Deciphering isozyme function: exploring cell biology with chemistry in the post-genomic era

Craig M Crews

Genome sequencing projects are identifying protein sequences faster than it is possible to discover their functions. Fortunately, combinatorial chemistry offers an opportunity to develop new biological reagents with which to determine the roles of related isozymes.

Address: Biology Department, Yale University, New Haven, CT 06520-8103, USA.

E-mail: craig.crews@yale.edu

Chemistry & Biology December 1996, **3**:961–965

 ISSN 1074-5521

Few defining moments in science can be predicted years in advance. It is clear, however, that the completion of the several genome-sequencing projects that are currently under way will have a significant impact on almost every aspect of biology. While this may be one of the slowest and most highly observed paradigm shifts in recent scientific history, it will be profound. Once these projects are completed, an enormous amount of information about new genes and gene families will become available. The first wave of this flood of data has already washed through; the first eukaryotic genome sequence, that of the yeast *Saccharomyces cerevisiae*, was completed in April [1]. This genome consists of 12 million base pairs containing an estimated 6000 genes. The analysis of these new sequences and the discovery of new gene families has just begun.

Coping with this flood of new information will be a challenge, and one of the most difficult aspects will be assigning functions to previously unidentified proteins. As the first genome projects come to completion, attention is already beginning to shift away from the collection of sequence data towards trying to understand the function of the many new gene products discovered. This task is made more difficult by the fact that a significant number of these gene products are highly related to each other. Analysis of the yeast genome has revealed that 70 % of yeast protein-coding sequences are related to known genes [1]. If this pattern holds for the estimated 80 000 genes in the human genome, then there will be far more protein families than we currently know about, greatly increasing the intricacy of the already complex picture emerging from cell biology. More and more emphasis will therefore need to be placed on discovering the exact functions of a set of highly related protein family members or isozymes, ascertaining how much the duties of each family member overlap and determining whether individual members have distinct functions. Here, I discuss the shortcomings of current approaches to this problem, and suggest that combinatorial chemistry may offer an excellent solution.

Identification is easy, determining function is hard

When the problem of deciphering the function of a new protein arises, an obvious approach is to compare its amino acid sequence with the sequences of proteins of known function. Such analyses often reveal that domains whose function, in general terms, is known, are present within the new protein. Several molecular biology techniques have been developed to aid in the isolation of related gene sequences (e.g. low stringency DNA hybridization and degenerate PCR), thus making broader comparisons possible. It is often useful to define consensus sequences that are always present in all members of a given protein family, and then to use this consensus motif to identify additional family members. For example, the seminal paper on the kinase motif [2] reported an alignment of 65 different protein kinase amino acid sequences that revealed 11 conserved subdomains as being critical for kinase function.

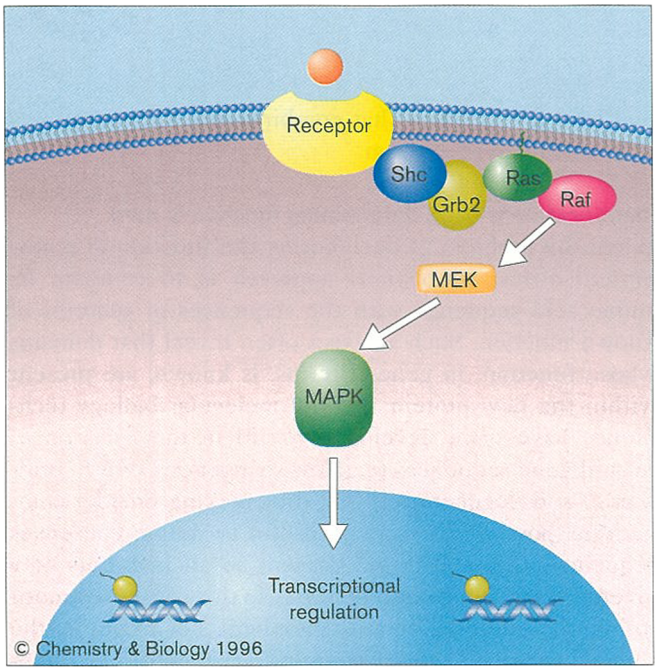
Not long ago, the identification of related gene products required weeks of work by a molecular biologist. Now, however, it is increasingly possible to find what one needs with a gene comparison search that takes only a few minutes at the computer terminal. For example, after the identification of a novel regulator of G-protein signaling (RGS) using a traditional genetic screen, a sequence comparison with known gene sequences in GenBank identified nine sequences related to RGS [3].

