

# Activity-based proteomics: enzyme chemistry redux

John W Kozarich

The principles of enzyme chemistry, mechanism of action and inhibitor design are being applied to proteomics by the development of activity-based probes. This approach suggests a potentially broad method for interrogating enzyme family members, both known and unknown, in cells and proteomic fractions without the need for individual assay development and isolation. The serine hydrolases and cysteine proteases have provided the proofs of concept for activity-based proteomics, and other studies are rapidly following. The result will be a proteomics technology of great value to drug discovery and development.

## Addresses

ActivX Biosciences, 11025 North Torrey Pines Road,  
La Jolla, CA 92037, USA

**Current Opinion in Chemical Biology** 2003, 7:78–83

This review comes from the themed issue on  
Proteomics and genomics  
Edited by Matthew Bogyo and James Hurley

1367-5931/03/\$ – see front matter  
© 2003 Elsevier Science Ltd. All rights reserved.

DOI 10.1016/S1367-5931(02)00013-3

## Abbreviation

ABP activity-based proteomics

## Introduction

With the human genome essentially complete [1,2], increasing attention has been focused on the proteome—the unique subset of genomically encoded proteins and their post-translational variants that is present in specific tissues and cells under specific physiological conditions [3]. While all cells of an organism are defined by the same genome, the prospect of thousands of unique proteomes defining the functional nuances of these cells adds a nightmarish dimension to complexity of proteomic analysis. Because >100 000 proteins are expected from the genome and millions of protein–protein interactions are possible in protein complexes and networks [4–7], fractionating the proteome into manageable and physiologically relevant subsets is critical. Moreover, the lack of a consistent correlation between individual mRNA abundance and protein levels does not permit an easy extrapolation from gene expression data [8,9].

Many methods for fractionating of the proteome are possible (e.g. by chromatographic properties—hydrophobicity, isoelectric point, molecular weight—by

subcellular organelle isolation, or by isolations of multi-protein complexes) [10,11]. Although these approaches have value, they do not, in general, fractionate proteins on the basis of common biochemical functions. This review analyzes the recent advances in activity-based proteomics, an approach whose goal is to fractionate the proteome into families of proteins that are united by common mechanistic functions in their active form. Since many of these functional families of proteins are the subjects of intense drug-discovery research by the pharmaceutical industry, we argue that this approach to proteomic simplification has the greatest potential to effectively integrate proteomics into the drug discovery and development process.

## Importance of activity-based proteomics

The establishment of proteomics as a high-value component of drug discovery and development has been problematic. The simple reason for this is that proteomics researchers have focused on the analytical methodology of cataloging proteins and measuring their abundance while pharmaceutical researchers need information of the functional relationships of proteins. This proteomics/pharmaceutical disconnect is understandable because proteomics evolved as an important discipline grounded in the development of sensitive analytical methods to separate, quantify and identify proteins. Thus, two-dimensional gel electrophoresis/mass spectrometry emerged as the gold standard of the first-generation of proteomic efforts, although limitations are apparent [11–15]. Drug discovery and development, on the other hand, is a functional science; drugs are designed largely to modulate protein function and, in many cases, the absolute abundance of a particular protein target is a secondary issue. The pharmaceutical researcher ideally desires a broad assessment of protein function from the standpoint of functionally related proteins (target discovery) and of the effects of drug leads on the target protein and functionally related proteins (drug selectivity). The ability to effectively deliver this type of information should be a major goal of any pharmaceutically integrated proteomics effort [16].

Activity-based proteomics (ABP) has emerged as a compelling approach to address the functional shortfall on other proteomics methods and constitutes a major strategy for functional proteomics [17,18]. The hypothesis underlying ABP is tantalizingly simple: design a specific chemical reagent that has the ability to recognize a conserved catalytic/functional/structural motif in a family of active proteins, to react with the target proteins in a covalent manner based on its conserved motif and to

