DRUG DESIGN

Discovering High-Affinity Ligands for Proteins

Philip J. Hajduk, Robert P. Meadows, Stephen W. Fesik

Drug discovery is a difficult and time-consuming process that involves the identification of potent therapeutic agents with good bioavailability, metabolic stability, and low toxicity. An important first step in this process is the identification of molecules that bind with high affinity and selectivity to the biological target. High-affinity ligands might be discovered by screening natural products or large chemical libraries. However, molecules may not exist that have functional groups in the correct spatial orientation to interact optimally with the binding site. To increase the chances of finding such a molecule, many compounds could be synthesized using recently developed approaches in combinatorial chemistry and parallel synthesis (*1*). Yet it is often difficult to develop the synthetic procedures required to assemble many different fragments with a variety of linkers. This may ultimately limit the diversity of the compounds that can be synthesized. Furthermore, the synthesis and testing of large numbers of molecules is both costly and time-consuming.

TECH VIEW

If small molecules that bind to the different pockets of the protein could be identified and one could determine how these fragments should be linked, fewer compounds would need to be synthesized to create high-affinity ligands. Although the molecular fragments may only bind weakly to the protein (as expected because of their small size), the free energy of binding of the linked compound is, in principle, the sum of the free energies of each fragment plus a term due to linking (*2*). Thus, linked compounds with submicromolar affinities could be obtained by linking two fragments that bind in the millimolar range.

A compound derived from molecular fragments could be designed rationally with molecular modeling based on the three-dimensional (3D) structure of the biological target. However, this still remains a challenging task because of possible conformational changes of the target or ligand upon complexation, the uncertainties in treating bound water molecules, and the use of relatively simple force fields (*3*). Ideally, experimental determination of the fragment molecules that bind to the protein and use of structural information to guide linker design is preferred. Toward this goal, we developed a method for producing high-affinity ligands in which small molecules that bind to proximal subsites of a protein are identified in an NMR-based screen and then linked together in their experimentally determined bound orientations (Fig. 1) (*4*). The method is called "SAR by NMR," which stands for "structure-activity relationships by nuclear magnetic resonance."

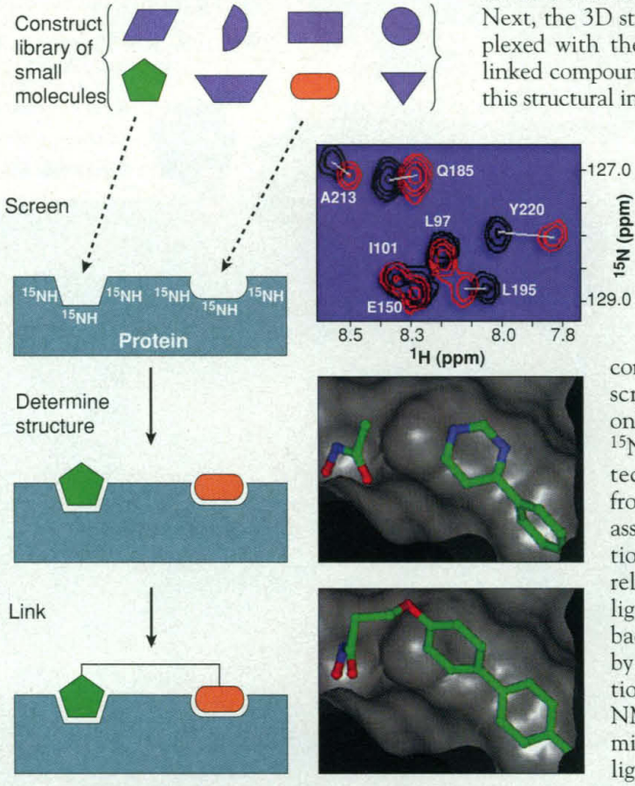

Fig. 1. The SAR-by-NMR method. In the first step, a library of small molecules is screened for binding to a protein. Binding is detected from the amide chemical shift changes observed in 2D HSQC spectra. A portion of the 2D HSQC spectra of ^{15}N-labeled stromelysin (top right) acquired in the absence (red) and presence (black) of a biaryl-containing ligand is shown. Once two ligands are identified that bind to the protein (green pentagon and orange ellipse), the structure of the ternary complex is determined (middle left). The NMR structure of stromelysin complexed with acetohydroxamic acid and 4-phenylpyrimidine is depicted (middle right). Based on structural information, linked compounds are synthesized (bottom left). Thus, a small, potent (IC_{50} = 25 nM), nonpeptide inhibitor of stromelysin (bottom right) was discovered with SAR by NMR by linking two fragments that only bind weakly to this enzyme.

