Supporting your pipeline with structural knowledge

Derek A Debe & Kevin Hambly
Eidogen, CA, USA

Structure-based drug design approaches have repeatedly demonstrated their value by making profound contributions at the difficult and costly lead optimization stage of drug discovery. The recent increase in the availability of structural information is enabling a variety of new structure-based techniques to support informed decision-making at all stages of the discovery pipeline.

Structure-based drug design (SBDD) is a powerful approach that is widely employed in our industry despite the significant resources required to obtain protein structure information via experimental methods. In the 10 years since the first structure-based drugs were developed to target HIV protease and influenza neuraminidase, inhibitors for more than 40 distinct targets have been discovered using SBDD approaches.

The latest data deluge

The structuring of the human proteome is a logical continuation of the sequencing of the human genome, and significant public and private structural genomics efforts are beginning to transform structural biology into biology's newest high-throughput, data-rich field. Early efforts have been instrumental in establishing high-throughput structural genomics platforms that employ automated protein expression, crystallization, data acquisition, and model refinement technologies. Continued research will ensure that 2004 will be the first year that more than 5000 new protein structures are deposited into the publicly available Protein Data Bank. Currently, the PDB is on course to contain approximately 30,000 structures by the end of 2004, more than 10 times the number just one decade ago.

Deriving value from structure

The data deluge produced by structural genomics projects is driving the need for structural informatics, just as the torrent of data created by combinatorial chemistry and high-throughput screening methods created the need for cheminformatics (Figure 1). This need is not addressed by existing molecular modeling packages, and there are a variety of technical barriers that must be overcome before a structural informatics platform can be used to attack real drug discovery problems. First and foremost, even though the number of experimentally determined structures is increasing at a tremendous rate, crystallographic coverage at the genome and gene-family level will remain far from complete over the next few decades. Thus, structural informatics platforms must be capable of supplementing existing experimental data with high quality homology-modeled structures. A second critical technical barrier arises from the fact that existing global protein structure classification systems such as FSSP, SCOP, and CATH do not adequately address the drug discovery relevant problem of understanding target similarity at the drug binding site level.

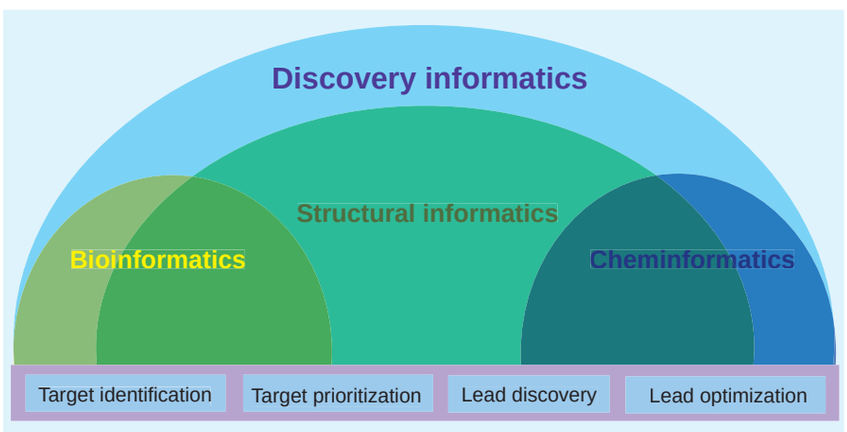

Figure 1. The recent increase in the availability of structural data is enabling structural informatics to join bioinformatics and cheminformatics as a data-rich discipline. Over the next few years, discovery informatics capabilities seamlessly integrating structural and cheminformatics approaches will simultaneously support high-throughput structure and lead-based drug discovery.

