Chemistry & Biology, Vol. 11, 593–597, May, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2004.05.001

Target Identification in Chemical Genetics: The (Often) Missing Link

Overview

Lyle Burdine and Thomas Kodadek*
Departments of Internal Medicine and Molecular Biology
Center for Biomedical Inventions
University of Texas Southwestern Medical Center
5323 Harry Hines Boulevard
Dallas, Texas 75390

Chemists and biologists have been using bioactive small molecules to manipulate biological pathways for a long time. Until quite recently, by far the most common source of molecules with potent bioactivity was nature. Things are changing, though. While natural products are, and will remain, an important source of drug leads and research tools, chemical genetics [1–3] is increasingly being employed to identify compounds with interesting activities from compound collections or combinatorial libraries. This is usually done by arraying compounds (typically a few hundred to a few thousand) in the wells of microtiter plates [4] along with cells or organisms engineered to report a desired phenotype that might be evoked by the molecule. For example, if the desired outcome is to turn on the expression of a particular gene or set of genes, the cells might be equipped with a reporter construct in which the relevant promoter drives expression of green fluorescent protein or some other convenient marker. Alternatively, one can use native cells if the desired phenotype can be detected using an antibody, such as the appearance of a cell surface protein modification [5]. In whole organism screens, such as abnormal organ formation in a zebrafish, screens can simply be assessed visually. Chemical genetics assays of this type are useful in providing molecules that can be used to probe the relevant biological pathways in the same way that mutations can be employed in genetically malleable model organisms such as yeast.

Despite its huge potential, chemical genetics is currently limited by (at least) two major technical hurdles. One is that compounds isolated from such screens are not terribly potent. Ideally, one would like to have molecules that evince their effect at low nM concentrations, since this connotes high affinity and specificity. This is almost never the case, and a low μM hit in such screens is generally considered to be a good result. The rapid maturation of these hits into more potent reagents without resorting to tedious medicinal chemistry is a major issue in developing a synthetic molecule whose utility as a pharmacological reagent rivals that of natural products.

The second issue is to identify the target(s) of the bioactive molecule. The reviews presented in this issue of *Chemistry & Biology* cover a number of different approaches to the target identification (ID) problem in detail. This overview is meant to complement these more detailed reviews and to provide a global perspective on efforts to solve this problem. In addition, we also focus on some important technical issues involved in carrying out such experiments, particularly biochemical assays, which are underappreciated but can have a significant impact on the success or failure of a protein "fishing expedition." It is our hope that some of the points raised here will stimulate and aid new research into this important problem.

Biochemical Approaches to Protein Target ID: Some Success Stories

One of the best examples of the power of small molecules (in this case a natural product) as discovery tools in biology is the trapoxin/histone deacetylases story [6, 7]. This work, carried out by Schreiber and colleagues, was a key contribution to our current understanding of the critical role of histone deacetylases in repression of gene transcription. At the time, it was known that trapoxin (Figure 1) inhibited histone deacetylation and caused phenotypic changes in cells, for example cell cycle arrest. But it was not understood why this arrest occurred or how it was connected to histone deacetylation. In an effort to probe this issue, Taunton et al. [7] synthesized an analog of the natural product, called K-trap, that facilitated the construction of an affinity matrix comprised of the small molecule linked covalently to a resin (Figure 1). Passage of a mammalian cell extract over this column, followed by thorough washing, resulted in the enrichment of two major proteins, as visualized by denaturing polyacrylamide gel electrophoresis and silver staining. Subsequent analysis revealed one of these proteins to be related to a yeast protein called Rpd3 [6], which was known to be involved in the repression of several genes in yeast, but was not known to be a histone deacetylase (HDAC). Thus, the identification of one of the trapoxin targets (we now know that there are at least nine human HDACs) provided a critical missing link that brought together many disparate observations into a coherent picture. An important point here is the fact that K-trap and trapoxin dissociate from target HDACs only under denaturing conditions. This frustrated attempts to identify radiolabeled bands on a denaturing gel when radiolabeled trapoxin was incubated with extract, but did allow for stringent washing in the affinity experiment to remove all nonspecific binders prior to eluting the HDAC from the column [7]. As will be elaborated below, the ability to wash such columns thoroughly is a critical issue in the success of such experiments.

