



**Bioinformatics
2009**

- * **DNA Analysis Tools**
- * **Protein Analysis Tools**
- * **Databases and Data Mining Tools**

Table of Contents

DNA Analysis Tools

| | |
|------------|---|
| SD2007-320 | A New Approach to Medical Diagnostics and Therapeutics |
| SD2007-022 | Method for Identifying Drug Targets Using Sequence Data |
| SD2002-853 | Genepalette: Software for Genome Sequence Navigation and Analysis |
| SD2001-814 | The Equalizer: DNA Array Analysis Software for Data Normalization |
| SD2000-835 | Family Pairwise Search V1.0 Software |

Protein Analysis Tools

| | |
|------------|--|
| SD2008-183 | Sequencing Cyclic Peptides |
| SD2007-293 | Method for Computing the Number of Peptide Reconstructions of Tandem Mass Spectra |
| SD2007-112 | Spectra |
| SD2007-022 | New Tools for Protein Functional Site Characterization |
| SD2006-834 | Method for Identifying Drug Targets Using Sequence Data |
| SD2004-148 | PepNovo Software e-licenses are available in OnLine Licensing |
| SD2003-801 | Predicted Ligands for Binding Orphan G-protein Coupled Receptors and Method |
| SD2003-027 | Thereof |
| SD2002-817 | Protein Kinase Resource (PKR) Software e-licenses are available in OnLine Licensing |
| SD2002-033 | Method for Fast Atomic Density Evaluation |
| SD2001-816 | Fast Atomic Density Evaluator (FADE) and Pairwise Atomic Density Reverse |
| SD2001-815 | Engineering (PADRE) |
| SD2001-034 | Whole-Proteome Interaction Mining: A Computational Approach |
| SD2000-835 | CE Software Software e-licenses are available in OnLine Licensing |
| | Biostar: Molecular Biology Annotation and Dissemination Tool |
| | Method for Predicting Protein Binding from Primary Structure Data |
| | Family Pairwise Search V1.0 Software |

Databases and Data Mining Tools

| | |
|--------------|--|
| SD2007-022 | Method for Identifying Drug Targets Using Sequence Data |
| SD2004-827 | Euler V2.0 Software e-licenses are available in OnLine Licensing |
| SD2003-801 | Protein Kinase Resource (PKR) Software e-licenses are available in OnLine Licensing |
| SD2002-826 | CKAAPS DB: A Conserved-Key-Amino-Acid-Positions Database |
| SD2002-814 | Open MMS |
| SD2002-812 | WhyWhere |
| SD2002-802 & | The MEME/MAST System Motif Discovery and Research, Version 3.0 |
| SD2001-804 | Software e-licenses are available in OnLine Licensing |
| SD2001-815 | Biostar: Molecular Biology Annotation and Dissemination Tool |
| SD2001-813 | Property Object Model (POM) Data Model, Associated Loaders and Query Tools |
| SD2000-827 | Software for QWB Score Analysis |

OnLine Licensing - <http://invent.ucsd.edu/onlinelicensing/index.htm>

DNA Analysis Tools

Predictive Biomarkers for Metastatic Potential in Breast Cancer

SD2008-084

Background: Within her lifetime, one of every eight women will confront a diagnosis of breast cancer. While we acknowledge that the term “cancer” is a “catch-all” for a number of different diseases with varying etiologies and probable outcomes, we have limited ability to discriminate and prescribe appropriate treatment. One fundamental differentiator for treatment options is the likelihood of metastasis. An improved understanding of the prognosis may enable appropriate therapies, which will be critical to improving survival rates and quality of life for women and physicians confronting this disease.

Technology: Using copyright software and novel methods, UC San Diego inventors have defined a proprietary set of biomarkers that enhance the ability to predict the likelihood that a tumor will metastasize within five years. The method is based in the reality that cancer is a ‘disease of interactive pathways’ and the assumption that keys for understanding the disease are encoded in protein networks. It is anticipated that these diagnostic biomarkers will improve the intelligence about a patient’s likely outcomes and most appropriate therapeutic options.

Advantages: While there are biomarkers in clinical use, this technology is differentiated by:

- Its basis in genes causing versus. correlating with cancer.
- The ability to predict outcomes as opposed to taking a snapshot diagnosis.
- Improved accuracy in diagnosis of metastasis on patients from a large independent study (Wang et al.).
- Use of “Network” biomarkers which, unlike classical and microarray-based diagnostic systems, tie marker genes to known genes causative for disease.

