Published in final edited form as:

Methods Enzymol. 2011 ; 493: 533–556. doi:10.1016/B978-0-12-381274-2.00021-2.

Fragment Screening of Infectious Disease Targets in a Structural Genomics Environment

Darren W Begley^{1,4}, Douglas R Davies^{1,4}, Robert Hartley^{1,4}, Thomas E Edwards^{1,4}, Bart L Staker^{1,4}, Wesley C Van Voorhis^{2,4}, Peter J Myler^{3,4}, and Lance J Stewart^{1,4,*}

¹ Emerald BioStructures, 7869 NE Day Road West, Bainbridge Island, WA 98110, USA, <http://www.emeraldbiostructures.com>

²University of Washington, Department of Medicine Box 356523, Seattle, WA 98195, USA <http://www.washington.edu>

³Seattle Biomedical Research Institute, 307 Westlake Avenue North, Suite 500, Seattle, WA 98109, USA <http://www.seattlebiomed.org>

⁴Seattle Structural Genomics Center for Infectious Disease, <http://www.ssgcid.org>

Abstract

Structural genomics efforts have traditionally focused on generating single protein structures of unique and diverse targets. However, a lone structure for a given target is often insufficient to firmly assign function or to drive drug discovery. As part of the Seattle Structural Genomics Center for Infectious Disease, we seek to expand the focus of structural genomics by elucidating ensembles of structures that examine small molecule-protein interactions for selected infectious disease targets. In this chapter, we discuss two applications for small molecule libraries in structural genomics: unbiased fragment screening to provide inspiration for lead development, and targeted, knowledge-based screening to confirm or correct the functional annotation of a given gene product. This shift in emphasis results in a structural genomics effort that is more engaged with the infectious disease research community, and one that produces structures of greater utility to researchers interested in both protein function and inhibitor development. We also describe specific methods for conducting high-throughput fragment screening in a structural genomics context by X-ray crystallography.

INTRODUCTION

Structural genomics of infectious disease proteins

Since its inception in 2000, the US-based Protein Structure Initiative has rapidly advanced high-throughput techniques for cloning, expression, purification, crystallization and structure determination (Fox et al., 2008). This pipeline approach has led to technological advances which dramatically reduce the per-target time and expense required to generate high-resolution protein structures. A decade later, the Protein Structure Initiative and other structural genomic efforts have generated nearly 5000 high-resolution structures, and

*Corresponding author: Lance J Stewart, lstewart@embios.com.

annually deposit the most unique structures in the Protein Data Bank (PDB) (Dessailly et al., 2009; Nair et al., 2009). Despite this impressive output, the Protein Structure Initiative has fallen short of its original aim to make "3D atomic-level structures of most proteins easily obtainable from knowledge of their corresponding DNA sequences" (Burley et al., 2008), leading external academic reviewers to question its overall goals and scientific value (Moore, 2007; Petsko, 2007; Service, 2008). One criticism often levied against structural genomics is a lack of engagement with the greater research community. High-resolution structure data for a well-characterized target can provide invaluable details on binding surfaces, allosteric and other biochemical behavior. However, biological function is rarely ascertained from structure alone, and protein structures whose function is uncharacterized can be seen as possessing limited value to biological research. Since the goal of the Protein Structure Initiative is to determine as many unique structures as possible, targets are often pushed through the pipeline long before functional characterization is validated, with putative function assigned based on sequence similarity. The reality is that target-specific functional assays are difficult and time-consuming endeavors which cannot keep pace with current structural genomics efforts (Gerlt, 2007).

The mission of the Seattle Structural Genomics Center for Infectious Disease (SSGCID) is to annually determine over 100 protein structures from NIAID Category A, B and C organisms (Myler et al., 2009). With this project, we seek to address some of the above-mentioned concerns regarding structural genomics by selecting potential drug targets which would directly benefit the infectious disease research community. To drive this effort, SSGCID target selection is guided by essential, known or putative function; this process provides a list of gene products which in some way could be the focus of drug design efforts. In pursuit of biologically relevant structures, we solicit requests directly from the scientific community. Most of our collaborators conduct research in academic laboratories, and many share our interest in generating novel leads for underserved infectious diseases. Today, there is an ever greater need for such research, as drug resistance continues to occur in pathogenic bacteria and other organisms, while the pipeline of future antimicrobial drugs dries up (Boucher et al., 2009). Our collaborative work includes a dedicated effort to solve small molecule-protein complexes for which there are proven or experimental inhibitors against homologous targets in other species (Myler et al., 2009; Van Voorhis et al., 2009). Structures of infectious disease proteins in complex with inhibitors can provide critical information in re-optimizing current drugs for maximum host-parasite selectivity when a structure of the human homolog is available. The structural information gained from protein-ligand complexes can also greatly accelerate functional assignment of targets for which sequence similarity is insufficient for annotation, as well as largely uncharacterized gene products with little available *in vitro* activity data (Hermann et al., 2007; Rakus et al., 2009; Song et al., 2007). Thus, in generating one or more ligand-bound structures for an infectious disease protein, we aim to fulfill our goal of providing a "blueprint for structure-guided drug design" (Myler et al., 2009).

Ligand-bound structure ensembles

In collaborating with biologists and other external researchers on SSGCID targets, we have found that *apo* structures, although valuable in and of themselves, are often not sufficient to

facilitate drug discovery efforts. At a minimum, infectious disease researchers would like information on the native substrate, or a close chemical analog, bound to the target in question. Such information offers scientific support confirming, or correcting, the annotated function of a given target and provides validated ligands which can serve as chemical probes for biological assays. Obtaining one or more inhibitor-bound structures can also guide the initial rounds of structure-based drug design, particularly for enzymes and proteins which act upon small molecule metabolites. Like other structural genomics initiatives, the SSGCID is ideally suited to generate *ensembles* of structures for a given target, employing the same structural genomics pipeline put in place to determine novel, *apo* protein structures.

When a new protein structure has been solved, the majority of the work – from sequence validation, expression, purification and crystallization to structure refinement – has been completed. By comparison, obtaining a series of ligand-bound structures requires little additional effort. Moreover, there is often a significant amount of protein remaining from purification that can be used for ligand binding studies by X-ray crystallography and other biophysical methods. With this asset of protein in hand, we believe that the additional work required to obtain one or more ligand-bound structures is well justified. For example, if there are known inhibitors of the human homolog, structures of those small molecules complexed with one or more microbial targets are highly desirable to the greater scientific community. Complexes with known inhibitors of homologous targets further support the functional characterization of the specific protein of interest, and can provide a basis for the first stages of drug design and development. This information can be of particular value in identifying point mutations or other structural mechanisms implicated in the development of observed drug resistance in pathogenic organisms (Yuvaniyama et al., 2003).

Fragment screening and X-ray crystallography

One strategy for generating an ensemble of structures for a single protein is fragment-based screening (Begley, manuscript in preparation; Brough et al., 2009; Card et al., 2005; Davies et al., 2009; Hartshorn et al., 2005). Fragment screening by X-ray crystallography with an all-purpose, unfocused compound library allows both the testing of proven and the discovery of new chemical moieties which bind to a protein. The size and reduced complexity of typical fragment molecules (< 300 Da) allow for a diverse sampling of chemical space with a compound library small enough for practical applications in crystallographic screening (Duarte et al., 2007; Hann et al., 2001; Makara, 2007; Rees et al., 2004). Due to their relatively small size, the overall affinity of fragments tends to be weaker, but their per-atom binding efficiency can be much higher than that of larger molecules (Abad-Zapatero and Metz, 2005; Congreve et al., 2003; Hajduk, 2006a; Hopkins et al., 2004). For most small molecules consisting of 15 or more non-hydrogen atoms, contributions to binding free energy increase linearly at best with atomic mass (Hajduk et al., 2005; Kuntz et al., 1999). This energetic correlation suggests that the binding energy of an initial, fragment-sized scaffold greatly determines the potential outcome of further optimization efforts, and that highly efficient fragment binders are therefore valuable starting points in drug design. Furthermore, fragments tend to bind protein "hot spots," localized areas which contribute the majority of binding energy to complex formation, even for large protein-protein interaction surfaces (DeLano, 2002; Hajduk et al., 2005; Wells and McClendon, 2007). Probing such

surfaces is a much more challenging prospect with larger compounds, such as those typically employed in high-throughput screening (HTS). The additional bulk and steric effects for such compounds often prevent hit detection by HTS, including those that embed or possess valuable chemical motifs within their scaffolds which could bind as individual fragments (Payne et al., 2007; Rees et al., 2004; Wells and McClendon, 2007). Thus, fragment-based screening can identify the most important regions of larger binding pockets which might be missed by screening larger compounds, and can be used to assess hits in the critical early stages of drug development (Bembenek et al., 2009; Hajduk, 2006a; Reitz et al., 2009).