In the first step of the SAR-by-NMR method, we screen a library of small molecules for binding to an ^{15}N-labeled protein. If a molecule binds to the protein, it will alter the local chemical environment and thereby cause changes in the chemical shifts of nuclei in the protein-binding site. These changes are detected in 2D heteronuclear single quantum correlation (HSQC) spectra acquired in the presence and absence of added compound. Using this NMR method, we screen for compounds that bind to proximal sites on a protein. Once initial ligands are identified, analogs are screened and binding constants are obtained in order to optimize the interactions with the protein. Binding constants are measured from the changes in the 2D HSQC spectra as a function of ligand concentration. Next, the 3D structure of the protein complexed with the ligands is obtained, and linked compounds are synthesized based on this structural information.

The key to the success of this technique is the use of NMR for identifying small molecules that bind to proteins. Even though small molecules typically bind weakly to the protein and have to be tested at high concentrations, NMR-based screening is reliable because only the amide signals of the ^{15}N-labeled protein are detected without interference from other components of the assay. In contrast, many traditional screening assays cannot reliably detect weakly bound ligands because of the large background signals produced by high compound concentrations. Another advantage of NMR is the ability to determine the location of the ligand-binding sites by analyzing the specific amide signals of the protein that change upon the addition of the molecule. This is especially important for small, weakly bound ligands that could interact with more than one site on the protein. By having the capability to reliably compare the structures of molecules that bind to the same site, the functional groups that are important for binding can be deter-

The authors are at Abbott Laboratories, Pharmaceutical Discovery Division, Abbott Park, IL 60064, USA. E-mail: fesik@steves.abbott.com

mined. In addition, since many analogs of such small molecule leads are often commercially available, useful structure-activity relationships can be typically obtained without chemical synthesis.

Is NMR practical for screening relatively large libraries of compounds? For the NMR-based screen, a 2D HSQC spectrum of an ^{15}N-labeled protein at a concentration of 0.3 mM can generally be acquired in less than 15 min on a 500-MHz NMR spectrometer. To efficiently screen our library of compounds, we test the molecules in groups of 10. This allows 1000 compounds to be screened in 1 day with an automatic sample changer. Currently, our library consists of a diverse set of more than 10,000 small molecules dissolved in perdeuterated dimethyl sulfoxide. Although this number of compounds may appear to be small relative to the hundreds of thousands typically contained in corporate databases or combinatorial libraries, these 10,000 small molecules only represent fragments of the linked compounds that can be prepared. For example, if assembled pairwise with only 10 linkers, our collection of 10,000 fragments represents a virtual library of one billion compounds (Fig. 2). Synthesis and analysis of a real library of this size would be an arduous and costly task.

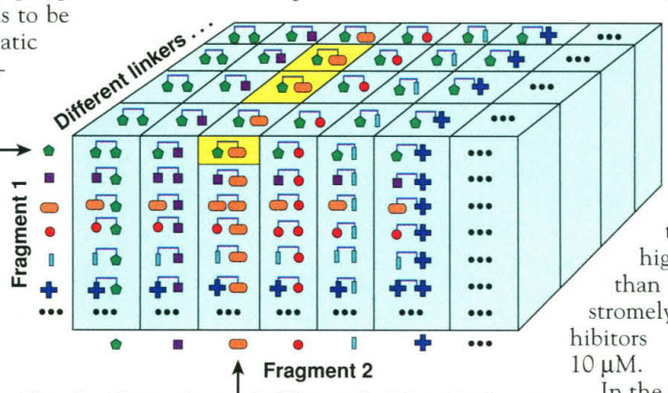

Fig. 2. Comparison of different building-block approaches. With combinatorial chemistry, many linked compounds [(fragment 1) × (fragment 2) × (different linkers)] are synthesized that contain all combinations of fragments and linkers. In contrast, with SAR by NMR, only a few compounds need to be synthesized (yellow highlighted boxes) because the fragments that bind to the protein are identified before linking (arrows), and the linkers are selected on the basis of structural information.