Target prioritization applications
Target drugability assessment
- Binding site property analysis
- Binding site selectivity analysis
- *Drug resistance mutation analysis*
Animal model suitability analysis
Broad spectrum anti-infective target analysis
Lead discovery applications
Binding site property analyses & HTS library selection
Novel lead fragment & scaffold discovery
Structure-based library design
Lead optimization applications
Structure-based ligand affinity optimization
Structure-based co-inhibitor affinity optimization
Structure-based selectivity optimization
Structure-based broad spectrum anti-infective optimization
'Off-target' discovery and analysis
'Off-target' assay panel design
Novel opportunity discovery
Novel drug rescue or redesign opportunities
- Due to drug-induced mutations
- Due to poor selectivity within a family
- Due to off-target interactions
Novel drug binding site opportunities
Novel co-inhibition opportunities
Novel anti-infective target discovery

Table 1. Structural informatics approaches can be used to solve a variety of problems at all stages of the drug discovery pipeline. The applications discussed in this article are shown in italics.

Hence, binding-site level classification approaches must be developed before structural informatics provides reliable information that can significantly impact the drug discovery process.

Recently, scientists at Eidogen developed a structural informatics technology platform that overcomes both of these technology barriers. Our platform fully integrates the algorithm and database components required for structure determination, binding site-determination, and binding site comparison and classification. Pairing these technologies with novel, comparative visualization approaches results in a structural informatics platform that allows users to understand binding site similarity as effortlessly as they are accustomed to understanding gene sequence similarities via bioinformatics approaches, and chemical structure similarities via cheminformatics approaches (Table 1).

Target prioritization

As drug discovery researchers work to translate vast amounts of uncharted genomic information into new drugs, strategies must be implemented to ensure that failed projects are terminated as early as possible. Targets must be correctly prioritized according to their probability of being potently and selectively drugged to reduce the high downstream attrition rates associated with drug development.

Structural informatics supports high-throughput target prioritization by enabling the systematic evaluation of a variety of factors that can strongly influence target drugability. For example, structure-based approaches can be extremely useful for understanding how a target's drugability is influenced by its pharmacogenetic and mutation profile.

One recently-developed drug whose efficacy is strongly influenced by its target's mutation profile is Gleevec from Novartis. This drug halts the progression of chronic myelogenous leukemia (CML) by binding to the ATP binding pocket of the Bcr-Abl fusion protein. Recently, a correlation was established between the efficacy of Gleevec and non-synonymous mutations within the Abl kinase domain. Furthermore, mapping these

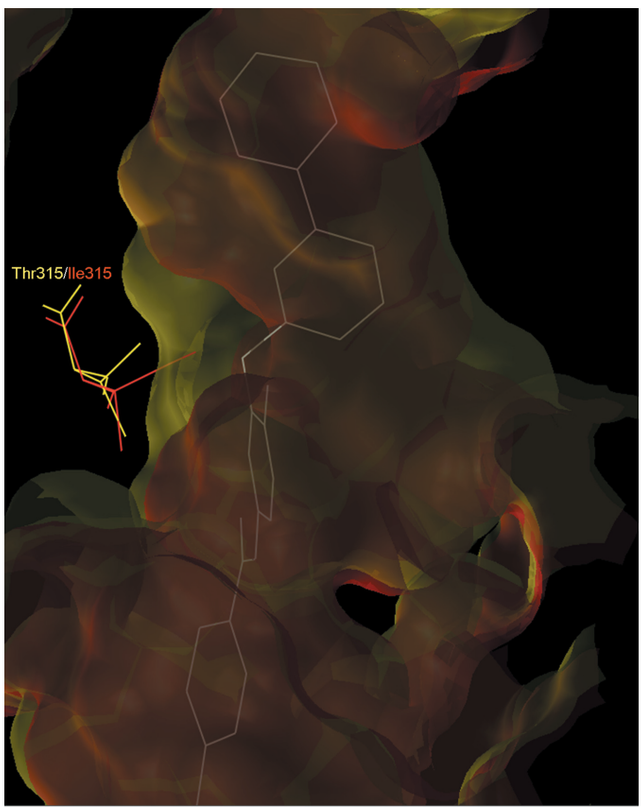

Figure 2. The binding sites from a mutant model of Abl kinase (red) and the Abl crystal structure binding site complexed with Gleevec (yellow). Substitution of isoleucine for threonine at position 315 significantly changes the geometry of the narrow Gleevec binding pocket. The loss of the threonine hydroxyl group in the mutant removes an important hydrogen bond and the bulk of the isoleucine sidechain causes a sterically unfavorable situation.