Protein-based nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography are two biophysical methods which can be used to conduct screening of macromolecular targets as well as structurally characterize the bound conformations of fragment hits (Nienaber et al., 2000; Shuker et al., 1996). Early indications of success by these methods have driven considerable investment in fragment screening technologies, which include ligand-based NMR (Assadi-Porter et al., 2008; Lepre et al., 2004; Mayer and James, 2004; Pellecchia, 2005), mass spectrometry (Swayze et al., 2002), surface plasmon resonance (Navratilova and Hopkins, 2010; Neumann et al., 2007), differential phase interferometry (Bornhop et al., 2007), fluorescence correlation (Hesterkamp et al., 2007), and calorimetry (Ladbury et al., 2010). Many of these solution-based methods are well-suited to the rapid screening and detection of weak binding interactions between fragments and a broad range of macromolecular targets. However, some ligand-observe methods are prone to generating false positives, while most require additional measures to confirm binding site specificity, and none provide detailed structural information on the complex. Fragment screening by X-ray crystallography has the advantage that hits are simultaneously identified and structurally characterized in their target-bound state. Although not entirely dependent upon structural information (Erlanson et al., 2004), the majority of successful fragment screening campaigns reported in the literature have utilized experimentally-determined binding modes to advance weak but efficient fragments into potent inhibitors (Brough et al., 2009; Card et al., 2005; Davies et al., 2009; Hartshorn et al., 2005; Murray and Blundell, 2010; Sanders et al., 2004; Shuker et al., 1996). Recent successes using fragment screening and structure-based methods against infectious diseases have also yielded novel leads for the inhibition of fatty-acid (Mochalkin et al., 2009), pantothenate (Ciulli et al., 2008) and folate (Mpamhanga et al., 2009) biosynthesis in pathogenic organisms. It is clear that complex structures obtained from fragment screening can propel novel lead synthesis and development, and therefore can have a much greater impact on the work of external researchers investigating new ways to combat infectious diseases than *apo* structures alone.

METHODS

Fragment library selection, pooling and storage

Much has previously been published on fragment library selection and the various determinants for suitable candidate molecules, all of which depend on the biophysical technique(s) that are used for screening (Bemis and Murcko, 1996; Duarte et al., 2007; Ertl et al., 2006; Fejzo et al., 1999; Hajduk, 2006b; Hubbard et al., 2007; Muller, 2003; Siegel

and Vieth, 2007; Smith and Morowitz, 2004; Zartler and Shapiro, 2008). We employ a diverse, metabolome-inspired fragment library that consists of a structurally diverse array of over 1,500 water-soluble metabolites, their bioisosteres and close chemical analogs (Davies et al., 2009). In striving to establish a compact fragment set, we sought the greatest diversity within this biological context among fragment-sized molecules which already exist in nature, such as substrates, natural products, allosteric regulators, and other associates of metabolic pathways. We took into account ligand efficiency (Abad-Zapatero and Metz, 2005; Andrews et al., 1984; Mochalkin et al., 2009), privileged scaffolds (Bemis and Murcko, 1996; Duarte et al., 2007; Ertl et al., 2006; Muller, 2003), shape diversity (Binkowski and Joachimiak, 2008; Hajduk, 2006b; Lipkus et al., 2008), physicochemical properties and other parameters in order to maximize coverage of chemical space. Among the fragments returned from these searches, those which would best represent their class while also facilitating various methods of chemical elaboration were prioritized. The result of this cross-parameter search guided by the natural metabolome was the Fragments of Life™ (FOL) small molecule screening collection, our in-house fragment compound library (Davies et al., 2009). In our view, application of this all-purpose library maximizes the likelihood of finding native substrates or substructures bound, while simultaneously probing unpredicted chemistry for binding interactions with our chosen targets.

Compounds selected for inclusion in our Fragments of Life™ library were purchased from various vendors and tested for 50 mM solubility in methanol. This initial level of solubility facilitates the use of ligands at millimolar concentrations needed to perform ligand-based NMR studies and crystal soaks. The use of methanol avoids precipitation issues associated with long-term storage of test compounds more commonly prepared in dimethyl sulfoxide solution (Waybright et al., 2009). We require compound purity to be 95% as confirmed by NMR and LCMS spectroscopy, prior to dissolution and preparation for screening. Compounds were rapidly dissolved in methanol and stored at -20°C to prevent concentration-altering evaporation. We then prepared mixtures of Fragments of Life™ compounds, to test a greater number of fragments for each crystal soaked. Several techniques, including NMR spectroscopy and X-ray crystallography, utilize mixtures of diverse fragments for screening which can later be deconvoluted from a single experiment with reasonable confidence. However, we follow up every mixture hit with individual ligand experiments to confirm fragment identity and binding.

To prepare our pools, we used computational distance metrics to randomly sort our fragments into groups of 8 diverse molecules (Davies et al., 2009). The spatial arrangement of small molecule atoms in each pool is intended to facilitate rapid identification of bound fragments based on shape differences in electron density blobs. Once pooled, solutions were visually inspected to ensure maximum solubility and stable long-term storage. Any signs of precipitation, aggregation or destabilization upon mixing meant fragments would be recombined into different pools, until we arrived at a stable set of fragment mixtures. The majority of our fragment pools consist of 8 ligands per mixture, dissolved in methanol at 50 mM total ligand. This results in a nominal concentration of 6.25 mM for each fragment in a pool during the soak. For individual fragment soaks, such as those conducted to confirm identities of binding molecules, we typically use stocks at 25 mM.

We intentionally selected an all-purpose, volatile solvent for our fragment stock solutions to provide optimal flexibility when setting up crystallographic experiments. We allow methanol-based fragment stocks to evaporate prior to preparing soaking drops, leaving virtually no solvent residue behind which might dissolve or interfere with the protein crystal. This allows the rapid preparation of crystallization plates with dry films of fragment mixtures that can be stored indefinitely. Methanol is also miscible with many crystallization buffers, allowing the direct addition of fragment stock solutions when necessary. Using an evaporative solvent further facilitates concentration scaling within the confines of the crystallization plate, allowing different fragment-protein ratios for target-specific experiments. However, methanol creates certain liquid handling challenges because of evaporation, and care must be taken to keep solutions in sealed vials as much as possible. In preparing our Fragments of Life™ library, we conducted informal evaporation studies and showed that a borosilicate vial with a Teflon cap containing 2.5 mL of methanol loses less than 14 mg of methanol per day when stored at room temperature. Storage of the same solution in the same vial at −20 °C reduces loss of solvent by evaporation to negligible levels (data not shown). Hence we have found borosilicate glass vials with plastic screw caps and Teflon inserts to be an acceptable medium for long-term storage of methanol stocks at −20 °C.

Target selection

Each target nominated for fragment studies within the SSGCID must first be assessed to ensure prioritization of the most relevant and potentially impactful targets. Targets are either nominated by the external scientific community and approved by our NIAID program officer, or proposed internally and then reviewed and approved by an external panel of scientific advisors. Approved targets are then evaluated for their feasibility in crystallographic fragment screening studies. In another chapter of this volume, we discuss and analyze characteristics of protein targets which are amenable to fragment screening. To date, the majority of targets selected for fragment screening studies have been evaluated using various experimental criteria to predict the likelihood of success in obtaining ligand-bound structures, as described below:

1. To begin, a protein must first be successfully crystallized and a structure determined, whether in *apo* form or with a targeted ligand bound.
2. Protein crystals are typically selected which diffract to 1.8 Å in order to reliably obtain 2.0 Å datasets using our in-house X-ray source. Resolution in this range permits the correct orientation of a bound ligand to be determined from the experimentally-observed electron density.
3. Crystals with small unit cells and high symmetry space groups are preferable to minimize data analysis time, but we have established no fixed limit on either criterion to be used as a requirement for fragment screening.
4. If any metals, co-factors, prosthetic groups, or other components are necessary to make relevant binding site(s) available on the protein, they are added at millimolar concentration to the crystallization drop.

