1. SiteSorter

SiteSorter is Eidogen-Sertanty's proprietary and patent pending binding site similarity determination algorithm. SiteSorter is capable of confidently detecting physicochemical similarities between drug binding sites in the absence of sequence or structure homology. In the lead discovery stage, SiteSorter is an invaluable tool for quickly identifying promising "first guess" lead scaffolds and compound classes. During lead optimization, SiteSorter allows the identification of potential secondary or cross-reactive targets. In conjunction with TIP's comprehensive structure database and SiteSeeker algorithm, SiteSorter gives Eidogen-Sertanty's customers an unprecedented proteome-wide view of the pharmaceutically relevant similarities among target binding sites.

1.1. Introduction & Significance

While numerous methods have been successfully developed for detecting sequence and structural similarities among proteins, a method for the detection of similarities among small molecule binding sites has proven much more challenging to develop, and is currently not part of the computational drug discovery toolset. Developing such a method is extremely important, since an approach that can successfully characterize the structural and physiochemical similarities between binding sites will enable a variety of new and unexplored applications at the early and late discovery stages:

Target Discovery & Validation
- Assigning ligand-binding functions to novel targets and orphan receptors

Lead Discovery
- Identifying novel lead scaffolds for a novel or established target
- Finding new targets that will bind to known scaffolds

Lead Optimization
- Resolving selectivity concerns within a target family
- Determining toxicity effects from off-target binding interactions

1.2. Limitations of Existing Approaches

The current state-of-the-art approach for detecting binding site similarities was recently developed by Stefan Schmitt from AstraZeneca in collaboration with Gerhard Klebe's group at the Institute of Pharmaceutical Chemistry in Marburg, Germany.[1] Schmitt and co-workers were the first to publish a general method for mapping functional group information onto binding site surfaces to estimate their physicochemical similarities. They demonstrated that these functional mappings dramatically reduce the number of false similarity relationships when compared to approaches which only measure shape complementarity. Their work is the first to demonstrate that meaningful binding site relationships between two proteins that do not share any sequence or structure homology can be ranked in the top 5% in a large group of false positives.

Schmitt and co-workers achieved their success over previous methods by introducing a scoring function that takes into account the important physicochemical characteristics of the binding site surfaces. Eidogen-Sertanty's SiteSorter method significantly extends this approach by supplementing Schmitt's physicochemical scoring function with conservation and shape complementarity in formation, resulting in a more complete scoring function that further reduces the incidence of false positives.

While the results obtained by Schmitt *et al.* are very exciting, a major drawback of their algorithm is its computational speed. The analytical graph-theory comparison approach they use scales very poorly with the size of the surfaces being compared (the number of points and edges in the graph representation of the surfaces). At Eidogen-Sertanty, we've overcome this problem by developing a novel hierarchical procedure that reduces the number of surface points and edges that must be considered during the initial stages of the binding site overlay optimization. This procedure leads to a significant increase in computational efficiency without suffering a loss in overlay precision or scoring accuracy.

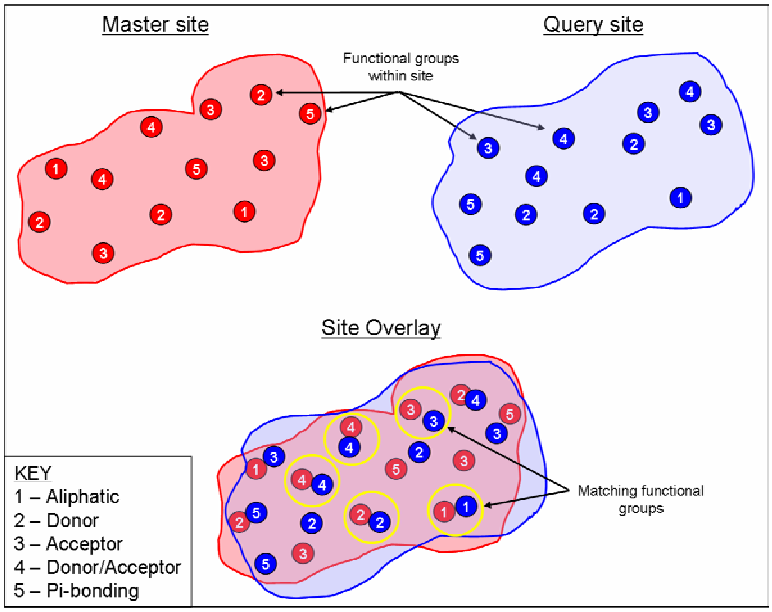

Figure 1. Diagram of SiteSorter's general approach to overlaying and scoring similar sites. Important functional groups at the surface of each site are optimally overlaid such that the maximum number of matching functional groups is obtained. SiteSorter then assigns a similarity score which is an indicator of the number of identical functional groups found in the two overlaid sites, within certain directional and distance criteria.

1.3. Examples and Implications for Important Drug Targets

In this section, we demonstrate SiteSorter's capabilities by showing multiple examples where SiteSorter detects important binding site similarities in the complete absence of sequence or structure similarity. These examples were chosen because they are pharmaceutically interesting cases that demonstrate SiteSorter's capabilities across a wide variety of binding pocket shapes, sizes and functions.

