



ELSEVIER

Obtaining and screening compound collections: a user's guide and a call to chemists

Paul J Hergenrother

Advances in genetics, proteomics and cell biology over the past 20 years have unearthed a multitude of potential macromolecular targets for the selective treatment of disease. The challenge remains to find appropriate small molecule ligands for these proteins (or nucleic acids), and to use these ligands to validate novel disease targets. The advent of low-cost instrumentation has made industrial-style high-throughput screening possible in academic settings. Unfortunately for many, access to large collections of compounds is still limited and limiting. This article is aimed at the user who has an interest in compound screening but does not have ready access to large collections of small molecules. High-throughput screening need not be the exclusive domain of institutions and centers with vast resources and NIH Roadmap-funded compound repositories. As it turns out, many of the most interesting compounds are probably within arm's reach, in our laboratory freezers and in those of our colleagues.

Addresses

Department of Chemistry, Roger Adams Laboratory, University of Illinois, Urbana, IL 61801, USA

Corresponding author: Hergenrother, Paul J (hergenro@uiuc.edu)

Current Opinion in Chemical Biology 2006, **10**:213–218

This review comes from a themed issue on
Combinatorial chemistry and molecular diversity
Edited by Scott K Silverman and Paul J Hergenrother

Available online 3rd May 2006

1367-5931/\$ – see front matter

© 2006 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.cbpa.2006.04.005

Introduction: the value of high-throughput screening

The importance of biologically active small molecules is unquestioned [1^{*}]. While the immediate utility of these compounds as drugs or drug leads is perhaps most obvious, the use of small molecules to interrogate biological systems of interest can take many forms, such as chemical genetic screens [2^{*},3,4], identification of protein targets for important small molecules [5^{*},6], and using compounds to validate novel targets for the treatment of disease [7^{*},8]. All of these areas, of course, are predicated on the ability to first identify such biologically active small molecules.

Traditional avenues for compound discovery, such as natural product isolation, synthesis and derivatization,

are still utilized today [9]; these routes are increasingly being augmented by modern techniques of *in silico* screening [10], and structure-based design/combinatorial chemistry [11]. However, a great number of initial small molecule 'hits' now arise from high-throughput screening (HTS) of collections of organic compounds. The success stories to emerge from HTS efforts are myriad and have been reviewed extensively [4,12]. In addition, the instrumentation and logistics required for an effective high-throughput screen have been covered in several up-to-date reviews and will not be repeated here [13]. This review seeks to inform those who are interested in HTS but have no compound collections to screen.

Everything old is new again: screens of FDA-approved drugs

A major bottleneck in the drug discovery process is gaining FDA approval for a new chemical entity. Depending on the therapeutic area, this process can extend over years and at considerable expense: current estimates put the cost of developing a novel therapeutic at \$800 million [14]. As a consequence, the fastest way to move a compound to market is not to develop a novel compound at all, but to obtain a new indication for an already-approved drug. Recognizing the inherent value in FDA-approved compounds, the majority of which are off-patent, a number of different high-throughput screens have been performed that exclusively use FDA-approved drugs [15^{**}]. The obvious advantage of this strategy is that a wealth of information is already available on the bioavailability, toxicity and dosing of any 'hit' molecules that arise from such screens.

To facilitate this process, collections of FDA approved drugs are available for purchase in formats suitable for screening, and several interesting stories are emerging from this line of research. Companies such as Prestwick Chemical Company (<http://www.prestwickchemical.com>), Microsource (<http://www.msdiscovery.com/spect.html>) and Sequoia (<http://www.seqchem.com>) sell collections of various on- and off-patent medications for screening purposes. The potential for β -lactam antibiotics to treat amyotrophic lateral sclerosis (ALS) [16], non-steroidal anti-inflammatory drugs to treat Alzheimer's disease, and the dopamine receptor ligand fluphenazine against multiple myeloma are some of the FDA-approved drugs that have been identified as having the potential for additional therapeutic indications [15^{**}].