5. Protein structures with well-ordered electron density for polyatomic anions or buffer components observed in the predicted or known active site(s) are generally undesirable, as they will compete with fragments for the binding pocket(s). In our experience, citrates, phosphates, sulfates, acetates, and other such charged groups which appear highly ordered in a crystal can be difficult to displace with weak-binding fragments. Crystallization conditions that require polyatomic anions are therefore to be avoided, if possible.
6. Strongly acidic or basic conditions are generally undesirable, as these can have adverse effects on the stability of the target. However, non-neutral pH is sometimes implicated in the physiological relevance of a target, such as with pH sensor proteins. The pH of the crystal soaking drop will affect the protonation state of fragments as well as protein side chains, and should therefore be taken into account when conducting specific pH-range screens.

Crystal growth plates

Once a candidate protein has been selected, 96-well sitting-drop "farm" trays are plated with the same condition in each well and reservoir. Crystallization conditions are developed from the random sparse matrix screening effort employed to obtain the original crystal and structure, prior to fragment studies. The precipitant chosen for fragment-soaking crystals is usually the same, or based on consensus between several successful conditions which led to crystal growth. Typically, drops consist of 0.4 μL protein solution mixed with 0.4 μL precipitant in the sitting drop well, with 80–100 μL of precipitant placed in the reservoir. Our standard protein solution contains approximately 20 mg/mL protein with 500 mM NaCl, 20 mM HEPES, 5% glycerol and 2 mM DTT at a pH = 7.0. Similar buffer conditions have been shown to allow flash-freezing in PCR-sized tubes with minimal effects on protein precipitation (Deng et al., 2004). Wells are sealed with transparent tape, and ideally generate useable, uniform crystals within 24 hrs to 7 days of sitting. If crystals take longer than a week to grow, or appear non-uniform between wells, further optimization and consideration of additives is done, depending on the target.

Fragment soak plates

To prepare soak plates, 1.0 μL drops of pooled fragments dissolved in methanol are added to each of several sitting-drop tray wells and allowed to evaporate overnight. A 1.0 μL drop of soak solution is then added to one well per pool to resolubilize the fragments, prior to adding protein crystals. The other wells containing fragment dry films are reserved for cryoprotecting and repeat experiments. Although methanol-based fragment drops can also be plated serially to create soaking mixtures, consistent plating is difficult for volumes < 0.25 μL with highly evaporative solvents, while plating > 1.0 μL runs the risk of fragment material deposition above the level of the soak drop. We typically use Emerald BioSystems Compact Jr. crystallization plates (EBS-XJR) for long-term storage of dry film and fragment pool solutions. Important factors for storage are chemical resistance to degradation, shape of wells for easy harvesting, proper sealing to prevent contamination, and accommodation of drop volumes up to 2.5 μL without wicking to the sealant.

Crystal soaking

Once the *apo* starting crystals of the target protein are grown, they are transferred to sitting-drop wells for fragment soaking. Typically, 1.0 μ L of the crystallization condition is used as the soak solution, and is first added to resolubilize dry film fragment pools. Additional target-specific components can also be added to the soak, such as catalytically important metal ions, or small molecules which facilitate fragment binding to specific sites through co-complexation. Then 2 or more protein crystals are transferred to each individual well using cryo loops, and the wells sealed using Crystal Clear sealing tape (Hampton Research, Aliso Viejo, CA). Prior to sealing, 20 μ L of the crystallization solution (without additives) is added to the reservoir to maintain vapor pressure and prevent the drop from drying out completely.

Ideal soaking solutions are those which contain cryoprotectants, and which solubilize the fragments without destroying the protein crystal. If one component of the crystal precipitant has cryoprotecting qualities, we often increase that component to 25% (v/v) so fragment-soaked crystals can be directly frozen in liquid nitrogen from the soak drop. To maximize fragment solubility, 5% (v/v) or more dimethyl sulfoxide is sometimes added for crystals previously proven to tolerate this co-solvent concentration without dissolving. Crystals are typically soaked for 24 hrs to 1 week or longer, depending on the target, to ensure maximum fragment binding. We have found that some targets with highly ordered buffer components or coordinated water molecules in the active site require longer soak periods in order for fragments to effectively displace them within the crystal. We usually perform trial soaks with 24 to 48 fragment pools, as well as any known substrates or binding ligands found through literature searches, before screening the entire Fragments of Life™ library against a new target.

Cryoprotection

After fragment soaking, one or more protein crystals per pool are harvested by looping and frozen in liquid nitrogen. Unless the soak condition was designed as a cryoprotectant, crystals are dipped into a cryogenic protecting drop immediately prior to freezing. It is important that all fragment molecules be included in the cryo drop at the same concentration as the soak drop to prevent concentration-dependent diffusion and loss of fragments from the crystal. Typical cryogenic conditions are 25 % (v/v) ethylene glycol, glycerol, or low molecular weight polyethylene glycol mixed with the soaking solution. This cryo drop is used to resolubilize fragment pools previously added as dry film to separate wells or mixed with fragments in solution, before dipping and freezing crystals.

Data collection, data processing and fragment soaking analysis

In most cases, collection parameters for a fragment screening target are automatically selected using d*TREK software with our Rigaku X-ray system (Pflugrath, 1999). These parameters are based on the symmetry order determined from previous experiments and a few diffraction images of the first crystal tested. Diffraction data are reduced with XDS/XSCALE and customized in-house processing scripts (Kabsch, 1993). Initial structure solutions are achieved by molecular replacement methods, followed by rigid body refinement using a previous structure of the protein stripped of waters and other ligands but

retaining any catalytic metals or prosthetic groups. Crystals must diffract to 2.5 Å or better resolution for a robust analysis of fragment screening; if crystals dissolve, or diffract to weaker than 2.5 Å resolution after initial testing with fragment soaks, we re-optimize for new conditions before pursuing a full library screen. Unmodeled blobs with $(F_o - F_c) > 2.5\sigma$ and $(2F_o - F_c) > 1.0\sigma$ are first verified against the model protein structure using Coot, the Crystallographic Object-Oriented Toolkit (Emsley and Cowtan, 2004). Blobs are ignored which are likely to be unmodeled protein density, or identified as small molecules associated with crystallization conditions, cryogens, or the protein buffer. The remaining regions of unmodeled density are then examined for shape complementarity with individual fragments in a given mixture. When a putative fragment hit is isolated, preliminary deconvolution is done to identify the ligand by modeling in the suspected fragment and refining the structure. The fragment identity is then confirmed by conducting additional soaks and data collection for protein crystals with individual fragments.

We utilize the Collaborative Computational Project Number 4 (CCP4) suite of programs, along with TLS motion determination, to refine crystallographic datasets into accurate structural models (CCP4, 1994; Painter and Merritt, 2006). Every protein residue, ligand, and water molecule in every SSGCID structure model that is deposited in the PDB is visually examined and compared to its local electron density. Every SSGCID structure solved by a crystallographer is then internally reviewed by another crystallographer for consistency between electron density and backbone, side-chain and solvent atoms in the model, prior to deposition. This procedure routinely catches small errors, such as side chain rotamer flips, poorly modeled waters, and novel ligand stereochemical or tautomeric centers. All structures released by the SSGCID must meet a series of internally set criteria prior to deposition with the PDB. Datasets must be 80% complete in the highest resolution shell and 90% over the entire dataset. Resolution must be 2.5 Å or better, with a signal-to-noise ratio $(I/\sigma) > 2.0$ across the top 20 resolution shells. All models must have > 95% of residues in favored Ramachandran regions as defined by MolProbity (Chen et al., 2010; Davis et al., 2007), with structural or experimental rationale provided for any outliers.

