Perspective

Gene Knockouts of c-*src*, Transforming Growth Factor β1, and Tenascin Suggest Superfluous, Nonfunctional Expression of Proteins

Harold P. Erickson

Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27710

THE technology of gene targeting is still in its infancy, but one astonishing theme is being seen over and over. Disruption of a supposedly important gene frequently produces a minimal phenotype. In some cases a phenotype is discovered in an unexpected tissue, but there is no disruption of the tissues where the protein is most highly expressed, and where its function was thought to be most important. A frequent explanation is that proteins are "redundant," in the sense that closely related proteins with duplicated functions might fill in for the one eliminated. However, the word redundant can also be used to mean superfluous, and this may be the more appropriate sense. Coexpression of proteins with duplicated functions is probably superfluous and wasteful but, as argued below, may be tolerated. The more extreme view of superfluous expression is that proteins may be prominently expressed in cells or tissues where they have no function at all.

Over the last decade many proteins have been discovered through cloning or antibodies, leaving the investigators to determine the function. The search for function frequently begins with immunocytochemistry, to determine where and when the protein is expressed. The interpretation, that the protein is playing an important role at the sites where it is most prominently expressed, is almost universal. If, however, proteins are expressed superfluously, where they have no function, these interpretations need to be re-evaluated.

Three Gene Knockouts Suggesting Superfluous Protein Expression

The c-*src* protein (9, 15) is a tyrosine kinase that is most prominently expressed in platelets, where it constitutes 0.2–0.4% of the total protein, and in neurons, where two splice variants are differentially expressed and concentrated in the growth cones. When the c-*src* gene was totally disrupted in mice (15), the expectation was that the homozygous null mice would have severe and probably lethal defects in platelets and the nervous system. Surprisingly, these tissues appeared completely normal. The mice without c-*src* did suffer a severe phenotype, osteopetrosis, and it was soon discovered that c-*src* was also prominently expressed in osteoclasts (4a, 9). Presumably c-*src* is playing an essential role in osteoclast function. But what is the role of c-*src* in platelets and neurons, tissues in which it is also prominently expressed, but apparently dispensable?

Transforming growth factor (TGF)[1]-β1, -β2, and -β3 are a family of closely related protein growth factors. These isoforms are prominently expressed in specific patterns in many embryonic tissues, largely overlapping but sometimes distinct (7, 10). TGF-βs are also expressed in many cell types and extracellular matrix locations of adults, but most attention has focused on the embryonic expression, with the expectation that each isoform is playing vital roles in development. Numerous studies in cell culture have demonstrated powerful stimulatory and inhibitory effects on cell function, building on the expectation that TGF-βs will have an important role wherever they are expressed. Surprisingly, the recent gene knockout of TGF-β1 (14) gave mice that were born with no apparent defect in development. They did develop severe, multifocal inflammatory disease after birth and die at ~20 d of age, confirming the essential role of TGF-β1 in modulating the inflammatory response. But what is the role of TGF-β1 in the several embryonic tissues where it is expressed prominently and in specific patterns?

Tenascin (3) is a large extracellular matrix protein expressed in specific patterns in developing brain, cartilage, smooth muscle, and in several tissues involving epithelial–mesenchymal interactions. Tenascin is prominently re-expressed in many tumors and in healing wounds. Two tenascin paralogues (related genes) are known, but their expression patterns show only limited overlap (4, 6, 8, 11; ref. 4 also demonstrates a recent duplication of tenascin gene X, producing a large tenascin XB protein and a severely truncated XA). The multi-domain structure of the tenascin subunit and the regulated expression of splice variants, suggested the possibility of multiple, independent functions, each of which could be essential in one or more tissues in which tenascin is expressed. Indeed, some of us thought that the gene targeting of tenascin should be approached piecemeal, knocking out one or a small cluster of domains at a time to determine separately the multiple functions. However, the bolder knockout of the entire tenascin gene was done first, with the amazing conclusion that "Mice develop normally without tenascin" (13). What is the role of the tenascin in the embryonic tissues where it is prominently expressed with such precise temporal and spatial regulation?

The answer in all three cases may be—no functional role at all. A key assumption in interpreting immunolocalization,

1. *Abbreviations used in this paper*: NGF, nerve growth factor; TGF, transforming growth factor.

that a gene is expressed only where the protein is needed, may be misleading.

