classification of different species? If yes, at what level is the conservation of phosphosite useful for understanding species provenance: DNA or protein, given the differing roles and importance of phosphorylation at the nucleotide and amino acid level? From another perspective, a different but pertinent question of interest concerns the evolutionary and natural selection forces that shape the selection of specific phosphosite over others? At a more granular level, this highlights deeper biochemical and biophysical aspects that afford a phosphosite greater evolutionary significance compared to other nearby sites; for example, propensity to binding to kinase motif and extent in which a phosphosite confer desired function to a protein.

Attempting to answer the second question in a discovery science paper, Villen and coworkers ("Evolution of protein phosphorylation across 18 fungal species", *Science*, Vol. 354, Issue 6309,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2966v1 | CC BY 4.0 Open Access | rec: 6 May 2017, publ: 6 May 2017

pp. 229 – 232, Link), used mass spectrometry to profile the phosphoproteome of 18 fungal species ranging from *Saccharomyces cerevisiae* to *K. lactis* and others. Looking for phosphosites, where phosphate groups are added or removed, the mass spectrometry data highlights specific phosphosites that enjoy high degree of conservation; for example, at protein interfaces necessary for protein-protein interactions.[1] Subsequent assessment of the extent of binding affinity between kinase motif and profiled phosphosite gave an understanding of the functional importance of individual phosphosite assayed. Finally, biochemical assays were employed to provide a view towards how different peptide libraries correlate with kinase activity level.

Overall, the research provides fresh details on the relative extent in which different phosphosites are conserved in cells. However, a holistic assessment of the research presented reveals that it does not answer the deep evolutionary question posed in the title of the paper. Specifically, while understanding the types and structural motifs of the different phosphosite helps provide a grounding of the relative importance of different phosphosite and their structural significance, they do not, on the whole, gave an understanding of the evolutionary forces that first engender phosphorylation in the first place. Secondly, in using H2 histone protein and eIF4E as model protein system for a systematic analysis of the presence of different phosphosites and their functional importance, coverage is only extended to the DNA replication and transcription processes using single representative protein important to the respective processes, which are helped by many accessory factors that give rise, as a unit, to observed phenomenological evolutionary underpinnings. In essence, by enlisting a larger set of proteins that each carry out differentiated functions in metabolism or signaling, a better understanding could be gleaned concerning the role of natural selection in enabling specific phosphosite to gain functional prominence over others in the cell's repertoire of phosphorylation sites useful for myriad cellular processes.

Hence, certain circumstances in the distant past must have laid down the environmental conditions that facilitate the selection of mutations that confer phosphorylation to the cell's functional toolkit. With the ability to store energy currency in phosphate groups for later use such as in enabling movement of different protein domains relative to each other, or as a marker for specific instructions to higher levels of biological function such as transcription and translation, phosphorylation is a foundational biological process whose secrets, particularly at the evolutionary and phylogenetic level, awaits elucidation. A basis question would be: could understanding phosphorylation help classify different species? The answer is likely no, given lack of diversity at the motif and sequence level since many of the phosphosites are well conserved across the different branches of the tree of life. On the other hand, do we know enough of the structural motif and sequence underpinnings that point to functional significance and evolutionary importance of particular phosphosite at the cellular level? Perhaps, more research is needed to provide a fine level understanding of the factors that potentiate the evolutionary conservation of specific phosphosite.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2966v1 | CC BY 4.0 Open Access | rec: 6 May 2017, publ: 6 May 2017

Another consideration is the accuracy and mass resolution of contemporary state of the art electrospray ionization time of flight mass spectrometry (ESI-MS) as well as matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS). Specifically, higher mass resolution and throughput are needed to allow systematic analysis of phosphosites at the cellular level across a larger population of cells. Doing so would allow more representative understanding of the heterogeneity of phosphosite in cells of specific species, knowledge of which facilitates a more thorough statistical analysis of relative abundance of different phosphosites in various proteins; thus, providing a more complete, but necessarily mixed and complicated picture of phosphorylation at the cellular level. Holistically, understanding such a high dimensional dataset for single cell represents a significant challenge in developing the requisite binning and classification algorithms for first sorting the diverse mass spectrometry data of varying statistical significance and error, as well as creative depiction of multi-dimensional data for easy visual understanding and interpretation. But, what about the data challenge at the population level where a large group of cells provide the most complete depiction of how phosphorylation works at the cellular level? Understanding the underlying working principles of phosphorylation would hopefully lend a lens to a fundamental question of biology: what role does evolution play in selecting phosphorylation as an energy storage mechanism for molecules, as well as an information repository for identifying specific proteins, the latter important for functional screening of proteins in other pathway steps?

Finally, an important point not clearly delineated or tackled in the paper concerns the apportionment of physiological significance to principal protein effectors of specific processes such as transcription. For a process as complicated as transcription, which is helped and aided by multiple accessory factors and protein subunits each playing nuanced roles in a choreographed dance for enabling an observed function, determining the evolutionary provenance of a specific phosphosite in a protein subunit relative to another closely-related one require a significant leap in our current understanding of why evolution chooses specific factors for endowing phosphorylation function. In essence, the question, reframed from the biochemical and biophysical perspectives is: what are the physical attributes that afford the selection of certain protein subunits or accessory factors as sites for storing cellular energy through phosphorylation? Cutting to the fundamentals of the issue, in a multiple effector process such as DNA replication and transcription, does the choice of protein for storing cellular energy and information through phosphosites affect the overall efficiency and effectiveness of the process, for example, on transcription error rates? And, more importantly, why and how evolution chooses specific proteins for encoding cellular information and storing energy through phosphorylation? Solutions to these questions and others will be sought by the academic community and could serve as a case example of how evolution works at the protein level in complex processes enabled by multiple molecular effectors.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2966v1 | CC BY 4.0 Open Access | rec: 6 May 2017, publ: 6 May 2017

Reference

1. Studer, R. A. *et al.* Evolution of protein phosphorylation across 18 fungal species. *Science* **354,** 229 (2016).

Conflicts of interest

The author declares no conflicts of interest.

Author's contribution

The author read the cited paper and thought deeply about the evolutionary forces that first give rise to phosphorylation as a method for storing cellular energy at the protein level for enabling movement in multi-domain proteins, as well as serving as a marker (i.e., information storage) for identifying specific proteins important for downstream cellular processes. He wrote the preprint.

Funding

No funding was used in this work.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2966v1 | CC BY 4.0 Open Access | rec: 6 May 2017, publ: 6 May 2017