CASE STUDIES

Fragment screening for lead generation in structural genomics

Much interest has been garnered in the past decade on the methyl-erythritol isoprenoid (MEP) biosynthetic pathway as a point of intervention for antimicrobials (Akerley et al., 2002; Brown and Parish, 2008; Campbell and Brown, 2002; Campos et al., 2001; Freiberg et al., 2001). This pathway is present and appears to be essential in *Plasmodium falciparum*, *Mycobacterium tuberculosis*, *Pseudomonas aeruginosa*, *Burkholderia pseudomallei*, *Escherichia coli* and other infectious diseases (Brown and Parish, 2008; Freiberg et al., 2001; Hunter, 2007; Jacobs et al., 2003; Jomaa et al., 1999). MEP enzymes are absent in mammalian species, all of which utilize the well-characterized mevalonate-dependent (MAD) pathway to synthesize isoprenoid precursors (Akerley et al., 2002; Brown and Parish, 2008; Campbell and Brown, 2002; Campos et al., 2001; Freiberg et al., 2001). Early studies on fosmidomycin indicate clinical utility and minimal side-effects in its use for treating *P. falciparum* infections by inhibiting IspC, one enzyme in the MEP pathway

(Borrmann et al., 2004a; Borrmann et al., 2004b; Jomaa et al., 1999; Wiesner et al., 2003; Wiesner et al., 2002). This observation suggests that enzymes within the MEP pathway, such as IspC, can be targeted to treat infectious agents with minimal host toxicity. Recrudescence after two weeks of treatment revealed fosmidomycin alone allows rapid development of drug resistance, but if given in combination, fosmidomycin appears to retain its clinical activity (Missinou et al., 2002; Wiesner et al., 2003). Therefore, novel chemotherapies which target one or more enzymes within the MEP pathway should minimize the chance for resistance mutations to develop and survive in variant strains of pathogenic species for which it is essential.

Another enzyme in the MEP pathway is 2C-methyl-D-erythritol-2,4-cyclo-diphosphate synthase (IspF). Sequence analysis shows over 30% identity for IspF proteins across prokaryotic and eukaryotic species, and structural studies reveal high conservation for residues essential in substrate-binding, catalysis and other aspects of IspF enzymatic function (Buetow et al., 2007; Eisenreich et al., 2004; Eoh et al., 2007; Hunter, 2007; Lange et al., 2000; Ni et al., 2004a; Ni et al., 2004b; Rohdich et al., 2001; Steinbacher et al., 2002; Ulaganathan et al., 2007). Preliminary analysis of IspF from *Burkholderia pseudomallei* (BpIspF) showed the purified enzyme to be active *in vitro*, stable in low salt buffers, and capable of forming relatively large, solid crystals that are amenable to soaking in fragment-friendly drop conditions. Screening with a subset of our library also revealed that approximately 15% of fragments bind to BpIspF in solution by ligand-observe NMR spectroscopy. These results and success with preliminary crystal trials prompted a full screen of BpIspF against our complete Fragments of Life™ library (Begley, manuscript in preparation).

Screening BpIspF with our Fragments of Life™ library generated an ensemble of fragment-bound structures (Figure 1). One subset of BpIspF structures generated from fragment screening contain cytosine-based fragments, including cytosine, cytidine, CMP, CDP and CTP. These nucleotide-derived fragments all bind to the same active site sub-pocket and mimic the cytidyl portion of the native substrate, 4-diphospho-cytidyl-2-C-methyl-D-erythritol 2-phosphate (Figure 1). Results for this subset of fragments with BpIspF were somewhat anticipated, and the binding modes of these cytosine-based molecules recapitulate those observed in IspF from other organisms (Kemp et al., 2005). Fragment screening generated another subset of structures which contain heteroaromatic fragments bound to a zinc ion in the BpIspF active site (Figure 1). Each of these complex structures reveals an aromatic nitrogen coordinated to zinc, displacing a crystallographically conserved water molecule that normally occupies the tetrahedral coordination site. These structures are unlike most ligand-bound complexes of IspF, as they do not require any substrate-derived moieties for binding. Finally, a third subset of structures resulted from a focused set of crystal soaks with NMR-derived fragment hits in the presence of cytidine (Figure 1). These fragments bind in a small hydrophobic cleft outside the active site formed by a protein loop just above the catalytic zinc ion. Soaking trials for these fragments alone and in combination with cytosine did not result in ligand-bound structures, suggesting that the ribose moiety of cytidine is required for high occupancy binding (Begley, manuscript in preparation). Combination soaks show that the binding modes for representative fragments from all three

subsites to be reproducible within a single crystal complex, as is seen in PDB #####. Thus for one target, fragment screening by two techniques with a single library has led to the deposition of over a dozen complexes, which currently comprise nearly a third of all IspF structures from any species currently available in the PDB.

The ensemble of fragment-bound BpIspF structures generated by SSGCID provides several avenues for structure-based, fragment-assisted inhibitor design. Novel fragments bound in the zinc site serve as starting points for elaboration into higher affinity metal-chelating compounds. Fragment pairs bound to adjacent sub-pockets of the active site provide accurate inter-molecular distances and orientations, information which can be used in merging or linking strategies to enlarge fragments into initial leads. Preliminary attempts to design and synthesize such leads led to a small series of cytidine derivatives (Begley, manuscript in preparation). These pseudo-bidentate ligands were designed from adjacent binding fragment pairs of cytidine and several heteroaromatic zinc-binding fragment motifs. Our first cytidine fusion molecule (PDB ID 3KE1) shows the pyridine nitrogen coordinated to zinc in the same manner as was observed in its fragment precursor (Figure 1). Our second cytidine derivative (PDB ID #####) recapitulates a different zinc-coordinated fragment, but its mode of binding does not reflect that of the fragment which inspired it. Instead of coordinating directly to zinc, this bicyclo-aryl octane moiety binds into a hydrophobic groove above the metal, on the opposite side of the external fragment-binding site loop (Figure 1). This hydrophobic pocket helps position the erythritol portion of the substrate for catalysis, and although computationally predicted to bind fragments, none were observed through experimental fragment screening. Utilizing fragment hits to develop and test larger molecules in this manner expands our available knowledge of the ligand-binding active site landscape. Moreover, conducting crystallographic trials with fragment-inspired leads reveal that such ligands do not necessarily recapitulate the binding modes of their fragment precursors when bound. Thus the acquisition and interpretation of these and future datasets on BpIspF complexes provide crucial details for each iterative cycle of our design efforts in pursuit of potent, novel inhibitors for this infectious disease target.

Investigations into the mechanism of IspF suggest that the loop above the catalytic zinc ion properly positions the 2-phosphate of the substrate for attack on its own β -phosphate, prior to cyclization and release of product (Hunter, 2007; Steinbacher et al., 2002). This loop is well-ordered in the *apo* structure of IspF (PDB ID 3F0E) but appears too disordered to accurately model when complexed with a bidentate ligand (PDB ID 3KE1), suggesting a degree of flexibility which may be required for catalytic activity. External site-binding fragments appear to stabilize this loop, binding to it outside the active site, and result in complexes which recapitulate the *apo* structure (Begley, manuscript in preparation). Currently, any real or potential biological effects of binding small molecules to this external site are unknown. Investigations with such external site fragment hits should always be undertaken cautiously, especially when there are no proven competitors for the site or any known or hypothesized allosteric mechanisms affecting enzymatic function. Nevertheless, this example of biophysical screening and structure ensemble generation reveal the potential to identify novel binding sites through fragment screening. Such complexes may also be of particular value in a structural genomics environment where annotated function is

unconfirmed and supply biological research with additional chemical probes for functional characterization.

Ligand screening for functional annotation in structural genomics

Several genes from the SSGCID pipeline under investigation are annotated as arginine/ornithine transport system ATPases from *Mycobacterium* genomes. Our initial efforts were focused on the ATPases from *M. tuberculosis*, the organism which causes tuberculosis and which is estimated to infect one-third of the world's population (Jasmer et al., 2002). Crystallization trials of this target in the *apo* form yielded needle clusters that failed to diffract better than 7 Å resolution, despite several attempts at gradient screening to optimize conditions. Due to the difficulty in obtaining well-diffracting *apo* crystals, this particular target was not a candidate for fragment screening. Our continued efforts to obtain any structures of this target then took a two-pronged approach: knowledge-based ligand additives to enhance crystallization conditions, and re-engineering of the construct using genetic data from closely related non-pathogenic *M. smegmatis* and *M. thermoresistibile* species (Edwards, manuscript in preparation).