A good example of successful target ID in the case of a library-derived compound is the recent work of Ding et al. [8]. They screened a library of purine-like molecules for compounds that would induce P19 stem cells to adopt a neuronal fate. One of the molecules that came out of this phenotypic screen was coupled to an affinity matrix and used to probe a P19 cell extract. A molecule that did not score as a positive in this assay was also coupled to a column and employed as a control. Two bands were highly enriched from the lysate. These were shown to be forms of GSK-3β, a protein kinase known

*Correspondence: thomas.kodadek@utsouthwestern.edu

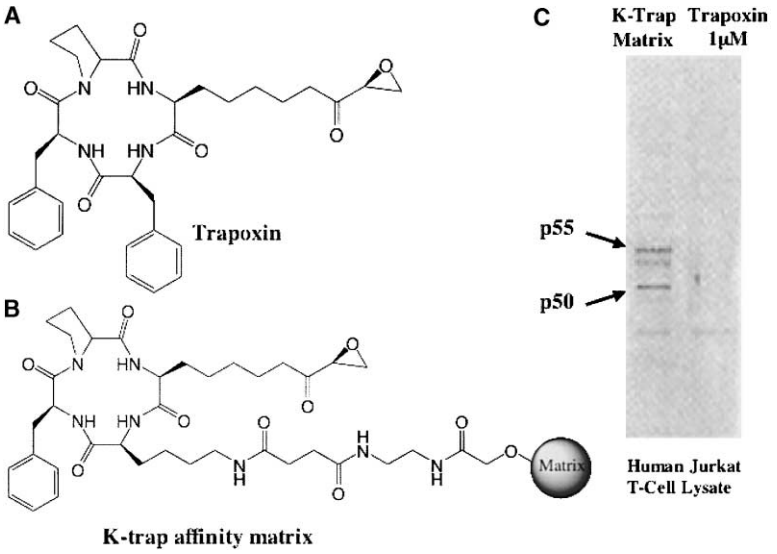

Figure 1. Natural Product Target Identification

(A) Chemical structure of trapoxin.

(B) Chemical structure of the K-trap affinity matrix.

(C) Silver-stained gel showing human trapoxin binding proteins, p55 = HDAC 1, p50 = RbAp48. Proteins bound to the K-trap affinity matrix in the presence or absence of soluble trapoxin were eluted by boiling in SDS loading buffer and analyzed by SDS-PAGE (9% gel). This panel is reprinted with permission from [6].

to be involved in important developmental pathways, including Wnt signaling [9]. A surface plasmon resonance experiment revealed that the K_D of the TWS 119•GSK-3β complex was 126 nM [8], a better than usual result for a hit from a combinatorial library. This is presumably the result of using a biased library based on the purine scaffold, which is rich in molecules disposed to bind protein kinases and other ATP binding proteins. The relatively high affinity undoubtedly was important in the success of this experiment (vide infra). Thus, GSK-3β is a strong candidate for the physiologically relevant target of the small molecule.

The Limitations of Classical Binding Experiments

The work described above shows clearly that affinity chromatography can be a powerful method for target identification. So why is this journal devoting so much space to this problem if it can be addressed in such a straightforward manner? Unfortunately, the truth is that this approach often doesn't work, and almost all successful affinity chromatography experiments have involved the happy combination of a high-affinity small molecule with a fairly abundant protein receptor.