Superfluous Expression of Nerve Growth Factor and EGF

Nerve growth factor (NGF, reviewed in 16) was first identified by its extremely potent activity in stimulating the growth of neurons. Research was greatly facilitated by the surprising discovery that NGF is highly concentrated in the submaxillary glands of adult male mice. It is specifically localized to the granular convoluted tubule cells, and is secreted in the saliva rather than in an endocrine fashion. EGF is also expressed at extremely high concentration in these submaxillary glands. Neither protein is expressed in female mice, nor in male or female rats, nor in the salivary glands of most other species. Curiously, NGF is also expressed in high concentration in the prostate gland of guinea pig, rabbit and bull, but not in rat, mouse, hamster, or human (16).

Despite many efforts, no function has been found for NGF in these male excretory glands. Indeed, it is difficult to imagine a function for NGF in male mouse saliva, a function not needed in female mice nor in rats, but somehow used in the semen of three unrelated mammals. I suggest that this may be another example of superfluous protein expression, with no function in the tissue of highest expression.

Several Possibilities for Apparent Nonfunctional Expression

There are several possible reasons why deletion of a prominently expressed protein might be benign. The first three listed here have already been discussed substantially (see especially ref. 2 for a brief but elegant analysis).

First, the protein could have a modest, rather than a vital function. A gene can become fixed in a population if it provides only a small survival advantage, which might be undetectable in a laboratory animal.

Second, it is possible that loss of a protein might up-regulate compensatory pathways, which could be affected by common feedback mechanisms. This has been argued in a recent gene knockout of the myo-D protein (12).

Third, the usual concept of redundancy: coexpression of proteins with duplicated functions. This has been convincingly argued as the reason for minimal phenotype in several cases of gene knockout (1, 2, 5, 12, 14, 15), and is probably important in many systems. If several proteins with similar functions are coexpressed in the same cell or tissue, knocking out any one of them may have no effect on that tissue.

Fourth, proteins may be expressed in tissues where they have no function at all. A rationale for this, applying also to the coexpression of related proteins, will be developed below.

Why Turn It Off If It Doesn't Hurt?

It is clearly essential that a gene must be turned on where the protein is needed, but the converse is not necessarily true. How important would it be to switch off a gene in cells where it is not needed? One might suggest economy, and it is natural for the biochemist to regard proteins as precious commodities, not to be wasted. But from the perspective of cellular metabolism, protein synthesis is very cheap. Thus the energy required to synthesize the c-*src* in platelets, where it constitutes up to 0.4% of the total cellular protein, is only a tiny fraction of the cell's overall metabolism, most of which goes to pumping ions. Even if c-*src* is totally useless in the platelet, the cell can easily afford to make it. Similarly, tenascin may be nonfunctional in the embryonic brain tissues where it is prominently expressed, but the expense of secreting it is minimal.

A truly important reason for turning off a gene would be if the protein were toxic or otherwise deleterious. Indeed, most proteins are probably deleterious in many places, and this is probably a major basis for the elaborate mechanisms for gene regulation. Also, some economy must be exercised not to have the protein synthesis machinery monopolized by thousands of useless proteins. But precise regulation need not apply to all proteins, even proteins that would appear to have a potent biological activity. Thus c-*src* might be tolerated in platelets and neurons, and TGF-β1 in embryonic tissues, if their receptors were already saturated by coexpressed proteins. NGF and EGF are probably innocuous in saliva and semen because the secreted proteins do not contact receptors. The still unknown functions of tenascin may be unimportant in at least some embryonic tissues.

The question remains, shouldn't useless proteins be turned off even if the expense of producing them is small? The answer probably lies in the economics of gene control mechanisms—they are expensive. Certainly each protein cannot have its own private control mechanism. The number of separate switches, i.e., transcription factors and the DNA elements they bind, is necessarily much smaller than the number of proteins that need to be controlled. Even with combinations of switches a unique control for every protein is probably impossible. Moreover, it is probably unnecessary.

It is no doubt much more expensive to provide a separate control mechanism than to make a little extra innocuous protein. The control mechanisms themselves require synthesis of separate proteins, and their invention by evolution is not a trivial process. Rather than devising a unique control for every gene, evolution may have generated a more limited set of shared switches. The rules governing the sharing would be that genes must have a switch to turn them on where the protein is needed, and they must be turned off (or not turned on) where they are toxic. Some overall economy is probably also needed, not to turn on too many useless proteins. But in any particular cell or tissue, one might expect to find proteins that are wastefully expressed, because one protein in a regulatory bank is needed and the others are not harmful.

The two types of superfluous protein expression discussed here may both be accidents of evolution, and closely related. When related proteins with duplicate functions are coexpressed, each of the proteins could exercise its function, but no one of them is essential. In the more extreme case a protein may be expressed where its function is not used at all. In both cases the superfluous protein expression is somewhat wasteful but apparently not a sufficient burden to generate a new control mechanism that would turn it off.