In an interesting twist on this strategy, a systematic screening of FDA-approved compounds — not as single

entity agents but in combinations — has also proved quite fruitful. As pioneered by the company CombinatoRx [17^{**},18], the logic is that a collection of 1000 FDA-approved drugs can give rise to >500 000 pairwise combinations. By explicitly screening combinations of compounds within the FDA-approved collection, several provocative discoveries were made, including a combination of a sedative (phenothiazine) and an anti-fungal (pentamidine) that together have anti-cancer activity [18]. Several of these combination therapies have now advanced to clinical trials for treatment of rheumatoid arthritis, psoriasis, asthma and cancer [19].

Although screens of known drugs have obvious utility and probably represent the most facile pathway from bench to bedside, there are limitations to such efforts. Given that there are only several thousand FDA-approved small molecules, the chemical space covered by these compounds is small when compared with the space covered by all compounds that are known. As therapeutic regimens progress toward the era of ‘personalized medicine’ (where the precise molecular defect in a diseased cell is known and exploited), which will require high-specificity compounds, it is unlikely that all the necessary chemotypes will be present in these small collections of FDA-approved molecules. Thus, there is still considerable utility in the screening of large compound collections in search of new biologically active compounds and novel pharmacophores.

Access to large compound collections

For many people who are interested in identifying a small molecule modulator of a given biological system, access to large collections of compounds remains limiting. Fortunately, options are available to those who wish to conduct high-throughput screens but do not have such facilities at their home institutions (Table 1).

Large collections of compounds (for free)

The Drug Testing Program (DTP) through the National Cancer Institute (NCI) has a large repository of compounds that have been gathered for evaluation as possible anti-cancer or anti-HIV agents. Numerous investigators have submitted these diverse small molecules over the past several decades. The DTP will send specially plated sets of these compounds to users, free of charge. The only

requirement is the completion of appropriate paperwork, consisting of a short proposal (with relevance to cancer or HIV) and other user information. Approved researchers can obtain the ‘Diversity set’ of 1990 compounds, the ‘Mechanistic set’ of 879 compounds, the ‘Challenge set’ of 57 cytotoxins with unknown mechanisms of action, and the ‘Natural product set’ of 235 compounds. All are sent as 1 or 10 mM stock solutions in DMSO, formatted in 96-well plates (see: http://dtp.nci.nih.gov/branches/dscb/repo_open.html).

Large collections of compounds (for purchase)

As the cost of instrumentation for HTS has dropped and as high-profile results have been generated from HTS, the number of laboratories applying these techniques has risen dramatically. To meet the demand for collections of pure, discrete small molecules, a cottage industry has arisen based on the sale of compound collections to industrial and academic users. There are multiple companies providing this service, including ChemBridge Corporation, Maybridge, Chemical Diversity Laboratories, Tripos, and Asinex Corporation; Sigma-Aldrich now also has a drug-like screening collection for sale (MyriaScreen Diversity Collection). In general, compounds are shipped as 1–10 mM solutions in DMSO in 96-well plates. While prices vary according to the amount of compound purchased, the company supplying the compound, and whether the user is in an academic or industry setting, most are priced at \$0.5–1.0 US dollars per compound. Although the benefits of screening commercial compound collections include ready access to more material for follow-up studies, a reasonable chance for high-purity compounds, and (in some cases) a bias toward ‘drug-like’ small molecules, the large cost and small compound quantities discourage many potential users.

Screening services (governmental and academic)

As part of the Roadmap initiative, the NIH has constructed a Small-Molecule Repository (SMR) that will ultimately house >1 million unique compounds (see: <http://grants2.nih.gov/grants/guide/notice-files/NOT-RM-04-003.html>). The vision is to have this SMR populated in part by compounds constructed through another NIH program, that of the Pilot-Scale Libraries for High-Throughput Screening (see: <http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-05-014.html>). Awards for the

Table 1

A spectrum of options is available to those interested in identifying small molecules from high-throughput screens^a

	Free	Fee-for-service
Compound collections (for screening by users)	NCI-DTP Collected compounds	Commercial vendors
High-throughput screens (performed by others)	NIH-SMR	Academic centers Commercial services

^a Users who wish to perform the screens themselves can obtain compound collections for free (from the NCI Drug Testing Program), can create their own collection, or can purchase compounds from a commercial supplier. Users who wish to have a screen performed by others can either submit it to the NIH through the Small Molecule Repository (SMR), or go through an academic center or commercial service.

construction of these libraries have already been granted to eight institutions. In the current incarnation, compounds in the SMR will not be available directly to users *per se*. Rather, interested parties with a developed and validated high-throughput screen can submit a proposal to the NIH (<http://grants.nih.gov/grants/guide/pa-files/PA05-147.html>). If accepted, the compounds in the SMR will be screened for free in the user-developed assay by one of the 10 Molecular Libraries Screening Centers Network (MLSCN). All screening data arising from this effort will be deposited in the PubChem database immediately after they have been verified for accuracy. A separate funding mechanism is available for the actual development of high-throughput screens (<http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-05-011.html>).

