Published in final edited form as:
Dev Cell. 2011 July 19; 21(1): 32–34. doi:10.1016/j.devcel.2011.06.004.

Protein Evolution in Cell and Tissue Development: Going Beyond Sequence and Transcriptional Analysis

Daniel J. Dickinson[1], **William I. Weis**[1,2,3,*], and **W. James Nelson**[1,2,4,*]

[1]Program in Cancer Biology, Stanford University, Stanford CA 94305, USA

[2]Departments of Molecular & Cellular Physiology, Stanford University, Stanford CA 94305, USA

[3]Structural Biology, Stanford University, Stanford CA 94305, USA

[4]Biology, Stanford University, Stanford CA 94305, USA

Abstract

Studies of animal evolution often focus on sequence and transcriptional analysis, based on an assumption that the evolution of development is driven by changes in gene expression. We argue that biochemical and cell biological approaches are also required, because sequence-conserved proteins can have different biochemical, cellular and developmental properties.

Introduction

A major goal of evolutionary and developmental biology is to understand how changes in DNA sequence and gene expression result in different cellular and developmental outcomes. Historically, developmental biology has focused on a few model organisms (mouse, fly, worm, sea urchin, zebrafish), whereas evolutionary biology has taken a broader view by comparing many more species to determine how phenotypes change over time. More recently, development has been studied in the context of its evolution over time ("evo-devo"), using the rapidly expanding number of sequenced genomes from diverse organisms.

A surprising finding from early work in the evo-devo field was that many genes important for development have homologs in a wide variety of animals. The fact that so-called "toolkit" genes can be found in many species led to the view that transcriptional regulation of protein machinery is more significant for evolution than changes to how the machinery itself works at the cellular level (Carroll, 2008). Thus there has been a strong focus on understanding gene regulatory networks, in which transcription factors control expression of target genes in the context of a complex developmental process. Also, in developmental biology, focusing on a small number of genetically tractable organisms is justified in part by the assumption that developmental mechanisms and transcriptional programs elucidated in one organism should be directly relevant across a broad range of animal species.

Although the idea that gene regulatory networks control the expression of common sets of functionally-conserved cytoplasmic proteins in all animals is appealing, very few studies have experimentally tested whether "toolkit" proteins actually have conserved functions in different organisms. On the other hand, there is considerable evidence that not all of the cellular processes that drive development are conserved among species. Genetic approaches in a number of organisms have identified proteins that are clearly essential for normal development, but lack homologs outside of closely-related species [e.g. *Drosophila* Nullo

*Correspondence: W. James Nelson wjnelson@stanford.edu William I. Weis bill.weis@stanford.edu.

(Hunter and Wieschaus, 2000)]. In other cases, proteins which are essential for development in most animals have been lost in certain clades [e.g, Dkk proteins, modulators of Wnt signaling in vertebrates and cnidarians, are absent in ecdysozoans (Guder et al., 2006)]. It is widely accepted that conserved pathways can gain or lose regulatory inputs over the course of evolution; the loss of Dkk in ecdysozoans is one example. However, even proteins that have obvious homologs in all animals can have different biochemical properties and cellular activities in different organisms (see below for examples).

Although sequence homology between two proteins can sometimes be indicative of conserved function (for example in metabolic enzymes), the relationship between sequence conservation, biochemical similarity and developmental function is not simple. For example, vertebrate and invertebrate E-cadherins have substantially different domain architectures and mediate cell-cell adhesion using a different molecular interface, but they appear to have similar roles in development (Shapiro and Weis, 2009). On the other hand, vertebrate E- and N-cadherins have >60% sequence similarity and mediate cell-cell adhesion using very similar molecular interfaces, but they have different developmental functions and cannot substitute for one another (Kan et al., 2007).