It is much easier to characterize high-affinity small molecule-protein complexes than those with modest affinity biochemically. Most bioactive synthetic molecules are somewhat hydrophobic, and in an aqueous buffer this "stickiness" will always result in the retention of many proteins from an extract in addition to the physiological relevant target(s). This is particularly true in an affinity chromatography experiment, where an extract is passed over a packed column consisting of the small molecule immobilized on agarose or some other suitable support. In this case, the proteins "see" a high local concentration of the small molecule that is sufficient to retain both specific and many nonspecific binding partners. Washing is inevitably required. A high-affinity small molecule-protein interaction will usually have a kinetic half-life sufficient to survive extensive washing, whereas low affinity complexes often will not. This differential between the lifetimes of nonspecific and specific complexes is a key determinant of the success or failure of an affinity chromatography experiment, since this determines how thoroughly one can wash the resin without losing all of the bound proteins. Recall that in the trapoxin work discussed above, the small molecule-HDAC complexes dissociated only under denaturing conditions, a highly favorable scenario. The AS-119-GSK-3β complex has a K_D in the 100 nM range, at least 10-fold better than most complexes that result from phenotypic screens of unbiased libraries.

Another obvious contributor to the success of an affinity binding experiment is the abundance of the target protein(s). This is because an abundant protein need not be enriched greatly in order to become the most prominent band on a stained gel. It is far more difficult for a low abundance protein, such as a transcription factor, to be detected above background in this way unless an enormous enrichment is achieved.

These statements may seem obvious now, but this was not always the case. In the early days of chemical biology, there was a naive assumption that the target ID problem would not be so difficult. In a way, the early chemical biology community was a victim of its own success, since this pathway for target ID worked beautifully in some early experiments. Perhaps the best example was the identification of FKBP as the target of the immunosuppressant FK-506, an effort in which affinity chromatography played a key role [10]. FKBP is an extremely abundant, highly soluble protein, and the FK-506-FKBP complex has a K_D in the low nM range. While not taking anything away from the brilliance of this pioneering work, in retrospect it is clear that one could hardly have hoped for a more favorable case. It was both lucky and unlucky that the stars aligned so nicely in this early work. The good news is that it allowed this fabulous story to be unraveled at a rapid pace. The impact of this work extended beyond understanding the mechanism of action of the drug in suppressing the immune system [11]. These experiments had a lot to do with convincing biologists that there really might be something to the idea of using chemistry as a key tool in molecular biol-

ogy, while also suggesting to a recalcitrant synthetic chemistry community that it might be OK to work with real biological systems. The bad news is that it created the false impression among chemists without much classical biochemical experience that the target identification problem could be solved in most cases using this simple affinity chromatography approach. This view soon crashed upon the rocks of reality. Indeed, some natural products chemists interested in target ID have focused entirely on molecules with reactive functional groups such as epoxides, the expectation being that these molecules will form covalent complexes with their protein targets so that extremely stringent washing conditions (including denaturing buffers) can be employed in biochemical fishing experiments. If the target ID problem is tough for natural products chemists, it is much worse for chemical geneticists, since, as stated above, they are usually working with far less potent compounds.

The Next Generation of Biochemical Solutions to the Target ID Problem

Having said this, we are optimistic about the future of biochemical approaches to the target ID problem in chemical genetics. It seems reasonable to suggest that the implementation of several new protocols and techniques combined with some clearer analysis of the problem (vide infra) will have a rejuvenating effect on the field.

First and foremost, as stated above, it would be desirable to carry out such "fishing expeditions" with potent, and therefore presumably high-affinity, molecules. There have been many reports in the last few years of clever strategies to mature modest potency leads into much more potent derivatives without resorting to classical and tedious medicinal chemistry approaches (for a review, see [12]). Many of these approaches involve creating bivalent protein ligands in which two or more modest affinity, noncompetitive binding elements are linked in such a way as to provide a high-affinity molecule [13–19] (M. Reddy, K. Bachhawatt-Sikder, and T.K., submitted). Most of this work has been done in the context of simple protein binding or enzyme inhibitor screens, but there is no reason that the same strategies could not be adapted to chemical genetics screens. Thus, we anticipate that much more potent compounds will begin to come out of chemical genetics screens in the near future, making the subsequent target ID problem more tractable.