High-throughput screening facilities have also sprung up at several universities and research institutes over the past five years. Some of these are at the Broad Institute, Johns Hopkins University, the University of Wisconsin, Scripps Research Institute and the University of Michigan. While many are in their embryonic stage of development and exact user policies are still being established, the Broad Chemical Biology screening facility was formerly the Harvard Institute of Chemistry and Cell Biology (ICCB) screening facility, and has been in operation since 1997. This institute is actively seeking compound contributions from the chemistry community (see the 'ANNOUNCEMENT', highlighted in yellow, at: <http://www.broad.harvard.edu/chembio/>).

Screening services (commercial)

Although most accessible HTS centers are affiliated with an academic institution, there do exist some fee-for-service commercial HTS facilities. For example, the Michigan High Throughput Screening Center (Kalamazoo, MI; <http://mhtsc.kvcc.edu/index.htm>) has compound collections and provides screening services to academic or governmental laboratories, as well as to pharmaceutical or biotechnology companies.

A call to chemists

Although the options described above may be adequate to meet the HTS needs of some, each possesses obvious disadvantages. Compound collections from the federal government or a commercial supplier come with certain logistical handcuffs. Most critically, the actual amount of each compound that is provided is small, typically 25 μ l of a 1 mM solution (25 nanomoles, or 7.5 μ g for a MW = 300 compound). While this may be plenty of material for a fully vetted and optimized high-throughput assay, at these low amounts the compounds are extremely precious and thus not appropriate for the experimentation often needed to develop a robust high-throughput screen. In addition, the diversity and structural complexity of such compounds is often less than what is found in typical natural products.

Synthetic chemists possess a unique and enviable skill set — the ability to create novel molecular entities. The past 50 years can rightly be labeled as the 'golden age' of chemical synthesis, and as such practitioners of the art of total synthesis and their disciples have populated industrial laboratories and chemistry departments. However, more often than not the products, intermediates and side products produced by synthetic chemists are not used in any further experiments — the goal of these studies is the synthetic *tour de force*, not the generation of material for biological testing. Typically, the highly purified and fully characterized compounds generated through these efforts sit idle on shelves or in freezers, a vast untapped resource.

With minimal effort, this culture could change and the availability of collections of such structurally complex compounds could have a tremendous impact on both the specificity and hit rate of compounds identified from HTS efforts. This requires cooperativity and a grass-roots-type effort on the part of individuals within chemistry departments. The task of measuring 2–10 mg of each synthetic compound is one that can be performed with minimal training. Creating stock solutions in DMSO or EtOH and formatting compounds in multi-well plates (384 well plates currently being the most common type) will allow for instant compatibility with a high-throughput screen. As an example, we have assembled a collection of small molecules that were culled from leftover compounds in the attic of Roger Adams Laboratory, the main chemical research building at University of Illinois at Urbana-Champaign. These are structurally diverse samples synthesized by numerous research groups over the past 50 years. While these compounds have no further synthetic utility (milligram quantities are present, and many are late-stage intermediates with limited sites for functionalization), they are perfect for screening. These compounds were made the 'old fashioned' way, one compound at a time, and as such they have diversity, complexity and carbon skeletons that are not generally available in screening libraries. Through this effort close to 10 000 compounds are now in our screening collection; see: <http://www.scs.uiuc.edu/~phgroup/comcollections.html> for structures and histories of the compounds. Having multiple milligrams of each compound in the collection is liberating for assay development and following up on screening hits. Entering compounds into the collection has also greatly facilitated the necessary disposal of old and degraded samples. Similar collections could be constructed in any chemistry department.