Copyright, Patents Pending

References and Supplemental Information:

Related Case: <http://invent.ucsd.edu/technology/cases/2006/SD2006-131.htm>

Inventor Interests: <http://chianti.ucsd.edu/idekerlab/index.html>

News Release: http://www.jacobsschool.ucsd.edu/news_events/releases/release.sfe?id=528

A New Approach to Medical Diagnostics and Therapeutics

SD2007-320

Description: UC San Diego inventors have come up with an approach to patient selection and assessment of response to therapy that is very different from traditional pharmacogenomic approaches. This novel approach allows doctors to make various health prognosis tailored to individuals. It can also be used to estimate the life span, predisposition to diseases, and reaction to various chemotherapies in a variety of patient groupings. This invention would allow a company to profitably leverage the large sets of genome data soon to be available.

DNA Analysis Tools (continued)
Method for Identifying Drug Targets Using Sequence Data
SD2007-022

Horizontal gene transfer is defined as the movement of genetic material between phylogenetically unrelated organisms by mechanisms other than parent to progeny inheritance. Any biological advantage provided to the recipient organism by the transferred DNA creates selective pressure for its retention in the host genome. Horizontal transmission is now considered a major factor in the process of environmental adaptation, for both individual species and entire microbial populations. These adaptations can include acquisition of new metabolic competencies, allowing survival under extreme conditions, resistance to drugs or environmental pollutants, defenses against external predators, and resistance to internal parasites. Horizontal transfer has also been proposed to play a role in the emergence of novel human diseases, as well as determining their virulence. The dynamics of horizontal gene transfer in pathogenic organisms may prove particularly useful in identifying novel targets for therapeutic agents, predicting the phylogenetic specificity of these agents, and estimating the likelihood of resistance development.

Scientists at UC San Diego have discovered a computer algorithm for the identification of genes that have been acquired by lateral gene transfer, rather than normal inheritance. The user inputs a file of protein sequence data from each of several test organisms, and receives as output a ranked list of genes with statistical data describing their likelihood of horizontal transfer. Genes that have been laterally acquired by more than one member of a group of pathogens can be used for identifying potential drug targets that would disrupt survival of an entire group of pathogenic organisms sharing a particular habitat.

Advantages and Commercial Uses:

- Identification and ranking (as to likelihood of success) of new targets for drug discovery;
- Prediction of effectiveness (or ineffectiveness) of drug targets previously identified by other methods, in cases where a disease is caused by interaction of multiple organisms;
- Use genomic sequence data to identify pathogenic species with a high (or low) propensity for incorporating foreign DNA (through either greater uptake or less internal resistance to foreign DNA integration, for example a lack of restriction enzyme defenses).
-

Provisional application filed.

DNA Analysis Tools (continued)

**GenePalette:
Software for Genome
Sequence
Navigation
and Analysis**

SD2002-853

GenePalette is a powerful cross-platform and cross-species bioinformatics tool for genome sequence visualization and navigation. Written in Java, this program allows users on Mac, PC, or UNIX platforms to access genome sequence data quickly and easily through a unified interface. Users can download from NCBI's GenBank database large or small segments of genome sequence from a variety of organisms (e.g., yeast, human, fly, worm, mouse, plants), preserving the gene annotation that is associated with that sequence. Sequence elements of interest (transcription factor binding sites and other regulatory motifs, restriction enzyme sites, primer sequences, SNPs, microsatellites, etc.) can be searched for and identified in the loaded sequence, and then clearly visualized within a colorful graphical representation of gene organization and intron/exon structure. Among the many features that GenePalette provides are intuitive clickable user interface elements; customizable libraries of user-defined sequence features; the ability to export both graphical and nucleotide views of genomic sequence data in GIF format; and a restriction enzyme site library containing more than 200 sites.

See <http://www.genepalette.org>

**The Equalizer:
DNA Array Analysis
Software for Data
Normalization**

SD2001-814

The Equalizer is a custom application which uses rank-order similarity in gene expression intensity to construct non-linear equalization vectors which are then used to linerize the data matrix to a consistent slope of 1. In addition, The Equalizer rationally eliminates "negative" values of expression seen in Affymetrix data. Use of The Equalizer greatly facilitates downstream data analysis.

