Final Report: 5P20HG003899-02 “Establishment of the Rensselaer Exploratory Center for Cheminformatics Research (RECCR)” – C. Breneman, PI.

Goal and Objectives

The Rensselaer Exploratory Center for Cheminformatics Research (RECCR) was (and remains) dedicated to advancing the field of Cheminformatics and increasing the availability of new methods within the Cheminformatics user community. Toward this goal RECCR members developed new multi-objective machine learning methods, high information-content descriptors, data fusion techniques and infrastructure for extending the reliability and applicability of informatics-based prediction techniques.

Advances in the generation, mining and analysis of chemical information being crucial to the development of new drug therapies, and to modern methods of bioinformatics and molecular medicine, RECCR brought together and stimulated collaborative pilot projects among experts in Cheminformatics-related fields ranging from methods of encoding and capturing molecular information, to machine learning and data mining techniques, to predictive model development, validation, interpretation and utilization. In addition, RECCR also brought together a set of domain specialists and application scientists to serve as both data generators and end-users of the knowledge provided by the molecular property models and modeling methods developed by RECCR, and too test the new cheminformatics software developed at RECCR. The many diverse project areas pursued at RECCR can be grouped into one or more overlapping categories:

- **Data Generation** – using theoretical or experimental methods for creating or extracting knowledge;
- **Machine Learning and Datamining** – model validation, feature selection, pattern recognition, generation of potentials of mean force and knowledge-based potentials;
- **Property-Prediction** – chemically-aware model building, molecular property descriptor generation, Quantitative Structure-Property Relationship modeling, validation, and interpretation;
- **Applications** – utilizing the information made available using the new tools and methods that are developed as part of RECCR.

RECCR emphasized the central role of Cheminformatics in modern biotechnology efforts, molecular design projects and bioinformatics programs, seeding new interdisciplinary projects and train graduate students in these areas, with the overall goal of continually advancing the field of Cheminformatics research and developing descriptors, machine learning methods and infrastructure for extending the reliability and applicability of informatics-based prediction techniques. ADME/Tox predictions, ligand/protein scoring, drug discovery, molecular fingerprint analysis, and bioinformatics methodologies benefited from advances in Cheminformatics at RECCR.

RECCR focused much of its efforts on developing and delivering novel, fast and state of the art descriptors, machine learning applications and model validation methodologies to the scientific community in an effort to address the most serious problems in prospective modeling, and to reduce the computational resources that might otherwise be required for first principles-based approaches for predicting chemical and biological activities. A strong software engineering and dissemination system exists today that extends this in-house technology to the greater community through web-based delivery systems as well as downloadable stand-alone software modules. The domains explored by RECCR members include a wide variety of behaviors and characteristics of small molecule, biopolymers and nanocomposite materials. Some of the major accomplishments in software, algorithms, dissemination and outreach are detailed below:

ACCOMPLISHMENTS

Software Engineering and Dissemination

•A variety of novel descriptor generation tools were created and made public:

- Graphical & data mining output implemented in DIXEL

- Graphical output implemented in PROTEIN RECON in addition to existing numerical output
- SIMIL peptide descriptor alignment incorporated into PROTEIN RECON (used in winning the CoEPrA QSAR regression competition)
- Ultra-high-throughput USR shape descriptors were incorporated into RECON for NCGC virtual screening
- QPEST – high-throughput shape/property hybrid descriptor method was developed
- PPEST - Protein Dissimilarity Analysis was created and tested

•Protein modeling suite development

- Development of CLEAN_PROT pdb pre-processing web tool
- Development of PROLICSS surface-based protein/ligand scoring function

•RECCR Online Modeling System for Machine Learning applications was launched with functionality for PLS, KPLS, SVM, Cross validation and Feature selection.

•Data Mining Template Library was rolled out on RECCR website – already downloaded by over 1800 researchers

Modeling methodology

•Calibration and benchmarking of existing technology and state of the art methods was performed and implemented on CoEPrA regression competition.

•Implemented **RANKING**, **Multi-Objective methods** and **Multi-task learning** – Kernel PLS and SVM QSPR models have shown that inference models can support discovery and understanding of bioseparations and protein/surface interactions (Breneman et al 2003). By developing extensions to these approaches targeted towards ranking and multi-task modeling, we can further accelerate the discovery process. Highly nonlinear ranking methods have been developed by simply changing the loss function used in SVM to a loss function appropriate for ranking. In the past, PLS and K-PLS could not be readily adapted to other loss functions. We have developed a novel dimensionality reduction method called Boosted Latent Factors (BLF) (Momma and Bennett 2005). For any given loss function, BLF creates latent variables or principal components similar to those produced by PLS and PCA. We have extended BLF to ranking loss-function with great success. BLF can use the kernel approach of SVM and K-PLS to construct highly nonlinear ranking functions. For the least squares loss, BLF reduces to PLS, but now we can rapidly create learning methods for any convex loss function that maintains the many benefits of PLS. Simultaneous modeling of a multi-task problem can improve insight into the causal model underlying the methods. PLS was developed for such multi-task and multi-response models but is limited to least squares regression loss functions. Multiple Latent Analysis (MLA) extends BLF to multi-task problems optimized using any convex loss function (Bennett 2006). With MLA, we can model the tasks as interrelated ranking problems in order to determine the experimental conditions likely to achieve a desired outcome. Recently, SVMs have also been extended to multi-task modeling (Evgeniou and Pontil 2004). We have developed and applied the multi-task learning methods for small-molecule chromatographic displacer property prediction as an exemplar of probe design problems. Multi-task modeling is applicable to many problems in cheminformatics (e.g. in drug discovery, we typically want to model and optimize several properties of small molecules related to efficacy, absorption, and toxicity).

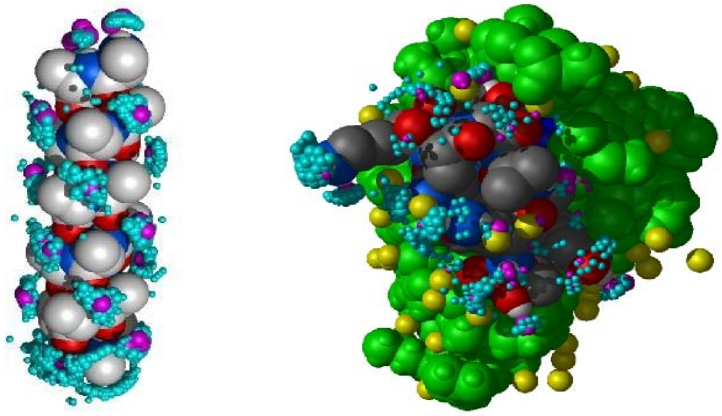
•Implemented **Data Fusion** – Integration of data from multiple sources. Data fusion was first introduced in the radar sensing community and refers to the process of combining multi-sensor data from different sources such that the resulting information/model is in some sense “better” than would be possible when these sources were used individually. We have extended the idea of data fusion to molecular property analysis and prediction, where rather than using different sensor sources, we use different descriptor fields for a set of molecules and apply data fusion techniques to improve the predictive performance of QSAR models for unknown cases. In this situation we use the term “auto-fusion” rather than data fusion, because the same molecules, and in certain cases the same descriptors are used, but different preprocessing techniques extract different features from the data - such as principal component analysis and independent component analysis (ICA). It has been shown that kernel partial-least squares (K-PLS) models in auto-fusion mode show a significant boost in performance compared to traditional K-PLS models. Note that this approach is distinct from the more familiar methods of consensus or bagged modeling, and performs better in prediction.

Descriptor technology

•**Protein Dissimilarity Analysis using Shape/Property Descriptors:** PPEST (Protein Property-Encoded Surface Translator) were developed for describing the shape and surface property distributions of proteins. This method uses a technique akin to ray-tracing to explore the volume enclosed by a protein. Probability distributions were derived from the ray-trace, based on the geometry of the reflecting ray and include joint dependence on properties such as the molecular lipophilicity potentials (MLP) and molecular electrostatic potential (MEP). These probability distributions, stored as histograms or wavelet coefficients, create a unique profile for each protein and are independent of molecular orientation. The profiles generated by PPEST can be rapidly compared to test for similarity between proteins. The triangulated protein surface subjected to internal ray-reflection can be derived from the GaussAccessible surface provided by MOE.