Although finding novel family members aids the recognition of novel conserved protein motifs, nowadays genes are often cloned before there is any information on the proteins they encode within the cell. Despite our ability to find sequence similarities between the new genes and those previously known, we cannot necessarily reach any useful conclusions about their functions. For example, the use of PCR and low stringency DNA hybridization has allowed the identification of a large number of putative steroid receptors based on sequence similarity, but the ligands for most of these orphan receptors are still unknown [4].

Making sense of signal transduction webs

One area of cell biology in which isozyme redundancy is particularly noticeable is that of signal transduction. Over the past 20 years, researchers have made significant progress in elucidating linear regulatory pathways (consisting of GTP-binding proteins, transcriptional activators

Figure 1



Growth-factor receptor signal transduction links the binding of a growth factor to the external face of a cell with events within the nucleus. Ligand binding causes autophosphorylation of the receptor, allowing Shc and the adapter protein Grb2 to bind. Ras is then activated through GDP to GTP exchange, enabling it to bind to the Raf protein and recruit it to the plasma membrane. Raf then activates MEK, which in turn phosphorylates MAP kinase, activating it and allowing it to phosphorylate a number of downstream targets.

and protein kinase cascades) that link the binding of a growth factor to the external face of a cell with events within the nucleus (Fig. 1). After many years of work, these different components have been arranged in a hierarchical order, but this simple picture of linear pathways is now giving way to the idea that many isozyme regulatory components are connected in a much more complex network or 'web'.

The threonine/tyrosine-specific kinase MEK exemplifies this increasing complexity. MEK links signaling events at the membrane with cytoplasmic protein phosphorylation. The first MEK protein was purified only four years ago based on a biochemical *in vitro* kinase assay [5]. Since then, five more related MEK sequences have been identified using a variety of approaches [6]. Although MEK1 was identified using its *in vitro* ability to phosphorylate and activate MAP kinase, it is becoming clear that different MEK proteins phosphorylate multiple MAP kinases, and that they are activated by different upstream kinases (Fig. 2). Deciphering the different roles of the various signaling components has taken many years and several false starts, partly because of the limitations of the approaches and tools that have been available.

Current biological approaches and their inadequacies

The number of enzymes that fall into closely related sets of isozymes is growing rapidly. Assigning functions to apparently redundant related protein family members is not easy. To determine the function of a single isozyme, what is needed is the ability to block (or increase) the action of just that one protein. All of the approaches currently available to do this have certain advantages, although each has its own limitations.

Gene knock-outs

One of the most straightforward ways to understand a protein's function is to eliminate it. This genetic approach is easy and informative in simple systems, but in mammalian systems it has proven non-trivial. An effective method to eliminate a gene product from an animal is to remove the functioning gene from the genome. In mice, these are known as 'knock-out' experiments: mouse embryonic stem cells are genetically manipulated *in vitro* and then implanted into pseudopregnant female mice. Using this process, one can definitively eliminate a gene's function. The procedure is costly and time consuming, however [7], and this technique therefore does not lend itself easily to a study of a large gene family. Moreover, expression of a given isozyme is often restricted to certain stages of development. Mice in which such an isozyme is knocked out may upregulate other isozymes to compensate for the loss, thereby complicating the interpretation of the experiment [8]. Thus, the generation of knock-out mice is clearly not the preferred approach to the study of large isozyme families.