This software has been used for Affymetrix chip analysis but has applications to other microarray systems as well (see Robert O. Stuart, Kevin T. Bush, and Sanjay K. Nigam. 2001. Changes in global gene expression patterns during development and maturation of the rat kidney. PNAS. 98:5649-54).

Visit Web Site: <http://organogenesis.ucsd.edu/software.html> for more information.



San Diego is home to the University of California, San Diego—one of the nation's most accomplished research universities, widely acknowledged for its local impact, national influence, and global reach. The region's third largest employer, UC San Diego is a significant economic engine for business growth: alumni and faculty have created at least 193 start-up companies, including many of our local biotech businesses that have helped to make San Diego home to one of the largest clusters of biotechnology companies in the world.

DNA Analysis Tools (continued)

Family Pairwise Search V1.0 Software

Description: Family Pairwise Search (FPS) is a method for scoring a single biological sequence against a family of sequences. FPS compares pairs of sequences and then combines the pairwise scores into an overall score for the match of the single sequence to the family of sequences.

SD2000-835

FPS operates in two modes. In the single-sequence mode, FPS compares a single query sequence to a library of sequence families. The query consists of a single sequence, and the target library consists of families of sequences. The result is a family classification for the query sequence. In family-query mode, FPS compares a query set of related sequences to a database of single sequences. In this mode, FPS outputs a set of sequences from the database that are related to the query family.

Further technical information on FPS may be found at the San Diego Supercomputer Center website address, fps.sdsc.edu.



UC San Diego boasts top-rated programs at the School of Medicine, Jacobs School of Engineering, the Division of Biological Sciences, the Division of Physical Sciences, Scripps Institution of Oceanography, the School of Pharmacy and Pharmaceutical Sciences, the Rady School of Management, and the UC San Diego Medical Center.

The UC San Diego Technology Transfer Office (TTO) evaluates, patents, markets, and licenses university-developed technologies for commercial applications in the global market. In addition, the office's activities promote the economic development of the Greater Southern California region by out-licensing the fruits of UC San Diego research. The products of this research are also disseminated throughout the world. Companies from North America, Europe, Asia, and Australia license and commercialize UC San Diego technologies in their business ventures.

The Technology Transfer Office organizes the Pipeline Events Series which provides an educational outreach to the greater UC San Diego community on intellectual property issues. This series includes informative seminars, symposia, and workshops that foster innovation flow from the benchtop to the marketplace.

Protein Analysis Tools

Sequencing Cyclic Peptides

SD2008-183

Description: University inventors have invented two new methods for automatically sequencing cyclic peptides in a high throughput manner. Each computational algorithm, complemented with certain experimental results, will aid the identification of natural products of special interest in the drug discovery field

<http://ucsdnews.ucsd.edu/newsrel/science/03-08DrugDiscoveryBottleneck.asp>

Method for Computing the Number of Peptide Reconstructions of Tandem Mass Spectra

SD2007-293

Description: The invention allows for the determination of the statistical significance of a peptide reconstruction. The algorithm takes as input a spectrum and a scoring function, and returns the number of peptide reconstructions at each score value. This allows for the computation of the statistical significance of the identified peptides without need of a decoy database by constructing the partition function of the spectrum, i.e., the number of peptide reconstructions at each value of the scoring function.

The invention is useful to anyone interested in identifying peptides, protein, and post-translational modifications on Tandem Mass Spectrometry data from biological samples.

New Tools for Protein Functional Site Characterization

SD2007-112

Description: UC San Diego inventors have come up with a novel method of determining the properties of a functional site of a protein structure and determining its alignment with other such sites which are independent of the sequence order of the protein (returning equivalent sites of sub-sites). For example, a ligand binding site identified in one protein can now be identified in a large database of proteins returning equivalent sites or sub-sites (a sub-site would accommodate part of the ligand). This has application in the pharmaceutical industry for target identification and validation, lead discovery and optimization and ADME/Tox prediction (since the competing targets are now known).

The method is fast and tolerant of flexibility of the protein. The method has been implemented in a software prototype.

