Journal of Cellular Biochemistry Supplement 37:42–52 (2001)

Signal Transduction Driving Technology Driving Signal Transduction: Factors in the Design of Targeted Therapies

Erica A. Golemis,[1]* Michael F. Ochs,[2] and Elena N. Pugacheva[1]

[1]Cell and Developmental Biology Working Group, Division of Basic Science, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111
[2]Bioinformatics Working Group, Division of Basic Science, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111

Abstract A significant number of human diseases can be attributed to defects in cellular signal transduction pathways. Large-scale proteomics projects now in progress seek to better define critical components of signal transduction networks, to enable more intelligent design of therapeutic agents that can specifically correct disease-specific signaling alterations by targeting individual proteins. A complicating factor in this endeavor is the fact that intracellular signaling involves many diverse mechanisms that in sum finely modulate the activity of individual proteins in response to different biological inputs. Ability to develop reagents that selectively correct disease-associated signaling activities, while leaving intact benign or essential activities, encompassed within a single protein requires an intimate knowledge of pathway-specific control mechanisms. To illustrate these points, we provide examples of some of the complex control mechanisms regulating the Cas proteins, which contribute to integrin-dependent biological response. We then discuss issues involved in systematically incorporating information related to complex control mechanisms in proteomic databases. Finally, we describe some recent instances in which protein interaction technologies have been specifically adapted to identify small molecule agents that regulate protein response in physiologically desirable ways, and discuss issues relevant to future drug discovery efforts. J. Cell. Biochem. Suppl. 37: 42–52, 2001.

Key words: signal transduction; complexity; protein interactions; informatics; databases; targeted drug discovery

Current development strategies for 21st century therapeutic agents involve the creation of biological intellectual assets parallel to those involved in providing effective phone service. For the last decade, governments and industry have sponsored immense "-ome" projects that enumerate the complete contents of DNA, RNA, and proteins within human and other organisms. Based on this global census, which for humans is expected to be completed within the next few years, the major challenges in therapy are first, to understand the intrinsic wiring connecting the different molecular components constituting the human body; second, to identify the points at which the wiring is disrupted in clinically significant disease states; and third, to direct therapeutic agents specifically and uniquely to the point of disruption, so as not to induce non-specific toxicities. Together these feats are comparable to pinpointing a single individual out of in excess of 6 billion with an 11-digit phone number. The vast majority of critical targeting information currently extent has been developed over many years and in many academic laboratories. On the one hand, the coherent integration of this information with the data assembled in the -ome projects presents a major computational challenge. On the other hand, the creation of small molecule agents that can precisely manipulate proteins and signaling pathways will require the use of approaches that closely couple design of screening technologies to the mechanisms they seek to target. This review will focus on these issues.

Grant sponsor: NIH; Grant number: CA06927; Grant sponsor: Pew Foundation (to Fox Chase Cancer Center); Grant sponsor: NIH (to E.A.G.); Grant number: CA63366.

*Correspondence to: Erica A. Golemis, Fox Chase Cancer Centre, 7701 Burholme Avenue, Philadelphia, PA, 19111. E-mail: EA_Golemis@fccc.edu

Received 13 September 2001; Accepted 14 September 2001

DOI 10.1002/jcb.10064

PROTEOMICS MEETS A COMPLEX WORLD: ISSUES IN SIGNAL TRANSDUCTION

How does a cell achieve fine control of its properties of division, metabolism, shape, and movement in response to dynamic internal and external cues? How are these control mechanisms subverted in disease? While we do not yet have the complete answer to these very large questions, a number of individual signaling systems within cells are becoming very well defined, allowing description of the mechanisms that are utilized to modulate signaling readout. A description of one such system, centered on the Cas family proteins [O'Neill et al., 2000], is useful in illustrating the complexity of issues to be considered in analyzing the function of biological signaling pathways, and correcting the dysfunction of these pathways in disease.

Although some evidence suggests additional functions for Cas proteins, the most well-documented activity of this group is in transducing signals through integrins (Fig. 1). Integrins are transmembrane receptor proteins located at structures termed focal adhesions that act as bridges between proteins of the extracellular