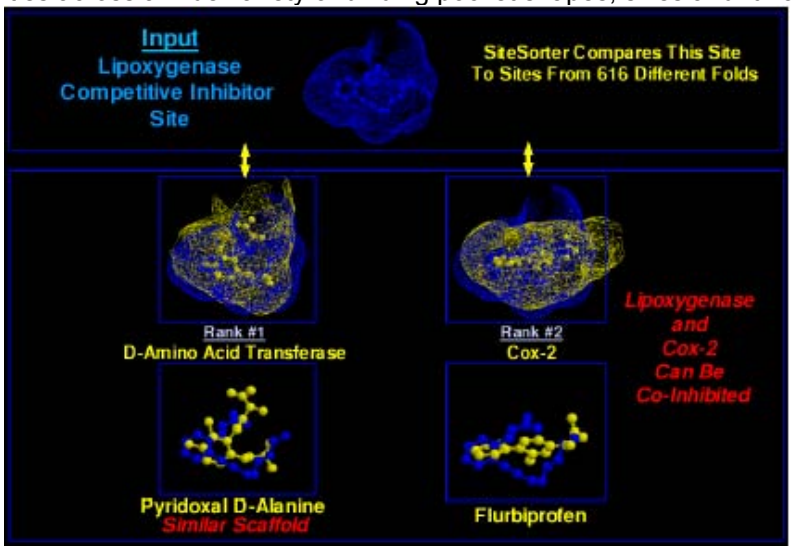

Figure 2. SiteSorter was used to query the arachidonic acid binding site of lipoxygenase with a set of PDB co-crystals from 616 different folds. The top two ranking co-crystal sites (yellow) are shown overlaid with the lipoxygenase site (blue) in the middle row and with their respective ligand overlays below them. Rank 1 is a pyridoxyl d-alanine binding site in D-amino acid transferase, and the overlay of the two sites shows that the ligands possess similar scaffolds. Rank 2 is the COX-2 NSAID-binding site, shown here complexed with flurbiprofen. Both lipoxygenase and COX-2 are in the arachidonic acid pathway, and it has been shown that they can be co-inhibited by a single ligand. Neither of these two hits possess any overall sequence or structural similarity to lipoxygenase, but SiteSorter successfully identifies the similarity between their small molecule binding sites.

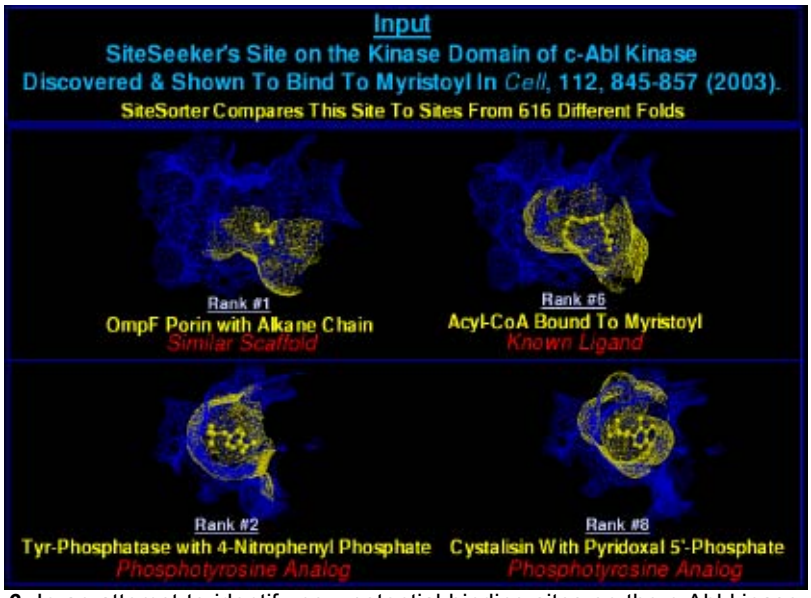

Figure 3: In an attempt to identify new potential binding sites on the c-Abl kinase, SiteSeeker was run on a c-Abl-Gleevec complex (PDB 1IEP). A high ranking site was found in the C-terminal lobe of the kinase domain, spatially distinct from the Gleevec-binding site. When this new site was used as the query for SiteSorter, it was found to overlay well with sites that bind to myristate and other alkane chains (2 hits shown in the middle row). Interestingly, there were also several high ranking hits that bound to phosphotyrosine analogs (2 hits shown in the bottom row). A recently determined co-crystal of an Abl-myristate complex (PDB 1opl) verifies that the SiteSeeker predicted site on Abl is indeed a myristate-binding site. The ability of this site to bind phosphotyrosine or phosphotyrosine analogs has not been experimentally verified, although it is an intriguing possibility that warrants further investigation.