The lack of strict functional conservation of individual proteins becomes even more pronounced at the cellular and developmental levels. For example, the cellular events associated with Hedgehog signaling appear to differ between vertebrates and protostomes: Hedgehog signaling requires primary cilia in mouse but not in *Drosophila* (Wilson and Chuang, 2010). At the organismal level, significant differences in developmental mechanisms can be found even between relatively closely-related organisms: the establishment of a segmented body plan requires Notch signaling in short germband insects such as the flour beetle *Tribolium castaneum*, but not in long germband insects including *Drosophila* (Damen, 2007). It is usually impossible to infer by comparing a small number of species whether a particular developmental mechanism is ancestral (and thus may be widely conserved) or derived.

Nevertheless, there are also clear cases in which conserved proteins function similarly in distantly-related organisms. In both *Drosophila* and *C. elegans*, pulsed contraction and relaxation cycles by actomyosin networks are required for many key steps in development (Kasza and Zallen, 2011). This behavior is probably an intrinsic property of the myosin II mechano-chemical cycle, whose basic kinetic properties are conserved between species. Since this occurs in different developmental contexts in these two organisms, pulsed actomyosin contraction appears to be a *bona fide* example of a common cellular mechanism which is deployed in response to different upstream signals in different species.

These examples show that it is not always possible to infer from sequence homology or expression pattern alone whether the biochemical or developmental properties of a particular protein will be conserved in a given species. Thus, such conservation should be treated not as an assumption, but as a hypothesis to be tested. The availability of genome sequence data provides an opportunity to experimentally address how protein functions have evolved over time and to understand the developmental significance of these changes. We offer several examples of experimental approaches that provide a paradigm for accomplishing these goals.

Biochemical and interaction studies

α-Catenin and vinculin are paralogous actin-binding proteins involved in cell adhesion in animals. In mammalian cells, vinculin interacts with talin and integrins at sites of cell-matrix adhesion, and α-catenin forms a complex with β-catenin and cadherin at cell-cell contacts. Whereas all animals have orthologs of both α-catenin and vinculin, the social amoeba

Dictyostelium discoideum has only a single member of the protein family. We characterized this protein *in vitro* and *in vivo* to determine whether it has biochemical and cellular properties of α-catenin and/or vinculin (Dickinson et al., 2011). The data showed that *D. discoideum* α-catenin is biochemically and functionally similar to metazoan α-catenin, and not to vinculin. Thus, metazoan α-catenin has retained properties that were present in the ancestor of this protein family, whereas vinculin has acquired novel properties that allow it to function in cell-matrix adhesion. These conclusions could not have been reached based on protein sequence analysis alone.

In a similar set of experiments we also characterized HMP-1, which is the *C. elegans* ortholog of α-catenin (Kwiatkowski et al., 2010). The results of that study, taken together with our results from *D. discoideum* (Dickinson et al., 2011), reveal an interesting and nuanced picture of α-catenin evolution. All three α-catenins bind β-catenin, influence actin organization and are necessary for epithelial cell polarity *in vivo*, suggesting that these functions are ancient and highly conserved. However, actin binding is regulated differently for each of these three α-catenins: actin binding appears to be constitutive in *D. discoideum*, autoinhibited in *C. elegans*, and autoinhibited but activated by α-catenin homodimerization in mammals. Thus it appears that α-catenin has acquired additional modes of regulation over the course of evolution, while retaining its core functions. Importantly, the observed biochemical differences between α-catenins from different species do not obviously correlate with any particular sequence features, and could only be discovered performing experiments directly on purified proteins.

As this example illustrates, *in vitro* biochemical experiments can reveal how proteins have acquired novel functions over the course of evolution. Importantly, because recombinant protein expression requires only knowledge of the sequence of a particular protein, biochemical analysis is useful even for proteins from organisms that cannot be manipulated in a laboratory. Thus, biochemistry allows us to compare protein family members across a wide range of clades and species to determine which properties are broadly conserved, and which are species- or clade-specific.

A complementary approach is the large-scale interrogation of protein-protein interaction networks. In principle, comparing the interaction partners of a protein in several species could reveal differences whose physiological and evolutionary significance could be further investigated. Such an approach is presently limited by the amount and quality of protein-protein interaction data and by the fact that the assays used in large-scale studies cannot yield quantitative information about the affinities and rates of binding, which are potentially important to understanding functional outcomes. Nevertheless, large-scale studies will certainly be important for understanding the evolution of protein-protein interaction networks in the future.

