

unfortunate given the magnitude of their biological significance. Although structural proteomics projects are currently underway to help solve the structures of unknown membrane proteins, the computational prediction of TMBBs has improved rapidly to compensate for the limits of empirical methods. To this end, our lab has developed one of the most accurate TMBB prediction programs available, which was used to build a database of TMBB predictions called TMBB-DB. This is a comprehensive database featuring TMBB prediction data from the proteomes of over 500 species of bacteria with nearly 1.9 million total sequences. Combining the TMBB prediction data with signal peptide prediction data generated using SignalP we predicted that more than 3% of the sequences encoded TMBBs, which on average is more than double the number of known or predicted TMBBs already annotated in the proteomes. Users will have access to an in-depth analysis of each sequence and be allowed to analyze sequences not included in the database using the same prediction tools. This database can be useful in directing the efforts of structural proteomics projects, antimicrobial drug therapy design, or vaccine development.

#### 1734-Pos Board B644

##### Validichro, a Software Tool for Validation of Circular Dichroism (CD) Spectroscopic Data

Benjamin Woollett, Lee Whitmore, Andrew J. Miles, Robert W. Janes, B.A. Wallace.

Circular Dichroism (CD) spectroscopy is a widely used technique in structural biology for examining protein structures, interactions and folding, and is used in industry for quality control of bioprocesses and pharmaceutical products. There currently exist a number of validation/checking programs for crystallographic and NMR data, such as PROCHECK, MolProbity and WHAT\_IF, but no such validation programs exist for CD data. We have recently created VALIDICHRO, a software suite for assaying the quality/validity of circular dichroism data. With the recent launch of the Protein Circular Dichroism Data Bank (PCDDDB), a public repository for CD spectral and associated metadata, the need for such validation software became paramount to ensure the integrity of the data bank and its utility for a wide range of bioinformatics and structural biology usages. In addition this software provides a means of establishing standards for "good practice" and quality control throughout the CD data-collecting community. The validation criteria were established through wide and open consultations with members of the CD user community and the PCDDDB Technical Advisory Board. VALIDICHRO does more than 20 types of checks for completeness, consistency and quality of CD spectra and their associated metadata and produces a user-friendly report. This can be employed as a guideline for users (either depositors or accessors), attached to a PCDDDB entry, or provided to journal reviewers as evidence of data quality. (Supported by grants from the U.K. BBSRC)

#### 1735-Pos Board B645

##### Docking Benchmark Set of Protein Models

Petras J. Kundrotas, Ivan Anishchenko, Alexander V. Tuzikov, Ilya A. Vakser.

In structure-based genome-wide studies of protein-protein interactions, most protein complexes have to be modeled by high-throughput computational approaches. The existing X-ray structures of protein complexes provide templates for ~ 20% of all known interactions. Thus, docking of the remaining complexes has to be based on independently modeled structures of the monomers. Application of the docking methodologies, both template-based and template-free, to inherently inaccurate models has its limitations. Thus, docking benchmarking on a large systematic set of protein models at different levels of structural accuracy of the monomers is important. Currently available benchmark sets of protein-protein complexes (DOCKGROUND unbound set, Boston University Benchmark set, etc.) are limited to the X-ray structures in the bound and unbound forms and thus are unsuitable for such studies. We present a set of models built for 99 binary complexes from the DOCKGROUND unbound set. For each monomer in the dataset, six models were generated with Ca RMSD between the native and the modeled structures in 1 - 6 Å range. The models were built by a combination of single-template homology modeling and Nudget Elastic Band (NEB) methodology. The dataset will be incorporated into the DOCKGROUND public resource ([dockground.bioinformatics.ku.edu](http://dockground.bioinformatics.ku.edu)).

#### 1736-Pos Board B646

##### Identification of hPXR Modulators via SAR-Based Virtual Screening

Yongmei Pan, Linhao Li, Gregory Kim, Sean Ekins, Hongbing Wang, Peter Swaan.

**Purpose.** Human pregnane X receptor (hPXR) has a key role in regulating metabolism of endogenous and exogenous substances. Identification of novel hPXR activators among commercial drugs aids in avoiding drug-drug interactions with future co-administered drugs.