•**Development of Simulation-derived Hydration-based descriptors for biological applications**: Traditional methods describe water in two distinct ways: (i) water as a continuum dielectric, and (ii) water described explicitly through classical molecular or quantum mechanical simulation. Significant recent work indicates that a continuum dielectric description of water is too coarse-grained to be useful in predicting the characteristics of the underlying solute surface. On the other hand, explicit inclusion of water in molecular or quantum mechanical simulations is computationally prohibitive given the thousands of water molecules that need to be included in such a calculation. Prototype water ensemble statistics generated in this way have provided molecular descriptors with unique content.

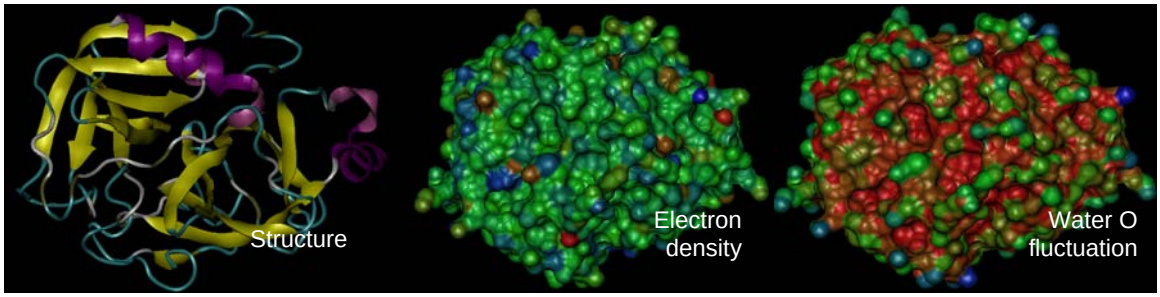
•**Development of enhanced descriptors based on Potentials of Mean Force-based hydration maps**: We have developed an efficient alternative to simulation-derived descriptors by means of a potentials-of-mean-force expansion (PMF) based method, employing a library of lower-order correlation functions derived from explicit simulations to predict the average equilibrium density and the orientation profile of water in the space surrounding biomolecules or ligands. The PMF code is currently available, and significant validation work has been performed by exhaustive comparisons of PMF results to those from detailed protein MD simulations, with several publications to support the approach.



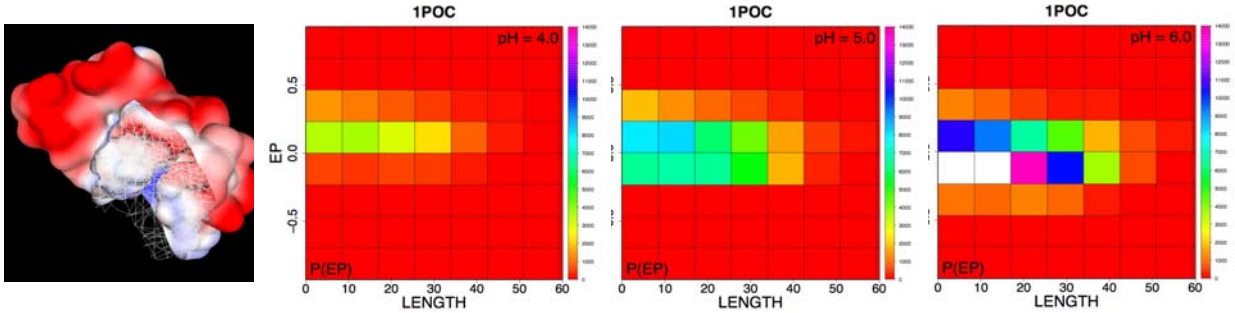
$$\rho^{(n,1)}(\mathbf{r}_W|\mathbf{r}_1,\ldots,\mathbf{r}_n)=\rho_0\prod_{i=1}^{n} g^{(2)}(\mathbf{r}_W,\mathbf{r}_i)\prod_{j=1}^{n-1}\prod_{k=j+1}^{n}\frac{g^{(3)}(\mathbf{r}_W,\mathbf{r}_j,\mathbf{r}_k)}{g^{(2)}(\mathbf{r}_W,\mathbf{r}_j)g^{(2)}(\mathbf{r}_j,\mathbf{r}_k)g^{(2)}(\mathbf{r}_k,\mathbf{r}_W)}\cdots$$

Figure 2. Water density values in space surrounding an alpha-helix (left) and a protein X (right) predicted using the PMF expansion (cyan) and obtained from exact simulation (magenta). The expression of the PMF-expansion is also shown .

•**Beyond hydration:** Developed efficient methods for calculation of free energy changes for biomolecular interactions, and on generalizing the PMF based methods for density and orientation calculation of other molecules (e.g., small-molecule ligands) and developing corresponding cheminformatics descriptors. Garde has one published paper demonstrating the utility of PMF-based methods for efficient calculations of free energies. Descriptor densities in the vicinity of two test proteins have been calculated.



•A novel set of pH-dependent protein molecular descriptors was created to represent protein charge and EP properties under real-world conditions. The use of this novel set of descriptors in concert with an SVM regression algorithm enabled the generation of integrated QSPR models which were successfully used to predict complex protein/surface binding behavior across a wide pH range. QSPR models were also developed for predicting water release values and gradient behavior for proteins interacting with hydrophobic surfaces, and for predicting the selectivity of mixed-mode chromatographic ligands.



•Developent and public dissemination of atom- and surface-based TAE-augmented scoring functions.

SOFTWARE

ROE Beta

This web based application enables automation of Rank Order Entropy utilizing Kendall Tau. This ROE process will examine an input dataset and give a recommendation for the user in terms of the stability of models created by the dataset as well as an evaluation of the ability of the dataset to create a predictive model. The procedure takes the dataset, divides it to training/testing sets, tests the training/testing sets for ability to model, and then models the data, and finally examines the data in terms of Rank Order (as Kendall Tau) and Rank Order Entropy (as Kendall Tau over several truncations of data).

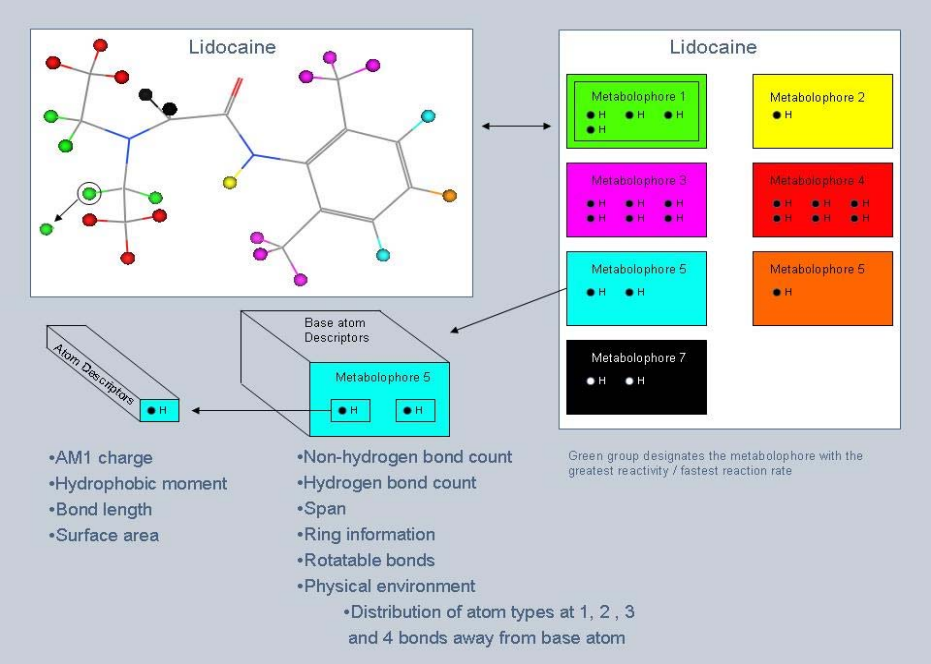
ROMS

The RECCR Online Modeling System (ROMS) is a general web-based machine learning system. By using the available learning methods, users can generate a model and visualize its performance by uploading their data set through the web client. Three learning methods provided are Partial Least Squares (PLS), Kernel-PLS and Support Vector Machine (SVM). In addition to basic modeling functionality, cross validation methods such as Leave-One-Out (LOO) and Monte Carlo Cross Validation (MCCV) are provided for model parameter selection.