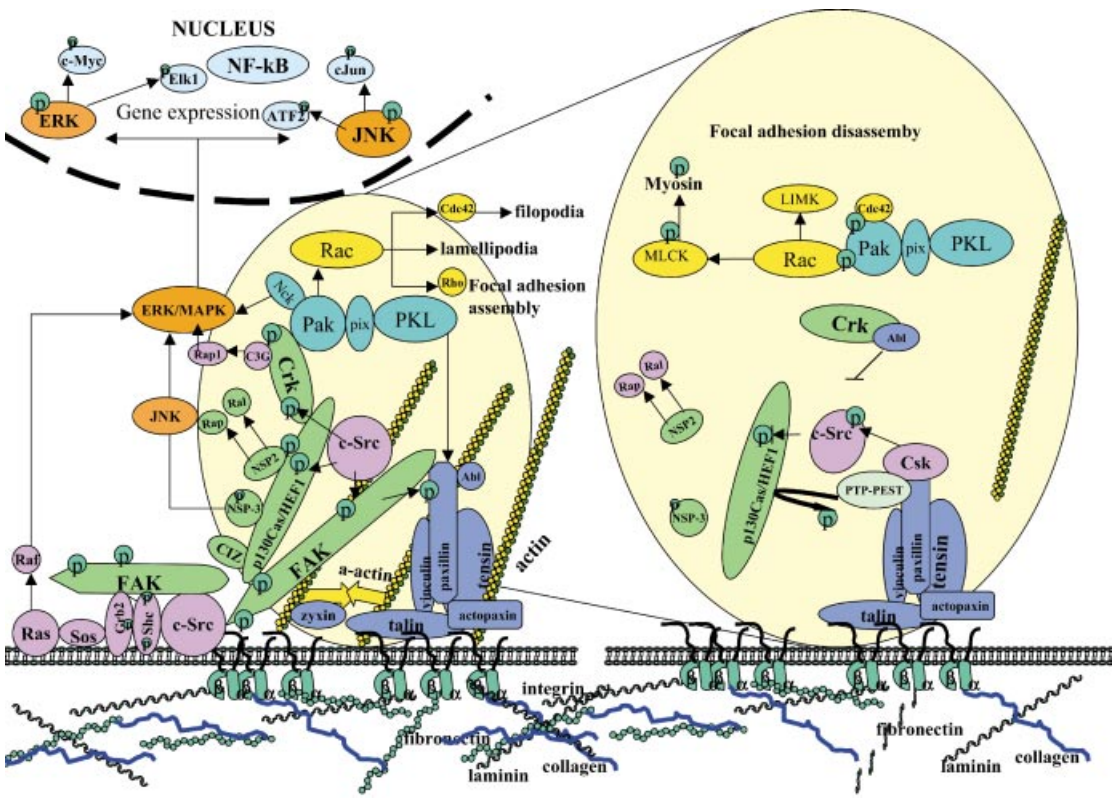

Fig. 1. Model depicting the role of Cas family proteins in focal adhesion assembly (left) and disassembly (right). **Left**: extracellular matrix (ECM) ligands bind and induce clustering of integrin receptor. Subsequently, cytoskeletal binding proteins including talin, vinculin, paxillin, and actopaxin are recruited for organization of the actin cytoskeleton at focal adhesions. Activated and autophosphorylated FAK(Tyr 397) in turn phosphorylates p130Cas/HEF1 (Cas), creating a Src binding site: Cas-Src binding contributes to activation of the Src tyrosine kinase activity. Reciprocally, Src-mediated phosphorylation of Cas results in the recruitment of a Crk-GEF signaling complex which in turn activates Rap1 and ERK. Some reports indicate as second possible Cas-activated pathway mediated through Shc-Ras-Raf-ERK interactions. Both FAK and Src can phosphorylate Shc at multiple sites to promote Grb2 binding: Grb2 association with the Sos GDP-GTP exchange protein can activate Ras and ERK1,2/MAP kinase. During focal adhesion turnover (critical for cell migration) transient complexes connect Cas to additional proteins, including PKL, PIX, and PAK, involving direct interaction of PAK with Nck, Crk, and paxillin. The binding of GTP-activated Cdc42 and Rac to PAK stimulates PAK kinase activity. **Right**: activated PAK modifies the actin cytoskeleton in a process involving the activation of LIM kinase and MLCK. Subsequently, paxillin-mediated recruitment to focal adhesions of the tyrosine kinase Csk (negative regulator of Src) and tyrosine phosphatase PTP-PEST, which dephosphorylates Cas and disrupts the Cas-Crk association, causes disassembly of the focal adhesions. Note, figure as shown presents highly simplified versions of actual signaling complexes. Cas and Cas-interacting proteins, green; Src-Shc-Ras-Raf signaling cascades, pink; cytoskeleton binding proteins, blue; common effector kinases, brown; transcription factors, light blue; focal adhesion disassembly components, aqua; PAK-effectors, yellow.

matrix and intracellular signal transduction proteins [reviewed in Giancotti and Ruoslahti, 1999]. Changes of integrin signaling are important in control of actin dynamics and cell movement, as well as for cellular growth and survival, and hence contribute to the execution of normal biological processes such as organismal development and immune function, and are disturbed in disease processes such as cancer cell metastasis [reviewed in Mercurio and Rabinovitz, 2001]. As shown (Fig. 1), the Cas family proteins reside in close proximity to integrins at focal adhesions, where they collaborate with a network of other proteins (see legend, Fig. 1) to transduce integrin-initiated signals. Many of the details of Cas protein action in a biological context have been recently reviewed [O'Neill et al., 2000]: the following summary is focussed specifically on the mechanisms that have been shown to govern Cas activity. Finally, although the Cas group is presented here as one model out of many that could have been used to illustrate signaling complexity, there are reasons to specifically consider the modulation of Cas protein function as a desirable goal in the context of treating specific human diseases (see Issue VII, below). The most critical issues that complicate understanding Cas function are listed below.

Signaling Issue I: Redundancy of Family Members: Reinforcing Versus Opposing Action

As with many other signaling protein families, the Cas family of proteins contains multiple members (Fig. 2). p130Cas, HEF1/Cas-L, and Efs/Sin are highly related in both protein structure and sequence. The two most closely related members of the group, p130Cas (870 amino acids) and HEF1 (834 amino acids), are 48% similar over their full length, with much of the conservation clusted within discrete domains (see below). The third member of the group, Efs/Sin, is more distantly related, and much less well characterized.