Protein Analysis Tools (continued)
Method for Identifying Drug Targets Using Sequence Data
SD2007-022

Description: Horizontal gene transfer is defined as the movement of genetic material between phylogenetically unrelated organisms by mechanisms other than parent to progeny inheritance. Any biological advantage provided to the recipient organism by the transferred DNA creates selective pressure for its retention in the host genome. Horizontal transmission is now considered a major factor in the process of environmental adaptation, for both individual species and entire microbial populations. These adaptations can include acquisition of new metabolic competencies, allowing survival under extreme conditions, resistance to drugs or environmental pollutants, defenses against external predators, and resistance to internal parasites. Horizontal transfer has also been proposed to play a role in the emergence of novel human diseases, as well as determining their virulence. The dynamics of horizontal gene transfer in pathogenic organisms may prove particularly useful in identifying novel targets for therapeutic agents, predicting the phylogenetic specificity of these agents, and estimating the likelihood of resistance development.

Scientists at UC San Diego have discovered a computer algorithm for the identification of genes that have been acquired by lateral gene transfer, rather than normal inheritance. The user inputs a file of protein sequence data from each of several test organisms, and receives as output a ranked list of genes with statistical data describing their likelihood of horizontal transfer. Genes that have been laterally acquired by more than one member of a group of pathogens can be used for identifying potential drug targets that would disrupt survival of an entire group of pathogenic organisms sharing a particular habitat.

Advantages and Commercial Uses:

- Identification and ranking (as to likelihood of success) of new targets for drug discovery.
- Prediction of effectiveness (or ineffectiveness) of drug targets previously identified by other methods, in cases where a disease is caused by interaction of multiple organisms.
- Use genomic sequence data to identify pathogenic species with a high (or low) propensity for incorporating foreign DNA (through either greater uptake or less internal resistance to foreign DNA integration, for example a lack of restriction enzyme defenses).

Provisional application filed.

PepNovo
SD2006-834

Description: PepNovo performs de Novo sequencing and tag generation for mass spectra. The software in this site implements novel algorithms for MS/MS interpretation. A description of the PepNovo de Novo sequencing algorithm can be found at: Frank, A. and Pevzner, P. "PepNovo: De Novo Peptide Sequencing via Probabilistic Network Modeling", *Analytical Chemistry* 77:964-973, 2005.

For more information see http://peptide.ucsd.edu/pepnovo_help.py

Protein Analysis Tools (continued)

Predicted Ligands for Binding Orphan G-protein Coupled Receptors and Method Thereof

SD2004-148

Summary: This technology consists of two parts:

1. New small-molecule ligands for 113 “orphan” G-protein-coupled receptors (GPCRs). None of these ligands were known previously to have activity at these receptors. The ligands were identified using a public domain chemical database in conjunction with the second part of this invention:
2. A trainable software system for data mining, which can be used to find any kind of ligand (protein, nucleic acid or small-molecule) for receptors of interest. This software can be used with virtually any database, including public-domain databases. It has already been shown to accurately predict protein-protein and protein-DNA binding across entire proteomes (see SD2002-033), and now has been improved to predict protein binding to small molecules.

Potential Commercial Applications:

1. GPCRs are one of the most widely screened classes of drug targets because of the roles they play in numerous diseases, and because of their excellent track record for therapeutic success, due to their availability at the cell surface. Drugs targeting GPCRs comprise 50 to 60 percent of currently marketed drugs, including 30 percent of the top-selling 100 drugs. Of the ~750 human GPCR genes, about 160 have not yet been functionally characterized nor have ligands been found for them. These “orphan” receptors therefore represent an as-yet-untapped source of new chemical entities (NCEs) for diseases involving GPCRs. This invention provides ligands for 113 of these orphan receptors.
2. This invention also provides a novel software tool for data mining, for small molecules as well as for proteins and nucleic acids, and can be used to ID ligands for virtually any protein target, including GPCRs. It is a turn-key system with a user-friendly interface. The investigators are willing to test any company’s data and compare the results with known data, to verify its predictive power.

This technology is available for licensing. Further information can be obtained under a confidentiality agreement.



- UC San Diego’s research engine generated over 300 innovations in each of the last 6 fiscal years.
- 1400+ technologies are available for licensing
- UC San Diego is credited with the formation of over 200 startups, of these, to date over 100 start-up companies were formed with licensed university innovations.
- Over 230 commercial products have been introduced to the market by licensed UC San Diego innovations.