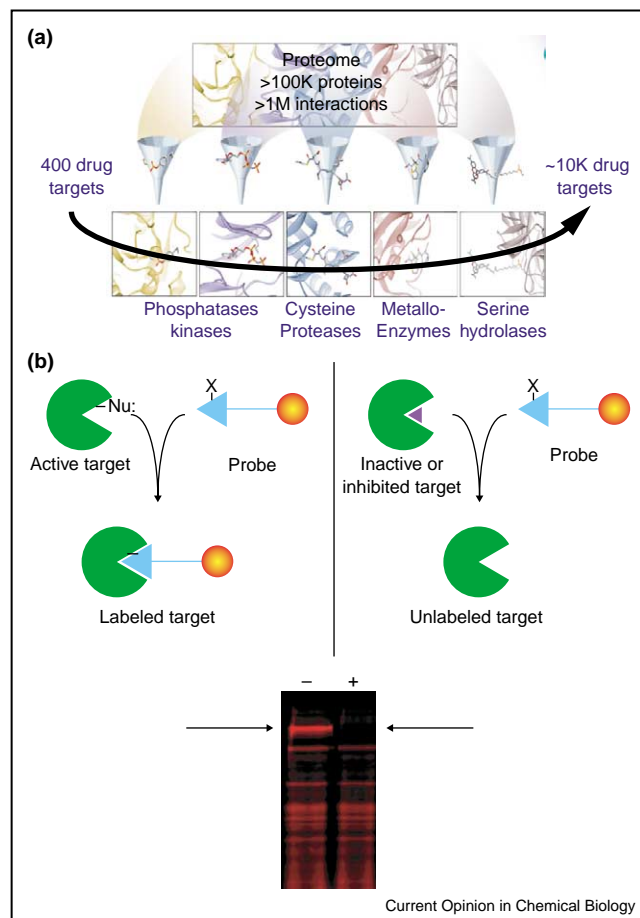
quantify the amount of each active protein. A series of chemical probes would permit a selective fractionation of a proteome on the basis of functional relatedness of proteins (Figure 1). The covalently tagged proteins could then be captured, quantified and identified by chromatographic, electrophoretic and mass spectrometric techniques. Thus, one could conceivably identify all of the active proteins of pharmaceutically important families such as serine hydrolases, kinases and phosphatases. In addition, a well-designed probe could, in principle, permit identification of novel family members that have eluded functional annotation by computational search algorithms because of low sequence homologies. Finally, the same probe would be useful across species since such motifs tend to be conserved in other organisms.

At first glance, the direct application of ABP to enzymes is obvious. Enzymes make and break chemical bonds as part of their function and contain conserved catalytic motifs that should be accessible by chemical probes. In addition, the use of chemical tools as a means to understand enzymes has been a mainstay of the field for more than half a century. Much of that research focused on designing highly selective chemical probes to understand the function and structure of a specific protein. The challenge for use of chemical tools in the setting of the proteome is to synthetically generalize the structure of the chemical probe so that it identifies potentially hundreds of family member instead of just one.

### Proofs-of-concepts for ABP

The serine and cysteine hydrolases have provided ideal platforms for proof-of-concept studies. Both classes effect catalysis by the transient formation of a covalent enzyme-substrate complex with an active site serine or cysteine that is a reactive nucleophile susceptible to electrophilic labeling by a chemical probe. In addition, the active site architecture is highly conserved across the family with respect to the catalytic center. The fluorophosphonate probe (Figure 2) for serine hydrolases (FP-biotin) pioneered by the Cravatt laboratory is a remarkable example of a near-perfect chemical probe for activity-based proteomics [19]. The fluorophosphonate group is an electrophilic group that has demonstrated high selectivity for the activated serine in many serine hydrolases, resulting in covalent phosphorylation of this serine. The linker region is designed to provide appropriate spacing for the biotin to permit unencumbered access to the active site of the functionally diverse members of the family. Variation of the linker group is useful in minimizing hydrophobic interactions found in straight-chain alkyl linkers. The biotin is useful for visualization and capture of covalently labeled proteins by standard avidin-based methods. There is nothing structurally obvious about FP-biotin that would suggest any selectivity for a specific member of the serine hydrolase family, which spans