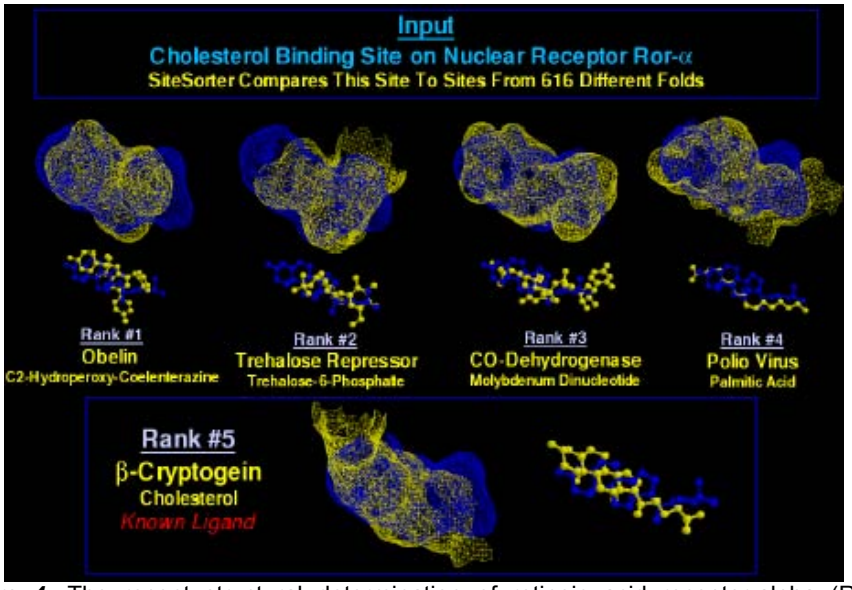

Figure 4: The recent structural determination of retinoic acid receptor-alpha (ROR-alpha) in complex with cholesterol (PDB 1n83) led to the initial identification of cholesterol as the physiological ligand for this orphan nuclear receptor. In order to test whether or not SiteSorter could have aided in the identification of cholesterol as the ligand for ROR-alpha, the cholesterol-binding site from the ROR-alpha structure was used as a SiteSorter query against the same set of co-crystal structures used in the previous examples. As shown in the figure, the sites ranked 1 through 4 are co-crystal sites which bind to ligands bearing some similarity to cholesterol (shown in blue), and the fifth ranking site is a cholesterol binding site from the sterol-transport protein beta-cryptogein. As in the previous example, all of the sites shown are found in structures having no

overall similarity to ROR-alpha. The observation that SiteSorter correctly identifies the physiological ligand for this orphan nuclear receptor within the top 5 hits highlights the algorithm's utility for determining the natural ligands of orphan nuclear receptors.

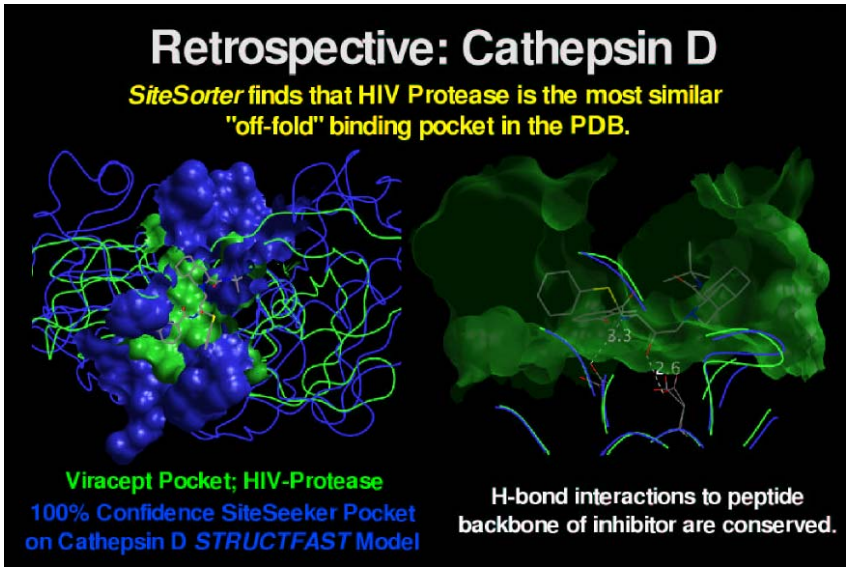

***Figure 5*:** Although Cathepsin D, a cancer target, and HIV protease are both members of the aspartic protease family of enzymes, there is no sequence or structural similarity between them. Despite this lack of sequence or structural similarity, SiteSorter finds that the HIVprotease and Cathepsin D substrate sites are highly similar. This is a retrospective case, since it is known that Cathepsin D is susceptible to inhibition by HIV-protease inhibitors.[2] Researchers have exploited their knowledge of this binding site similarity to design novel nonpeptidic inhibitors of Cathepsin D,[3] as well as HIV protease inhibitors with decreased affinity for human Cathepsin D.

---

[1] S. Schmitt *et al*., *J.Mol.Biol*. 323:387 (2002).
[2] L. Hoegle *et al*., *Pharmazie*. 1999 May;54(5):319-29.
[3] Kick *et al*. *Chem Biol*. 1997 Apr:4(4):297-307.