Protein Analysis Tools (continued)

Protein Kinase Resource (PKR)

SD2003-801

The Protein Kinase Resource (PKR) aims to become a web accessible compendium of information on the protein kinase family of enzymes. This resource includes tools for structural and computational analyses as well as links to related information maintained by others.

Software e-licenses are available: <http://pkr.sdsc.edu/>

Method for Fast Atomic Density Evaluation

SD2003-027

University researchers have invented a powerful method for the molecular modeling of proteins. This invention can rapidly calculate the distribution of atomic neighbors. A primary advantage of this atomic density method is its computational efficiency, especially over previous generation methods that use Fast Fourier Transforms. Software based on this algorithm can analyze molecular shape in seconds, while other methods may take minutes or hours.

This algorithm can be used to deduce surface shape features, such as crevices and protrusions. It can also do detailed analysis of shape complementarity for docked complexes. The ability to determine regions of strong shape match or mismatch in an interface is very useful to computer-aided drug design. In addition to research, atomic density methods offer an ideal tool for learning about the shape features of molecules. The basic ideas underlying density methods can be understood intuitively, and integration within existing packages for molecular visualization would be a great aid to the study of protein structure-function relationships.

Fast Atomic Density Evaluator (FADE) and Pairwise Atomic Density Reverse Engineering (PADRE)

SD2002-817

The Fast Atomic Density Evaluation (FADE) and Pairwise Atomic Density Reverse Engineering (PADRE) programs deduce molecular shape using the local density of atoms at points within a few Angstroms of the molecular surface. FADE uses Fast Fourier Transforms and convolution integrals to rapidly calculate the distribution of atomic neighbors. PADRE poses the question of atomic density as an inverse problem based on a one-dimensional integral of Lennard-Jones potentials. A primary advantage of atomic density methods is their computational efficiency. FADE can analyze molecular shape in seconds, while other methods may take minutes or hours.

FADE and PADRE can deduce surface shape features, such as crevices and protrusions. FADE is also able to do detailed analysis of shape complementarity for docked complexes. The ability to determine regions of strong shape match or mismatch in an interface is very useful to computer-aided drug design. In addition to research, atomic density methods offer an ideal tool for learning about the shape features of molecules. The basic ideas underlying density methods can be understood intuitively, and integration within existing packages for molecular visualization would be a great aid to students studying protein structure-function relationships.

See <http://www.sdsc.edu/CCMS/FP/> for more information.

Protein Analysis Tools (continued)

Whole-Proteome Interaction Mining: A Computational Approach

SD2002-033

Summary: UC San Diego researchers have developed a novel algorithm for the prediction of the positive interactions of proteins with other proteins, nucleic acids, small molecules and biopolymers. The system is first trained to recognize patterns that characterize positive interactions of proteins within a proteome, based on primary protein sequences and associated physiochemical information available in databases. Known molecular pathways and structural information relating to chemical interactions are then used to make predictions of the abilities of individual proteins to interact with other molecules in a pair-wise fashion (Whole-proteome interaction mining). The accuracy of these predictions in yeast is over 80 percent.

Recently, the method has been substantially improved such that the algorithm can make these predictions across proteomes, with a success rate higher than is available with other methods. This characteristic enables one to make protein interaction predictions for proteomes for which little experimental data are available.

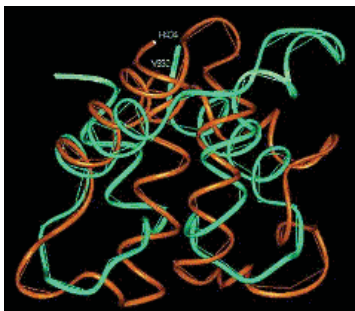
Potential Commercial Applications: This software is useful for virtual screening and validation of drug candidates that target proteins or protein/DNA interactions.

One can also use the algorithm to predict protein epitopes of functional domains, and binding sites for designer antibodies and biocompatible polymers, as well as a tool for basic research, such as for cell signaling pathways.

This technology is available for licensing or sponsored research. More information can be obtained under a Confidentiality Agreement.