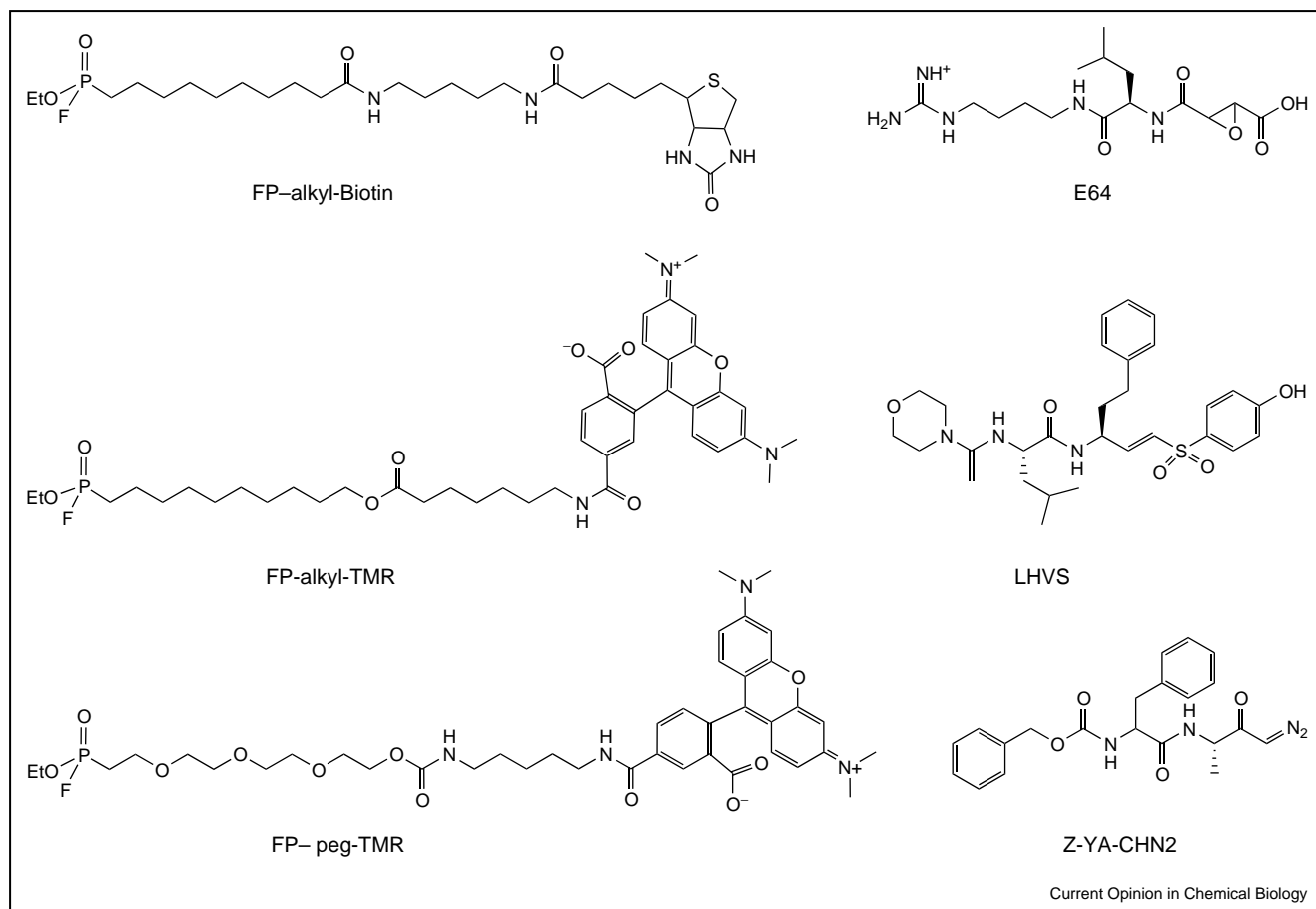
Figure 1



Principles of activity-based proteomics. (a) Fractionating the proteome into simple and useful subsets is a goal of many proteomic technologies. The underlying principle of activity-based proteomics is the design of chemical probes that recognize a broadly conserved functional motif in a family of active proteins and then covalently tag those proteins for subsequent detection and identification. Chemical probes that can fractionate the proteome into functional families of high pharmaceutical impact, such as the families shown, are highly desirable. The probes are derived from synthetic small-molecule inhibitors, natural products and endogenous substrates. Ultimately they may accelerate the target discovery from the ~400 known drug targets to the speculated thousands that may exist. (b) The chemical probes are generally designed to contain an electrophilic group (X) that is positioned to react with a conserved nucleophilic amino acid residue (Nu:) that is important for family function and/or binding of the active proteins. The probe also contains a fluorescent group or capture group, such as biotin, that is attached to the electrophilic recognition group via a linker. The covalently labeled target may be analyzed by gel electrophoresis and the fluorescent proteins visualized and quantified. Since inhibitors and drugs against specific proteins usually bind to the same site that the chemical probes bind to, kinetic competition with the inhibitor will result in a decrease in probe binding as evidenced by a decrease in fluorescence in the appropriate protein band (see arrows). Other protein family members showing a decrease in probe binding signify off-target binding activity of the inhibitor.

at least 1000 proteins including proteases, amidases, esterases and lipases. Indeed, it is the relatively weak non-covalent recognition of the probe for serine hydro-