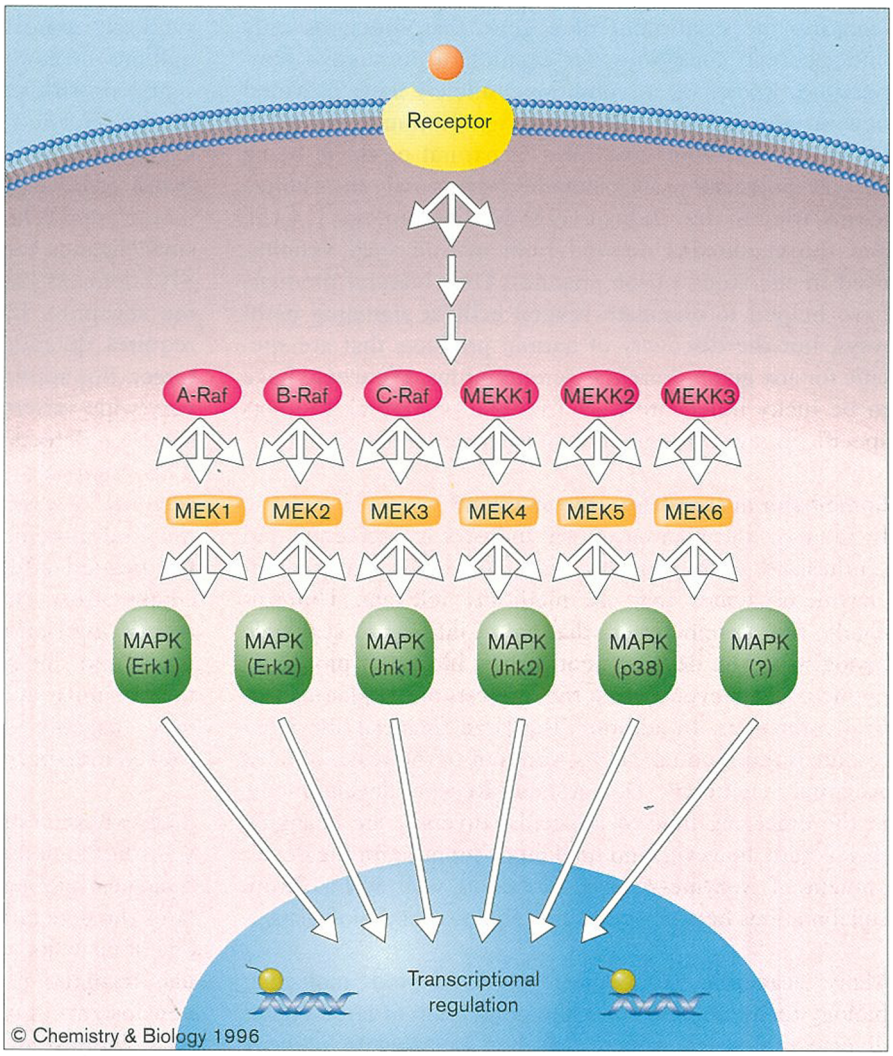
Gene transfection experiments

Overexpression of genes introduced into cell cultures is widely used in cell biology to study the intracellular consequences of their expression. Genes can be introduced into cells either transiently (in which case the gene is not integrated into the cell's genome and the protein product is therefore expressed only for a few days) or stably (which requires the gene to be stably integrated into the genome; this is accomplished by selecting for drug resistance). Stable integration has the advantage that all the cells selected from a single integration event express the same amount of the protein and can be expected to have stable gene expression for many cell generations. The process of gene integration, however, requires several weeks during which mutations can accumulate in the cell to compensate for any deleterious effects that result from overexpression of the introduced gene. In addition, since gene integration is essentially random, some genes may be integrated close to strong gene promoters or transcriptional silencers, giving unpredictable expression levels. Thus, several transfectants must be tested to ensure an adequate level of expression.

Since selection of stable transfectants is time-consuming and may result in cell lines that have adapted to the

Figure 2

The increasing complexity of the regulatory pathways is exemplified by MEK. Six MEK proteins have now been identified. The MEK proteins are each phosphorylated by different upstream kinases and have themselves been shown to phosphorylate multiple MAP kinases. These patterns of phosphorylation overlap; for example MEK1/MEK2 are activated by MEKK3 and B-Raf [30,31], so the pathway forms a network instead of a simple linear path.