Even with more effective "bait" in hand, another important improvement will be better experimental design. An underappreciated detail that has a tremendous impact on the level of nonspecific interactions is the molar ratio of the bait molecule to the protein input. Ideally, one would like to have a huge excess of input protein relative to the bait molecule. Under these conditions, potential binding proteins must compete for limiting bait, and only the highest-affinity interactions would be picked up. At the other end of the spectrum, if the bait is used in excess, then any protein with enough affinity to survive the washing steps employed will be detected as a potential binding partner. Unfortunately, most experiments of this type are done under conditions closer to the latter extreme than the former, with the predictable result that a lot of "junk" is retained by the immobilized molecule (and/or the support itself). This is true not only for experiments using small molecules as bait. Often protein-protein interactions are monitored using a similar protocol, in which the bait protein is immobilized on a resin (often in the form of a GST fusion protein bound to glutathione-agarose) and exposed to an extract. Rarely is a large excess of total protein used in these experiments. These are particularly bad experiments because Western blotting is usually used to probe for the presence of a suspected binding partner in comparison to a control sample displaying the background proteins retained by GST alone. Rarely is the total retained protein visualized using stained gels to see whether or not the band visualized by Western blotting is one of a host of bound factors. These inherently biased experiments have polluted the literature in many areas of biology. Many of the interactions identified in this way may be physiologically relevant, but it is hard to tell unless there are corroborating data for the interaction, such as genetic interactions, etc.

There is a reason that this type of experiment is usually done under suboptimal conditions. If one is interested in mammalian cells, it is tedious and expensive to produce large amounts of cell extract. In theory, this would be fine if the amount of bait were reduced, but a certain minimal level is needed in order to isolate detectable levels of protein binding partners. A simple solution that is often used in our laboratory is to mix the mammalian cell extract of interest with a large excess of concentrated extract obtained from *Escherichia coli* cells [20]. It is simple and cheap to produce large amounts of bacterial extract, and this highly heterogeneous mixture of proteins does an excellent job of soaking up nonspecific binding sites, thus competing away nonspecific eukaryotic proteins that might bind and confuse the issue. This approach is especially effective if one can label the bacterial and/or the mammalian proteins in such a way that they can be easily distinguished. For electrophoretic analysis, one can grow the mammalian cells in the presence of [^{35}S]methionine, but not label the bacterial proteins, which will then be invisible in an autoradiogram.

Another approach to distinguishing nonspecific binders from true target candidates is to compare the proteins retained by the bioactive molecule of interest and those retained by a structurally similar but inactive molecule (for example, see Figure 2). Ideally, the control molecule would be the enantiomer of the bioactive species if the enantiomer were inactive. If this is not the case or if the molecule is achiral, then the closer the control resembles the "hit," the better. In such experiments, mass spectrometry is the preferred method of analysis. One can label the proteins from one sample with a heavy reagent and the proteins from the other with a light reagent (see Figure 3). This can be done at the protein level by growing the cells in labeled media [21] or by labeling specific side chains (such as cysteinyl sulfurs) [22]. The proteins retained by the hit and by the control compound can then be compared by mixing the two samples prior to analysis, then examining the ratio of peaks in the mass spectrometer (Figure 3). Pairs of sig-

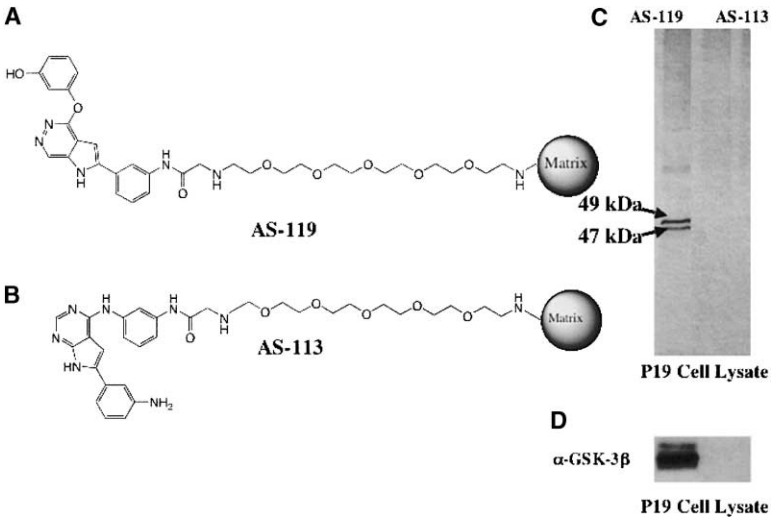

Figure 2. Target Identification with a Library-Derived Compound

Portions of this figure were reprinted with permission from [8].