Because of the sequence similarity observed between HEF1 and p130Cas, these proteins might be predicted to have congruent function. Based on the biological functions for p130Cas and HEF1 established to date, it appears that they can display opposing versus reinforcing activity depending on biological context. For example, it has been shown that overexpression

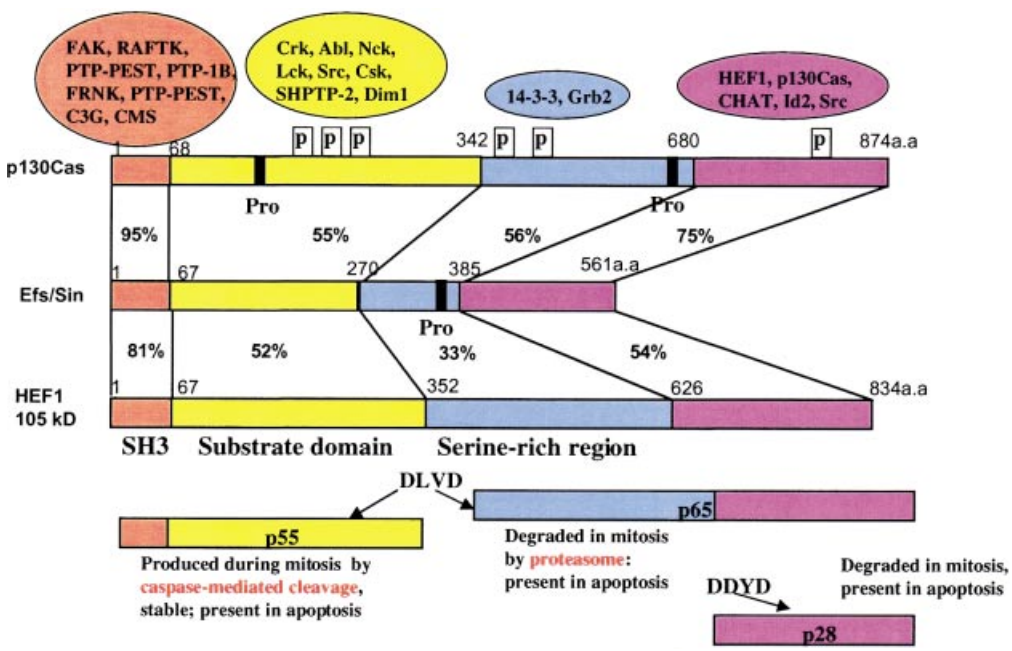

Fig. 2. Schematic representation of the conserved domain structure of Cas family proteins:p130Cas, Efs/Sin, and HEF1/Cas L. The amino acid (a.a) identity of the domains is noted: domains shown are (1) the Src homology 3(SH3)domain (red), (2) the substrate domain (yellow), (3) the serine-rich region (SRR) (blue), and (4) C-terminal conserved domain (pink). p130Cas and Efs additionally contain poly-proline regions, indicated as Pro. Regions of phosphorylation are marked P (for simplicity, only p130Cas is marked). Summary of Cas-family interacting proteins presented in ovals, adjacent to domain with which interaction has been assigned. Below, DLVD and DDYD sites of caspase cleavage identified in HEF1, and expression of resulting HEF1 cleavage products during different growth conditions, are shown.

or FAK-dependent activation of either p130Cas or HEF1 (see Fig. 1) induces cellular migration and may contribute to metastasis. In this process, the two proteins function in parallel. In contrast, while enhanced action of p130Cas has been shown to be protective against detachment-induced apoptosis (anoikis), enhanced expression of HEF1 has been shown to induce apoptosis. The contribution of Efs/Sin to migration and apoptosis remains to be determined, while the action of the three family members in other biological processes in which they have been implicated is also as yet unclear.

Signaling Issue II: Multiplicity of Domains and Interactive Motifs

Each individual Cas-family member has a complex sequence structure (Fig. 2). All Cas proteins have an amino-terminal SH3-domain, conferring association with partner molecules containing poly-proline motifs; multiple SH2-binding sites, which when phosphorylated confers association with partner molecules containing SH2-domains; and a carboxy-terminal region which remains poorly defined as to function, but which, as with the preceding sequence elements, is highly conserved within the group. Some Cas proteins contain additional predicted interaction motifs: for example, p130Cas and Efs/Sin contain poly-proline motifs, while HEF1 does not.

For each defined sequence element, multiple interactive protein partners have been identified [discussed in O'Neill et al., 2000]. Significantly, the mode of interaction of protein partners with Cas family members differs. Some partners are competitive for specific interactive domains. For example, PTP-PEST, PTP-1B, and FAK have all been shown to interact with the SH3 domain of p130Cas, and hence compete with each other for occupancy of the available binding motif. Other proteins bind co-operatively or sequentially to distinct sequences within the protein. For example, an initial FAK-Cas SH3 interaction initiates a sequence of events which culminates with the assembly of Cas in complex with CrkI, Dock180, Jnk, and other proteins. The ability of Cas proteins to act as a scaffold for other signaling molecules, which is thought to be necessary for Cas protein biological functions, is hence dependent in part on the initial success of FAK in competing with other protein partners for binding to a single Cas domain.

Signaling Issue III: Dynamic Post-Translational Modification

A striking feature about the Cas family group is the possession of a very large number of potential sites of phosphorylation. Both Cas and HEF1 contain at least 15 tyrosines in the context of SH2 binding sites, which contribute to protein-protein interactions, and which contribute to stability of Cas association with the cytoskeleton. Discrete pools of differentially tyrosine-phosphorylated Cas exist in cells. Factors contributing to changes in Cas protein phosphorylation include cellular attachment, growth factor stimulation, cell cycle progression, oncogenic transformation, and initiation of apoptosis, among others, and it has been proposed that dynamic oscillation between different phosphorylation states is necessary for Cas function.