MIRank

Multiple Instance Ranking (MIRank) is a novel machine learning model that enables ranking to be performed in a multiple instance learning setting. The motivation for MIRank stems from the hydrogen abstraction problem in computational chemistry, that of predicting the group of hydrogen atoms from which a hydrogen is abstracted (removed) during metabolism. The model predicts the preferred hydrogen group within a molecule by ranking the groups, with the ambiguity of not knowing which hydrogen atom within the preferred group is actually abstracted. The paper formulates MIRank in its general context and

proposes an algorithm for solving MIRank problems using successive linear programming. The method outperforms multiple instance classification models on several real and synthetic datasets. This website freely distributes the datasets and source codes used in this first study.



PEST

PEST Shape/Property hybrid descriptor technology, developed in DDASSL, allows better representation of the kinds of intermolecular interactions that are dependent on molecular shape. The inclusion of PEST descriptors has been found to significantly improve QSPR models where intermolecular interactions play an important role in the chemical effects being modeled. PEST descriptors are generated using TAE molecular surface representations to define property-encoded boundaries similar to the Zauhar "Shape Signature" ray-tracing approach to shape/property convolution.

PESD

Appropriate representation and accurate characterization of protein-ligand interactions is critical for developing a fundamental understanding of the chemistry of biomolecular recognition, as well as for the successful application of prospective potency and ADMET models in drug design. Relating biological activities to calculable properties of molecular structures in a quantitative manner has long been recognized as one of the grand challenges in chemistry and molecular biology. A fervent hope – even an implicit assumption – of most practitioners has been the hypothesis that similar molecules would tend to display similar activities in biological assays. Though often realized in practice, many departures from this guiding principle have been observed in recent years, even with increasing sophistication of machine learning methods designed to extract pertinent information from a proliferation of molecular descriptors. One reason for the departure from this similarity principle is the complex nature of the activity landscape within the existing descriptor space associated with any given biological assay (“activity cliffs”). This landscape complexity is inextricably linked with the appropriateness of the chemical space representation (molecular descriptors) and similarity assessment metric employed. Biomolecules have a hierarchy of structures, and thus a hierarchy of associated representations. Often they possess a wide range of protonation states depending upon the pH, and are generally associated with several molecules of bound water and counter-ions that critically influence their structure, dynamics and biological functions. A useful set of molecular descriptors needs to encode all these molecular characteristics pertinent to observed properties, including interactions with other molecules that result in modulation of specific metabolic pathways. Appropriate representation of these interactions is thus critical to successful application of a predictive statistical model and to the ultimate goal of developing a fundamental understanding of protein-ligand interaction, laying the foundation for a new generation of therapeutics and drug design. We have been thus developed the PESD computational technology for appropriate representation, visualization and scoring of protein-ligand interactions.

The first step in computational analysis of a binding site and its interactions with ligands is to encode the structure into a machine readable format. The encoding should be rich in information content such that the maximum amount of relevant knowledge can be extracted. The method should also be sufficiently fast to enable high-throughput analysis of protein-ligand interactions. The Property Encoded Shape Distributions (**PESD**) method was developed to represent the shape and surface property distribution of protein binding sites as a compact signature free of atom and residue labels, encoding the surface

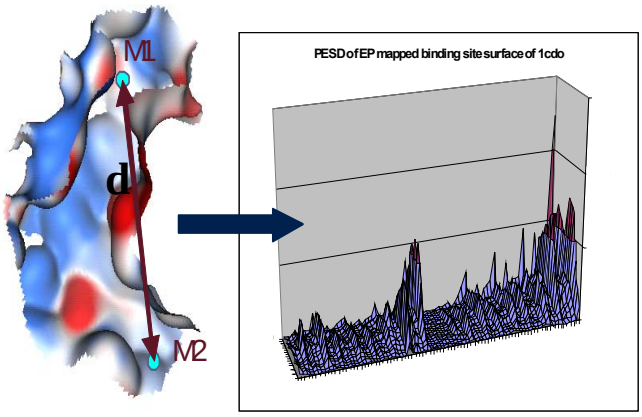


Figure 1. PESD signature of the electrostatic potential mapped binding site of 1cdo showing two points having magnitudes M1 and M2 of electrostatic potential values and the distance d between them. A large pool of such pairs of points is binned to generate a PESD signature.

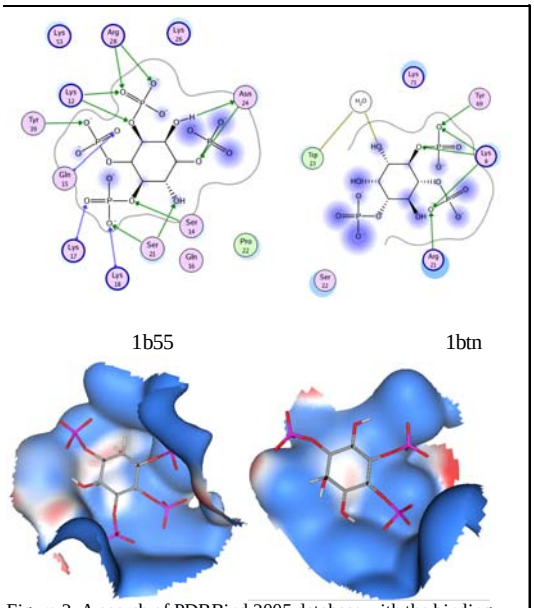
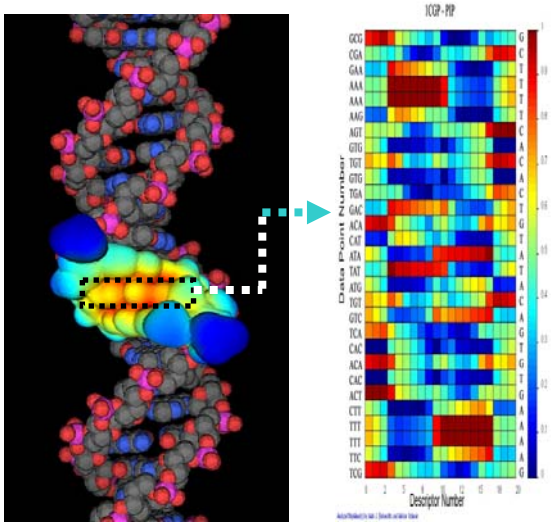


Figure 2. A search of PDBBind.2005 database with the binding site of 1b55 returns 1btn in the top 1% of ranked matches.

chemistry in the form of electrostatic potentials and polar, hydrophobic and hydrogen-bonding regions. Through rapid comparison of binding site shape/property patterns, cross-reacting ligands can be identified, thus making the method suitable for a prospective multi-objective virtual screening for biological reactivity patterns of drug-like molecules. Since the search is protein-centric, relevant alternate ligand substructures can be obtained from bound ligands that cannot be captured by ligand-based similarity analysis methods. The method works by random sampling of property values of pairs of points on the molecular surface and binning by both property distribution and the distance ‘d’ between the points. PESD provides a fast global similarity search procedure for binding sites and easy-to-compute signatures for capturing three-dimensional shape and mapped properties on molecular surfaces, making it possible to detect similarities in binding sites when bound ligand similarity is low. Figure 1 shows an example of a PESD signature. Binding site comparison algorithms have been actively investigated for over a decade (Moodie et al., 1996; Gold et al., 2006). However, application of binding site comparison techniques to predict ligand cross-reactivity has only been recently reported (Schalon et al., 2008; Kinnings et al., 2009). This is because of the restrictive nature of structure-based comparison techniques. Subtle partial similarities in shape and property distributions at the surface level of binding sites can lead to binding of the same ligand. Consequently it is important to develop new methods specially geared towards identification of potential sites for ligand cross-reactivity. A surface property and shape comparison based approach such as PESD will overcome the restrictions imposed by current amino-acid, atom or pseudo-atom based similarity analysis techniques. An example where the surface based PESD method can be effective is the similarity in binding sites of complexes 1b55 and 1btn (Figure 2). Although they bind to very similar ligands, the binding sites have low conservation of interacting residues. This results in algorithms such as SitesBase (Gold et al., 2006) and MultiBind (Shulman-Peleg et al., 2008) not detecting significant similarities at the binding site. However the shape and environment of the two sites are very similar and this similarity is effectively captured by PESD. On screening the PDBBind database with 1b55, 1btn appears in the top 1% of the ranked hits. It is also experimentally known that the ligand of 1b55 binds to the binding site of 1btn (Hyvönen et al., 1995).