Although the *Mycobacteria* genes which code for this target family are annotated as ATPases, a BLAST search (Altschul et al., 1990) against existing PDB entries yielded four hits, all of which belong to the RAS-like GTPase superfamily of proteins. Of these four entries, two structures were in the *apo* form, while the other two contained molecules of GDP bound to the protein. To maximize our chances for success, we attempted co-crystallization trials with the *M. tuberculosis* transporter protein in the presence of ATP, ADP, GTP, and GDP. Crystals grown in the presence of ATP or ADP yielded the same weakly-diffracting needle-shaped clusters that were observed for the *apo* protein samples. In contrast, crystals grown in the presence of either GTP or GDP yielded block shaped crystals. From a crystal grown in the presence of GDP, we determined a 2.45 Å resolution structure (PDB ID 3MD0). In this structure, the guanine moiety of GDP forms an N1/N3 interaction with the side chain carboxylate of Asp207, and stacks between the side chains of Lys205 and Val145. It is highly unlikely that an adenosine could form similar N6/N1 interactions with Asp207 without significantly shifting the nucleobase out of the binding pocket (Figure 2). Using optimization grid screening to further enhance crystallization conditions, we then obtained a 1.9 Å resolution structure from a crystal grown in the presence of GTP (PDB ID #####). Upon refining the model against the data, it became clear that a single molecule of GDP, not GTP, was bound in the G-domain active site. Apparently, the enzymatic activity of this target protein is sufficient to hydrolyze GTP into GDP during crystallization, with enough efficiency to effectively remove the phosphate by-product from the binding pocket. This finding strongly suggests that the original ATPase annotation for this transporter is incorrect and demonstrates the power of structural genomics to revise such information when searching for and prioritizing ligand-bound structures (Edwards, manuscript in preparation).

In further attempts to characterize this target family, we conducted co-crystallization trials of the transporter protein with GDP, arginine and ornithine. Despite obtaining several 1.8 Å resolution or better datasets, we did not observe electron density for either arginine (50 mM

in the crystallization drop) or ornithine (5 mM in the crystallization drop) in the *M. tuberculosis* target. We then determined crystal structures of the homologous transporter systems from *M. smegmatis* (2.3 Å resolution, PDB ID 3NXS) and *M. thermoresistible* (2.5 Å resolution, PDB ID #####). In each case, we observed GDP in the highly conserved active site of each protein, but no arginine or ornithine in the unit cell. The sum total of this effort has led us to conclude that these proteins from closely related *Mycobacterium* species are not likely arginine/ornithine transport system ATPases. Not only do these proteins fail to bind to ATP, ADP, arginine, or ornithine in crystallization trials, they require the presence of GDP or GTP to obtain crystals capable of high-resolution X-ray diffraction. Furthermore, the target from *M. tuberculosis* is capable of hydrolyzing GTP to GDP in the crystallization drop, demonstrating enzymatic activity. More detailed sequence analyses show these proteins to align well with metallochaperones from the G3E family of P-loop GTPases, such as MeaB from *Methylobacterium extorquens* and human MMAA, a protein associated with methylmalonic aciduria (Dobson et al., 2002; Korotkova et al., 2002; Korotkova and Lidstrom, 2004). Therefore, these proteins, which are currently annotated as arginine/ornithine transport system ATPases, are most likely metallochaperone GTPases similar to MeaB.

Conclusions

To have the biggest impact on biological research, we believe that structural genomics must continue to increase and improve the diversity of available ligand-bound structures. We believe that structural genomics of infectious diseases garners more significance and better engages the scientific community by producing *ensembles* of complex structures, rather than *apo* structures alone. Such structures have already generated interest within the scientific community, whose requests for ligand-bound structures and fragment screening at SSGCID continue to grow in number. Utilization of novel fragments complexed with infectious disease proteins may also provide a much-desired avenue for the development of novel antimicrobial drugs in the wake of recent failures to generate such leads by other means. Historically, the great challenge in antibacterials is the generation of new chemical scaffolds. The path to registration has less risk when compared with other disease areas, once an effective scaffold is recognized which targets a particular bacterium. A recent survey noted over 125 antibacterial high-throughput screening (HTS) campaigns at 34 different pharmaceutical companies in the past 15 years yielded no promising drug candidates (Chan et al., 2004; Payne et al., 2007; Payne et al., 2004). HTS studies typically employ larger, more developed small molecules in the hopes of immediate identification of a highly potent inhibitor, and insufficient chemical diversity across such libraries is a likely reason for this failure. The HTS assay format requires low micromolar activity for detection, which means that inherently tight-binding leads must already be present within the library before screening has begun. (Keseru and Makara, 2006; Murray and Blundell, 2010; Payne et al., 2007; Rees et al., 2004). Conversely, fragment molecules are smaller and less complex than those used in HTS technologies, and therefore are inherently less biased towards any specific protein family. Thus, academic and industrial researchers alike can benefit from utilization of small molecular fragments in structural investigations of rare and structurally diverse drug targets, such as those present in pathogenic organisms.

Not every structural genomics protein is suitable for ligand-binding studies, as the biological functions of many proteins cannot be influenced through small molecule binding. However, any structural genomics pipeline equipped to rapidly generate protein structures by X-ray crystallography has the ability to generate ensembles of structures for targets which do bind small molecules, with minimal added effort. One reason for this is that the majority of backbone and side chain residues across such an ensemble (for most targets) will be identical or highly similar; in most cases, only a small portion of target atoms are altered to accommodate the binding entity. But it is precisely this structural data at the interface of protein-ligand binding interactions which can be of greatest value to researchers investigating their biological function. Adding the known (or putative) native substrate to a target for structural purposes can provide evidence of functional activity and help correctly annotate function when homology or literature assignment is incomplete. For chemical biologists, fragments which have been biophysically characterized can provide soluble starting points for experimental investigations. And perhaps the greatest impact fragment-bound structures can make is on the development of novel chemotherapeutics, whether to treat infectious diseases or to guide other structure-based drug design endeavors. We believe fragment screening furnishes a valuable resource to the infectious disease community by providing blueprints for structure-based drug design which would not be discovered by any other methodology.

ACKNOWLEDGEMENTS

We would like to thank Alex S. Kiselyov and John M. McCall for helpful comments in reading this book chapter. We further acknowledge and greatly appreciate the efforts of the entire team at SSGCID, without whom our accomplishments would not be possible. Support for this research was funded by NIAID under Federal Contract No. HHSN272200700057C.

REFERENCES

- Abad-Zapatero C, Metz JT. Ligand efficiency indices as guideposts for drug discovery. *Drug Discov Today*. 2005; 10:464–469. [PubMed: 15809192]
- Akerley BJ, Rubin EJ, Novick VL, Amaya K, Judson N, Mekalanos JJ. A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. *Proc Natl Acad Sci U S A*. 2002; 99:966–971. [PubMed: 11805338]
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990; 215:403–410. [PubMed: 2231712]
- Andrews PR, Craik DJ, Martin JL. Functional group contributions to drug-receptor interactions. *J Med Chem*. 1984; 27:1648–1657. [PubMed: 6094812]
- Assadi-Porter FM, Tonelli M, Maillet E, Hallenga K, Benard O, Max M, Markley JL. Direct NMR detection of the binding of functional ligands to the human sweet receptor, a heterodimeric family 3 GPCR. *J Am Chem Soc*. 2008; 130:7212–7213. [PubMed: 18481853]
- Begley DW, Edwards TE, Hartley R, Staker BL, Davies D, Myler P, Stacy R, Stewart L. Iterative fragment binding for sub-pocket preorganization in MECP synthase from *B pseudomallei* (manuscript in preparation). manuscript in preparation.
- Bembenek SD, Tounge BA, Reynolds CH. Ligand efficiency and fragment-based drug discovery. *Drug Discov Today*. 2009; 14:278–283. [PubMed: 19073276]
- Bemis GW, Murcko MA. The properties of known drugs. 1. Molecular frameworks. *J Med Chem*. 1996; 39:2887–2893. [PubMed: 8709122]
- Binkowski TA, Joachimiak A. Protein functional surfaces: global shape matching and local spatial alignments of ligand binding sites. *BMC Struct Biol*. 2008; 8:45. [PubMed: 18954462]