(A) Chemical structure of AS-119, the linker derivative of TWS-119, a compound selected from a purine-like library for its ability to induce neuronal differentiation in P19 stem cells.

(B) Chemical structure of AS-113, the linker derivative of TWS-113, an inactive library member.

(C) Silver-stained gel showing proteins retained by AS-119.

(D) Proteins retained by AS-119 were shown to be GSK-3β by Western blotting.

nals whose ratio is tipped in favor of the retentate of the bioactive molecule are candidate target proteins, since they were retained selectively. This comparative method provides better sensitivity and higher throughput than gel-based methods.

While the above approach has focused on affinity chromatography or pulldown-type assays, the same basic concepts apply to any biochemical assay. As reviewed by Prestwich in this issue, a promising alternative to these methods is to employ chemical crosslinking. Crosslinking experiments do not require the agent of interest to be immobilized. Thus, it is easier to control important experimental parameters such as the concentration of the small molecule probe. We believe that chemical crosslinking has a very bright future in protein target identification. However, in order to realize the full power of this approach, more effective crosslinking chemistries would be very helpful. In particular, a system that could be used efficiently inside living cells would be a tremendous breakthrough. As crosslinking approaches mature, it would be a good idea to incorporate the crosslinkable moieties into the core structure of the molecules that one screens. This would eliminate the possibility that appending a crosslinkable (or any other) tag to a given molecule will compromise its activity.

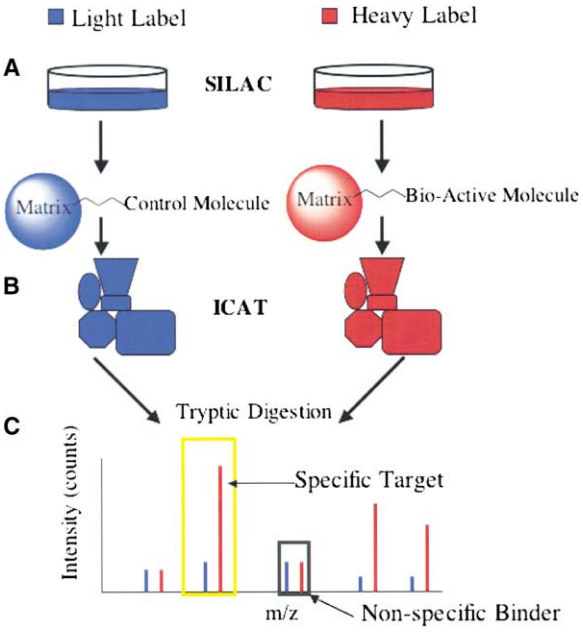

Figure 3. Isotopic Detection of Enriched Protein Targets

(A) SILAC, stable isotope labeling with amino acids in cell culture, a technique pioneered by Mann and coworkers to quantify protein abundance data. A ^{12}C or ^{13}C amino acid, e.g., ^{13}C-labeled arginine (+6), is supplied to cells in culture [21, 26].

(B) ICAT, isotope coded affinity tagging. Aebersold and others have developed chemical methods for incorporating isotope tags on amenable protein side chains, e.g., cysteine, after protein complexes are eluted from affinity matrices [27].

(C) After sample mixing and trypsin digestion, specific protein targets exhibit a Heavy:Light ratio much greater than 1 in the mass spectrum in contrast to nonspecific interactors [28].