The assembly of compound collections need not be limited to those who are synthesizing organic compounds. Indeed, there is a renewed interest in using metal complexes in medicine, with recent examples of organoruthenium complexes as kinase inhibitors being representative of the potential of this approach [20–22]. In fact, high-throughput screening efforts in which inorganic

complexes were included have led to serendipitous discoveries; for example, through HTS it was discovered that iron salts will ablate *Pseudomonas aeruginosa* biofilms, a bacterial state that is highly refractory to conventional antibiotics [23*].

Credit where credit is due

As with any multi-investigator initiative, the issues of who drives the project, who gets credit, and authorship will arise and need to be addressed head on. In this respect, organic compounds that are donated by individual PIs are not merely 'reagents', akin to a plasmid or an antibody, but rather are a precious, finite resource and need to be recognized as such. One model for such an initiative is as follows.

The laboratory of a synthetic organic chemist is producing compounds at a rate of three distinct compounds per student per month \times 10 students \times 12 months = 360 compounds per year. Every month each student weighs out 2–10 mg of each of the compounds he/she has created and donates them to the central compound repository at that institution with an informational sheet containing the chemical structure, molecular weight, solubility information, laboratory notebook page, etc. The compounds are then entered into a database, made up as 10 mM stock solutions in DMSO; part of that solution is transferred to a 384-well plate, while the rest is frozen. A chemical biologist or biologist performs a high-throughput screen for enzymatic inhibition, and one of these compounds is scored as a 'hit' in this primary assay. The investigator then conducts the proper dose-dependent secondary assays with the compound from the original stock solution. Once the hit is confirmed and the potency is of interest to the biologist, the chemist is contacted and told of the exciting news. Even if the appropriate follow up experiments can be performed with only the initial hit, authorship invitation should still be extended to the PI and the student who made the compound. In such cases, an 'author attribution' stating explicitly who did what may be in order; such statements are encouraged by several journals, including *Nature* [24]. Of course, in most cases, more compound will need to be made, and analogues and tagged versions (isotopically substituted, biotinylated, or fluorescent) of the compound will be required; this could lead to a full-fledged collaboration.

The model described above offers a balance between the chemist dictating what screens can be performed with their compounds and the biologist viewing the precious compounds as just another reagent. As with any collaboration, a certain level of trust is needed on both sides; however, the risks are relatively small and the potential rewards are very large. Indeed, several such collaborations following the model described above have already sprung forth. This type of collaboration between synthetic chemists and chemical biologists was nicely demonstrated in

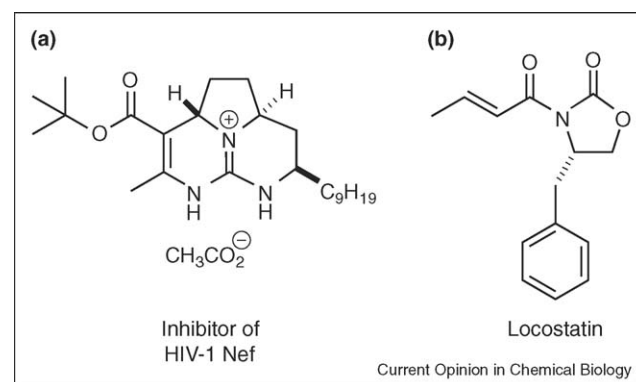
the identification of an inhibitor of the HIV-1 Nef protein through the screening of synthetic derivatives of guanidine alkaloid natural products (Figure 1a) [25]. In another example, a compound containing an oxazolidinone chiral auxiliary was found to inhibit epithelial cell sheet migration and was used to uncover the role of Raf kinase inhibitor protein in cell migration (Figure 1b) [26,27]. Of course, those investigators who are constructing diversity-oriented [28] or focused combinatorial libraries generally synthesize these with a downstream HTS in mind, and as such these libraries should be immediately compatible with a central compound repository.

Ideally, multiple chemistry departments would create such compound repositories, and in certain circumstances these could be shared among various institutions. Efforts along these lines are underway at some institutions, including University of California at Irvine, Yale University, and University of Illinois at Urbana-Champaign. Highly complex, structurally unique small molecules will populate these collections, and in most cases several milligrams of each compound will be immediately on hand, greatly facilitating follow-up experiments. The screening of collections of modest size (<50 000 members) need not be performed with robotics and automation; minimally, only a multi-well plate reader and a hand-operated apparatus for pin transfer of compounds are required to test several thousands of compounds in a single day.