- Bornhop DJ, Latham JC, Kussrow A, Markov DA, Jones RD, Sorensen HS. Free-solution, label-free molecular interactions studied by back-scattering interferometry. *Science*. 2007; 317:1732–1736. [PubMed: 17885132]
- Borrmann S, Adegnika AA, Matsiegui PB, Issifou S, Schindler A, Mawili-Mboumba DP, Baranek T, Wiesner J, Jomaa H, Kremsner PG. Fosmidomycin-clindamycin for *Plasmodium falciparum* Infections in African children. *J Infect Dis*. 2004a; 189:901–908. [PubMed: 14976608]
- Borrmann S, Issifou S, Esser G, Adegnika AA, Ramharter M, Matsiegui PB, Oyakhirome S, Mawili-Mboumba DP, Missinou MA, Kun JF, Jomaa H, Kremsner PG. Fosmidomycin-clindamycin for the treatment of *Plasmodium falciparum* malaria. *J Infect Dis*. 2004b; 190:1534–1540. [PubMed: 15478056]
- Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, Scheld M, Spellberg B, Bartlett J. Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin Infect Dis*. 2009; 48:1–12. [PubMed: 19035777]
- Brough PA, Barril X, Borgognoni J, Chene P, Davies NG, Davis B, Drysdale MJ, Dymock B, Eccles SA, Garcia-Echeverria C, Fromont C, Hayes A, Hubbard RE, Jordan AM, Jensen MR, Massey A, Merrett A, Padfield A, Parsons R, Radimerski T, Raynaud FI, Robertson A, Roughley SD, Schoepfer J, Simmonite H, Sharp SY, Surgenor A, Valenti M, Walls S, Webb P, Wood M, Workman P, Wright L. Combining hit identification strategies: fragment-based and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine inhibitors of the Hsp90 molecular chaperone. *J Med Chem*. 2009; 52:4794–4809. [PubMed: 19610616]
- Brown AC, Parish T. Dxr is essential in *Mycobacterium tuberculosis* and fosmidomycin resistance is due to a lack of uptake. *BMC Microbiol*. 2008; 8:78. [PubMed: 18489786]
- Buetow L, Brown AC, Parish T, Hunter WN. The structure of *Mycobacteria* 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase, an essential enzyme, provides a platform for drug discovery. *BMC Struct Biol*. 2007; 7:68. [PubMed: 17956607]
- Burley SK, Joachimiak A, Montelione GT, Wilson IA. Contributions to the NIH-NIGMS Protein Structure Initiative from the PSI Production Centers. *Structure*. 2008; 16:5–11. [PubMed: 18184575]
- Campbell TL, Brown ED. Characterization of the depletion of 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase in *Escherichia coli* and *Bacillus subtilis*. *J Bacteriol*. 2002; 184:5609–5618. [PubMed: 12270818]
- Campos N, Rodriguez-Concepcion M, Sauret-Gueto S, Gallego F, Lois LM, Boronat A. *Escherichia coli* engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate: a novel system for the genetic analysis of the 2-C-methyl-d-erythritol 4-phosphate pathway for isoprenoid biosynthesis. *Biochem J*. 2001; 353:59–67. [PubMed: 11115399]
- Card GL, Blasdel L, England BP, Zhang C, Suzuki Y, Gillette S, Fong D, Ibrahim PN, Artis DR, Bollag G, Milburn MV, Kim SH, Schlessinger J, Zhang KY. A family of phosphodiesterase inhibitors discovered by cocrystallography and scaffold-based drug design. *Nat Biotechnol*. 2005; 23:201–207. [PubMed: 15685167]
- CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*. 1994; 50:760–763. [PubMed: 15299374]
- Chan PF, Holmes DJ, Payne DJ. Finding the gems using genomic discovery: antibacterial drug discovery strategies - the successes and the challenges. *Drug Discovery Today: Therapeutic Strategies*. 2004; 1:519–527.
- Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*. 2010; 66:12–21. [PubMed: 20057044]
- Ciulli A, Scott DE, Ando M, Reyes F, Saldanha SA, Tuck KL, Chirgadze DY, Blundell TL, Abell C. Inhibition of *Mycobacterium tuberculosis* pantothenate synthetase by analogues of the reaction intermediate. *ChemBiochem*. 2008; 9:2606–2611. [PubMed: 18821554]
- Congreve M, Carr R, Murray C, Jhoti H. A 'rule of three' for fragment-based lead discovery? *Drug Discov Today*. 2003; 8:876–877. [PubMed: 14554012]
- Davies DR, Mamat B, Magnusson OT, Christensen J, Haraldsson MH, Mishra R, Pease B, Hansen E, Singh J, Zembower D, Kim H, Kiselyov AS, Burgin AB, Gurney ME, Stewart LJ. Discovery of

- leukotriene A4 hydrolase inhibitors using metabolomics biased fragment crystallography. *J Med Chem.* 2009; 52:4694–4715. [PubMed: 19618939]
- Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 2007; 35:W375–W383. [PubMed: 17452350]
- DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. *Curr Opin Struct Biol.* 2002; 12:14–20. [PubMed: 11839484]
- DeLano, WL. The PyMOL Molecular Graphics System. Palo Alto, CA, USA: DeLano Scientific LLC; 2008.
- Deng J, Davies DR, Wisedchaisri G, Wu M, Hol WG, Mehlin C. An improved protocol for rapid freezing of protein samples for long-term storage. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:203–204. [PubMed: 14684931]
- Dessailly BH, Nair R, Jaroszewski L, Fajardo JE, Kouranov A, Lee D, Fiser A, Godzik A, Rost B, Orengo C. PSI-2: structural genomics to cover protein domain family space. *Structure.* 2009; 17:869–881. [PubMed: 19523904]
- Dobson CM, Wai T, Leclerc D, Kadir H, Narang M, Lerner-Ellis JP, Hudson TJ, Rosenblatt DS, Gravel RA. Identification of the gene responsible for the cblB complementation group of vitamin B12-dependent methylmalonic aciduria. *Hum Mol Genet.* 2002; 11:3361–3369. [PubMed: 12471062]
- Duarte CD, Barreiro EJ, Fraga CA. Privileged structures: a useful concept for the rational design of new lead drug candidates. *Mini Rev Med Chem.* 2007; 7:1108–1119. [PubMed: 18045214]
- Edwards TE. Structural Genomics of RAS-like GTPase Transport Proteins. *Acta Crystallogr F.* (manuscript in preparation). manuscript in preparation.
- Eisenreich W, Bacher A, Arigoni D, Rohdich F. Biosynthesis of isoprenoids via the non-mevalonate pathway. *Cell Mol Life Sci.* 2004; 61:1401–1426. [PubMed: 15197467]
- Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:2126–2132. [PubMed: 15572765]
- Eoh H, Brown AC, Buetow L, Hunter WN, Parish T, Kaur D, Brennan PJ, Crick DC. Characterization of the Mycobacterium tuberculosis 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase: potential for drug development. *J Bacteriol.* 2007; 189:8922–8927. [PubMed: 17921290]
- Erlanson DA, McDowell RS, O'Brien T. Fragment-based drug discovery. *J Med Chem.* 2004; 47:3463–3482. [PubMed: 15214773]
- Ertl P, Jelfs S, Muhlbacher J, Schuffenhauer A, Selzer P. Quest for the rings. In silico exploration of ring universe to identify novel bioactive heteroaromatic scaffolds. *J Med Chem.* 2006; 49:4568–4573. [PubMed: 16854061]
- Fejzo J, Lepre CA, Peng JW, Bemis GW, Ajay, Murcko MA, Moore JM. The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem Biol.* 1999; 6:755–769. [PubMed: 10508679]
- Fox BG, Goulding C, Malkowski MG, Stewart L, Deacon A. Structural genomics: from genes to structures with valuable materials and many questions in between. *Nat Methods.* 2008; 5:129–132. [PubMed: 18235432]
- Freiberg C, Wieland B, Spaltmann F, Ehlert K, Brotz H, Labischinski H. Identification of novel essential Escherichia coli genes conserved among pathogenic bacteria. *J Mol Microbiol Biotechnol.* 2001; 3:483–489. [PubMed: 11361082]
- Gerlt JA. A Protein Structure (or Function ?) Initiative. *Structure.* 2007; 15:1353–1356. [PubMed: 17997960]
- Hajduk PJ. Fragment-based drug design: how big is too big? *J Med Chem.* 2006a; 49:6972–6976. [PubMed: 17125250]
- Hajduk PJ. Puzzling through fragment-based drug design. *Nat Chem Biol.* 2006b; 2:658–659. [PubMed: 17108979]
- Hajduk PJ, Huth JR, Tse C. Predicting protein druggability. *Drug Discov Today.* 2005; 10:1675–1682. [PubMed: 16376828]
- Hann MM, Leach AR, Harper G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J Chem Inf Comput Sci.* 2001; 41:856–864. [PubMed: 11410068]

- Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H. Fragment-based lead discovery using X-ray crystallography. *J Med Chem.* 2005; 48:403–413. [PubMed: 15658854]
- Hermann JC, Marti-Arbona R, Fedorov AA, Fedorov E, Almo SC, Shoichet BK, Raushel FM. Structure-based activity prediction for an enzyme of unknown function. *Nature.* 2007; 448:775–779. [PubMed: 17603473]
- Hesterkamp T, Barker J, Davenport A, Whittaker M. Fragment based drug discovery using fluorescence correlation: spectroscopy techniques: challenges and solutions. *Curr Top Med Chem.* 2007; 7:1582–1591. [PubMed: 17979769]
- Hopkins AL, Groom CR, Alex A. Ligand efficiency: a useful metric for lead selection. *Drug Discov Today.* 2004; 9:430–431. [PubMed: 15109945]
- Hubbard RE, Davis B, Chen I, Drysdale MJ. The SeeDs approach: integrating fragments into drug discovery. *CurrTop Med Chem.* 2007; 7:1568–1581.
- Hunter WN. The non-mevalonate pathway of isoprenoid precursor biosynthesis. *J Biol Chem.* 2007; 282:21573–21577. [PubMed: 17442674]
- Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, Will O, Kaul R, Raymond C, Levy R, Chun-Rong L, Guenther D, Bovee D, Olson MV, Manoil C. Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A.* 2003; 100:14339–14344. [PubMed: 14617778]
- Jasmer RM, Nahid P, Hopewell PC. Clinical practice. Latent tuberculosis infection. *N Engl J Med.* 2002; 347:1860–1866. [PubMed: 12466511]
- Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, Hintz M, Turbachova I, Eberl M, Zeidler J, Lichtenthaler HK, Soldati D, Beck E. Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science.* 1999; 285:1573–1576. [PubMed: 10477522]
- Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography.* 1993; 26:795–800.
- Kemp LE, Alphey MS, Bond CS, Ferguson MA, Hecht S, Bacher A, Eisenreich W, Rohdich F, Hunter WN. The identification of isoprenoids that bind in the intersubunit cavity of *Escherichia coli* 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase by complementary biophysical methods. *Acta Crystallogr D Biol Crystallogr.* 2005; 61:45–52. [PubMed: 15608374]
- Keseru GM, Makara GM. Hit discovery and hit-to-lead approaches. *Drug Discov Today.* 2006; 11:741–748. [PubMed: 16846802]
- Korotkova N, Chistoserdova L, Kuksa V, Lidstrom ME. Glyoxylate regeneration pathway in the methylotroph *Methylobacterium extorquens* AM1. *J Bacteriol.* 2002; 184:1750–1758. [PubMed: 11872727]
- Korotkova N, Lidstrom ME. MeaB is a component of the methylmalonyl-CoA mutase complex required for protection of the enzyme from inactivation. *J Biol Chem.* 2004; 279:13652–13658. [PubMed: 14734568]
- Kuntz ID, Chen K, Sharp KA, Kollman PA. The maximal affinity of ligands. *Proc Natl Acad Sci U S A.* 1999; 96:9997–10002. [PubMed: 10468550]
- Ladbury JE, Klebe G, Freire E. Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat Rev Drug Discov.* 2010; 9:23–27. [PubMed: 19960014]
- Lange BM, Rujan T, Martin W, Croteau R. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci U S A.* 2000; 97:13172–13177. [PubMed: 11078528]
- Lepre CA, Moore JM, Peng JW. Theory and applications of NMR-based screening in pharmaceutical research. *Chem Rev.* 2004; 104:3641–3676. [PubMed: 15303832]
- Lipkus AH, Yuan Q, Lucas KA, Funk SA, Bartelt WF 3rd, Schenck RJ, Trippe AJ. Structural diversity of organic chemistry. A scaffold analysis of the CAS Registry. *J Org Chem.* 2008; 73:4443–4451. [PubMed: 18505297]
- Makara GM. On sampling of fragment space. *J Med Chem.* 2007; 50:3214–3221. [PubMed: 17567120]
- Mayer M, James TL. NMR-based characterization of phenothiazines as a RNA binding scaffold. *J Am Chem Soc.* 2004; 126:4453–4460. [PubMed: 15053636]

- Missinou MA, Borrmann S, Schindler A, Issifou S, Adegnik AA, Matsiegui PB, Binder R, Lell B, Wiesner J, Baranek T, Jomaa H, Kremsner PG. Fosmidomycin for malaria. *Lancet*. 2002; 360:1941–1942. [PubMed: 12493263]
- Mochalkin I, Miller JR, Narasimhan L, Thanabal V, Erdman P, Cox PB, Prasad JV, Lightle S, Huband MD, Stover CK. Discovery of antibacterial biotin carboxylase inhibitors by virtual screening and fragment-based approaches. *ACS Chem Biol*. 2009; 4:473–483. [PubMed: 19413326]
- Moore PB. Let's call the whole thing off: some thoughts on the protein structure initiative. *Structure*. 2007; 15:1350–1352. [PubMed: 17997959]
- Mpmahanga CP, Spinks D, Tulloch LB, Shanks EJ, Robinson DA, Collie IT, Fairlamb AH, Wyatt PG, Frearson JA, Hunter WN, Gilbert IH, Brenk R. One scaffold, three binding modes: novel and selective pteridine reductase 1 inhibitors derived from fragment hits discovered by virtual screening. *J Med Chem*. 2009; 52:4454–4465. [PubMed: 19527033]
- Muller G. Medicinal chemistry of target family-directed masterkeys. *Drug Discov Today*. 2003; 8:681–691. [PubMed: 12927511]
- Murray CW, Blundell TL. Structural biology in fragment-based drug design. *Curr Opin Struct Biol*. 2010; 20:497–507. [PubMed: 20471246]
- Myler PJ, Stacy R, Stewart L, Staker BL, Van Voorhis WC, Varani G, Buchko GW. The Seattle Structural Genomics Center for Infectious Disease (SSGCID). *Infect Disord Drug Targets*. 2009; 9:493–506. [PubMed: 19594426]
- Nair R, Liu J, Soong TT, Acton TB, Everett JK, Kouranov A, Fiser A, Godzik A, Jaroszewski L, Orengo C, Montelione GT, Rost B. Structural genomics is the largest contributor of novel structural leverage. *J Struct Funct Genomics*. 2009; 10:181–191. [PubMed: 19194785]
- Navratilova I, Hopkins AL. Fragment Screening by Surface Plasmon Resonance. *ACS Medicinal Chemistry Letters*. 2010; 1:44–48. [PubMed: 24900174]
- Neumann T, Junker HD, Schmidt K, Sekul R. SPR-based fragment screening: advantages and applications. *Curr Top Med Chem*. 2007; 7:1630–1642. [PubMed: 17979772]
- Ni S, McAteer K, Bussiere DE, Kennedy MA. Crystallization and preliminary crystallographic analysis of an *Enterococcus faecalis* repressor protein, CylR2, involved in regulating cytolysin production through quorum-sensing. *Acta Crystallogr D Biol Crystallogr*. 2004a; 60:1145–1148. [PubMed: 15159583]
- Ni S, Robinson H, Marsing GC, Bussiere DE, Kennedy MA. Structure of 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase from *Shewanella oneidensis* at 1.6 Å: identification of farnesyl pyrophosphate trapped in a hydrophobic cavity. *Acta Crystallogr D Biol Crystallogr*. 2004b; 60:1949–1957. [PubMed: 15502301]
- Nienaber VL, Richardson PL, Klighofer V, Bouska JJ, Giranda VL, Greer J. Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat Biotechnol*. 2000; 18:1105–1108. [PubMed: 11017052]
- Painter J, Merritt EA. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr*. 2006; 62:439–450. [PubMed: 16552146]
- Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov*. 2007; 6:29–40. [PubMed: 17159923]
- Payne DJ, Gwynn MN, Holmes DJ, Rosenberg M. Genomic approaches to antibacterial discovery. *Methods Mol Biol*. 2004; 266:231–259. [PubMed: 15148422]
- Pellecchia M. Solution nuclear magnetic resonance spectroscopy techniques for probing intermolecular interactions. *Chem Biol*. 2005; 12:961–971. [PubMed: 16183020]
- Petsko GA. An idea whose time has gone. *Genome Biol*. 2007; 8:107. [PubMed: 17608958]
- Pflugrath JW. The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr*. 1999; 55:1718–1725. [PubMed: 10531521]
- Rakus JF, Kalyanaraman C, Fedorov AA, Fedorov EV, Mills-Groninger FP, Toro R, Bonanno J, Bain K, Sauder JM, Burley SK, Almo SC, Jacobson MP, Gerlt JA. Computation-facilitated assignment of the function in the enolase superfamily: a regiochemically distinct galactarate dehydratase from *Oceanobacillus iheyensis*. *Biochemistry*. 2009; 48:11546–11558. [PubMed: 19883118]