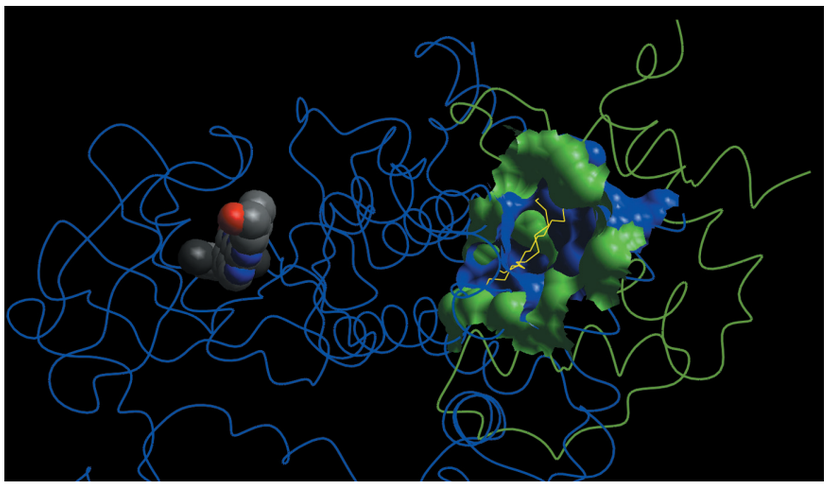

Figure 3. An overlay of the Abl kinase myristate site (PDB 1opl, blue), and the acyl-CoA binding protein (ACBP) myristoyl site (PDB 1hbk, green). The myristoyl ligands are shown in yellow on the right, and the ATP competitive inhibitor of Abl kinase is shown on the left. VTS was used to screen the Abl myristoyl binding site against all known co-crystal sites in the Protein Data Bank.

mutations to the Abl kinase structure reveals that Gleevec's efficacy is most significantly reduced when the mutations occur in the binding pocket (Figure 2). Clearly, in this instance, structural pharmacogenomics information is necessary to accurately post-rationalize the differing patient population responses to Gleevec. With the amount of reliable SNP and mutation data increasing rapidly, structural pharmacogenomics approaches, supported by structural informatics, can be applied preemptively to uncover discovery and development hurdles at the target prioritization stage.

Lead discovery

Gene family-based drug discovery is widely employed in the pharmaceutical industry, since many of the lessons learned during the discovery and optimization of inhibitors for one target in a family can be applied to other targets within the same family. With the recent explosion in the number of proteins that have been co-crystallized with small molecule compounds, it becomes increasingly important to develop approaches that enable researchers to apply information from this growing knowledgebase to their own targets.

"Structure-based approaches to lead optimization have proven consistently effective at improving the potency, selectivity profile, or pharmacokinetic properties of drug candidates."

Virtual Target Screening (VTS) is a new in silico approach for applying information from the co-crystal database to lead discovery on new targets where ligands are not yet known. Unlike virtual ligand screening (VLS), which entails screening a target structure against a database of small molecules, VTS screens a target's binding site against a database of small molecule co-crystal binding sites. The similarities between small molecule binding sites detected by VTS can provide drug discovery researchers with critical information about the classes of molecular scaffolds that are likely to bind to their target of interest. This information can be used to guide the selection and design of focused compound libraries to improve screening success rates for specific targets or target families.

Eidogen researchers have developed a novel VTS algorithm capable of detecting the similarities between two binding sites even if these sites are found within targets that do not share any sequence or structure homology. Figure 3 shows an example of how VTS can be used to discover novel lead scaffolds for an emerging target.