In this context, some kinases possess initiating function for subsequent interaction cascades, and hence are more critical for controlling Cas action. For example, the initial interaction of FAK with the SH3-domain of Cas proteins allows FAK to phosphorylate a single SH2-binding site in the carboxy-terminal region of Cas. Src kinase is recruited to this site, and then in turn phosphorylates the Cas SH2-binding site region, allowing recruitment of other SH2-domain containing proteins. Specific phosphatases, including LAR, PTP-PEST, and PTP-1B, have been shown to dephosphorylate the Cas family proteins, with this action associated with changes in cellular viability and motility. The Cas proteins are also known to be phosphorylated on serine and threonine residues, but the scope and biological significance of this is as yet unclear. Finally, HEF1 has been shown to be a target of hyper-ubiquitination in response to biological stimuli leading to degradation. Whether specific forms of HEF1 are preferentially ubiquitinated and degraded is currently under investigation.

Signaling Issue IV: Dynamic Substitution of Protein Forms

The gross physical composition of at least one Cas-family member, HEF1, is dynamically altered in response to changes in cell cycle and adhesion. At the beginning of mitosis, the protein is cleaved at a single caspase site (DLVD, at amino acids 360–363). The carboxy-terminal fragment released is rapidly degraded

through action of the proteasome, resulting in elimination of the full length form and substitution of a 55 kDa amino-terminal fragment. This fragment retains the SH3 domain and part of the SH2-binding site region, but lacks other conserved motifs. At the end of mitosis, the 55 kDa form of HEF1 is degraded, and replaced by resynthesized full length HEF1. Although p130Cas is not cleaved at mitosis, one group has reported that mitosis changes in phosphorylation of p130Cas and FAK render the two proteins unable to associate during this phase of cell cycle. Together, the cleavage of HEF1 and altered phosphorylation of p130Cas would be expected to eliminate the scaffolding function of these molecules, hence altering cellular signaling.

A still more complex situation is observed during the detachment of epithelial cells. In this case, part of the population of HEF1 is cleaved at two caspase sites—the DLVD motif utilized in mitosis and a DDYD motif located at amino acids 627–630. In this situation, proteasome clearance of HEF1 fragments also differs, so that full length HEF1 coexists with both amino-terminal (55 kDa) and carboxy-terminal (28 kDa) fragments. In this situation, one possibility is that full length HEF1 and its fragments might compete for binding partners, resulting in a naturally induced dominant negative situation: in fact, overexpression of the carboxy-terminal 28 kDa fragment in cells induces focal adhesion disassembly and cell rounding. Finally, the DDYD motif targeted for caspase cleavage exactly coincides with the unique site phosphorylated by FAK (DDYDYVHL) and bound by Src (issue III), suggesting crosstalk between HEF1 phosphorylation and cleavage.

Signaling Issue V: Dynamic Intracellular Localization

As noted in the previous section, HEF1 exists in a full length form in interphase cells, and as a truncated 55 kDa form in mitotic cells. These two forms exhibit strikingly different patterns of intracellular localization. While the full length form resides at focal adhesions, the truncated form is specifically associated with the mitotic spindle. The mechanism governing the translocation is not yet clear, nor is the physiological function, if any, of HEF1 associated with the spindle. However, given that the p55 form retains discrete protein interaction domains, one possibility is that it associates with a different set of partner molecules in this distinct intracellular milieu. Hence, truncated HEF1 may possess a second activity that is either distinct from, or integrated with, its action at focal adhesions.

Signaling Issue VI: Cell Type Specificity and Increasing Complexity

Each of the three Cas family members displays a different pattern of cell-type specific expression. For example, while p130Cas is expressed in many different cell types, HEF1 expression is abundant in epithelial cells derived from breast and lung, and in lymphoid cells, but not readily detectable in other cell types. The localization of Efs has been examined in less detail, but it appears that this protein is most abundant in the brain. Similarly, proteins that have been shown to interact with and/or modify the individual Cas proteins frequently belong to protein families and also demonstrate cell type-specific expression. For these reasons, it is to be expected that any individual cell type is going to possess a discrete population of possible Cas proteins, and Cas-associated proteins, that is not identical to that observed in other cell types. Further, the proteins that interact with Cas proteins in some contexts do not solely interact with Cas family members. As one example, while FAK interacts with Cas proteins, FAK also interacts with Src, phosphatidylinositol 3-kinase, Grb2, paxillin, ezrin, and other proteins. These additional interactions may compete with, complement, or synergize with the ability of FAK to interact with Cas proteins.

Signaling Issue VII: Defining the Root Molecular Cause of Pathology

Given the difficulty in developing a drug with defined specificity, the ultimate issue is identifying the most critical molecular lesion associated with a disease state. Taking cancer as an example, although in some cases, such as BCR-ABL induced chronic myelogenous leukemia, the root cause of the disease is clear, most cancers arise via the sequential addition of discrete genetic or regulatory changes, which collaborate to induce the disease. Specific genetic changes (altered proteins) may play an important role at one stage of tumor growth (for instance, survival as a single cell during anchorage-independent metastasis), then

become irrelevant or counterselected at a later stage (colonization of a remote site, and growth as a solid secondary tumor). Because different combinations of genetic alteration can function to achieve relatively similar tumor endpoints, and because early tumor stages where initiating changes occur are rarely detected, it is very difficult to make rigorous comparisons between clinical analyses and basic analysis of controlled signal transduction model systems.

Hence, while many of the properties identified as Cas-related (effects on motility, adhesion, and apoptosis, and interaction with oncogenes) imply a causal role in cancer, clear establishment of a role in in vivo human cancer or other disease states remains to be established. Recently, Cas proteins have been implicated as promoters of tumor metastasis, and their misexpression has been shown to confer resistance to anti-estrogens [Brinkman et al., 2000]. Cas proteins have been shown to complex with polycystin-1, encoded by the PKD-1 gene mutated in autosomal dominant polycystic kidney disease, and have been proposed to be an important component of a pathway perturbed in this syndrome [Wilson, 2001]. As signaling pathways continue to be elucidated, it is likely that Cas connections to further disease-related control networks will be identified. But are Cas proteins the critical protein to target for drug discovery aimed at blocking cancer? Would other proteins in the Cas-signaling network be preferred? Must multiple proteins in a Cas-centered network be blocked? These are not trivial questions.