DIXEL

This web-enabled DNA descriptor generator provides a TAE-based electronic representation of the chemical properties of either the major or minor grooves of DNA. DIXEL descriptors represent electron density-based features such as electrostatic potential (EP) as well as local average ionization potential (PIP) on the accessible surfaces of the major or minor groove by a grid of rectangles -- the "Dixel" coordinate system. These features can be displayed graphically and/or employed as input to data mining algorithms. Dixel descriptors have been evaluated as positionally-aware DNA “bar code” descriptors for characterizing protein/DNA interaction sites for gene regulation.



PROTEIN RECON

A version of the RECON/TAE program optimized for use with proteins, allowing users to rapidly produce a set of descriptors that can characterize protein behavior. Protein Recon is an algorithm for the rapid reconstruction of molecular charge density-based electronic properties of proteins, using peptide fragments precomputed from ab initio wavefunctions. These properties can be displayed graphically and/or employed as input to data mining algorithms.

WebPDB

WebPDB is a web-based workflow system that is flexible and capable of semi-automatic protein structure cleaning activities. The protein data may be provided by the user, but can also be directly downloaded from the PDB archive as part of the automated workflow. In its next generation, WebPDB will produce pH-sensitive protein surface descriptors that take into account appropriate protonation states and fractional protonation/deprotonation of basic and acidic side chain groups. WebPDB prepares proteins for use in virtual screening and predictive modeling. It removes gaps (through self-homology with FASTA information), heteroatoms and ligands (for re-use). Coupled with other modeling tools, WebPDB can be useful in probe development and the interpretation of secondary screening results through docking and scoring computations.

OUTREACH

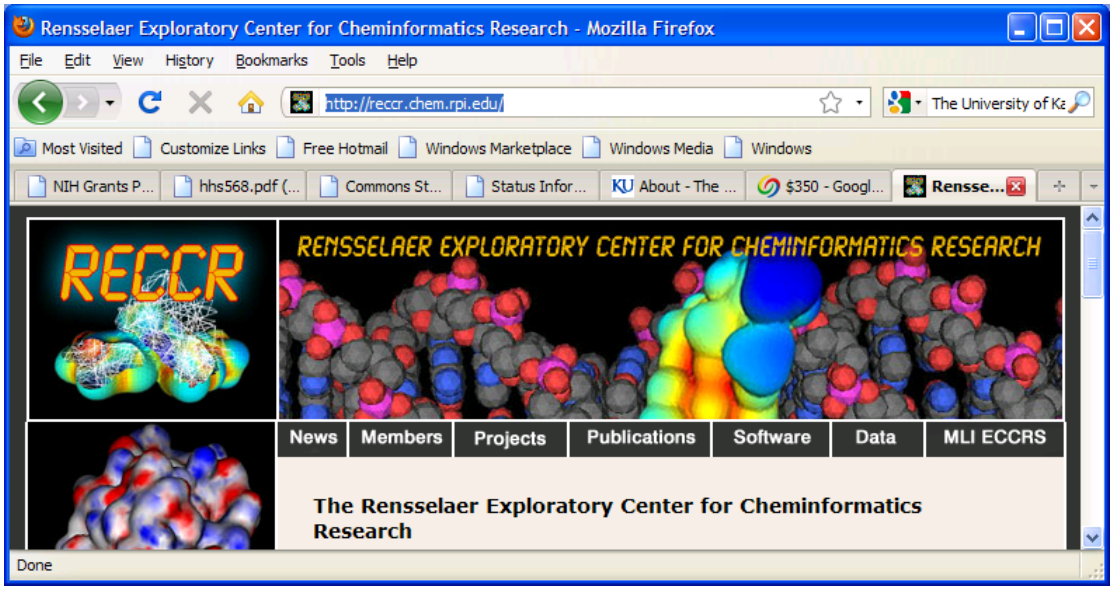
National Chemistry Day 2007

The Rensselaer Exploratory Center for Cheminformatics Research joined 15 other organizations from the capital region in participating in the National Chemistry Day on Oct. 21, 2007, organized by the Eastern New York section of the American Chemical Society at the New York State Museum. The event saw 450-500 kids and 900-1000 people in total attendance and received wide coverage in the local television and print news media.

National Chemistry Day 2008

The Rensselaer Exploratory Center for Cheminformatics Research again participated in the National Chemistry Day on Oct. 19, 2008, organized by the Eastern New York section of the American Chemical Society at the New York State Museum. The Eastern New York Section of ACS was chosen as winner of the ChemLuminary Award for Outstanding Performance by a Local Section in the Medium Large Size Category at the 238th National ACS Meeting in Washington, DC.

DISSEMINATION



The RECCR Website has been designed to support effective dissemination of the software, datasets and algorithms developed during the project period. See: http://reccr.chem.rpi.edu.

RECCR PUBLICATION LIST (76 ARTICLES)

1. P. Aguis, K. Bennett, and M. Zuker, **Comparing RNA secondary structures using a Relaxed Base Pair Score** *RNA journal*, under review, 2009
2. S. Das, A. Kokardekar, C. Breneman **Property-Encoded Shape Distributions (PESD) for Binding Site Comparison in Proteins** *Journal of Chemical Information and Modeling (JCIM)*, Accepted, 2009
3. C. Bergeron, K. Bennett, G. Moore, J. Zaretzki and C. Breneman **Fast Bundle Algorithm for Multiple Instance Ranking** *Proceedings of the International Conference on Machine Learning*, Accepted, 2009
4. Reeder PJ, Huang Y-M, Bystroff C, & Dordick JS **Re-wiring: Rational Design and Analysis of Protein Topology in Green Fluorescent Protein**. Proteins, Structure, Function and Bioinformatics (in revision, 2009).
5. N. Sukumar, Sourav Das, Michael Krein, Curt M. Breneman, Qiong Luo, Rahul Godawat, Shekhar Garde, Inna Vitol and Kristin P. Bennett, **Molecular Descriptors for Biological Systems** in "Computational Approaches in Cheminformatics and Bioinformatics" Rajarshi Guha and Andreas Bender, Eds. (Wiley, in press, 2009)
6. C. Bergeron, T. Hepburn, M. Sundling, N. Sukumar, W. P. Katt, K. P. Bennett, C. M. Breneman, **Prediction of peptide bonding affinity: Kernel methods for nonlinear modeling** Special Issue on COEPRA, Protein Peptide Lett, in press, 2009.
7. Cole B & Bystroff C. **Alpha helical crossovers favor right-handed supersecondary structures by a kinetic trapping mechanism. The phone cord effect in protein folding.** Protein Science 18(8) 1602 – 1608 (2009)
8. Huang Y-M & Bystroff C. **Specific complementation and reconstitution of flourescence from a circularly permuted, truncated green fluoresent protein.** Biochemistry 48(5):929-40. (2009)
9. Chung, W.K.; Hou, Y.; Freed, A; Holstein, M.; Makhatadze, G., and Cramer, S.M. Salem S, Zaki MJ, Bystroff C **Iterative non-sequential protein structural alignment**. J Bioinform Comput Biol 7(3):571-96 (2009)
10. **Investigation of protein binding affinity and preferred orientations in ion exchange systems using a homologous protein library**, Biotech. Bioeng. Volume 102, Issue 3 (p 869-881) (2009). (cover art featured on journal cover).
11. N. Sukumar, **The Chemist's Concept of Molecular Structure** *Foundations of Chemistry*, **11** (1), 7 (2009)
12. Pankaj Kulshrestha, N. Sukumar, Jane S. Murray, Rossman F. Giese and Troy D. Wood, **Computational Prediction of Antibody Binding Sites on Tetracycline Antibiotics: Electrostatic Potentials and Average Local Ionization Energies on Molecular Surfaces** *J. Phys. Chem. A*, **113** (4), 756-766 (2009)
13. Dong Wan Han and J. A. Moore, **Synthesis and Characterization of Adamantane-containing Poly(enaminonitriles)**, , Polymer, 50, 2551-2557 (2009).
14. Christopher J. Morrison, Curt M. Breneman, J.A. Moore, Steven M. Cramer, **Evaluation of Chemically Selective Displacer Analogues for Protein Purification**, , Analytical Chemistry, 81(15), 6186-6194, (2009).