Comparative studies in established model systems

Animal epithelial cells exhibit apical-basal polarity, in which the apical plasma membrane faces the lumen of the organ and the basolateral membrane contacts neighboring cells and the underlying extracellular matrix. This polarized organization is maintained in part by a number of “polarity proteins” (St Johnston and Ahringer, 2010), which have a complex set of interactions that maintain distinct apical and basolateral plasma membrane domains. Comparisons of mammalian and invertebrate epithelia have raised interesting questions about how this network of interactions evolved.

One important difference in epithelial organization between mammals and invertebrates is the positioning of cell-cell adhesion complexes (St Johnston and Ahringer, 2010). Whereas in *Drosophila* and *C. elegans* the adherens junction is located at the apical edge of cell-cell

contacts and defines the boundary between apical and lateral membranes, in mammalian cells the adherens junction is located more basally and the apical/lateral boundary is defined by the tight junction. The polarity protein network has evolved along with the different junctional morphologies in these organisms. For example, in mammals the Crumbs protein complex localizes to the tight junction and is essential for tight junction formation (St Johnston and Ahringer, 2010), but Crumbs does not appear to be essential for cell-cell junctions in *Drosophila* (Campbell et al., 2009). Thus, the mammalian Crumbs complex has gained the ability to interact with tight junction proteins. It is unknown whether these novel interactions affect other properties of Crumbs, including its binding to other polarity proteins.

Another well-studied polarity protein is Par-3 (called *Bazooka* in *Drosophila*). In insect epithelial cells Par-3 colocalizes with the adherens junction and appears to define a distinct "sub-apical" plasma membrane domain (Morais-de-Sá et al., 2010). The sub-apical localization of Par-3 depends upon two factors: phosphorylation of Par-3 by aPKC, and competition with Crumbs for binding to more apically-localized Par-6 (Morais-de-Sá et al., 2010). Par-3 is also phosphorylated by aPKC in mammalian cells (Morais-de-Sá et al., 2010), but this phosphorylation may have a different outcome given the different cellular role of mammalian Crumbs (see above). More generally, insect Par-3 is localized by a complex set of interactions between proteins whose abundances and binding affinities may be different in different species. To understand how Par-3 and its binding partners accommodate the different junctional morphology in vertebrates, it will be necessary to determine affinities of the relevant protein-protein interactions and to develop techniques that allow measurement of *in vivo* protein concentrations in different model systems.

Developmental mechanisms in new model organisms

Although experiments *in vitro* and in model organisms provide a starting point for characterizing the evolution of a protein's properties and activities, a deep understanding of how protein functions and developmental mechanisms have changed over time will require *in vivo* studies using a new generation of model organisms, rather than the current focus on a few established systems. A general principle is that the more conserved a protein, the broader the range of species that must be examined to understand its evolution.

For rapidly evolving proteins or traits, the most useful insights may be gained by comparing very closely related species, an approach that has been widely used in recent years (Carroll, 2008). However, proteins that are more broadly conserved will require detailed studies of a wider variety of organisms, including non-bilaterians. The advantages of studying early-diverging animals are twofold. First, protein functions that are conserved in bilaterians and non-bilaterians can be robustly inferred to have been present in the common ancestor of all animals, while properties that are unique to one clade or another can be identified. Second, since non-bilaterians generally have fewer tissues types and a simpler developmental program, they offer a simpler system in which to study basic cell biological processes that may be important in higher organisms. A promising non-bilaterian model system is the anemone *Nematostella vectensis*, which can be grown in the laboratory and is amenable to transgenesis and gene knockdown using morpholino oligonucleotides (Genikhovich and Technau, 2009). Most studies of *Nematostella* to date have focused on genome sequence analysis but, as we have argued above, this is insufficient to understand the evolution of conserved proteins at the biochemical, cellular and organismal levels. Future studies should make use of the genome sequence to derive proteins for biochemical experiments, and use the experimental tools available in *Nematostella* to test whether functions of interesting proteins are conserved between *Nematostella* and bilaterians.