CE Software

SD2001-816



Description: University inventors have implemented a new algorithm which builds an alignment between two protein structures. The algorithm involves a combinatorial extension (CE) of an alignment path defined by aligned fragment pairs (AFPs) rather than the more conventional techniques using dynamic programming and Monte Carlo optimization. AFPs, as the name suggests, are pairs of fragments, one from each protein, which confer structure similarity. AFPs are based on local geometry, rather than global features such as orientation of secondary structures and overall topology. Combinations of AFPs that represent possible continuous alignment paths are selectively extended or discarded thereby leading to a single optimal alignment. The algorithm is fast and accurate in finding an optimal structure alignment and hence suitable for database scanning and detailed analysis of large protein families. The method has been tested and compared with results from Dali and VAST using a representative sample of similar structures.

Software e-licenses are available: CE Site License: Send an e-mail of interest to Invent@ucsd.edu if you are interested in the software at this link: www.sdsc.edu/pb/Software.html

Protein Analysis Tools (continued)
Biostar: Molecular Biology Annotation and Dissemination Tool
SD2001-815

Description: Biostar is a software annotation tool that combines sequence, structure, function and experimental data into an integrated whole, using diverse presentation forms such graphics, images, figures, sequence alignments, and animation. This unique tool combines database lookup and traditional literature-style description to provide an enhanced learning and dissemination approach.

Method for Predicting Protein Binding from Primary Structure Data
SD2001-034

Description: An ambitious goal of proteomics is to elucidate the structure, interactions, and functions of all proteins within cells and organisms. The expectation is that this will provide a fuller appreciation of cellular processes and networks at the protein level, ultimately leading to a better understanding of disease mechanisms and suggesting new means for intervention. This technology provides a means to predict protein-protein interactions directly from primary structure and associated data. The system consistently predicts protein interactions at an accuracy rate above 80 percent.

The methodology may have applications in cell-map proteomics, genetic regulatory circuit modeling, epitope discovery, target-based drug design, cellular signaling and in blood-contacting biomaterials design. Future proteomics studies may benefit from this technology by proceeding directly from the automated identification of a cell's gene products to prediction of protein interaction pairs.

Family Pairwise Search V1.0 Software
SD2000-835

Description: Family Pairwise Search (FPS) is a method for scoring a single biological sequence against a family of sequences. FPS compares pairs of sequences and then combines the pairwise scores into an overall score for the match of the single sequence to the family of sequences.

FPS operates in two modes. In the single-sequence mode, FPS compares a single query sequence to a library of sequence families. The query consists of a single sequence, and the target library consists of families of sequences. The result is a family classification for the query sequence. In family-query mode, FPS compares a query set of related sequences to a database of single sequences. In this mode, FPS outputs a set of sequences from the database that are related to the query family. <http://fps.sdsc.edu>

Database and Data Mining Tools

Method for Identifying Drug Targets Using Sequence Data

SD2007-022

Description: Horizontal gene transfer is defined as the movement of genetic material between phylogenetically unrelated organisms by mechanisms other than parent to progeny inheritance. Any biological advantage provided to the recipient organism by the transferred DNA creates selective pressure for its retention in the host genome. Horizontal transmission is now considered a major factor in the process of environmental adaptation, for both individual species and entire microbial populations. These adaptations can include acquisition of new metabolic competencies, allowing survival under extreme conditions, resistance to drugs or environmental pollutants, defenses against external predators, and resistance to internal parasites. Horizontal transfer has also been proposed to play a role in the emergence of novel human diseases, as well as determining their virulence. The dynamics of horizontal gene transfer in pathogenic organisms may prove particularly useful in identifying novel targets for therapeutic agents, predicting the phylogenetic specificity of these agents, and estimating the likelihood of resistance development.

Scientists at UCSD have discovered a computer algorithm for the identification of genes that have been acquired by lateral gene transfer, rather than normal inheritance. The user inputs a file of protein sequence data from each of several test organisms, and receives as output a ranked list of genes with statistical data describing their likelihood of horizontal transfer. Genes that have been laterally acquired by more than one member of a group of pathogens can be used for identifying potential drug targets that would disrupt survival of an entire group of pathogenic organisms sharing a particular habitat.