The concept of "junk DNA" is widely accepted. We should perhaps now consider the possibility of "junk protein." This should not imply that there are proteins without function (these would not be preserved by evolution), but refers to expression where the function is not needed. Thus a protein might be considered junk in any cell or tissue where its functions are duplicated by coexpressed relatives, or where its functions are not used.

The essential function of c-*src* in osteoclasts was only discovered after the src-deficient mice demonstrated a defect in bone resorption. The knockout of TGF-β1 suggests that its expression in embryos is nonvital, but dramatically confirms its essential role in modulating the inflammatory response in adults. Defects in tenascin-deficient mice have not yet been found, and the functions of tenascin are still unknown. Perhaps the search for function will be more fruitful as we turn our attention from the sites of prominent expression in embryos, and look for more subtle roles in a small set of embryonic or adult tissues.

Received for publication 27 October 1992 and in revised form 27 November 1992.

References

1. Braun, T., M. A. Rudnicki, H.-H. Arnold, and R. Jaenisch. 1992. Targeted inactivation of the muscle regulatory gene Myf-5 results in abnormal rib development and perinatal death. *Cell.* 71:369–382.
2. Brookfield, J. 1992. Can genes be truly redundant? *Curr. Biol.* 2:553–554.
3. Erickson, H. P., and M. A. Bourdon. 1989. Tenascin: an extracellular matrix protein prominent in specialized embryonic tissues and tumors. *Annu. Rev. Cell Biol.* 5:71–92.
4. Gitelman, S. E., J. Bristow, and W. L. Miller. 1992. Mechanism and consequences of the duplication of the human C4/P450c21/Gene X locus. *Mol. Cell. Biol.* 12:2124–2134.

4a. Horne, W. C., L. Neff, D. Chatterjee, A. Lomri, J. B. Levy, and R. Baron. 1992. Osteoclasts express high levels of $pp60^{c\text{-}src}$ in association with intracellular membranes. *J. Cell Biol.* 119:1003–1013.

5. Joyner, A. L., K. Herrup, B. A. Auerbach, C. A. Davis, and J. Rossant. 1991. Subtle cerebellar phenotype in mice homozygous for a targeted deletion of the En-2 homeobox. *Science (Wash. DC).* 251:1239–1243.
6. Matsumoto, K., N. Ishihara, A. Ando, H. Inoko, and T. Ikemura. 1992. Extracellular matrix protein tenascin-like gene found in human MHC class III region. *Immunogenetics.* 36:400–403.
7. Millan, F. A., F. Denhez, P. Kondaiah, and R. J. Akhurst. 1991. Embryonic gene expression patterns of TGFβ1, β2 and β3 suggest different developmental functions in vivo. *Development (Camb.).* 111:131–143.
8. Nörenberg, U., H. Wille, J. M. Wolff, R. Frank, and F. G. Rathjen. 1992. The chicken neural extracellular matrix molecule restriction: similarity with EGF-, fibronectin type III-, and fibrinogen-like motifs. *Neuron.* 8:849–863.
9. Nowak, R. 1991. *c-src*'s cellular role astounds researchers. *J. NIH Res.* 3:54–58.
10. Pelton, R. W., B. Saxena, M. Jones, H. L. Moses, and L. I. Gold. 1991. Immunohistochemical localization of TGFβ1, TGFβ2, and TGFβ3 in the mouse embryo: expression patterns suggest multiple roles during embryonic development. *J. Cell Biol.* 115:1091–1105.
11. Pesheva, P., E. Spiess, and M. Schachner. 1989. J1-160 and J1-180 are oligodendrocyte-secreted nonpermissive substrates for cell adhesion. *J. Cell Biol.* 109:1765–1778.
12. Rudnicki, M. A., T. Braun, S. Hinuma, and R. Jaenisch. 1992. Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. *Cell.* 71:383–390.
13. Saga, Y., T. Yagi, Y. Ikawa, T. Sakakura, and S. Aizawa. 1992. Mice develop normally without tenascin. *Genes Dev.* 6:1821–1831.
14. Shull, M. M., I. Ormsby, A. B. Kier, S. Pawlowski, R. J. Diebold, M. Yin, R. Allen, C. Sidman, G. Proetzel, D. Calvin, N. Annunziata, and T. Doetschman. 1992. Targeted disruption of the mouse transforming growth factor-β1 gene results in multifocal inflammatory disease. *Nature (Lond.).* 359:693–699.
15. Soriano, P., C. Montgomery, R. Geske, and A. Bradley. 1991. Targeted disruption of the c-*src* proto-oncogene leads to osteopetrosis in mice. *Cell.* 64:693–702.
16. Thoenen, H., and Y. -A. Barde. 1980. Physiology of nerve growth factor. *Physiol. Rev.* 60:1284–1335.