presence of the foreign gene, many laboratories routinely use transient transfections instead. Transient overexpression experiments are relatively easy to perform and give high levels of gene expression. For precisely this reason, however, these assays are inadequate for the study of the subtle functional differences expected between members of a family of isozymes. Overexpression of a given protein may result in isozyme–substrate interactions that do not normally occur *in vivo*. The functional differences between related isozymes may derive from small differences in their K_M values for different substrates, and overexpression of one component may lead to activation of the wrong signaling cascade, resulting in spurious conclusions. This approach is, therefore, not suitable for the study of a signaling web consisting of various related isozymes. An unfortunate example that makes this clear is the story of the enzyme named MEK kinase (MEKK1). This kinase was cloned using sequence similarity with a yeast protein that activates an MEK homolog, and *in vitro* transient overexpression assays showed it to be able to phosphorylate MEK1. Based on these results, this gene was named MEKK, for MEK kinase [9]. Although MEKK can activate MEK1/MEK2 *in vitro*, subsequent experiments have shown that the most likely downstream target of MEKK is not MEK1 but a related homolog, MEK4 [10]. It is clear that although transient overexpression experiments have been instrumental in determining signal hierarchies, they are inadequate for the investigation of isozyme function.

Natural products as probes of cell biology

Most cell biologists long for a 'magic bullet' that would specifically activate or inactivate their protein of interest, allowing them to shed light on the complexities of the cellular machinery. New approaches are needed to develop such reagents. They should allow transient perturbation of a cellular system and should distinguish clearly between related isozymes. Natural products provide one effective way to probe intracellular mechanisms [11]. Recently,

immunosuppressive agents (such as FK506, rapamycin and cyclosporin A) that inactivate intracellular proteins (creating the equivalent of a gene that functions only under certain conditions, for example a permissive temperature, known as a 'conditional allele') have provided new insights into our understanding of immune cell regulation [12,13]. Other examples of natural products being used as molecular probes include the pertussis and cholera toxins, used in the study of GTP-binding proteins [14,15], and the conotoxins derived from marine snail venoms, used in the study of ion channels [16]. Natural products have helped to elucidate several cellular signaling pathways, but the discovery of natural products that are specific for any given protein is serendipitous. We would have to be lucky indeed to find a palette of natural products specific for every member of an isozyme family.

Combinatorial chemistry – the solution?

In the past, the pharmaceutical industry has screened and synthesized many specific inhibitors and activators of enzyme isozymes that are medically relevant. Unfortunately, the protein targets that are of interest to academic biologists in the development of new biological probes do not necessarily overlap with the interests of the pharmaceutical companies. In addition, the large financial and labor resources required takes this work out of the reach of most academic chemistry laboratories. Recent developments in the emerging field of molecular diversity are changing this picture, however, and hold much promise for the development of isozyme-specific probes as well as for future collaborations between academic chemists and biologists.

Many academic laboratories have adapted molecular biology techniques to allow the generation of large sets of diverse molecules, which can then be used to identify novel ligands. For example, oligopeptide presentation on the surface of bacteriophage (phage display) has been successfully used to identify ligands for a variety of receptors and antibodies (for review see [17]). Similarly, expressed peptides have been used to identify ligands for cyclin-dependent kinase isoforms [18]. This was a novel approach to the selection of peptide ligands (now sometimes known as peptide aptamers) from random peptide sequences *in vivo*. Nucleic acid aptamers that bind selectively to given isozyme targets have also been reported [19,20]. The usefulness of these reagents in addressing questions in cell biology is in doubt, however, due to their inherent *in vivo* susceptibility to degradation and their lack of cell-membrane permeability. In addition, the diversity represented by the 20 natural amino acids or 4 natural bases is not enormous, and does not take advantage of the many small compounds that are common in modern organic chemistry.