Genetic Approaches to Target Identification

It is a truism that it is easier to work with DNA than proteins. Therefore, most molecular biologists would argue that if one can solve a problem using genetics or some clever molecular biology trick, then this is preferable to spending several days in the cold room doing chromatography. Molecular biology approaches are especially powerful in genetically manipulable organisms, such as the yeast *Saccharomyces cerevisiae*. As is laid out in the review by Zheng et al. in this issue, some of the most impressive advances in the target ID area have come through the application of clever cell- and genetic-based screens. These include three-hybrid systems [23], looking for increased drug sensitivity in heterozygote collections [24, 25], and a host of other schemes. What makes genetic and cellular approaches so attractive is that these provide a tight connection between phenotype and genotype. In other words, a hit in any of these assays provides the cloned gene for the target protein directly. This is not the case in biochemical assays. On the other hand, some of these genetic assays require that target protein candidates be employed as artificial fusion constructs that may or may not behave as the

native proteins do, particularly if the protein is normally part of a multiprotein complex. The three-hybrid system is a good example of this. Finally, most of these approaches are limited to yeast or other unicellular organisms. Thus, they tend to be most effective for things like anticancer compounds or other chemicals that affect the most fundamental pathways of a cell. If one is interested in compounds that affect the nervous system, it may not be possible to carry out screens in yeast, depending on the method employed. Help may be on the way in the increasing use of more complex, but still genetically manipulable, organisms, such as fruit flies and zebrafish, in primary screens.

The challenging problem of identifying the target of bioactive small molecules has spawned a great deal of imaginative work that has advanced the field considerably. As mentioned above, genetic methods in particular have scored some impressive successes of late. Biochemical methods continue to improve and have been aided tremendously by advances in protein mass spectrometry. However, it is fair to say that this issue probably remains the rate-limiting step in applying the full power of chemical genetics and related small molecule approaches to unraveling interesting biological pathways. As more and more people recognize this as a central problem and focus on novel solutions, this will likely change. We look forward to seeing (and perhaps participating in) the breakthroughs in this important area that are yet to come.

Acknowledgments

We thank Profs. S.L. Schreiber and P.G. Schultz for permission to reprint previously published data.