Interface with screening databases

The sheer volume of data generated from any high-throughput screen presents both challenges and

Figure 1



Intermediates produced during the course of natural product synthesis can be valuable biological probes. (a) The guanidine alkaloid shown inhibits the interaction of the HIV-1 Nef protein with various cellular ligands. This compound was originally made as part of the synthesis of a guanidine alkaloid natural product and was discovered through a high-throughput screen. (b) HTS revealed the synthetic intermediate locostatin as an inhibitor of epithelial cell sheet migration. Locostatin was later found to target Raf kinase inhibitor protein (RKIP) and was used to unveil the role of RKIP in the regulation of cell motility.

opportunities. PubChem (<http://pubchem.ncbi.nlm.nih.gov/>), a component of the NIH Molecular Libraries Roadmap Initiative, will greatly facilitate the sharing and dissemination of screening data. Modeled after the popular PubMed, PubChem allows users to search chemical structures and to view full data sets from high-throughput screens. Other common databases, such as SciFinder and MDL Crossfire, also allow direct structure searches and refinement of results by biological activity.

Summary

The identification of biologically active small molecules is at the heart of modern chemical biology research. Several options are available for those who wish to perform a high-throughput screen, including purchasing collections of compounds or allowing others (who have collections of compounds) to perform the screen. However, other options exist, including one where compounds that have already been synthesized and that are present in every chemistry department are assembled into a collection for screening. These compounds constitute a unique opportunity and an immense, largely unexplored resource. Harnessing this potential in a systematic fashion, through assembly of compound collections, will greatly enable the next generation of compound discovery.

Acknowledgements

The author thanks the National Institutes of Health (R01-GM68385), the Office of Naval Research (N00014-02-1-0390), and the National Science Foundation (CHE-0134779, CTS-0120978) for funding various compound screening efforts.