Figure 2



Prototypic probe structures for serine hydrolases and cysteine proteases. FP-alkyl-biotin, FP-alkyl-TMR and FP-peg-TMR have been extensively studied as probes for analyzing serine hydrolases in complex proteomes. E-64 is a natural peptide epoxide that inhibits most cysteine proteases by alkylation of the catalytic cysteine and has served as the model for other epoxide-based chemical probes. The vinyl sulfone LHVS and the diazoketone Z-YA-CHN2 are synthetic probes using alternative electrophiles to covalently label cysteine proteases.

lases in general that ensures broad family coverage by the probe. The substitution of biotin with a fluorescent group as well as trifunctional probes incorporating both moieties are proving useful for direct visualization and quantification of active proteins, as well as capture of tagged proteins [20\*,21]. Bogoy and co-workers [22\*] have successfully explored a similar design approach using epoxide electrophiles (Figure 2) for profiling cysteine proteases.

Given the sheer size and diversity of the serine hydrolase family and the pivotal role these enzymes play in developmental, physiological and disease processes, the family has provided a rich environment for highlighting the power of activity-based profiling. Using FP-biotin, Cravatt and co-workers [19] demonstrated both tissue and expression level changes of serine hydrolases in soluble rat proteomes of liver, brain, testis and prostate. Consistent with the requirement for an accessible and activated serine for probe reactivity, either the inclusion of a serine

protease inhibitory protein to occlude the active site or heat denaturation resulted in loss of covalent labeling. In addition, the identification of novel hydrolases was observed, such as potential functional (not necessarily high sequence-related) orthologues of human prostate specific antigen (PSA) in the rat prostate.