15. Chen J., Yang T. and Cramer S.M. **Prediction of Protein Retention Times in Gradient Hydrophobic Interaction Chromatographic Systems** J. Chromatogr. A, 1177 (2008) 207–214.
16. Christopher J. Morrison, Sun Kyu Park, Chester Simocko, Scott A. McCallum, Steven M. Cramer, J.A. Moore, **Synthesis and Characterization of Fluorescent Displacers for Online Monitoring of Displacement Chromatography**., ,Journal of the American Chemical Society, 130 (50), 17029-17037 (2008).
17. K. P. Bennett, **Discussion: Evidence Contrary to the Statistical View of Boosting by David Mease and Abraham Wynar**, *Journal of Machine Learning Research*, 2008.
18. Daniel L. Silver, and K. P. Bennett, **Guest Editorial: Inductive Transfer Learning**,*Machine Learning Journal*, 2008.
19. J. Hu, J. E. Mitchell, J.-S. Pang, K. P. Bennett and G. Kunapuli **On the Global Solution of Linear Programs with Linear Complementarity Constraints,** *SIAM Journal on Optimization*, 2008.
20. Buck P. & Bystroff C. **Simulating protein folding initiation sites using an alpha-carbon-only knowledge-based force field** Proteins, Structure, Function and Bioinformatics 76(2):331-342. (2008)
21. Bulent Yener, Evrim Acar, Phaedra Agius, Scott L. Vandenberg, Kristin P Bennett and George E Plopper, **Multiway Modeling and Analysis in Stem Cell Systems**, BMC Systems Biology, 2:63. 2008.
22. N. Sukumar, Michael Krein and Curt M. Breneman, **Bio- and Chem-Informatics: Where do the twain meet?** *Curr. Opinion Drug Disc. Devel.* **11** (3) 311-319 (2008)
23. Chen J., Yang T. and Cramer S.M. **Prediction of Protein Retention Times in Gradient Hydrophobic Interaction Chromatographic Systems** J. Chromatogr. A, 1177 (2008) 207–214.
24. C. Bergeron, Jed Zaretzki, Curt Breneman and Kristin Bennett **Multiple Instance Ranking** *ACM International Conference Proceeding Series; Vol. 307, pp 48-55 Proceedings of the 25th International Conference on Machine learning*, 2008
25. Shentu, Z., al Hasan, M., Bystroff, C., Zaki, MJ. **Context Shapes: Efficient Complementary Shape Matching for Protein-Protein Docking.** Proteins: Structure, Function and Bioinformatics 70(3):1056-1073 (2008)
26. Shinde A, Bystroff C Wang C, Vogelezang MG, Vincent PA, Hynes RO, van de Water, **L. EIIIA (ED-A), the alternatively spliced segment of fibronectin exhibits cryptic cell adhesive activity to integrin α9β1**. J Biol Chem J Biol Chem. 283(5):2858-70 (2008)
27. Bystroff C. & Webb-Robertson, B.J. **Pairwise covariance adds little to secondary structure prediction but improves the prediction of non-canonical local structure** BMC Bioinformatics 9:429 (2008)
28. Demiriz, K. P. Bennett, and P. S. Bradley, **Using Assignment Constraints to Avoid Empty Clusters in k-means Clustering**. *Constrained Clustering: Advances in Algorithms, Theory, and Applications*, S. Basu, I. Davidson, and K. Wagstaff, CRC Press, pg. 203-219, 2008. (Refereed compilation)
29. K. P. Bennett, G. Kunapuli, J. Hu, and J.-S. Pang. **Optimization and Machine Learning**, in *Computational Intelligence: Research Frontiers : IEEE World Congress on Computational*

Intelligence, WCCI 2008, Hong Kong, China, June 1-6, 2008 : Plenary/invited Lectures. Edited By Jacek M. Zurada, Gary G. Yen, Jun Li Jim Wang, Springer, pg. 29-48, 2008.

30. G. Kunapuli, K. P. Bennett, J. Hu, and J. S. Pang, **Bilevel Model Selection for Support Vector Machines** *Data Mining and Mathematical Programming*, P. M. Pardalos and P. Hansen, CRM Proceedings and Lecture Notes, American Mathematical Society, volume : 45, pg. 129-158, 2008. (Refereed compilation)
31. G. Kunapuli, K. Bennett, J. Hu, and J.-S. Pang, **Classification Model Selection via Bilevel Programming** *Computational Optimization and Applications*, 2008.
32. R. Klees, R. Salasznyk, S. Vandenberg, K. Bennett, and G. Plopper, **Laminin-5 Activates Extracellular Matrix Production and Osteogenic Gene Focusing in Human Mesenchymal Stem Cells**, *Matrix Biology*, 2007.
33. P. Agius, B. Kreiswirth, N. Kurepina, and K. P. Bennett, **Typing *Staphylococcus aureus* using the *spa* gene and novel distance measures**, *IEEE Transactions on Computational Biology and Bioinformatics*, 2007.
34. Roberto J. Bayardo, Kristin P. Bennett, Gautam Das, Dimitrios Gunopulos, Johannes Gunopulos: I**ntroduction to special issue ACM SIGKDD 2006** ACM *Transactions on Knowledge Discovery and Data Mining,* 1(3): (2007)
35. K. Bennett, C. Bergeron, E. Acar, R. Klees. S. Vandenberg, B. Yener, and G. Plopper, **Proteomics reveals multiple routes to the osteogenic phenotype in mesenchymal stem cells**, *BMC Genomics*, 8:380, 2007.
36. Ekins, Sean; Embrechts, Mark J.; Breneman, Curt M.*; Jim, Kam; Wery, Jean-Pierre. **Novel applications of kernel-partial least squares to model a comprehensive array of properties for drug discovery.** *Computational Toxicology* (2007), 403-432.
37. N. Sukumar and Curt M. Breneman, “**QTAIM in Drug Discovery and Protein Modeling**” in “The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design” C.F. Matta & R.J. Boyd, Eds. (Wiley-VCH, 2007)
38. L. Han, M. Embrechts and K. Bennett, **Kernel-based Multiple Response Latent Analysis with Entropy Loss Function**, submitted to International Joint Conference on Neural Networks, 2007.
39. Chen, Jie; Luo, Qiong; Breneman, Curt M.; Cramer, Steven M. **Classification of protein adsorption and recovery at low salt conditions in hydrophobic interaction chromatographic systems.** Journal of Chromatography, A (2007), 1139(2), 236-246.
40. Ladiwala, Asif; Breneman, Curt M.; Cramer, Steven M. **A priori prediction of chromatographic separations from protein structure data.** Biotechnology and Bioprocessing (2007), 31 (Process Scale Bioseparations for the Biopharmaceutical Industry), 245-275.
41. **Embrechts**, Mark J.; Ekins, Sean. **Classification of metabolites with kernel-partial least squares (K-PLS).** Drug Metabolism and Disposition (2007), 35(3), 325-327.
42. K. Bennett, C. Bergeron, E. Acar, R. Klees. S. Vandenberg, B. Yener, and G. Plopper, **Multiple routes to the osteogenic phenotype in mesenchymal stem cells**, BMC Genomics, 2007.
43. G. Kunapuli, K. Bennett, J. Hu, and J.-S. Pang, **Classification Model Selection via Bilevel Programming, Computational Optimization and Applications**, 2007.

44. P. Agius, B. Kreiswirth, N. Kurepina, and K. P. Bennett, **Typing Staphylococcus aureus using the spa gene and novel distance measures**, IEEE Transactions on Computational Biology and Bioinformatics, 2007.

45. Changjian Huang, Mark J. Embrechts, N. Sukumar and Curt M. Breneman, **Data Fusion and Auto-fusion for Quantitative Structure-Activity Relationship (QSAR)** Lecture Notes in Computer Science, 4668/2007, p.628-637 (Springer, Berlin / Heidelberg, 2007)

46. R. Klees, R. Salasznyk, S. Vandenberg, K. Bennett, and G. Plopper, **Laminin-5 Activates Extracellular Matrix Production and Osteogenic Gene Focusing in Human Mesenchymal Stem Cells**, Matrix Biology, 2007.