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. Schreiber SL: **Small molecules: the missing link in the central dogma.** *Nat Chem Biol* 2005, **1**:64-66.
An overview and description of the notion that small molecules, defined genotypes, and resulting phenotypic readouts comprise a powerful matrix of interconnected information that arises from cell-based screening.
2. Spring DR: **Chemical genetics to chemical genomics: small molecules offer big insights.** *Chem Soc Rev* 2005, **34**:472-482.
An up-to-date primer on forward and reverse chemical genetic strategies.
3. Lokey RS: **Forward chemical genetics: progress and obstacles on the path to a new pharmacopeia.** *Curr Opin Chem Biol* 2003, **7**:91-96.
4. Stockwell BR: **Exploring biology with small organic molecules.** *Nature* 2004, **432**:846-854.
5. Sinha S, Chen JK: **Purmorphamine activates the hedgehog pathway by targeting smoothened.** *Nat Chem Biol* 2006, **2**:29-30.
In this manuscript a series of detailed cell biology and biochemical experiments are used to identify the target of a potent activator of the Hedgehog pathway.
6. Falsey RR, Marron MT, Gunaherath GM, Shirahatti N, Mahadevan D, Gunatilaka AA, Whitesell L: **Actin microfilament aggregation induced by withaferin A is mediated by annexin II.** *Nat Chem Biol* 2006, **2**:33-38.
7. Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z, Kong N, Kammlott U, Lukacs C, Klein C *et al.*: **In vivo activation of the p53 pathway by small-molecule antagonists of MDM2.** *Science* 2004, **303**:844-848.
The discovery of the nutilins, a class of compounds that bind to MDM2 and disrupt the MDM2-p53 interaction, is described.
8. Oltsersdorf T, Elmore SW, Shoemaker AR, Armstrong RC, Augeri DJ, Belli BA, Bruncko M, Deckwerth TL, Dinges J, Hajduk PJ *et al.*: **An inhibitor of Bcl-2 family proteins induces regression of solid tumours.** *Nature* 2005, **435**:677-681.
9. Ortholand JY, Ganesan A: **Natural products and combinatorial chemistry: back to the future.** *Curr Opin Chem Biol* 2004, **8**:271-280.
10. Ghosh S, Nie A, An J, Huang Z: **Structure-based virtual screening of chemical libraries for drug discovery.** *Curr Opin Chem Biol* 2006, **10**:194-202.
11. Rupasinghe CN, Spaller MR: **The interplay between structure-based design and combinatorial chemistry.** *Curr Opin Chem Biol* 2006, **10**:188-193.
12. Golebiowski A, Klopfenstein SR, Portlock DE: **Lead compounds discovered from libraries: part 2.** *Curr Opin Chem Biol* 2003, **7**:308-325.
13. Archer JR: **History, evolution, and trends in compound management for high throughput screening.** *Assay Drug Dev Technol* 2004, **2**:675-681.
14. Rawlins MD: **Cutting the cost of drug development?** *Nat Rev Drug Discov* 2004, **3**:360-364.
15. O'Connor KA, Roth BL: **Finding new tricks for old drugs: an efficient route for public-sector drug discovery.** *Nat Rev Drug Discov* 2005, **4**:1005-1014.
This review nicely highlights the benefits of screening FDA-approved compounds for novel biological activities.
16. Rothstein JD, Patel S, Regan MR, Haenggeli C, Huang YH, Bergles DE, Jin L, Dykes Hoberg M, Vidensky S, Chung DS *et al.*: **Beta-lactam antibiotics offer neuroprotection by increasing glutamate transporter expression.** *Nature* 2005, **433**:73-77.
17. Borisy AA, Elliott PJ, Hurst NW, Lee MS, Lehar J, Price ER, Serbedzija G, Zimmermann GR, Foley MA, Stockwell BR *et al.*: **Systematic discovery of multicomponent therapeutics.** *Proc Natl Acad Sci USA* 2003, **100**:7977-7982.
This paper describes the systematic screening of combinations of FDA-approved compounds and the discovery of potential therapies for a multitude of disease states.
18. Keith CT, Borisy AA, Stockwell BR: **Multicomponent therapeutics for networked systems.** *Nat Rev Drug Discov* 2005, **4**:71-78.
19. Wadman M: **The right combination.** *Nature* 2006, **439**:390-391.
20. Bregman H, Carroll PJ, Meggers E: **Rapid access to unexplored chemical space by ligand scanning around a ruthenium center: discovery of potent and selective protein kinase inhibitors.** *J Am Chem Soc* 2006, **128**:877-884.
21. Williams DS, Atilla GE, Bregman H, Arzoumanian A, Klein PS, Meggers E: **Switching on a signaling pathway with an organoruthenium complex.** *Angew Chem Int Ed Engl* 2005, **44**:1984-1987.
22. Bregman H, Williams DS, Atilla GE, Carroll PJ, Meggers E: **An organometallic inhibitor for glycogen synthase kinase 3.** *J Am Chem Soc* 2004, **126**:13594-13595.
23. Musk DJ, Banko DA, Hergenrother PJ: **Iron salts perturb biofilm formation and disrupt existing biofilms of pseudomonas aeruginosa.** *Chem Biol* 2005, **12**:789-796.
Screening of the compounds collected at the University of Illinois at Urbana-Champaign revealed ferric ammonium citrate as an agent that disrupts bacterial biofilm formation.
24. Anon.: **Policy on papers' contributors.** *Nature* 1999, **399**:393.
25. Olszewski A, Sato K, Aron ZD, Cohen F, Harris A, McDougall BR, Robinson WE Jr, Overman LE, Weiss GA: **Guanidine alkaloid analogs as inhibitors of HIV-1 Nef**

- interactions with p53, actin, and p56lck.** *Proc Natl Acad Sci USA* 2004, **101**:14079-14084.
26. Mc Henry KT, Ankala SV, Ghosh AK, Fenteany G: **A non-antibacterial oxazolidinone derivative that inhibits epithelial cell sheet migration.** *Chem Bio Chem* 2002, **3**:1105-1111.
27. Zhu S, Mc Henry KT, Lane WS, Fenteany G: **A chemical inhibitor reveals the role of Raf kinase inhibitor protein in cell migration.** *Chem Biol* 2005, **12**:981-991.
28. Burke MD, Schreiber SL: **A planning strategy for diversity-oriented synthesis.** *Angew Chem Int Ed Engl* 2004, **43**:46-58.