Lead optimization

Structure-based approaches to lead optimization have proven consistently effective at improving the potency, selectivity profile, or pharmacokinetic properties of drug candidates. Comparative analysis of all the structures within a gene family is a particularly valuable approach to improve drug selectivity. Structural informatics capabilities are now enabling large-scale comparative structural analyses, dramatically simplifying the process of lead optimization for gene family-based drug discovery.

One area where optimization of compound selectivity is of paramount importance is in the development of small molecule kinase antagonists. Nearly all kinase inhibitors act by competing for the ATP binding site, and because of the highly conserved nature of this binding site it is often quite difficult to develop selective inhibitors. An example of a potent, but non-selective kinase inhibitor is staurosporine, shown in Figure 4 bound to both glycogen synthase kinase-3 β (GSK-3 β) and cyclin-dependent kinase 2 (CDK2), two well-studied kinase drug targets involved in diverse human pathologies (CDK2 in cancer, GSK-3β in diabetes and Alzheimer's disease). Despite years of research, there has been little success in the development of inhibitors that are selective for only one of these two enzymes, with nearly all inhibitors designed for GSK-3β also inhibiting the various CDK enzymes, and vice versa. Indeed many of the adverse effects of GSK-3β and CDK inhibitors have been attributed to this 'off-target' inhibition. With the appropriate comparative visualization and analysis tools, strategies for achieving selectivity can be developed for these highly related enzymes.

Novel opportunity

In addition to making contributions at each stage of the discovery pipeline, structural informatics can be used to make new discoveries that historically have only been made serendipitously or via costly and time-consuming experimental efforts.

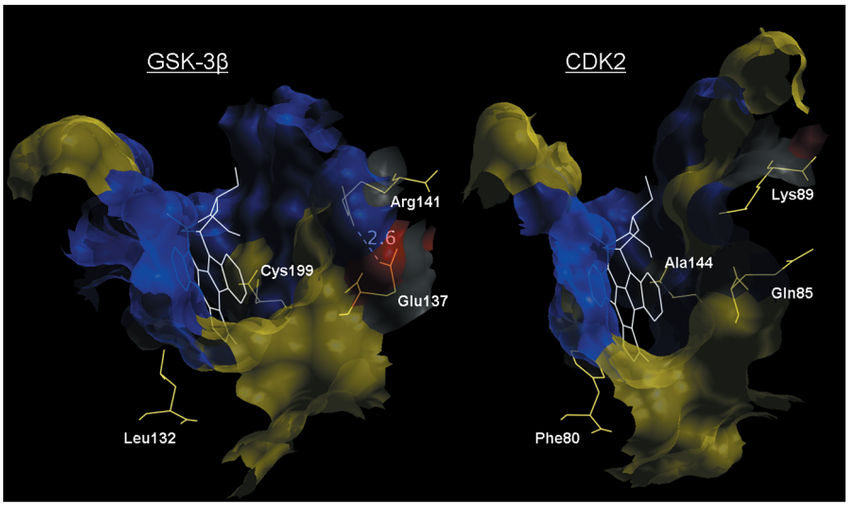

Figure 4. View of the ATP binding sites of GSK-3β (left) and CDK2 (right) in complex with staurosporine. The binding site is colored according to sequence conservation between the two enzymes (blue = conserved residues, yellow = non-conserved residues). GSK-3βpossesses a more elongated pocket along the hinge region, whereas CDK2 has a more compact center and a wider opening at the mouth of the pocket. Residue differences that may contribute to the selectivity of these two enzymes include GSK-3β Leu132/CDK2 Phe80 at the base of the pocket, GSK-3β Cys199/CDK2 Ala144 along the lining of the pocket, and at the entrance of the pocket the Glu137-Arg141 salt bridge in GSK-3β versus the corresponding residues Gln85-Lys89 in CDK2, which are not properly oriented to form a salt bridge at this position.