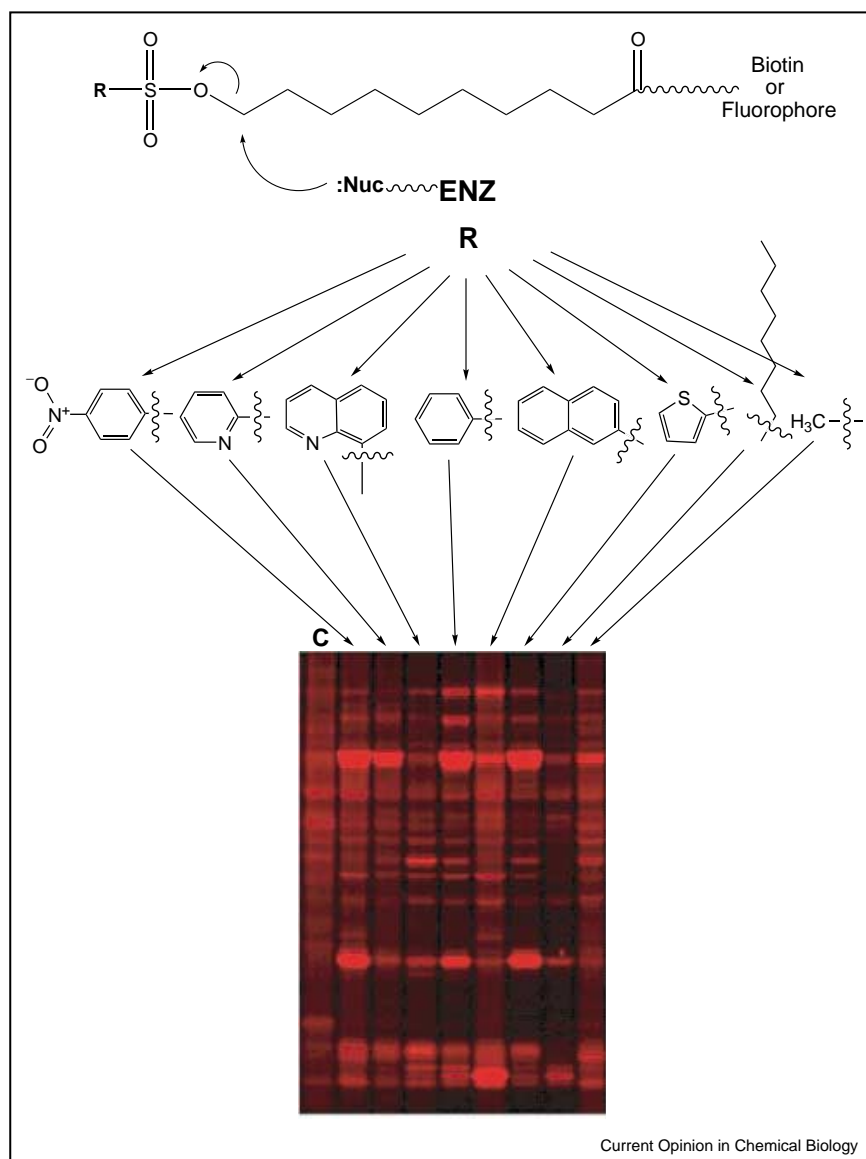
The issues of probe kinetics and of linker structure to modulate probe reactivity and the competition between the probe and a reversible enzyme inhibitor have also been explored by the Cravatt group [23\*\*]. Using FP-biotin and the corresponding FP-peg-biotin (Figure 2) containing the more hydrophilic poly(ethylene glycol) linker, it was demonstrated that the ability to detect competition between the probe and a reversible serine hydrolase inhibitor, a trifluoromethylketone, was directly affected by the rate of probe binding. This is entirely consistent with the kinetics of competition between an irreversible inhibitor (the probe) and a reversible one (the trifluor-

omethylketone): optimal competition is observed in the linear kinetic regime for the irreversible probe. In addition, the inhibition of probe labeling could be observed in multiple serine hydrolases by the putatively specific trifluoromethylketone. This observation has great potential for drug selectivity analysis. Since probes are designed to target sites that drugs are also designed to bind to, the pan-substrate properties of the probes can be used to broadly interrogate drugs for specific protein family mem-

bers against the proteomic repertoire of all family members present in a proteome under appropriate kinetic conditions. The opportunities for exploiting this phenomenon to determine off-target potential of drug candidates is enormous.

Recent elegant work from the Cravatt laboratory has attempted to integrate ABP profiling with cluster analysis methodologies to establish correlates to pathogenicity of

Figure 3



Non-directed activity-based probe strategy in proteomic profiling. Design of probes that explore combinations of electrophilic groups and recognition elements to access protein families has been nicely exploited by Cravatt and Sorensen [29\*\*,30\*\*]. They prepared a small library of sulfonate esters that would be susceptible to displacement by an enzyme nucleophile (:Nuc) resulting in covalent attachment of the linker and fluorescent group to the protein. A series of hydrophobic groups (R) was used to explore any proteomic recognition by the corresponding sulfonate esters. Analysis of the individual probes in proteomes by gel electrophoresis reveals kinetic and specificity differences among the probes and individual proteins. Identification of proteins tagged by specific probes can suggest underlying functional protein motifs that may be interrogated family-wide by a probe of specific structure. Thus, the probes were evaluated for utility by their behavior in proteomes of interest. The fluorescent gel is of the soluble heart proteome and is reproduced courtesy of *Nature Biotechnology* with permission [30\*\*].

various breast and melanoma cells lines [24\*\*]. They were able to demonstrate subtypes of these cell lines based on higher-order cellular properties such as tissue of origin and state of invasiveness, clustered significantly on the basis of secreted and membrane serine hydrolases but not cytosolic family members. Thus, the potential to derive meaningful diagnostic and prognostic markers from ABP profiling is high.