BROADENING THE DISCUSSION: WHY CONSIDERATION OF COMPLEXITY MIGHT GENERALLY ENHANCE THERAPEUTIC APPROACHES

The preceding section makes the point that analyzing the function of Cas proteins is complex. In fact, it is likely to be still more complex, as the above tally does not include several other factors that have recently been implicated as regulating Cas protein function. In contemplating these issues as a precursor to designing a targeted drug strategy intended to modulate the function of specific disease-associated proteins, two questions arise. First, do the Cas proteins represent a worst-case scenario, and is the regulation of other signaling pathways more simple? Second, even if the biological regulation of most disease proteins is complex, is it really necessary to take this complexity into account in designing drugs, or does consideration of the full regulatory control networks for a target protein of interest simply confuse the issue?

In answer to the first question, it does not appear that the complexity of control mechanisms noted above are particularly unusual. Many proteins belong to protein families composed of members with partially overlapping function. Many are engaged in highly complex patterns of protein interaction, as described in microcosm for Ras in [Kolch, 2000], and in macrocosm for the yeast proteome in [Ito et al., 2000; Schwikowski et al., 2000]. Selective cleavage leading to translocation and/or degradation of proteins occurs in many disease-related pathways, including IkappaB/NF-kappaB signaling, SRE-BP signaling, and APC trafficking. Complex phosphorylation patterns associated with changes in protein–protein association, localization, stability, and activity are very commonly observed, as are the other control mechanisms noted. In fact, the Cas proteins are not (known to be) controlled by still other mechanisms important in different signaling systems. These other mechanisms include but are not limited to biological activity differences dependent on a protein being in a liganded versus non-liganded state (as for G-protein coupled receptors, or hormone receptors); intracellular retargeting due to prenylation (as for Ras), or mono-ubiquitination [Hicke, 2001]; surface masking due to glycosylation; surface charge alteration due to acetylation (as for histones); and regulation at the level of mRNA stability and translation (many examples). The list continues to expand.

Granted that signaling control mechanisms are generally complex, the degree to which information about signaling control mechanisms needs to be incorporated in the development of novel therapeutics is currently a critical issue. A number of public consortia and proteomics companies have vested extensive capital in the development of protein interaction maps, on the theory that this information will be valuable in the future. In considering these points, it is useful to note the current status of gene- or protein-targeted therapeutics: these fall into a number of different functional classes, which include as major strategies.

Drugs Which Alter the Stability of a Protein-Specified Structure

These include classic chemotherapeutic antimitotic agents such as taxol, which suppresses the turnover (dynamic instability) of tubulin at the ends of microtubules of the mitotic spindle; and destabilizing agents such as nocodazole and vinblastine, that block tubulin polymerization.

Catalytic Inhibitors

Many examples exist, targeted in particular on kinases, and increasingly, phosphatases. These usually function by binding and blocking the catalytic site of the targeted enzyme; Gleevec (STI571) is a particularly successful recent example of a chemotherapeutic agent directed against the BCR-ABL tyrosine kinase. Iressa (ZD1839), targeted against the intracellular kinase domain of the epidermal growth factor receptor (EGF-R), is another example of this class of agent.

Allosteric Inhibitors

Through binding to a surface of an enzyme, a small molecule can induce a conformational change that restricts the catalytic activity of the enzyme. Examples include polycitone A, as a reverse transcriptase inhibitor, or galantamine, as an acetylcholinesterase inhibitor, for Alzheimer's disease.

Agonists and Antagonists

Most growth factor or hormone receptors display altered activity based on binding to a peptidyl or small molecule ligand. Drugs, peptidomimetics, and in some cases monoclonal antibodies, can be designed that mimic the ligand, but have enhanced or opposing effects; bind the protein target together with the ligand, and modulate its activity; or alternatively, that restrict the access of an activating ligand to its binding site. The ImClone monoclonal antibody C225 is an example of an EGF-R antagonist that prevents ligand binding.

Protein "Removal" Approaches—Antisense and Antichaperones

Finally, an alternative approach is to try to reduce levels of proteins that through misexpression or mutationally altered activity are inducing a disease state. The large field of antisense biology works on this paradigm, targeting disease-inducing genes at the level of mRNA [Crooke, 2000]. Alternatively, some agents, such as radicicol or geldanomycin, target and restrict the activity of the chaperone protein hsp90 [Neckers et al., 1999]. This leads to loss of hsp90-mediated stabilization of structurally "fragile" proteins such as ErbB2 and Raf, causing these proteins to be degraded.

Certainly, each of these therapeutic approaches is promising, and has merit. However, upon consideration of the organization of signaling networks, some of these strategies also possess some intrinsic limitations. For example, a kinase of interest as a drug target is (1) likely to be highly related to other family members, particularly in the catalytic region, and (2) responsible for phosphorylating multiple different substrates, for different cellular functions. Hence, a drug discovery strategy based on kinase inhibition may lead to the identification of inhibitors that are not specific for their intended target, which may lead to general toxicity. Even if highly specific for a single target, such drugs may fail to result in a biological effect because other family members of the targeted kinase possess redundant function. Alternatively, strategies based on protein removal may have distant and non-specific signaling repercussions, as the removal of single targeted protein causes changes in extended protein–protein interaction networks.