Advantages and Commercial Uses:

- Identification and ranking (as to likelihood of success) of new targets for drug discovery.
- Prediction of effectiveness (or ineffectiveness) of drug targets previously identified by other methods, in cases where a disease is caused by interaction of multiple organisms.
- Use genomic sequence data to identify pathogenic species with a high (or low) propensity for incorporating foreign DNA (through either greater uptake or less internal resistance to foreign DNA integration, for example a lack of restriction enzyme defenses).

Patent Application: Provisional application filed

Technology Transfer is Beneficial



UC San Diego Technology Transfer Office (TTO), established in 1994, promotes and facilitates the transfer of UC San Diego innovations for the benefit of the University community and the public. TTO manages the intellectual property developed at UC San Diego by filing patents to protect intellectual property rights, marketing and licensing available technologies to businesses – locally and around the globe, and assisting with the formation of new start-up companies. In addition, TTO provides educational and information services on intellectual property matters to the UC San Diego community. <http://invent.ucsd.edu>

Database and Data Mining Tools (continued)

Euler V2.0

SD2004-827

This approach abandons the classical “overlap-layout -consensus” approach in favor of a new graph approach that, for the first time, resolves the problem of repeats in fragment assembly. The graph approach, in contrast to the Celera assembler, does not mask repeats but uses them instead as a powerful fragment assembly tool. The software also works with 454 contigs.

For general information about the EULER project see Pevzner, et al, PNAS, 98, 2001 and <http://nbc.sdsu.edu/euler/>.

Software e-licenses are available:

Protein Kinase Resource (PKR)

SD2003-801



The Protein Kinase Resource (PKR) aims to become a web accessible compendium of information on the protein kinase family of enzymes. This resource includes tools for structural and computational analyses as well as links to related information maintained by others. Software e-licenses are available: <http://pkr.sdsu.edu/>

CKAAPS DB: A Conserved-Key-Amino-Acid-Positions Database

SD2002-826

A Protein Structural Alignment-Based Sequence Analysis that covers 1295 representative structures from FSSP and 1803 representative structures from CE. <http://ckaap.sdsu.edu/>

Open MMS

SD2002-814

Description: Researchers at UC San Diego have produced bioinformatics middleware for working with protein and nucleic acid macromolecular structure data stored in the new standard mmCIF format now being supported by the Protein Data Bank. The software will be very useful to researchers in basic and applied research in biology, biochemistry, pharmacology, medical research, molecular biology, and protein chemistry. The toolkit contains software for parsing mmCIF files, loading the molecular structure data into a relational database, translating the data into XML formatted files, and running an OMG standard LSR/MMS CORBA server. The software is written entirely in Java.

Database and Data Mining Tools (continued)
WhyWhere

Description: Researchers at UC San Diego have come up with efficient software for data mining and statistical modeling of biodiversity data. This new technique for interrogating large datasets is scalable, parallelizable, and easily adaptable to other fields of use.

SD2002-812

See also http://biodi.sdsc.edu/ww_home.html and http://biodi.sdsc.edu/Visual/ww_poster.pdf

The MEME/MAST System Motif Discovery and Research Version 3.0

MEME is a software tool useful for discovering highly conserved regions or motifs in a group of related DNA or protein sequences. Version 3.0 differs from version 2.2 (formerly Case # SDC98-010) in several ways, including improved sensitivity, E-values and improved output format. www.sdsc.edu/Software (Click on “Molecular Biology Software”.)

SD2002-802

A software tool useful for searching biological sequence databases for sequences that contain one or more of a group of known motifs. MAST takes as input a file containing the descriptions of one or more motifs. A file from MEME can be used as input into MAST. Version 3.0 differs from version 2.2 (formerly Case # SDC98-011) in several ways, including:

- wCombining DNA strands: MAST now combines the score of a site in a DNA sequence with the score of the corresponding site on the reverse-complement strand. (The final score for the site is the maximum of the two scores.)
- wScoring DNA strands separately: MAST still allows each strand of a DNA sequence to be treated as a separate sequence at the user’s request.
- wIgnoring reverse-complement DNA: MAST also allows the user to score only the given DNA strand, not scoring the reverse complement strand at all.
- wBackground model: MAST now allows the user to specify the background residue frequencies used for computing E-values of scores.
- wComposition-adjusted statistics: MAST can now use a different random model for each target sequence, based on the letter composition of that sequence. This can greatly reduce erroneous matches due to biased sequence composition. <http://www.sdsc.edu/Software/>. (Click on “Molecular Biology Software.”)