47. Yang T., Breneman C., and Cramer S.M., **Investigation of multi-modal high-salt binding ion-exchange chromatography using quantitative structure–property relationship modeling** Journal of Chromatography A, Volume 1175, Issue 1, Pages 96-105 (2007).

48. Yang T, Sundling M, Freed A, Breneman C. and Cramer, S.M. **Prediction of pH-Dependent Chromatographic Behavior in Ion-Exchange Systems**, Anal. Chem.; 79(23); 8927-8939. (2007).

49. Chen J., Yang T., Luo ., Breneman C., and Cramer S.M. **Investigation of protein retention in hydrophobic interaction chromatographic (HIC) systems using the preferential interaction theory and quantitative structure property relationship models** Reactive & Functional Polymers 67: 1561–1569 (2007).

50. Xia K, Manning M, Hesham H, Lin Q, Bystroff C & Colon W. **Identifying the Subproteome of Kinetically Stable Proteins Via Diagonal 2D SDS-PAGE**. Proc Nat Acad Sci U S A. 104(44):17329-34 (2007)

51. Yang, Lu; Garde, Shekhar. **Modeling the selective partitioning of cations into negatively charged nanopores in water.** Los Alamos National Laboratory, Preprint Archive, Physics (2007), 1-22, arXiv:physics/0702065.

52. Athawale, Manoj V.; Goel, Gaurav; Ghosh, Tuhin; Truskett, Thomas M.; Garde, Shekhar. **Effects of length scales and attractions on the collapse of hydrophobic polymers in water.** Los Alamos National Laboratory, Preprint Archive, Condensed Matter (2007), 1-24, arXiv:cond-mat/0701676.

53. Kristin P. Bennett, Emilio Parrado-Hernández, **The Interplay of Optimization and Machine Learning Research** *Journal of Machine Learning Research*,7(Jul):1265--1281, 2006.

54. Oloff, Scott; Zhang, Shuxing; Sukumar, Nagamani; Breneman, Curt; Tropsha, Alexander. **Chemometric analysis of ligand receptor complementarity: identifying complementary ligands based on receptor information (CoLiBRI).** Journal of Chemical Information and Modeling (2006), 46(2), 844-851.

55. Liu, Jia; Yang, Ting; Ladiwala, Asif; Cramer, Steven M.; Breneman, Curtis M.. **High throughput determination and QSER modeling of displacer DC-50 values for ion exchange systems.** Separation Science and Technology (2006), 41(14), 3079-3107.

56. Ladiwala, Asif; Xia, Fang; Luo, Qiong; Breneman, Curt M.; Cramer, Steven M. **Investigation of protein retention and selectivity in HIC systems using quantitative structure retention relationship models.** Biotechnology and Bioengineering (2006), 93(5), 836-850. Kristin P. Bennett, Emilio Parrado-Hernández, **The Interplay of Optimization and Machine Learning Research**, Journal of Machine Learning Research,7(Jul):1265--1281, 2006.

57. J. Liu, S.-K. Park, J. A. Moore and S.M. Cramer **The Effect of Displacer Chemistry on Displacer Efficacy for a Sugar-based Anion Exchange Displacer Library**,, Industrial Engineering & Chemistry Research, 45, 9107-9114 (2006).

58. Liu, Jia; Park, Sun K.; Moore, J. A.; Cramer, Steven M.. **Effect of Displacer Chemistry on Displacer Efficacy for a Sugar-Based Anion Exchange Displacer Library.** Industrial & Engineering Chemistry Research (2006), 45(26), 9107-9114.

59. Sundling, C. Matthew; Sukumar, Nagamani; Zhang, Hongmei; Embrechts, Mark J.; Breneman, Curt M. **Wavelets in chemistry and cheminformatics.** Reviews in Computational Chemistry (2006), 22 295-329.

60. Vitol, J. Driscoll, B. Kreiswirth, N. Kurepina, and K. P. Bennett, Identifying **Mycobacterium tuberculosis complex Families using Spoligotypes, Infection**, Genetics and Evolution, Nov;6(6):491-504, 2006.

61. K. P. Bennett, J. Hu, G. Kunapuli, and J.-S. Pang, **Model Selection via Bilevel Optimization,** Proceedings of the International Joint Conference in Neural Networks, Vancouver, 2006.

62. Zhiwei Zhu, Qiang Ji, Kristin P. Bennett: **Nonlinear Eye Gaze Mapping Function Estimation via Support Vector Regression**. ICPR (1) 2006: 1132-1135

63. Tikhonov, Aleksey M.; Patel, Harshit; Garde, Shekhar; Schlossman, Mark L. **Tail Ordering Due to Headgroup Hydrogen Bonding Interactions in Surfactant Monolayers at the Water-Oil Interface.** Journal of Physical Chemistry B (2006), 110(39), 19093-19096.

64. Jain, Sandeep; Garde, Shekhar; Kumar, Sanat K. **Do Inverse Monte Carlo Algorithms Yield Thermodynamically Consistent Interaction Potentials?** Industrial & Engineering Chemistry Research (2006), 45(16), 5614-5618.

65. Pereira, Brian; Jain, Sandeep; Garde, Shekhar. **Quantifying the protein core flexibility through analysis of cavity formation.** Journal of Chemical Physics 124(7), 074704/1-074704/5. 2006

66. Pereira Brian; Jain Sandeep; **Garde**, Shekhar **Quantifying the protein core flexibility through analysis of cavity formation**. The Journal of chemical physics (2006), 124(7), 74704.

67. Zaki, Mohammed J.; Sequeira, Karlton. **Data mining in computational biology.** Handbook of Computational Molecular Biology (2006), 38/1-38/26.

68. Zhang Yongqiang; **Zaki**, Mohammed J **EXMOTIF: efficient structured motif extraction**. Algorithms for molecular biology : AMB (2006), 1 21.

69. Zhang Yongqiang; Zaki, Mohammed J **SMOTIF: efficient structured pattern and profile motif search.** Algorithms for molecular biology : AMB (2006), 1 22.

70. M. Momma and K. Bennett, **Constructing Orthogonal Latent Features for Arbitrary Loss**, *Feature Extraction, Foundations and Applications,* Isabelle Guyon, Steve Gunn, Masoud Nikravesh, and Lofti Zadeh, editors, Springer, 2006. (Refereed compilation)

71. Rege K, Ladiwala A, Hu S, Breneman CM, Dordick JS, Cramer SM **Investigation of DNA-binding properties of an aminoglycoside-polyamine library using quantitative structure-activity relationship (QSAR) models**, J. of Chem. Information and Modeling. 45 (6): 1854-1863 (2005).

72. Ladiwala A, Rege K, Breneman CM, Cramer SM, **A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems**, PNAS 102 (33): 11710-11715 (2005)

73. Robert Grossman, Roberto Bayardo, Kristin P. Bennett, ***Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining***, Chicago, Illinois, USA, August 21-24, 2005 ACM 2005

74. Z. Xiang and K. P. Bennett, **Multitask Transfer Using Multiple Latent Analysis, workshop on Inductive Transfer: 10 Years Later** Workshop at Neural Information Processing Systems Conference, December 2005, Whistler, Canada.