Anti-infective target discovery is an area that has traditionally been dominated by serendipitous rather than rational discovery. Many of the most successful antibacterial and antifungal drugs have been brought to market long before their molecular targets were even known, much less structurally characterized. Figure 5 illustrates how comparative structural informatics can be used to drive the discovery of target selectivity of the commercially successful azole class of antifungal drugs.

Future advances

The use of protein structure will continue to broaden throughout the drug discovery pipeline as the number of solved structures increases. Structural informatics approaches will support rational drug design just as cheminformatics currently supports small molecule drug discovery and in time, these fields will merge into a unified framework that simultaneously supports high-throughput in silico structure and lead-based drug discovery.

Derek A Debe
Co-founder, President & CEO
Eidogen
055 East Colorado Blvd
Suite 550
Pasadena
CA 91106
USA

Email: derek@eidogen.com
www.structuresoflife.com

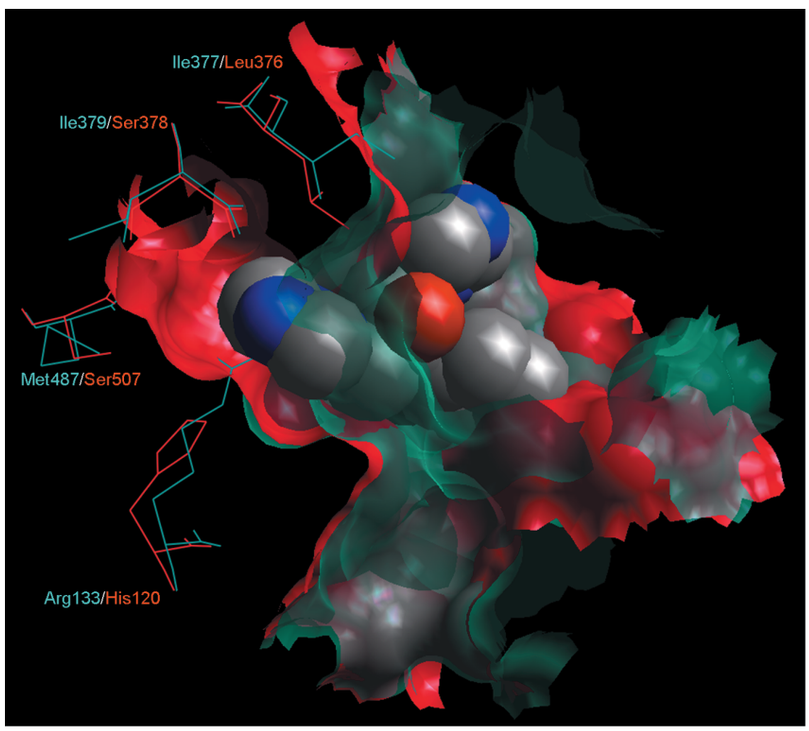

Figure 5. The discovery that azoles inhibit the enzyme CYP51 was not established until decades after the drugs were first used to treat fungal infection. Although CYP51 is highly conserved in fungi and humans, the azoles were serendipitously found to inhibit only the fungal enzyme. An overlay of the azole-binding sites in homology models of the CYP51 enzyme from the pathogenic fungi Candida albicans (red), and human (cyan). A fluconazole molecule is shown in its predicted binding mode. A number of important residue substitutions in the binding pockets of these orthologous enzymes lead to substantial differences in the geometry of the binding pocket. In particular, the bulky Met487 and Arg133 residues in the human enzyme appear to close off a portion of the binding pocket that is accessible in the fungal enzyme, which have serine and histidine residues in the corresponding positions. The shallower binding pocket cannot accommodate the binding of fluconazole or other multiply substituted azole compounds. This observation rationalizes the selectivity of the azoles, and illustrates the ability of structural informatics approaches to discover novel anti-infective targets by binding site selectivity analysis.