Finally, for proteins with homologs outside metazoans, investigation of the non-metazoan family members may shed light on the evolutionary events that accompanied the transition to multicellularity. Choanoflagellates are the closest living unicellular relatives of metazoans, and recent sequencing of a choanoflagellate genome has revealed that they have many homologs of proteins that are important for multicellular development in animals (King et al., 2008). Biochemical and cell biological studies in choanoflagellates will shed light on how the functions of these proteins evolved to allow multicellularity. Additionally, the social amoeba *Dictyostelium* has been mostly used as a model of single cell chemotaxis and motility, but we recently found that it forms a structurally and functionally polarized epithelial tissue similar to animal epithelia during the multicellular phase of its life cycle (Dickinson et al., 2011). Thus, developmental studies in *Dictyostelium* may uncover basic cellular mechanisms of multicellularity, and *Dictyostelium* would be a more appropriate model than yeast for understanding the cellular functions of some proteins in the context of multicellular tissue organization and differentiation.

Conclusions

We have argued in favor of a broad experimental approach to evolutionary studies that goes beyond genome sequence comparison and the study of transcriptional regulatory networks to include biochemistry, cell biology, and *in vivo* functional studies in a wide range of organisms. Such an approach is motivated by the observation that so-called “conserved” proteins can have different functions in different systems. Clearly, this approach cannot be carried out by any single research group, but involves contributions from many laboratories with different expertise.

We emphasize that protein properties that are not widely conserved are not necessarily uninteresting; indeed, if our goal is to understand how different species achieve different developmental outcomes, we may ultimately be more interested in differences between homologs than similarities. Thus, a multidisciplinary approach that determines how protein structures, activities, cell biology and developmental mechanisms have changed over time will identify core mechanisms of animal development and will reveal adaptations that have allowed cells and organisms to achieve a wide variety of developmental outcomes.

References

Campbell K, Knust E, Skaer H. Journal of Cell Science. 2009; 122:2604–2612. [PubMed: 19567473]

Carroll SB. Cell. 2008; 134:25–36. [PubMed: 18614008]

Damen WGM. Dev Dyn. 2007; 236:1379–1391. [PubMed: 17440988]

Dickinson DJ, Nelson WJ, Weis WI. Science. 2011; 331:1336–1339. [PubMed: 21393547]

Genikhovich, G.; Technau, U. Cold Spring Harbor Protocols *2009*, pdb.emo129.. 2009.

Guder C, Pinho S, Nacak TG, Schmidt HA, Hobmayer B, Niehrs C, Holstein TW. Development. 2006; 133:901–911. [PubMed: 16452091]

Hunter C, Wieschaus E. The Journal of Cell Biology. 2000; 150:391–401. [PubMed: 10908580]

Kan NG, Stemmler MP, Junghans D, Kanzler B, de Vries WN, Dominis M, Kemler R. Development. 2007; 134:31–41. [PubMed: 17138661]

Kasza KE, Zallen JA. Current Opinion in Cell Biology. 2011; 23:30–38. [PubMed: 21130639]

King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I, et al. Nature. 2008; 451:783–788. [PubMed: 18273011]

Kwiatkowski AV, Maiden SL, Pokutta S, Choi H-J, Benjamin JM, Lynch AM, Nelson WJ, Weis WI, Hardin J. Proc Natl Acad Sci USA. 2010; 107:14591–14596. [PubMed: 20689042]

Morais-de-Sá E, Mirouse V, St Johnston D. Cell. 2010; 141:509–523. [PubMed: 20434988]

Shapiro L, Weis WI. Cold Spring Harb Perspect Biol. 2009; 1:a003053. [PubMed: 20066110]

St Johnston D, Ahringer J. Cell. 2010; 141:757–774. [PubMed: 20510924]
Wilson CW, Chuang P-T. Development. 2010; 137:2079–2094. [PubMed: 20530542]