References

1. Schreiber, S.L. (1998). Chemical genetics resulting from a passion for synthetic organic chemistry. Bioorg. Med. Chem. *6*, 1127–1152.
2. Schreiber, S.L. (2003). The small molecule approach to biology. Chem. & Eng. 1199 News *81*, 51–61.
3. Crews, C.M., and Splittgerber, U. (1999). Chemical genetics: exploring and controlling cellular processes with chemical probes. Trends Biochem. Sci. *24*, 317–320.
4. Clemons, P.A., Koehler, A.N., Wagner, B.K., Sprinings, T.G., Spring, D.R., King, R.W., Schreiber, S.L., and Foley, M.A. (2001). A one-bead, one-stock solution approach to chemical genetics: part 2. Chem. Biol. *8*, 1183–1195.
5. Stockwell, B.R., Haggarty, S.J., and Schreiber, S.L. (1999). High-throughput screening of small molecules in miniaturized mammalian cell-based assays involving post-translational modifications. Chem. Biol. *6*, 71–83.
6. Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science *272*, 408–411.
7. Taunton, J., Collins, J.L., and Schreiber, S.L. (1996). Synthesis of natural and modified trapoxens, useful reagents for exploring histone deacetylase function. J. Am. Chem. Soc. *118*, 10412–10422.
8. Ding, S., Wu, T.Y.H., Brinker, A., Peters, E.C., Hur, W., Gray, N.S., and Schultz, P.G. (2003). Synthetic small molecules that control stem cell fate. Proc. Natl. Acad. Sci. USA *100*, 856–861.
9. Doble, B.W., and Woodgett, J.R. (2003). GSK-3: tricks of the trade for a multi-tasking kinase. J. Cell Sci. *116*, 1175–1186.
10. Harding, M.W., Galat, A., Uehling, D.E., and Schreiber, S.L. (1989). A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. Nature *341*, 758–760.
11. Schreiber, S.L., and Crabtree, G.R. (1992). The mechanism of action of cyclosporin A and FK506. Immunol. Today *13*, 136–142.
12. Kodadek, T. (2002). Protein-detecting microarrays. Trends Biochem. Sci. *27*, 295–300.
13. Shuker, S.B., Hajduk, P.J., Meadows, R.P., and Fesik, S.W. (1996). Discovering high-affinity ligands for proteins: SAR by NMR. Science *274*, 1531–1534.
14. Maly, D.J., Choong, I.C., and Ellman, J.A. (2000). Combinatorial target-guided ligand assembly: identification of potent subtype-selective c-Src inhibitors. Proc. Natl. Acad. Sci. USA *97*, 2419–2424.
15. Sem, D.S., Bertolaet, B.L., Baker, B., Chang, E., Costache, A.D., Coutts, S., Dong, Q., Hansen, M., Hong, V., Huang, X., et al. (2004). Systems-based design of bi-ligand inhibitors of oxidoreductases: Filling the chemical proteomic toolbox. Chem. Biol. *11*, 185–194.
16. Profit, A.A., Lee, T.R., and Lawrence, D.S. (1999). Bivalent inhibitors of protein tyrosine kinases. J. Am. Chem. Soc. *121*, 280–283.
17. Green, N.S., Palaninathan, S.K., Sacchettini, J.C., and Kelly, J.W. (2003). Synthesis and characterization of potent bivalent amyloidosis inhibitors that bind prior to transthyretin tetramerization. J. Am. Chem. Soc. *125*, 13404–13414.
18. Kitov, P.I., Shimizu, H., Homans, S.W., and Bundle, D.R. (2003). Optimization of tether length in nonglycosidically linked bivalent ligands that target sites 2 and 1 of a Shiga-like toxin. J. Am. Chem. Soc. *125*, 3284–3294.
19. Morken, J.P., Kapoor, T.M., Feng, S., Shirai, F., and Schreiber, S.L. (1998). Exploring the leucine-proline binding pocket of the Src SH3 domain using structure-based, split pool synthesis and affinity-based selection. J. Am. Chem. Soc. *120*, 30–36.
20. Alluri, P.G., Reddy, M.M., Bacchawat-Sikder, K., Olivos, H.J., and Kodadek, T. (2003). Isolation of protein ligands from large peptoid libraries. J. Am. Chem. Soc. *125*, 13995–14004.
21. Ong, S.E., Foster, L.J., and Mann, M. (2003). Mass spectrometric-based approaches in quantitative proteomics. Methods *29*, 124–130.
22. Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. *17*, 994–999.
23. Jaeger, S., Eriani, G., and Martin, F. (2004). Results and prospects of the yeast three-hybrid system. FEBS Lett. *556*, 7–12.
24. Pek, Y.L., Armour, C.D., Stepaniants, S.B., Cavet, G., Wolf, M.K., Butler, J.S., Hinshaw, J.C., Garnier, P., Prestwich, G.D., Leonardson, A., et al. (2004). Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. Cell *116*, 121–137.
25. Giaever, G., Flaherty, P., Kumm, J., Proctor, M., Nislow, C., Jaramillo, D.F., Chu, A.M., Jordan, M.I., Arkin, A.P., and Davis, R.W. (2004). Chemogenomic profiling: Identifying the functional interactions of small molecules in yeast. Proc. Natl. Acad. Sci. USA *101*, 793–798.
26. Blagoev, B., Kratchmarova, I., Ong, S.E., Nielsen, M., Foster, L.J., and Mann, M. (2003). A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. Nat. Biotechnol. *21*, 315–318.
27. Ranish, J.A., Yi, E.C., Leslie, D.M., Purvine, S.O., Goodlett, D.R., Eng, J., and Aebersold, R. (2003). The study of macromolecular complexes by quantitative proteomics. Nat. Genet. *33*, 349–355.
28. Oda, Y., Owa, T., Sato, T., Boucher, B., Daniels, S., Yamanaka, H., Shinohara, Y., Yokoi, A., Kuromitsu, J., and Nagasu, T. (2003). Quantitative chemical proteomics for identifying candidate drug targets. Anal. Chem. *75*, 2159–2165.