Software e-licenses are available: <http://invent.ucsd.edu/technology/cases/2002/SD2002-802.htm>

Faculty Winners of the Nobel Prize

Currently, we have nine UC San Diego faculty members who are recipients of the Nobel Prize.

| Year Won | Recipient | Department |
|----------|-----------------|-------------------------------------|
| 2008 | Roger Tsien | Pharmacology |
| 2003 | Robert Engle | Economics |
| 2003 | Clive Granger | Economics |
| 2002 | Sydney Brenner | School of Medicine |
| 1995 | Paul Crutzen | Scripps Institution of Oceanography |
| 1995 | Mario Molina | Chemistry and Biochemistry |
| 1990 | Harry Markowitz | Economics |
| 1975 | Renato Dulbecco | School of Medicine |
| 1974 | George Palade | School of Medicine |

Database and Data Mining Tools (continued)

Biostar: Molecular Biology Annotation and Dissemination Tool

SD2001-815

Biostar is a software annotation tool that combines sequence, structure, function and experimental data into an integrated whole, using diverse presentation forms such as graphics, images, figures, sequence alignments and animation. This unique tool combines database lookup and traditional literature-style description to provide an enhanced learning and dissemination approach.

Property Object Model (POM) Data Model, Associated Loaders and Query Tools

SD2001-813

POM is a data model and contains associated tools to instantiate, populate and query the data model. It is useful for describing the abstract properties of a system and enables very fast query of properties described using the data model. The software has been validated using protein data but has wider general applications.

Software for QWB Score Analysis

SD2000-827

The Quality of Well-being (QWB) Scale is a preference-weighted measure of individual general health status that may be summed to produce an average population health status. It combines a measure of Symptom/ Problem Complexes with three Scales of functioning (Mobility, Physical Activity and Social Activity) to produce an expression of Well-being on a scale of 0.0 (for death) to 1.0 (for asymptomatic full function). The QWB instrument uses interviewer-administered questions, with standardized follow-up probes to verify initial responses, for obtaining data. From these data, a single QWB score for an individual for a 6 day time period is computer-generated for analysis. These individual scores can then be combined into general or specific population scores.

Interviewer-administered questionnaires have been shown, overall, to provide a better correlation with actual state of health (determined by established procedures) than do self-administered questionnaires. QWB scores have been used, for example, for monitoring change in individual health status over time or evaluating the effect of a particular health treatment, policy or program on the health of a patient population.

The QWB package includes the questionnaire, an interviewer's manual, examples of training interviews and the software required for entry of QWB data into data files and the derivation of QWB Scale scores. The software is available as QWB-DOS Application, as well as SPSS-PC, SPSS-MAC and SPSS Windows programs for running on standard PC or MAC desktop computer.

**Mission:**

Promote and facilitate the transfer of UC San Diego innovations for the benefit of the University community and the public.

Jane Moores, PhD

Assistant Vice Chancellor, Technology Transfer
858-822-3144
jmoores@ucsd.edu

William J. Decker, PhD

Associate Director
858-822-5128
wjdecker@ucsd.edu**Life and Biomedical Sciences**

Donna Shaw, PhD, CLP

Assistant Director, Life Science Licensing
858-534-6086
donnashaw@ucsd.edu

Denise Lew, PhD

Senior Licensing Officer
858-822-5279
dlew@ucsd.edu

April Kilburn, PhD

Senior Licensing Officer
858-822-1858
akilburn@ucsd.edu

Graí Andreason, PhD, MBA

Senior Licensing Officer
858-822-5138
gandreason@ucsd.edu

Dominic Montisano, PhD

Senior Licensing Officer
858-822-3833
dfmontisano@ucsd.edu

Rose Murphy, MBA

Senior Licensing Officer
858-534-7125
rmmurphy@ucsd.edu**Engineering and Physical Sciences**

David Gibbons, PE, MBA

Assistant Director, Physical Science Licensing
858-534-0175
dgibbons@ucsd.edu

Victoria Cajipe, PhD

Senior Licensing Officer
858-822-2304
vbcajipe@ucsd.edu

Wendy Shih, PhD

Licensing Officer
858-822-2595
wendyshih@ucsd.edu