75. B.K. Lavine, C.E. Davidson, W. Katt, C.M. Breneman*, **Analysis of odor structure relationships using electronic Van der Waals surface property descriptors and genetic algorithms** ACS Symposium Series, 894 *Chemometrics and Cheminformatics*, 127-143. (2005)

76. Curt M. Breneman, Minghu Song, Jinbo Bi, N. Sukumar, Kristin P. Bennett, Steven Cramer and Nihal Tugcu, **Prediction of Protein Retention Times in Anion-exchange Chromatography Systems using Support Vector Regression** in Chemometrics and Chemoinformatics, ACS Symp. Ser. 894, Barry K. Lavine (Ed.), American Chemical Society, pp. 111-125 (2005)

RECCR PRESENTATIONS (64 ORAL PRESENTATIONS)

1. C.M. Breneman "Nanomaterials Informatics and Materials QSPR" Office of Naval Research, Arlington, VA, June 15, 2009)
2. C.M. Breneman "Nanomaterials Informatics" Lockheed Martin Advanced Technologies Group, Cherry Hill, NJ (Feb 29, 2008)
3. C.M. Breneman "The Role of Cheminformatics Innovation and Development in the MLI/MLSCN/NHGRI Program", NIH MLI/ECCR Joint Workshop, Bethesda Hyatt, Bethesda, MD. May 16, 2008.
4. C.M. Breneman "Visual Analytics: A Cheminformatics Perspective" Rensselaer Research Retreat, Red Lion Inn, Stockbridge, MA (May 27, 2008)
5. C.M. Breneman "A Hard Look at Predictive Modeling – How much data is enough?" Cheminformatics Conference / Computational Toxicology Session, InnovationWell - eCheminfo community of practice, Bryn Mawr College (Oct 17, 2008)
6. C.M. Breneman "Nanomaterials Informatics & MQSPR: Review and Update" Lockheed Martin Aero Plant, Fort Worth, TX. (Dec 8, 2008)
7. C.M. Breneman "The Creation and Use of Novel Molecular Descriptors and Data Fusion Methods", University of North Texas, Denton, TX, Nov 2007
8. C.M. Breneman "The Creation and Use of Novel Molecular Descriptors and Data Fusion Methods", Southern Research Institute (SRI), Sept 2007

9. C.M. Breneman “Predictive Cheminformatics: QSAR Applications of Machine Learning”, University of Cambridge, Unilever Cheminformatics Centre, Cambridge, England, June 2007
10. C.M. Breneman “Predictive Cheminformatics: Towards Fulfilling the Promise of Virtual Screening”, Featured Speaker, 90th Canadian Chemical Conference, Winnipeg, Manitoba, May 2007
11. C.M. Breneman “Predictive Cheminformatics: Best Practices for determining Model Domain Applicability”, Plenary Talk, Sanibel Conference, St. Simons Island, GA Feb 2007
12. C.M. Breneman, “Predictive ADME : How do I know if my predictions will be useful?”, as part of the Predictive Toxicology eCheminfo Program, Bryn Mawr College, Bryn Mawr, PA, October 2006
13. C.M. Breneman, “Advances in Protein QSPR and Surface Analysis”, FACSS Symposium in Honor of Peter Jurs, Orlando, Florida, September 2006
14. C.M. Breneman, “Advances in Cheminformatics: Applications in Biotechnology, Drug Design and Bioseparations” Perspectives in Chemistry, RPI, September 2006
15. C.M. Breneman, “New Cheminformatics Tools for PubChem”, NIH MLSCN Steering Committee, Washington, D.C., July 2006
16. C.M. Breneman, “RECCR Organization and Capabilities” NIH ECCR Group, Washington, D.C., July 2006
17. C.M. Breneman, “PPEST: The use of Property-Encoded Solvent Accessible Surfaces for Protein Characterization and Classification” CCG MOE UGM, Montreal, CA June 2006
18. C.M. Breneman, “Advances in Cheminformatics at the Rensselaer Exploratory Center for Cheminformatics Research” RPI CBIS Seminar Series, March 2006
19. C.M. Breneman, “Ab-initio prediction of the Energetics of reversible Catalytic Hydrogenation in Carbazoles” GE CR&D, Niskayuna, NY, February 2006
20. C.M. Breneman, “Prediction of Protein-DNA Interactions”, CMS/Axelrod Institute, Albany, NY, January, 2006
21. C.M. Breneman, “Virtual Screening of potential p53 Inhibitors” telepresence seminar, transmitted from VCC/RPI to Lerner Institute, Cleveland Clinic, January 2006
22. Curt M. Breneman “Intelligent Design of Nanocomposites vis Informatics” ACS National Meeting, Salt Lake City, UT, Spring 2009
23. Curt M. Breneman “How much computation is enough?” ACS National Meeting, Philadelphia, PA, Fall 2008
24. Curt M. Breneman and Mark Embrechts “Text Mining for Cheminformatics Applications” ACS National Meeting, Philadelphia, PA Fall 2008
25. Curt M. Breneman “Cheminformatics developments at RECCR: New tools, collaborations and outreach” ACS National Meeting, New Orleans, LA, April 2008
26. C. Matthew Sundling, Curt M. Breneman, Mark Embrechts, Jack Chang, Xaiohua Wu and N. Sukumar “Testing the limits of a QSAR model: How many cases are actually

needed to develop a reliable predictive model?” ACS National Meeting, New Orleans, LA, April 2008

27. M. Dominic Ryan, Curt M. Breneman, Maggie McLellan and Mike Krein “QSAR model stability: How much information is in the data?” ACS National Meeting, New Orleans, LA, April 2008
28. N. Sukumar, Curt M. Breneman, Kristin P. Bennett, Charles Bergeron, Theresa Hepburn, C. Matthew Sundling, Shekhar Garde, Rahul Godawat, Ishita Manjrekar, Margaret McLellan, Mike Krein “Bio- and Chem-Informatics: Where do the twain meet?” ACS National Meeting, Boston, MA Sept 2007
29. C. M. Breneman, N Sukumar, Mark J. Embrechts, Kristin P. Bennett, C. Matthew Sundling, Mike Krein, Theresa Hepburn “Realizing Prospective QSAR through Data Fusion and Modern Descriptors” ACS National Meeting, Boston, MA Sept 2007
30. Jed Zaretzki, , Curt M. Breneman, , Charles Bergeron, N. Sukumar, and Mike Krein “A reactivity and recognition component-based methodology for computational prediction of likely sites of CYP 450 3A4-mediated metabolism”, ACS National Meeting, Boston, MA Sept 2007
31. M. Dominic Ryan, Theresa Hepburn, N. Sukumar, Sourav Das, Curt M. Breneman “TAE Augmented scoring functions: Two approaches, atom and surface based” ACS National Meeting, Boston, MA Sept 2007
32. C. M. Breneman, Qiong Luo, and Matt Sundling “Design, Development and Evaluation of Novel Proetin Property-encoded Surface Translator (PPEST) Descriptors for Protein Similarity Comparison”, ACS National Meeting, San Francisco, CA, Sept 2006
33. C.M. Breneman, Matt Sundling and Dominic Ryan “PROLICss: Analysis and Applications” ACS National Meeting, San Francisco, CA, Sept 2006
34. C.M. Breneman, N. Sukumar, Matt Sundling “New Tools for Virtual High-Throughput Screening”, ACS National Meeting, San Francisco, CA, Sept 2006
35. C.M. Breneman, Dominic Ryan, Wei Deng, and Mark Embrechts “Application of Novel Refined Distance-Dependent Descriptors, Transferable Atom Equivalent (TAE) Techniques and Machine-Learning Methods in Predicting Protein-Ligand Binding Affinity”, ACS National Meeting, San Francisco, CA, Sept 2006
36. Jie Chen, Matt Sundling, C.M. Breneman, Abigail Laurent, Jace Fogle, Steve Cramer, Erik Fernandez and Todd Przybycien “Investigation of Protein Adsorption Behavior and Structural Changes in Hydrophobic Interaction Chromatographic (HIC) Systems”, ACS National Meeting, San Francisco, CA, Sept 2006
37. Ting Yang, C.M. Breneman, Matt Sundling, and Steve Cramer “Investigation of Mobile Phase pH Effect on Protein Binding Affinity and Selectivity in Cation Exchange Systems”, ACS National Meeting, San Francisco, CA Sept 2006
38. Jia Liu, C.M. Breneman, Min Li, N. Sukumar and Steve Cramer “Identification of Novel Chemically-selective Displacers from Commercial Chemical Databases using an SVM Classification Model and Parallel Batch Screening Experiments”, ACS National Meeting, San Francisco, CA Sept 2006

39. N. Sukumar, Curt M. Breneman, Steven M. Cramer, Kristin P. Bennett, Matthew Sundling, Qiong Luo and Dechuan Zhuang, **“**Intelligent Data Mining for Modeling and Prediction of Protein-Protein Protein-Surface and Protein-DNA Interactions”, *Pacifichem 2005 - International Chemical Congress of Pacific Basin Societies, Honolulu, Hawaii,* December 2005