The application of ABP to the important, albeit smaller, family of cysteine proteases, such as the cathepsins, has been the focus of important studies from the Bogoy group. In an extension of the earlier work of Bogoy and Ploegh in the proteasome, a series of radiolabeled and biotinylated vinyl sulfone and epoxide substrate analogs were shown to target lysosomal cysteine proteases, such as cathepsins B, L and S [22\*,25]. Selective probes for cathepsin B revealed significant increases in this enzyme in metastatic tumors. Inhibitor specificity was also studied using the competition approach discussed above. Recently, these studies have expanded to include fluorescently labeled probes, a more comprehensive analysis of other cathepsin family members and the analysis of positional scanning libraries of epoxide inhibitors to optimize cathepsin B inhibition in rat proteome lysates [26\*\*,27\*\*]. Overall, this work clearly demonstrates the utility of ABP in the cysteine hydrolase family.

A recent technological variation has been reported using selective cysteine protease probes that are individually coded via tethering to a peptide nucleic acid sequence [28\*]. Capture of the protein-probe complexes can be effected on an oligonucleotide microarray containing complementary oligonucleotide sequences, thus permitting spatial identification and quantification of individual family members. Because the optimal exploitation of the array requires many probes of high selectivity for individual proteins, the approach seems problematic for broadly reactive probes, such as the fluorophosphonates for serine hydrolases that can label many family members in a single sample. Other capture and separation methods will be more useful for these types of probes.

### Non-directed strategies for probe design

While the design of the serine and cysteine hydrolase probes were founded in specific knowledge of the types of family-based inhibitors that could translate into useful activity-based probes, non-directed strategies for capturing less predictable protein families have been beautifully exploited by Cravatt and Sorensen [29\*\*]. They prepared a small library of biotinylated sulfonate esters from the corresponding aryl and alkylsulfonyl chlorides (Figure 3). Thus, the binding effects of the aryl and alkyl substituent could be evaluated in proteomes to establish individual probe utility. They found that a pyridylsulfonate probe effected an activity-based labeling of a class I aldehyde dehydrogenase. Recent work using rhodamine-tagged

sulfonate esters for protein detection in conjunction with the biotinylated analogs have resulted in several additional protein families reacting with this library, including thiolases, aldehyde dehydrogenases, NAD/NADP-dependent oxidoreductases, enoyl CoA hydratases, epoxide hydrolases and glutathione *S*-transferases [30\*\*]. Although it is unclear whether these enzymes share some common structural or functional motif, they are united by the occurrence of probe-susceptible amino acid nucleophiles, such as cysteine and aspartate, in the active site, which may be the target of these sulfonate esters. Significantly, these enzymes had not been previously identified by other activity-based probes, suggesting that the non-directed approach has considerable potential to interrogate new regions of proteomic space.

### Future directions

The promising work discussed above has stimulated efforts in several laboratories to develop probes for other enzyme families. Recent creative strategies in probe design for the deubiquitinating enzymes and for the receptor tyrosine phosphatases have been reported [31\*\*,32,33,34\*]. In addition, numerous unpublished reports of progress in the development of probes for kinases and other nucleotide-dependent proteins [35,36] and metalloproteinases suggest that the approach has many avenues left to explore. Indeed, the current focus on enzymes for obvious mechanistic reasons may obscure a broader vision for the use of such probes in other protein families where small molecule interactions are functionally important, such as receptors and ion channels. The strategy may ultimately be limited only by the creativity of probe design.

### Acknowledgements

I am grateful for numerous important discussions with Professor Ben Cravatt and with the talented scientists of ActivX Biosciences who have stimulated my interest in this exciting area of proteomics research.