- Rees DC, Congreve M, Murray CW, Carr R. Fragment-based lead discovery. *Nat Rev Drug Discov*. 2004; 3:660–672. [PubMed: 15286733]
- Reitz AB, Smith GR, Tounge BA, Reynolds CH. Hit triage using efficiency indices after screening of compound libraries in drug discovery. *Curr Top Med Chem*. 2009; 9:1718–1724. [PubMed: 19929837]
- Rohdich F, Eisenreich W, Wungsintaweekul J, Hecht S, Schuhr CA, Bacher A. Biosynthesis of terpenoids. 2C-Methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) from *Plasmodium falciparum*. *Eur J Biochem*. 2001; 268:3190–3197. [PubMed: 11389720]
- Sanders WJ, Nienaber VL, Lerner CG, McCall JO, Merrick SM, Swanson SJ, Harlan JE, Stoll VS, Stamper GF, Betz SF, Condroski KR, Meadows RP, Severin JM, Walter KA, Magdalinos P, Jakob CG, Wagner R, Beutel BA. Discovery of potent inhibitors of dihydroneopterin aldolase using CrystaLEAD high-throughput X-ray crystallographic screening and structure-directed lead optimization. *J Med Chem*. 2004; 47:1709–1718. [PubMed: 15027862]
- Service RF. Structural biology. Protein structure initiative: phase 3 or phase out. *Science*. 2008; 319:1610–1613. [PubMed: 18356501]
- Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. Discovering high-affinity ligands for proteins: SAR by NMR. *Science*. 1996; 274:1531–1534. [PubMed: 8929414]
- Siegel MG, Vieth M. Drugs in other drugs: a new look at drugs as fragments. *Drug Discov Today*. 2007; 12:71–79. [PubMed: 17198975]
- Smith E, Morowitz HJ. Universality in intermediary metabolism. *Proc Natl Acad Sci U S A*. 2004; 101:13168–13173. [PubMed: 15340153]
- Song L, Kalyanaraman C, Fedorov AA, Fedorov EV, Glasner ME, Brown S, Imker HJ, Babbitt PC, Almo SC, Jacobson MP, Gerlt JA. Prediction and assignment of function for a divergent N-succinyl amino acid racemase. *Nat Chem Biol*. 2007; 3:486–491. [PubMed: 17603539]
- Steinbacher S, Kaiser J, Wungsintaweekul J, Hecht S, Eisenreich W, Gerhardt S, Bacher A, Rohdich F. Structure of 2C-methyl-d-erythritol-2,4-cyclodiphosphate synthase involved in mevalonate-independent biosynthesis of isoprenoids. *J Mol Biol*. 2002; 316:79–88. [PubMed: 11829504]
- Swayze EE, Jefferson EA, Sannes-Lowery KA, Blyn LB, Risen LM, Arakawa S, Osgood SA, Hofstadler SA, Griffey RH. SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. *J Med Chem*. 2002; 45:3816–3819. [PubMed: 12190303]
- Ulaganathan V, Agacan MF, Buetow L, Tulloch LB, Hunter WN. Structure of *Staphylococcus aureus* 1,4-dihydroxy-2-naphthoyl-CoA synthase (MenB) in complex with acetoacetyl-CoA. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2007; 63:908–913.
- Van Voorhis WC, Hol WG, Myler PJ, Stewart LJ. The role of medical structural genomics in discovering new drugs for infectious diseases. *PLoS Comput Biol*. 2009; 5:e1000530. [PubMed: 19855826]
- Waybright TJ, Britt JR, McCloud TG. Overcoming Problems of Compound Storage in DMSO: Solvent and Process Alternatives. *Journal of Biomolecular Screening*. 2009; 14:708–715. [PubMed: 19531665]
- Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*. 2007; 450:1001–1009. [PubMed: 18075579]
- Wiesner J, Borrmann S, Jomaa H. Fosmidomycin for the treatment of malaria. *Parasitol Res*. 2003; 90(Suppl 2):S71–S76. [PubMed: 12937969]
- Wiesner J, Henschker D, Hutchinson DB, Beck E, Jomaa H. In vitro and in vivo synergy of fosmidomycin, a novel antimalarial drug, with clindamycin. *Antimicrob Agents Chemother*. 2002; 46:2889–2894. [PubMed: 12183243]
- Yuvaniyama J, Chitnumsub P, Kamchonwongpaisan S, Vanichthanankul J, Sirawaraporn W, Taylor P, Walkinshaw MD, Yuthavong Y. Insights into antifolate resistance from malarial DHFR-TS structures. *Nat Struct Biol*. 2003; 10:357–365. [PubMed: 12704428]
- Zartler, E.; Shapiro, M. *Fragment-based drug discovery : a practical approach*. Hoboken, N.J: John Wiley & Sons; 2008.

Figure 1.

Small molecules from ligand-bound structures super-imposed onto the *apo* structure of IspF from *Burkholderia pseudomallei*, with key side chains (lines) and the catalytic zinc ion (sphere) depicted. Fragment screening by NMR spectroscopy and X-ray crystallography led to three distinct subsets of ligands which bind in the cytosine sub-pocket (above left), coordinate to catalytic zinc (above right), and bind a hydrophobic cleft external to the active site (bottom left). Two lead compounds were designed by linking cytidine to zinc site-binding fragments (bottom right). While their cytosine moieties bind in the same manner, only one lead compound recapitulates the binding mode of its fragment precursor by coordinating to zinc through a heteroaromatic nitrogen. Figure generated using PyMol (DeLano, 2008) using PDB IDs 3F0E, 3IKE, 3IEQ, 3F0F, 3IEW, 3IKF, 3JVH, 3K14, 3KE1, ##### and ##### (Begley, manuscript in preparation).

Figure 2. GDP-bound structure of a protein monomer from *Mycobacterium tuberculosis* annotated as an arginine/ornithine transporter with ATPase activity. Diffraction-quality crystals were not obtainable in the presence of ADP or ATP, while hydrolysis of GTP to GDP was observed crystallographically. The N1/N3 interaction between GDP and the Asp207 side chain confirm a highly specific guanine binding pocket. Figure generated using PyMol (DeLano, 2008) from PDB ID 3MD0 (Edwards, manuscript in preparation).