In general, it would be useful to have a drug discovery strategy that took into account the facts that (1) proteins are involved in multiple interactions, and multiple signaling pathways; (2) that their natural function requires oscillation between different activity levels, in concert with different partners, and (3) that blockade of a signaling pathway may require simultaneous "blockade" of different parallel effector systems. Ideally, this strategy would enable the targeting of disease-related, but not essential, activities of the protein. Ideally, a pathway focused strategy would also develop multiple independent ways to target protein pathways of interest: thus, if mutational changes lead to resistance to a first agent (as is beginning to be observed, for example, with Gleevec, [Gorre et al., 2001]), then a second agent can be used to block signaling by binding elsewhere on the target protein, or by working further downstream on the same pathway. One way in which these goals might be facilitated is in targeting the dynamic regulation and protein–protein interactions of targets of interest.

DEVELOPING RESOURCES TO MANAGE RELEVANT INFORMATION

It is clear from what is known about the Cas proteins, Ras/Raf, and other proteins involved in signaling, that the interactions which need to be understood during drug development are myriad and complex. In addition to maintaining a complete list of interaction partners, domains and motifs, cellular location, and timing of expression, it is also necessary to maintain knowledge about how each specific protein modification such as phosphorylation and cleavage changes interactions and affects downstream events. In addition, the complexity of interaction leads inevitably to a high probability that some interactions will be misinterpreted, and that results which may later be found to be erroneous will enter the literature. All of these issues have major implications for the databases which track protein interaction data.

Currently, there are a number of databases which maintain information on protein interactions (see Table I), as well as many databases which link directly to them (e.g., the Saccharomyces Genome Database). For example, the KEGG database permits the user to search for specific genes or look at specific pathways graphically, with each interaction indicated by visual attachment of protein names. The pathway images have each protein linked back to genomic information. The Pathways database from Curagen provides similar information, though not in a graphical form, but with references to the source of the information in the primary literature. Of the databases in Table I, only the TRANSPATH and PROTEOME databases include phosphorylation information. Even in these cases, display is limited to showing when a protein is phosphorylated during its interaction. However, the arguments above show that it will be necessary to maintain much more detailed information in order to understand how proteins interact in complex pathways.

Presently all databases are designed for individual searches exploring in detail the interactions of a single protein. However, a different approach will become more important as the need to model the interactions increases. The importance of modeling becomes clear when we begin looking at pathway proteins and their interactions globally. Metabolic and signaling networks within cells are highly interconnected [Jeong et al., 2000]. The resulting systems form nonlinear networks in which specific stimuli and modifications result in unpredictable changes due to the interactions between subsystems in the organism (often within the same cell) [Bailey, 1999; Stephanopoulos and Kelleher, 2001]. However, network analysis [Albert et al., 2000], interaction modeling [Fussenegger et al., 2000], and the more ambitious cellular modeling [Tomita, 2001] can provide tools for constructing the biological models. The problem then becomes how to allow the models or even simpler automated analytic tools to use the databases in an efficient manner. There are multiple methods in which this could be accomplished. Ideally, a core data model could be used allowing uniform access methods, perhaps based on a minimal structural vocabulary [Kazic, 2000]. This will take some time to put into place, so a less ideal solution allowing access to existing databases is needed. A key feature for allowing access to the databases will be the provision of well-defined, stable interfaces (application programming

TABLE I. Websites Incorporating Proteomics Information

Database	Website
KEGG	http://www.genome.ad.jp/kegg/
SPAD	http:/www.grt.kyushu-u.ac.jp/spad/
CSNDB	http:/geo.nihs.go.jp/csndb/
MIPS	http:/www.mips.biochem.mpg.de/
Interactive fly	http://sdb.bio.purdue.edu/fly/aimain/laahome.htm
TRANSPATH	http://transpath.gbf.de/
Wnt homepage	http://www.stanford.edu/~rnusse/wntwindow.html
BRITE	http://www.genome.ad.jp/brite/
DIP	http://dip.doe-mbi.ucla.edu/
BIND	http://bioinfo.mshri.on.ca/
Pathcalling	http://portal.curagen.com/extpc/com.curagen.portal.servlet.Yeast
PROTEOME	http://www.proteome.com/
PIM	http://www.hybrigenics.fr/

interfaces, or APIs) which will provide other systems (modeling systems, web-based access systems) guidelines and methods for retrieving information quickly and with high throughput. Such a system would establish a so-called N-tier architecture, where data is gathered as needed from databases (tier 1) and brought to a computationally powerful system for modeling or analysis (tier 2), with the results displayed at an individual's computer (tier 3), perhaps in a Java-based web browser. The stability of the interfaces is critical, since constantly changing interfaces would cause retrieval systems to fail and become unmaintainable. With stable interfaces, an N-tier system allows for the inclusion of disparate data storage systems, so that the large number of databases under development can be utilized, and also allows data owners to grant limited access for cases where intellectual property matters or privacy concerns apply. In addition, the use of N-tier architecture will allow the inclusion in the modeling and analysis of data from other data sources, including drug discovery data, clinical data, pharmacogenomic data, genomics data, phenotypic data, genotyping data, etc. Over time, it is likely that a thorough understanding of cellular function, protein interactions, and disease development and prevention will require inclusion of these disparate data sources in the analysis.

