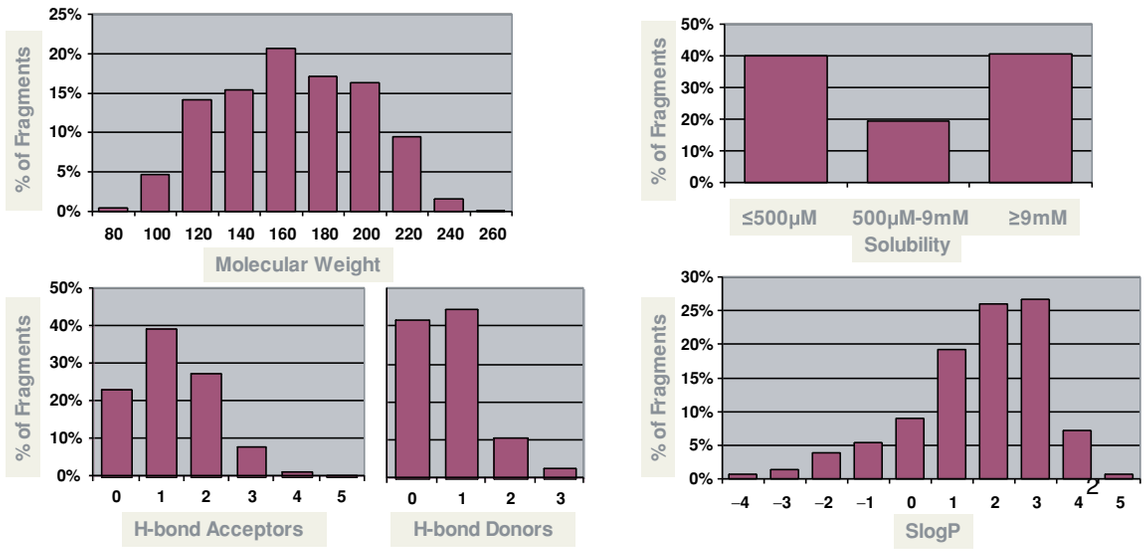

Figure 3.2. Properties of the SGX core fragment library.

into 150 shape-diverse mixtures of 10 compounds each. As discussed earlier, the estimated number of possible druglike molecules that could be included in an HTS library is $\sim 10^{60}$.[8] In contrast, the estimated number of all possible leadlike molecules with MW $<$ 160 is only about 14,000,000.[23] A fragment library containing a modest number of compounds (i.e., ~1,500) can, therefore, be used to sample leadlike compound space much more efficiently than an HTS library samples the space of druglike molecules. A fragment library of 1,000–10,000 small compounds (MW $<$ 160) represents ~0.001–0.01% of all possible leadlike compound spaces. In stark contrast, a typical HTS library, containing 10^5–10^6 compounds, encompasses $\sim 10^{-55}$ of the total estimated druglike space. Empirically, we have observed that hit rates for x-ray screening of our core fragment library fall in the range of 1–5%, thereby providing fifteen to seventy-five possible hits for subsequent fragment optimization. Taken together, these arguments/observations document that our ~1,500-compound core fragment library represents an efficient means of generating a reasonable number of starting points for fragment elaboration from a not insubstantial fraction of the total leadlike chemical space.

The potential chemical diversity of the SGX core fragment library is estimated to fall in the range of 10^8–10^{17} compounds. Intrinsic to most of the ~1,500 fragments are two to three sites for R-group addition or chemical handles. For each chemical handle, the number of commercial reagents available for chemical modification ranges from a minimum of ~400 to a maximum of ~40,000. In the most pessimistic scenario (i.e., use of only two chemical handles with only 400 possible independent modifications at each handle), ~1,500 fragments can be elaborated into $\sim 2.4 \times 10^8$ distinct compounds. In the most optimistic scenario (i.e., use of all three chemical handles with 40,000 possible independent modifications at each handle), ~1,500 fragments can be elaborated into $\sim 10^{17}$ distinct compounds, which is comparable to accepted estimates for the age of the universe in minutes.

Enabling the target, fragment x-ray screening, complementary biophysical screening, SAR optimization, and the end game

Enabling the target

At SGX, de novo protein crystal structures are determined using a gene-to-structure platform that supports prosecution of multiple protein samples in parallel. Our platform consists of modular robotics and a comprehensive laboratory information management system (SGX LIMS) that facilitates data entry and electronic data capture at all stages of the process. The SGX LIMS system also permits comprehensive data mining for troubleshooting and project management. The SGX target-to-structure platform has facilitated high-resolution (typically better than ~2Å resolution) structure determinations for a large number of drug discovery targets, including more than sixty unique human protein kinases, more than twenty unique human and pathogen protein phosphatases, a large number of nuclear hormone receptor ligand binding domains, and many bacterial and viral proteins. Successes have included many targets not represented in the public domain Protein Data Bank (PDB; www.pdb.org), some of which have been regarded as being extremely difficult if not "impossible" to express, purify, and crystallize (e.g., the IκB kinases). Modular SGX platform robotics encompasses gene cloning, protein expression and purification, crystallization, and structure determination. Most of this work is conducted using ninety-six-well-format liquid-handling robotics to process multiple expression constructs for many protein targets in parallel. Multiple constructs for a given target