**Methods.** Virtual screening with Structure-Activity Relationship (SAR) models has been applied for identification of novel hPXR activators. Ligand-based modeling was conducted with Discover Studio (DS) 2.1. Bayesian classification models were generated with a training set comprising 177 compound, and were validated with a test set with 145 compounds. The activities of commonly prescribed drugs from SCUT database were predicted with one of the Bayesian models. Cell-based luciferase reporter assay was used for evaluation of chemical-mediated hPXR activation. HepG2 cells were co-transfected with PXR expression vector and CYP2B6-luciferase reporter construct. 0.1% DMSO solution was used as vehicle control while rifampicin as the positive control. The binding mode between an experimental validated hPXR modulators and hPXR were studied by docking a ligand into hPXR binding domain (PDB ID: 1NRL) with programs FlexX and SurfFlex.

**Results.** The Bayesian models showed specificity and overall prediction accuracy up to 0.92 and 0.69 for test set compounds. One Bayesian model with specificity of 0.92 was selected to screen the SCUT database and retrieved 113 hits. 17 compounds were chosen for in vitro testing. The luciferase reporter assay confirmed that seven drugs, i.e., fluticasone, nimodipine, nisoldipine, beclomethasone, finasteride, flunisolide, and megestrol were previously unidentified potent or moderate hPXR activators, with 2.7 to 18.5-fold increase in luciferase activity compared to vehicle control.

**Conclusion.** In this study, virtual screening based on SAR models successfully identified seven novel hPXR activators among FDA approved and commonly prescribed drugs. The same approach could be used for identification of activators or inhibitors of other protein targets.

#### 1737-Pos Board B647

##### Using Entropy Leads to a Better Understanding of Biological Systems

Chih-Yuan Tseng, Jack A. Tuszynski.

In studying biological systems, conventional approaches based on laws of physics almost always require introducing appropriate approximations. We argue that a comprehensive approach that integrates laws of physics and principles of inference provides a better conceptual framework than these approaches. The crux of this comprehensive approach hinges on entropy. Entropy is not merely a physical quantity. It also is a reasoning tool to process information with the least bias. Here, we review three distinctive examples in biology and drug discovery to demonstrate the developments, applications and advantages of the approach based on this comprehensive route.

In the first example, we will show that the incorporation of laws of physics and principles of inference provides a more comprehensive method to investigate and reveal protein folding dynamics compared to conventional approaches. Second, a maximum entropy method is developed to predict tubulin isotype expression levels in a cell when the cell is exposed to various cytotoxic derivatives of the anti-cancer drug, colchicine. The last example discussed aims to provide a theoretical method based on the maximum entropy approach to design short RNA/DNA sequences, aptamers that specifically bind to bio-molecular targets of interest. These three examples will provide strong evidence that entropy plays a crucial and fundamental role in conceptual development rather than being merely involved in either a measurement of randomness or a tool for processing information.

**Acknowledgement.** This research was funded by the Alberta Cancer Foundation, the Allard Foundation, NSERC, the Canadian Breast Cancer Foundation and Alberta Advanced Education and Technology.

#### 1738-Pos Board B648

##### Comparison of the Charge Distribution of Protein Molecules via Multipole Parameterizations

Apostol Gramada, Thomas Evangelidis, Philip E. Bourne.

The electrostatic potential is the essential component for the long-range interaction of proteins with other biological molecules. Therefore, similar functions often derive from similar configurations of the electrostatic field which, typically, denotes similar charge distribution inside proteins. This has been clearly established for large families of proteins. However, tools that can provide a measure of similarity in terms of charge profile similarity are not readily available. We present a methodology that provides a quantitative measure of charge profile similarity between protein molecules in terms of a collective parameterization of this property in terms of multipole moments. We demonstrate the performance of this approach with a well characterized enzyme protein set, for which it has been shown that the charge distribution around functional sites has been an essential evolutionary factor that allows discrimination between functional families. Our methodology provides now quantitative evidence of these findings.