The limitations of the SAR-by-NMR method are that it requires high-field NMR spectrometers, large amounts of pure ^{15}N-labeled protein (>200 mg), and biomolecular targets with a molecular mass less than 40 kD that are water soluble at concentrations of ~0.2 mM. Fortunately, many proteins that serve as drug targets are small or contain small functional domains and can be overexpressed in bacteria, ^{15}N-labeled, and prepared in sufficient quantities. For those target proteins that do not meet these criteria, we have developed 1D NMR methods that exploit the changes in relaxation or diffusion rates of small molecules that occur upon complexation (5). Because these methods depend on the detection of uncomplexed molecules rather than the ^{15}N-labeled protein, they can be used with very large biomolecules. Furthermore, the 1D technique may be used to directly identify ligands from a complex mixture of compounds, reducing the amount of time needed for NMR-based screening. Another requirement of SAR by NMR (or any method used for detecting weakly bound ligands) is that the compounds must be soluble at millimolar concentrations in aqueous solution. This requirement is easily met, because many water-soluble compounds with diverse functional groups are available from commercial sources.

Does the SAR-by-NMR method work in practice? We now have several examples that demonstrate the utility of this technique. In our first example, a molecule that binds tightly (K_d = 19 nM) to the FK506 binding protein (FKBP) was discovered by tethering two fragments that only bind weakly (K_d = 2 μM and 100 μM) to this protein (4). The two fragments were rapidly identified and optimized (less than 2 months), and only five linked molecules were synthesized. All five compounds exhibited nanomolar affinities for FKBP. The SAR-by-NMR technique was also used to discover potent, nonpeptide inhibitors of stromelysin (6). This enzyme is a member of the family of matrix metalloproteinases which, when overexpressed or disregulated, are associated with pathological conditions such as arthritis and tumor metastases. Acetohydroxamic acid and a biphenyl-containing compound were found to bind to adjacent sites on stromelysin with affinities of 17 mM and 0.02 mM, respectively. Guided by the NMR structure of two ligands bound to stromelysin, fragments were linked together to produce nanomolar inhibitors of this enzyme. Only 6 months were required for the entire process, since the chemistry was highly focused on linking together molecules that were optimized for binding to the protein. The structure-activity relationships observed for the unlinked molecules were found to correlate well with those observed for the linked compounds (6). In addition to inhibiting stromelysin, the linked compounds are potent inhibitors of gelatinase A, an enzyme that is thought to play an important role in tumor metastases (7). Thus, the inhibitors discovered using SAR by NMR (or analogs of these compounds) may prove useful in the fight against cancer. In our most recent example, an NMR-based screen was used to identify compounds that bind to the DNA-binding domain of human papillomavirus E2, a protein that is required for viral replication (8). Biphenyl and biphenyl ethers containing a carboxylic acid bind to a site near the DNA recognition helix and inhibit the binding of the E2 protein to DNA. By combining information on the structure-activity relationships for both series of molecules, a compound was synthesized that represents a new lead for the development of an antiviral agent against the human papillomavirus. It is important to note that conventional high-throughput screening of more than 100,000 compounds against stromelysin and E2 failed to produce inhibitors with potencies better than 10 μM.

In the relatively short time since its inception, we have demonstrated that SAR by NMR is a useful method for rapidly discovering high-affinity ligands for proteins. In the future, it should be possible to extend the applicability of the technique to larger proteins and other biomolecular targets such as RNA. It may also be possible to use SAR by NMR to help overcome problems in bioavailability and metabolic stability by suggesting new fragments with improved properties as replacements for those that cause unwanted effects. Our experience indicates that SAR by NMR is emerging as a useful tool that complements existing methods for tackling the challenging problem of drug discovery.

References

1. M. A. Gallop, R. W. Barrett, W. J. Dower, S. P. A. Fodor, E. M. Gordon, *J. Med. Chem.* **37**, 1233 (1994); D. J. Ecker and S. T. Crooke, *Biotechnology* **13**, 351 (1995).
2. W. P. Jencks, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4046 (1981).
3. C. L. M. J. Verlinde, G. Rudenko, W. G. J. Hol, *J. Comput. Aided Mol. Des.* **6**, 131 (1992); H.-J. Böhm, *ibid.*, p. 61; I. D. Kuntz, *Science* **257**, 1078 (1992).
4. S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science* **274**, 1531 (1996).
5. P. J. Hajduk, E. T. Olejniczak, S. W. Fesik, *J. Am. Chem. Soc.*, in press.
6. P. J. Hajduk *et al.*, *ibid.* **119**, 5818 (1997); E. T. Olejniczak *et al.*, *ibid.*, p. 5828.
7. G. Giannelli, J. Falk-Marzillier, O. Schiraldi, W. G. Stetler-Stevenson, V. Quaranta, *Science* **277**, 225 (1997).
8. P. J. Hajduk *et al.*, *J. Med Chem.* **40**, 3144 (1997).

TechWire Forum:
www.sciencemag.org/dmail.cgi?53371