### References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
  - of outstanding interest
1. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA *et al.*: **The sequence of the human genome.** *Science* 2001, **291**:1304-1351.
  2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
  3. Pandey A, Mann M: **Proteomics to study genes and genomes.** *Nature* 2000, **405**:837-846.
  4. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P *et al.*: **A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*.** *Nature* 2000, **403**:623-627.
  5. Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S, Sakaki Y: **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* 2000, **97**:1143-1147.

6. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y: **A comprehensive two-hybrid analysis to explore the yeast protein interactome.** *Proc Natl Acad Sci USA* 2001, **98**:4569-4574.
7. Mrowka R, Patzak A, Herzel H: **Is there a bias in proteome research?** *Genome Res* 2001, **11**:1971-1973.
8. Anderson L, Seilhamer J: **A comparison of selected mRNA and protein abundances in human liver.** *Electrophoresis* 1997, **18**:533-537.
9. Gygi SP, Rochon Y, Franza BR, Aebersold R: **Correlation between protein and mRNA abundance in yeast.** *Mol Cell Biol* 1999, **19**:1720-1730.
10. Issaq HJ: **The role of separation science in proteomics research.** *Electrophoresis* 2001, **22**:3629-3638.
11. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC: **The dynamic range of protein expression: a challenge for proteomic research.** *Electrophoresis* 2000, **21**:1104-1115.
12. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R: **Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology.** *Proc Natl Acad Sci USA* 2000, **97**:9390-9395.
13. Mann M, Hendrickson RC, Pandey A: **Analysis of proteins and proteomes by mass spectrometry.** *Annu Rev Biochem* 2001, **70**:437-473.
14. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R: **Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.** *Nat Biotechnol* 1999, **17**:994-999.
15. Washburn MP, Wolters D, Yates JR III: **Large-scale analysis of the yeast proteome by multidimensional protein identification technology.** *Nat Biotechnol* 2001, **19**:242-247.
16. Drews J: **Drug discovery: a historical perspective.** *Science* 2000, **287**:1960-1964.
17. Adam GC, Sorensen EJ, Cravatt BF: **Chemical strategies for functional proteomics.** *Mol Cell Proteomics* 2002, **1**:781-790.  
This is an excellent review of the broader functional proteomics landscape and the use of a variety of chemical approaches to fractionate the proteome in functionally useful ways.
18. Cravatt BF, Sorensen EJ: **Chemical strategies for the global analysis of protein function.** *Curr Opin Chem Biol* 2000, **4**:663-668.
19. Liu Y, Patricelli MP, Cravatt BF: **Activity-based protein profiling: the serine hydrolases.** *Proc Natl Acad Sci USA* 1999, **96**:14694-14699.
20. Patricelli MP, Giang DK, Stamp LM, Burbaum JJ: **Direct visualization of serine hydrolyase activities in complex proteomes using fluorescent active site-directed probes.** *Proteomics* 2001, **1**:1067-1071.  
This paper describes the useful development of fluorescent serine hydrolyase probes that are valuable for high-throughput protein detection and quantification.
21. Adam GC, Sorensen EJ, Cravatt BF: **Trifunctional chemical probes for the consolidated detection and identification of enzyme activities from complex proteomes.** *Mol Cell Proteomics* 2002, **1**:828-835.
22. Bogoyo M, Verhelst S, Bellingard-Dubouchaud V, Toba S, Greenbaum D: **Selective targeting of lysosomal cysteine proteases with radiolabeled electrophilic substrate analogs.** *Chem Biol* 2000, **7**:27-38.  
This paper demonstrates the utility of a variety of cathepsin reactive probes albeit in the limited setting of radioisotopic analysis.
23. Kidd D, Liu Y, Cravatt BF: **Profiling serine hydrolase activities in complex proteomes.** *Biochemistry* 2001, **40**:4005-4015.  
This important paper introduces several specificity and kinetic issues that must be carefully considered in the evaluation of activity probes for proteomic analysis. In addition, the breadth of detection and identification of serine hydrolases in cytosol and membranes is nicely highlighted, as is the potential for drug selectivity analysis by kinetic modulation of the probe reactivity.