40. N. Sukumar, Curt M. Breneman, Steven M. Cramer, James A. Moore, Kristin P. Bennett, Mark J. Embrechts, Min Li, Jia Liu and Long Han, **“**Closing the Loop: From High-throughput Screening to Synthesis of Novel Protein Displacers”, *231st National Meeting American Chemical Society*, Atlanta, March 2006

41. Curt M. Breneman, Matthew C. Sundling, N. Sukumar, Kristin P. Bennett, Mark J. Embrechts and Steven Cramer, **“**Beyond PEST descriptors: Binding site and ligand shape/property fingerprints”, *231st National Meeting American Chemical Society*, Atlanta, March 2006

42. N. Sukumar, Curt M. Breneman and C. Matthew Sundling, **“**New tools for virtual high-throughput screening”, *232nd National Meeting American Chemical Society*, San Francisco, September 2006

43. Jia Liu, Min Li, N. Sukumar, Curt M. Breneman and Steven M. Cramer, **“**Identification of novel chemically selective displacers from commercial chemical databases using an SVM classification model and parallel batch screening experiments”, *232nd National Meeting American Chemical Society*, San Francisco, September 2006

44. Curt M. Breneman, N Sukumar, Mark J. Embrechts, Kristin P. Bennett, C. Matthew Sundling, Mike Krein and Theresa Hepburn, **“**Realizing Prospective QSAR through data fusion and modern descriptors”, *234th National Meeting American Chemical Society*, Boston, August 2007

45. Jed Zaretzki, Curt M. Breneman, Charles Bergeron, N. Sukumar and Mike Krein, **“**A reactivity and recognition component-based methodology for computational prediction of likely sites of CYP 450 3A4-mediated metabolism”, *234th National Meeting American Chemical Society*, Boston, August 2007

46. M. Dominic Ryan, Theresa Hepburn, N. Sukumar, Sourav Das and Curt M. Breneman, **“**TAE Augmented scoring functions: Two approaches, atom and surface based”, *234th National Meeting American Chemical Society*, Boston, August 2007

47. N. Sukumar, “Analyzing fleas on an elephant: Molecular similarity analysis and scoring through features of the electron density", invited lecture, *Eastern New York Section, American Chemical Society*, Albany, NY, 2008

48. Sourav Das, Curt M. Breneman, N. Sukumar and M. Dominic Ryan, **“**TAE augmented scoring functions: Application to enzymatic and nonenzymatic proteins”, *235th National Meeting American Chemical Society*, New Orleans, April 2008

49. C. Matthew Sundling, Curt M. Breneman, Mark J. Embrechts, Changjian Huang, Xiaohua Wu and N. Sukumar, **“**Testing the limits of a QSAR model: How many cases are actually needed to develop a reliable predictive model?”, *235th National Meeting American Chemical Society*, New Orleans, April 2008

50. Curt M. Breneman and N. Sukumar, **“**Cheminformatics developments at RECCR: New tools, collaborations and outreach”, *235th National Meeting American Chemical Society*, New Orleans, April 2008
51. Mark J. Embrechts, Curt M. Breneman, Changjian Huang and N. Sukumar, **“**Testing the validity range of QSAR models using one-class support vector machines”, *235th National Meeting American Chemical Society*, New Orleans, April 2008
52. Sunanda Sukumar, Benjamin Woo, N. Sukumar, Arshad S. Kokardekar, Judith Klein-Seetharaman and Kalyan C. Tirupula, **“**Docking Studies of Dipeptides to Metabotropic Glutamate Receptors”, *37th Northeast Regional Meeting American Chemical Society*, Burlington, VT, June 2008
53. Curt M. Breneman, N. Sukumar, Mike Krein, Margaret McLellan, Jed Zaretzki, Sourav Das and Arshad Shirish Kokardekar, “How much computation is enough?”, *236th National Meeting American Chemical Society*, Philadelphia, August 2008
54. L. C. Brinson, Linda S. Schadler, Curt M. Breneman, N. Sukumar, M. Kreim and R. Qiao, “Intelligent design of nanocomposites via informatics”, *237th National Meeting American Chemical Society*, Salt Lake City, March 2009
55. N. Sukumar and Mike Krein, **“**Mapping the Network Topology of Chemical Spaces” *Central Regional Meeting of the American Chemical Society*, Cleveland, May 2009
56. N. Sukumar, C. M. Breneman, S. M. Cramer, J. A. Moore, K. P Bennett, M. J. Embrechts, M. Li, J. Liu, L. Han, "Closing the Loop: from High-throughput-screening to Synthesis of Novel Protein Displacers", Abstracts of Papers, 231st ACS National Meeting, Atlanta, GA, March, 2006, Division of Chemical Information, Abstr. # 0028.
57. J. A. Moore, "Dead Leaves & Lawn Clippings: Waste or Opportunity?", Abstracts of Papers, IUMACRO 2007, 2nd Strategic Polymer Symposium, Brooklyn, NY, June, 2007, Abstr. #S5.
58. Chris Morrison, Scott McCallum, Rahul Godawat, J. A. Moore, Shekhar Garde, and Steven M. Cramer, "Investigation of Chemically Selective Displacers Using Robotic High Throughput Screening, SPR, NMR and MD Simulations", Abstracts of Papers, 234th ACS National meeting, Boston, MA, August, 2007, Division of Biochemical Technology, Abstr.# 301.
59. J. A. Moore, "Mother Nature as a Source of New Materials: Everything Old is New Again", Abstracts of Papers, 234th ACS National meeting, Boston, MA, August, 2007, Division of Polymer Chemistry, Abstr.# 254.
60. J. A. Moore, "Mother Nature as a Source of New Materials: Everything Old is New Again", Abstracts of Papers, 40th ACS Middle Atlantic Regional Meeting, Bayside, Queens, NY, May, 2008, Abstr. # 167.
61. J. A. Moore, "Functional Polycarbonates", Abstracts of Papers, 40th ACS Middle Atlantic Regional Meeting, Bayside, Queens, NY, May, 2008, Abstr. # 569.
62. Christopher J. Morrison, Sun Kyu Park, Scott McCallum, J. A. Moore and Steven M. Cramer, "The Development of Chemically Selective Displacement Processes for

Industrial Applications", Abstracts of Papers, 236th ACS National meeting, Philadelphia, PA, August, 2008, Division of Biochemical Technology, Abstr. # 253.

63. Christopher J. Morrison, Sun Kyu Park, Scott McCallum, Rahul Godawat, Shekhar Garde, J. A. Moore and Steven M. Cramer, "Chemically Selective Displacement Chromatography: Development and Application for Bioseparations", Abstracts of Papers, Am. Inst. Of Chem. Eng., National Meeting, Philadelphia, PA, November, 2008, Food, Pharmaceutical & Bioengineering Division, Abstr. # 571bo.
64. J. A. Moore, "Functional Polycarbonates", Abstracts of Papers, 238th ACS National meeting, Washington, D. C., August, 2009, Division of Polymer Chemistry, Abstr. # 450.

Colon U.S. Patents

Patent Number 7,217,348 (2007) Methods of identifying kinetically stable proteins. Wilfredo Colón and Marta Manning, Rensselaer Polytechnic Institute.

Patent Number 7,393,443 (2008) Methods of identifying kinetically stable proteins. Wilfredo Colón and Marta Manning, Rensselaer Polytechnic Institute.

Breneman Patents and Copyrights

2005 RPI Case 833 / Copyright Issued / “DIXELS” (A Fuzzy Bar Code Representations of DNA using Surface Pixels)

2005 RPI Case 834 / Copyright Issued / “EDDFA” (Electron Density-Derived Field Analysis)

2008 PEST Patent issued (November 4th) “System and method of computing and displaying property-encoded surface translator descriptors” 07446777 Cl. 345-581.

Moore Patents

"High-affinity, Low-molecular-mass Displacers for Ion-exchange Chromatography", Cramer; Steven M.; Moore; James A.; Park**;** Sun Kyu; Tugcu; Nihal, U. S. Patent #6,929,747, August 16, 2005.

"High-Affinity, Low-Molecular-Mass Displacers for Ion-Exchange Chromatography", Tugcu Nihal; Park, Sun Kyu; Moore, James A.; Cramer, Steven M., U. S. Pat.# 7,189,3240, March 13, 2007.