The databases themselves will need to include a number of pieces of information to enhance the usefulness of the stored data. Most obviously it is clear that detailed information on phosphorylation states, cleavage sites, and the changes in interactions when such sites are modified, must be included. In addition, since errors are inevitable within the data itself, it is important that references to the primary literature for interactions be maintained. This is particularly critical since as the number of databases grow, the errors are likely to propagate from database to database. A future search tool may well look at all the databases, and while trying to determine the complex interactions for a protein, may try to intelligently weight the conflicting evidence. Without reference back to the original report of the interaction, a single erroneous report may be given excessive weight in an analysis and modeling as it can appear in numerous databases. Since the biological systems are nonlinear, small errors can lead to large problems in the models, so source tracking needs to be included.

Finally, a feature which will need to be implemented either within the databases or more likely in the middletier modeling and analysis tools is some form of visualization tool. What form that tool will take is unclear, although attempts to visualize the complex data are being made [Schwikowski et al., 2001]. The problem of visualizing inherently complex data and models has been faced before, and generally requires a dedication on the part of the users of the systems to learning a complex tool (e.g., Einstein himself did not follow Feynman diagrams when first faced with them). The interactions present between proteins in a cell are far more complex than those between virtual and real subatomic particles, so learning how to interpret visualizations will be an integral part of developing the visualizations themselves. Future users will need to accept a learning curve in using the new tools, as the data and models will be far more complex than those being generated today.

DECONVOLUTION AND PRACTICAL APPLICATION: FUTURE STRATEGIES

After acknowledging the complexity of signaling networks, and after assembling the relevant information in a searchable form that will accurately profile regulatory variables for key signaling proteins—then what? The ability to effectively utilize information tabulated in these database resources will require the development of reagents that can modulate the properties thus tabulated. For example, for the Cas proteins described above, there are several points that might be expected to specifically and selectively regulate function. One approach would be to block the access of the FAK protein to the YDYVHL motif whose phosphorylation initiates the signaling cascade. A second approach would be to develop reagents that block the mitosis and/or apoptosis-specific cleavages of HEF1, based on binding HEF1 in the vicinity of the cleavage site. A third approach might be to identify agents that modulate the interaction of HEF1—but not p130Cas—with shared partner molecules such as FAK or Crk.

What are the key issues? Certainly, given the number of potential modifications and interactions pertaining to a protein target, a starting place must be the determination of which

specific properties govern critical physiological activities of the target. The site of the identified critical modifications must be suitable for targeting with small molecule or peptide/antibody reagents (e.g., must present a "druggable" surface). Coupled with this functional analysis, it is critical to have the technological ability to create reagents that allow targeting of a site of interest. Finally, while the exact targeting of individual protein modifications or interactions might offer unprecedented specificity, depending on the level of signaling redundancy, this may undesirably limit drug efficacy. For example, if the interaction of FAK with p130Cas, or FAK with HEF1, promotes motility, a compound which only disrupted the FAK-p130Cas interaction might be inadequate to block the induction of motility induced by FAK. In such a case, either a less specific compound (which blocks the interaction of FAK with both p130Cas and HEF1) or a combination of two or more compounds separately targeting interactions must be utilized. In every case, a protein-targeted strategy requires a comprehensive knowledge of the signaling pathways related to the target protein either directly or indirectly. Finally, given that "downstream" components of signaling pathways are generally activated by and effectors of multiple discrete "upstream" initiating proteins, logically, targeting the upstream proteins seems far more likely to generate a limited specific result than targeting the downstream effectors.

Are protein interaction or modification targeted approaches likely to be generally feasible? At this point, it is difficult to know, as it is hard to generate accurate figures concerning how much effort has gone into screening for small molecules that disrupt protein–protein interactions. The steep curve of development of both the enabling technologies and an adequate information base for signal transduction has only made selective screening plausible within the last ~5 years. The preponderance of such screening is likely to have occurred in industrial settings, where publication is frequently delayed, rather than in academic labs: in either setting, unsuccessful results are rarely publicly aired.

Against this background of necessary cautions, a number of reports have begun to show successes in adapting protein engineering technologies to specifically manipulate the cell signaling machinery. For example, the yeast two-hybrid system, originally designed to identify protein–protein interactions, has been increasingly modified to serve as a means to study and perturb the interaction of proteins with small molecules and peptides [reviewed in Serebriiskii et al., 2001a]. Several recent papers have reported the identification of small molecules or peptides that disrupt [Huang and Schreiber, 1997; Kolonin and Finley, 1998; Young et al., 1998; Degterev et al., 2001; Sharma et al., 2001; Tilley et al., 2001] or promote [Zhang et al., 1999] protein–protein interactions, using the two-hybrid system and other screening technologies. Other reports have described peptides capable of controlling target protein localization [Colas et al., 2000] or cleavage [Zhang et al., 2001]. Supporting such efforts, we and others have found that the two-hybrid system can be modified [for example, Serebriiskii et al., 1999] for use as a tool to deconvolute disease-related signaling, in applications including the detection of proteins that interact with mutated vs. wild type forms of target proteins [Serebriiskii et al., 2001b], or in generating and selecting novel mutations that can uncouple a protein from one of two or more partner molecules [Reeder et al., 2001]. The information on signaling specificity thus obtained can in turn be used to guide the development of small molecules which might target specific protein interactions. Although the effects described in these reports are quite specific and encouraging, the mode of action of identified compounds and mutations at the level of protein molecular structure remains to be determined. This information will undoubtedly clarify the key question of whether similar results and screening strategies are generalizable to multiple protein targets, as discussed elsewhere in this issue. At present, there is cause for optimism that just as a better understanding of signal transduction has provided increased clarity to the process of drug development, the future development of targeted drugs will be able to intelligently reprogram signal transduction, and ultimately reduce the burden of human disease.