24. Jessani N, Liu Y, Humphrey M, Cravatt BF: **Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness.** *Proc Natl Acad Sci USA* 2002, **99**:10335-10340.  
This impressive paper interprets the serine hydrolase activity profiles of a panel of breast and melanoma cell lines by cluster analysis. The broad-based analysis integrates activity, subcellular distribution and glycosylation state to identify a cluster of serine hydrolases that distinguish cell line based upon tissue of origin. A unique set of hydrolases was determined to be highly correlated to invasiveness. The potential for the diagnosis and treatment of cancers by an activity-based approach is clearly demonstrated.
25. Bogoyo M, Shin S, McMaster JS, Ploegh H: **Substrate binding and sequence preference of the proteasome revealed by active-site-directed affinity probes.** *Chem Biol* 1998, **5**:307-320.
26. Greenbaum D, Medzhiradzky KF, Burlingame A, Bogoyo M: **Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools.** *Chem Biol* 2000, **7**:569-581.  
This paper nicely demonstrates the utility of libraries of epoxide affinity probes for cysteine proteases for obtaining detailed information on the function and inhibitor specificity without the need for prior purification or identification of proteins.
27. Greenbaum D, Baruch A, Hayrapetian L, Darula Z, Burlingame A, Medzhiradzky KF, Bogoyo M: **Chemical approaches for functionally probing the proteome.** *Mol Cell Proteomics* 2002, **1**:60-68.  
This paper continues a detailed study of the cathepsins using activity-based fluorescent probes and introduces their use with positional scanning libraries of peptide inhibitors to assess target protein specificity.
28. Winssinger N, Ficarro S, Schultz PG, Harris JL: **Profiling protein function with small molecule arrays.** *Proc Natl Acad Sci USA* 2002, **99**:11139-11144.  
This paper reports an interesting application of microarray technology with small libraries of specific probes to cysteine proteases for the capture, quantitation and identification of target proteins.
29. Adam GC, Cravatt BF, Sorensen EJ: **Profiling the specific reactivity of the proteome with non-directed activity-based probes.** *Chem Biol* 2001, **8**:81-95.  
This is the first report of the use of libraries of probes that explore recognition and electrophilic design features to identify probes with interesting and specific proteomic targeting capabilities.
30. Adam GC, Sorensen EJ, Cravatt BF: **Proteomic profiling of mechanistically distinct enzyme classes using a common chemotype.** *Nat Biotechnol* 2002, **20**:805-809.  
This excellent paper further develops the concepts underlying the non-directed approach to probe design.
31. Borodovsky A, Ovaa H, Kolli N, Gan-Erdene T, Wilkinson KD, Ploegh HL, Kessler BM: **Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family.** *Chem Biol* 2002, **9**:1149-1159.  
This paper reports on a novel application of intein-based chemical ligation to generate a library of electrophilic probes for the family of deubiquitinating enzymes, demonstrating the breadth of design approaches and protein families that are under investigation with activity-based probes.
32. Born TL, Myers JK, Widlanski TS, Rusnak F: **4-(Fluoromethyl)phenyl phosphate acts as a mechanism-based inhibitor of calcineurin.** *J Biol Chem* 1995, **270**:25651-25655.
33. Myers JK, Widlanski TS: **Mechanism-based inactivation of prostatic acid phosphatase.** *Science* 1993, **262**:1451-1453.
34. Lo L-C, Pang T-L, Kuo C-H, Chiang Y-L, Wang H-Y, Lin J-J: **Design and synthesis of class-selective activity probes for protein tyrosine phosphatases.** *J Proteom Res* 2002, **1**:35-40.  
This paper reports on an interesting first attempt to apply the design concepts articulated in [33] and [34] to activity-based probes for these important enzymes.
35. Colman RF: **Chemical arrows for enzymatic targets.** *FASEB J* 1997, **11**:217-226.
36. Colman RF: **Affinity labels for NAD(P)-specific sites.** *Methods Enzymol* 1997, **280**:186-203.