Fortunately, recent developments in combinatorial chemistry offer much hope that new reagent development will soon become much easier. Combinatorial chemistry is not limited to a natural pool of monomers. Using the split/pool synthetic approach pioneered by Furka *et al.* [21], literally millions of compounds can be prepared by solid-phase synthesis with relative ease. These libraries of compounds can be screened for a variety of novel biological activities using a number of assays. 'On-bead', solution- and cell-based screening of libraries are becoming routine [22]. These approaches have been used to successfully identify small ligands for protein domains such as SH2 [23] and SH3 domains [24], opioid receptors [25] and benzodiazapine receptors [26]. Unlike rational drug design, which requires detailed structural information about a protein target, this approach of 'irrational ligand design' presents a very wide variety of possible ligands to the protein target and then detects the optimal ligand selected by the target. This approach is analogous to a genetic screen, in that selective pressure is applied to a pool of randomly mutated proteins to identify the few mutations that have the desired effect on the function of the protein. This combinatorial chemistry approach has been successful in identifying isozyme-selective inhibitors of carbonic anhydrases and zinc endopeptidases [27,28], further suggesting that a similar approach would have the potential to yield new reagents that could distinguish between closely related members of a signaling family.

The explosion of combinatorial chemistry in the past few years has fueled interest in the use of chemistry to address biological problems. Since combinatorial chemistry facilitates the generation of many ligands for any given protein, one of its major applications has been in the field of molecular recognition. Although this has produced important new insights in protein–ligand interactions and has been of great interest to structural biologists, the opportunity to develop these ligands into practical biological tools is too often overlooked. More collaborations are now needed between chemists and biologists to take advantage of the many biological opportunities offered by combinatorial chemistry. In particular, the next phase in the developing story of molecular diversity generation should bring chemistry into areas of biology (such as cell biology) that are not currently at the centre of the chemistry/biology interface [29]. The emerging field of chemical biology will need to address problems in many biological areas in the post-genomic era. Combinatorial chemistry may come to the rescue by developing the wide range of novel reagents needed to address questions of isozyme function.

Acknowledgements

I thank the Burroughs Wellcome Fund for financial support (New Investigator Award in the Pharmacological Sciences).

References

1. Dujon, B. (1996). The yeast genome project: what did we learn? *Trends Genet.* **12**, 263–270.
2. Hanks, S.K., Quinn, A.M. & Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic

domains. *Science* **241**, 42–51.
3. Koelle, M.R. & Horvitz, H.R. (1996). EGL-10 regulates G-protein signaling in the *C. elegans* nervous system and shares a conserved domain with many mammalian proteins. *Cell* **84**, 115–125.
4. Ortiz, M.A., Piedrafita, F.J., Pfahl, M. & Maki, R. (1995). TOR: a new orphan receptor expressed in the thymus that can modulate retinoid and thyroid hormone signals. *Mol. Endocrinol.* **9**, 1679–1691.
5. Crews, C.M., Alessandrini, A. & Erikson, R.L. (1992). The primary structure of MEK, a protein kinase that phosphorylates the ERK gene product. *Science* **258**, 478–480.
6. Stein, B., Brady, H., Yang, M.X., Young, D.B. & Barbosa, M.S. (1996). Cloning and characterization of MEK6, a novel member of the mitogen-activated protein kinase kinase cascade. *J. Biol. Chem.* **271**, 11427–11433.
7. Brandon, E.P., Idzerda, R.L. & McKnight, G.S. (1995). Targeting the mouse genome: a compendium of knock-outs, I. *Curr. Biol.* **5**, 625–634.
8. Erickson, H.P. (1993). Gene knock-outs of c-*src*, transforming growth factor β1, and tenascin suggest superfluous, nonfunctional expression of proteins. *J. Cell Biol.* **120**, 1079–1081.
9. Lange-Carter, C.A., Pleiman, C.M., Gardner, A.M., Blumer, K.J. & Johnson, G.L. (1993). A divergence in the MAP kinase regulatory network defined by MEK kinase and Raf. *Science* **260**, 315–319.
10. Yan, M.H., *et al.*, & Templeton, D.J. (1994). Activation of stress-activated protein kinase by MEKK1 phosphorylation of its activator SEK1. *Nature* **372**, 798–800.
11. Hung, D., Jamison, T.F. & Schreiber, S.L. (1996). Understanding and controlling the cell cycle with natural products. *Chemistry & Biology* **3**, 623–639.
12. Brown, E.J., *et al.*, & Schreiber, S.L. (1994). A mammalian protein targeted by G1-arresting rapamycin–receptor complex. *Nature* **369**, 756–758.
13. Liu, J., Farmer, J.D.J., Lane, W.S., Friedman, J., Weissman, I. & Schreiber, S.L. (1991). Calcineurin is a common target of cyclophilin–cyclosporin A and FKBP–FK506 complexes. *Cell* **66**, 807–815.
14. Ramirez, I., Tebar, F., Grau, M. & Soley, M. (1995). Role of heterotrimeric G-proteins in epidermal growth factor signalling. *Cell. Signal.* **7**, 303–311.
15. Moss, J. & Vaughan, M. (1993). ADP-ribosylation factors, 20 000 M(r) guanine nucleotide-binding protein activators of cholera toxin and components of intracellular vesicular transport systems. *Cell. Signal.* **5**, 367–379.
16. McIntosh, J.M., *et al.*, & Olivera, B.M. (1995). A new family of conotoxins that blocks voltage-gated sodium channels. *J. Biol. Chem.* **270**, 16796–16802.
17. Cortese, R., *et al.*, & Sollazzo, M. (1995). Identification of biologically active peptides using random libraries displayed on phage. *Curr. Opin. Biotech.* **6**, 73–80.
18. Colas, P., Cohen, B., Jessen, T., Grishna, I., McCoy, J. & Brent, R. (1996). Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* **380**, 548–550.
19. Schneider, D.J., Feigon, J., Hostomsky, Z. & Gold, L. (1995). High-affinity ssDNA inhibitors of the reverse transcriptase of type 1 human immunodeficiency virus. *Biochemistry* **34**, 9599–9610.
20. Conrad, R., Keranen, L.M., Ellington, A.D. & Newton, A.C. (1994). Isozyme-specific inhibition of protein kinase C by RNA aptamers. *J. Biol. Chem.* **269**, 32051–32054.
21. Furka, A., Sevestyen, F., Asgedom, M. & Dibo, G. (1991). General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Peptide Res.* **37**, 487–493.
22. Lebl, M., *et al.*, & Lam, K.S. (1995). One-bead-one-structure combinatorial libraries. *Biopolymers* **37**, 177–198.
23. Muller, K., *et al.*, & Baumann, G. (1996). Rapid identification of phosphopeptide ligands for SH2 domains – screening of peptide libraries by fluorescence-activated bead sorting. *J. Biol. Chem.* **271**, 16500–16505.
24. Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W. & Schreiber, S.L. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76**, 933–945.
25. Dooley, C.T., Kaplan, R.A., Chung, N.N., Schiller, P.W., Bidlack, J.M. & Houghten, R.A. (1995). Six highly active μ-selective opioid peptides identified from two synthetic combinatorial libraries. *Pept. Res.* **8**, 124–137.
26. Bunin, B.A., Plunkett, M.J. & Ellman, J.A. (1994). The combinatorial synthesis and chemical and biological evaluation of a 1,4-benzodiazepine library. *Proc. Natl. Acad. Sci. USA* **91**, 4708–4712.
27. Burbaum, J.J., *et al.*, & Baldwin, J.J. (1995). A paradigm for drug discovery employing encoded combinatorial libraries. *Proc. Natl. Acad. Sci. USA* **92**, 6027–6031.
28. Jiracek, J., *et al.*, & Dive, V. (1995). Development of highly selective phosphinic peptide inhibitors of zinc endopeptidase 24–15 using combinatorial chemistry. *J. Biol. Chem.* **270**, 21701–21706.
29. Mitchison, T.J. (1994). Towards a pharmacological genetics. *Chemistry & Biology* **1**, 3–6.
30. Blank, J.L., Gerwins, P., Elliot, E.M., Sather, S. & Johnson, G.L. (1996). Molecular cloning of mitogen-activated protein/ERK kinase kinases (MEKK) 2 and 3. Regulation of sequential phosphorylation pathways involving mitogen-activated protein kinase and c-jun kinase. *J. Biol. Chem.* **271** 5361–5368.
31. Reuter, C.W., Catling, A.D., Jelinek, T. & Weber, M.J. (1995). Biochemical analysis of MEK activation of B-Raf and other activators. *J. Biol. Chem.* **270** 7644–7655.