ACKNOWLEDGMENTS

We thank Adam Arkin, Gary Hudes, and Ilya Serebriiskii for critical commentary and helpful discussions. The authors are supported by core grant NIH CA06927 and Pew Foundation

Funds to Fox Chase Cancer Center. Work related to HEF1/Cas signaling is supported by NIH CA63366 (to E.A.G.).

REFERENCES

Albert R, Jeong H, Barabasi AL. 2000. Error and attack tolerance of complex networks. Nature 406:378–382.

Bailey JE. 1999. Lessons from metabolic engineering for functional genomics and drug discovery. Nat Biotechnol 17:616–618.

Brinkman A, van Der Flier S, Kok EM, Dorssers LC. 2000. BCAR1, a human homologue of the adapter protein p130Cas, and antiestrogen resistance in breast cancer cells. J Natl Cancer Inst 92:112–120.

Colas P, Cohen B, Ferrigno PK, Silver PA, Brent R. 2000. Targeted modification and transportation of cellular proteins. Proc Natl Acad Sci USA 97:13720–13725.

Crooke ST. 2000. Potential roles of antisense technology in cancer chemotherapy. Oncogene 19:6651–6659.

Degterev A, Lugovskoy A, Cardone M, Mulley B, Wagner G, Mitchison T, Yuan J. 2001. Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-xL. Nat Cell Biol 3:173–182.

Fussenegger M, Bailey JE, Varner J. 2000. A mathematical model of caspase function in apoptosis. Nat Biotechnol 18:768–774.

Giancotti FG, Ruoslahti E. 1999. Integrin signaling. Science 285:1028–1032.

Gorre ME, Mohammed M, Ellwood K, Hsu N, Paquette R, Rao PN, Sawyers CL. 2001. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. Science 293:876–880.

Hicke L. 2001. Protein regulation by monoubiquitin. Nat Rev Mol Cell Biol 2:195–201.

Huang J, Schreiber SL. 1997. A yeast genetic system for selecting small molecule inhibitors of protein–protein interactions in nanodroplets. Proc Nat Acad Sci USA 94:13396–13401.

Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S, Sakaki Y. 2000. Toward a protein–protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc Natl Acad Sci USA 97:1143–1147.

Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. 2000. The large-scale organization of metabolic networks. Nature 407:651–654.

Kazic T. 2000. Semiotes: a semantics for sharing. Bioinformatics 16:1129–1144.

Kolch W. 2000. Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions. Biochem J 351:289–305.

Kolonin MG, Finley RL Jr. 1998. Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. Proc Natl Acad Sci USA 95:14266–14271.

Mercurio AM, Rabinovitz I. 2001. Towards a mechanistic understanding of tumor invasion—lessons from the alpha6beta 4 integrin. Semin Cancer Biol 11:129–141.

Neckers L, Schulte TW, Mimnaugh E. 1999. Geldanamycin as a potential anti-cancer agent: its molecular target and biochemical activity. Invest New Drugs 17:361–373.

O'Neill GM, Fashena SJ, Golemis EA. 2000. Integrin signaling: a new Cas(t) of characters enters the stage. Trends Cell Biol 10:111–119.

Reeder MK, Serebriiskii IG, Golemis EA, Chernoff J. 2001. Analysis of small GTPase signaling pathways using Pak1 mutants that selectively couple to Cdc42. J Biol Chem 276:40606–40613.

Schwikowski B, Uetz P, Fields S. 2000. A network of protein–protein interactions in yeast. Nat Biotechnol 18:1257–1261.

Schwikowski B, Ideker T, Uetz P. 2001. Chapter 19: Visualization and integration of protein–protein interactions. In: Golemis EA, editor. Protein Interactions. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Serebriiskii I, Khazak V, Golemis EA. 1999. A two-hybrid dual bait system to discriminate specificity of protein interactions. J Biol Chem 274:17080–17087.

Serebriiskii IG, Khazak V, Golemis EA. 2001a. Redefinition of the yeast two-hybrid system in dialogue with changing priorities in biological research. BioTechniques 30:634–655.

Serebriiskii IG, Mitina OV, Chernoff J, Golemis EA. 2001b. Use of a two-hybrid dual bait system to discriminate specificity of protein interactions in small GTPases. Methods Enzym 332:277–300.

Sharma SV, Oneyama C, Yamashita Y, Nakano H, Sugawara K, Hamada M, Kosaka N, Tamaoki T. 2001. UCS15A, a non-kinase inhibitor of Src signal transduction. Oncogene 20:2068–2079.

Stephanopoulos G, Kelleher J. 2001. Biochemistry. How to make a superior cell. Science 292:2024–2025.

Tilley JW, Kaplan G, Rowan K, Schwinge V, Wolitzky B. 2001. Imide and lactam derivatives of N-benzylpyroglutamyl-L-phenylalanine as VCAM/VLA-4 antagonists. Bioorg Med Chem Lett 11:1–4.

Tomita M. 2001. Whole-cell simulation: a grand challenge of the 21st century. Trends Biotechnol 19:205–210.

Wilson PD. 2001. Polycystin: new aspects of structure, function, and regulation. J Am Soc Nephrol 12:834–845.

Young K, Lin S, Sun L, Lee E, Modi M, Hellings S, Husbands M, Ozenberger B, Franco R. 1998. Identification of a calcium channel modulator using a high throughput yeast two-hybrid screen. Nat Biotechnol 16:946–950.

Zhang Z, Murphy A, Hu JC, Kodadek T. 1999. Genetic selection of short peptides that support protein oligomerization in vivo. Curr Biol 9:417–420.

Zhang Z, Ly T, Kodadek T. 2001. An inhibitor of sequence-specific proteolysis that targets the substrate rather than the enzyme. Chem Biol 8:391–397.