

Preface

Bioinformatics and computational biology are quickly advancing fields over the past decade, thanks to the Human Genome Project and the development of many high-throughput technologies. Biological and medical sciences have blended their traditional boundaries with many other disciplines, such as computer science, mathematics and statistics, physical sciences, engineering, etc., in becoming a more quantitative and precise science. Establishing a mathematical model in describing a life phenomenon and using statistical tools in inferring probable outcome of living systems or guiding future experiments are becoming common practice around the world. Nonetheless, despite such exciting progress of these fields, how our genome is organized, how our genes perform their pertinent functions, how diseases develop, etc., are all complex and hard questions that remain to be solved.

Therefore, we believe it is of great significance and importance to summarize and present some of the most recent research carried out in these fields to encourage and guide future research. The book *Interdisciplinary Research and Applications in Bioinformatics, Computational Biology, and Environmental Sciences* is created with this exact purpose.

This book is a collection of conference submissions to the 1st Annual International Conference on Computational and Systems Biology, Shanghai, China, October 2009. Among over 40 submissions, 27 manuscripts were accepted to appear in this book. These 27 chapters are contributed by world-leading researchers from a broad range of disciplines, such as biological and medical sciences, chemistry, mathematics, environmental sciences, etc. The topics discussed in the book are in bioinformatics, computational biology, environmental sciences, and their related interdisciplinary fields.

The chapters, including 26 research articles and a short review, are written to highlight the most recent advances and breakthroughs in these subject fields. In order to improve the readability of the book, each chapter includes an "Additional Reading" section and a "Key Terms and Definitions" section.

TARGETS AND OVERALL OBJECTIVES OF THE BOOK

This book aims to provide a platform of communication for researchers in the fields of bioinformatics and computational biology and to foster the continuing development of these interdisciplinary fields. We hope that this book may serve the research community in bioinformatics and computational biology by presenting some of the most exciting and frontier research work in these fields. The prospective audience of the book would be scientists, researchers, and students actively working in these fields and related disciplines. We hope by reading this book, the readers may get a sense of where these fields are and what limitations there may be in existing work. Interested readers may then be empowered to formulate new ideas for future research and push the fields forward by embarking on such innovative work.

ORGANIZATION OF THE BOOK

The 27 chapters of the book are divided into 7 sections:

- Method Development and Application in Bioinformatics
- Biological Networks and Systems Biology
- Computational Predictions of Drug Properties
- Medical Signal Processing and Analysis
- Computational Biology
- Structure and Modeling
- Problems and Solutions in Environmental Sciences

A broad range of topics in bioinformatics and computational biology are covered by the book, including protein functional classification by a novel software tool (Chapter 1), a novel method for microarray data integration (Chapter 2), a novel representation of genomes by chaos game theory (Chapter 3), protein fold classification and recognition (Chapters 4 and 5), DNA sequence similarity analysis (Chapter 6), a novel feature for microRNA representation and classification (Chapter 7), a combined experimental and bioinformatics study of lipid enzymes (Chapter 8), biological networks and pathways in systems biology and their computational modeling (Chapters 9, 10, and 11), drug properties and delivery (Chapters 12, 13, and 14), medical signal analysis (Chapter 15), standardization and modernization of traditional Chinese medicine diagnostic methods (Chapters 16 and 17), computational models for ecological system (Chapters 18 and 19), neural science (Chapter 20) and the evolution of proteins (Chapter 21), molecular modeling of bio-materials (Chapter 22), and a short review on the structures of bacteria hydrogenases (Chapter 23). The book also contains four chapters in the field of environmental sciences, covering several important topics ranging from plant adaptation to water shortage, plant water uptake model, China's energy risk analysis (Chapter 26), and novel methods for the recycling of electronic waste (Chapter 27).

A brief description of each of the chapters follows:

Section 1: Method Development and Application in Bioinformatics

Chapter 1: JFeature: A Java Package for Extracting Global Sequence Features from Proteins for Functional Classification, by Xin Chen and Hangyang Xu

Prediction of various functional properties of proteins has long been a central theme of bioinformatics in the post-genomic era. Statistical learning, in addition to analysis based on sequence similarity, was proven successful to detect complex sequence-function associations in many applications. JFeature is an integrated Java tool to facilitate extraction of global sequence features and preparation of example sets, in statistical learning studies of sequence-function relationships. With a user-friendly graphical interface, it computes the composition, distribution, transition and auto-correlation features from sequence. It also helps to assemble a negative example set based on the most-dissimilar principle. The Java package and supplementary documentations are available at <http://www.cls.zju.edu.cn/rlibs/software/jfeature.html>.

Chapter 2: Cross-Platform Microarray Data Integration Combining Meta-Analysis and Gene Set Enrichment Analysis, by Jian Yu, Jun Wu, Miaoxin Li, Yajun Yi, Yu Shyr, Yixue Li, and Lu Xie

Integrative analysis of microarray data has been proven as a more reliable approach to deciphering molecular mechanisms underlying biological studies. Traditional integration such as meta-analysis is

usually gene-centered. Recently, gene set enrichment analysis (GSEA) has been widely applied to bring gene-level interpretation to pathway-level. GSEA is an algorithm focusing on whether an a priori defined set of genes shows statistically significant differences between two biological states. However, GSEA does not support integrating multiple microarray datasets generated from different studies. To overcome this, the improved version of GSEA, ASSESS, is more applicable, after necessary modifications. By making proper combined use of meta-analysis, GSEA, and modified ASSESS, this chapter reports two workflow pipelines to extract consistent expression pattern change at pathway-level, from multiple microarray datasets generated by the same or different microarray production platforms, respectively. Such strategies amplify the advantage and overcome the disadvantage than if using each method individually, and may achieve a more comprehensive interpretation towards a biological theme based on an increased sample size. With further network analysis, it may also allow an overview of cross-talking pathways based on statistical integration of multiple gene expression studies. A web server where one of the pipelines is implemented is available at: <http://lifecenter.sgst.cn/mgsea/home.htm>.

Chapter 3: *Chaos Game Representation of Mitochondrial Genomes: Markov Chain Model Simulation and Vertebrate Phylogeny*, by Guo-Sheng Han, Zu-Guo Yu, Bo Li, Vo Anh, and Yi-Quan Li

The mitochondrial genomes have provided much information on the evolution of this organelle and have been used for phylogenetic reconstruction by various methods with or without sequence alignment. In this paper, we explore the mitochondrial genomes by means of the chaos game representation (CGR), a tool derived from the chaotic dynamical systems theory. If the DNA sequence is a random collection of bases, the CGR will be a uniformly filled square; on the other hand, any pattern visible in the CGR contains information on the DNA sequence. First we use the Markov chain models to simulate the CGR of mitochondrial genomes. Then we model the noise background in the genome sequences by a Markov chain. A simple correlation-related distance approach without sequence alignment based on the CGR of mitochondrial genomes is proposed to analyze the phylogeny of 64 selected vertebrates.

Chapter 4: *A Two-layer Learning Architecture for Multi-Class Protein Folds Classification*, by Ruofei Wang and Xieping Gao

Classification of protein folds plays a very important role in the protein structure discovery process, especially when traditional sequence alignment methods fail to yield convincing structural homologies. In this chapter, we have developed a two-layer learning architecture, named TLLA, for multi-class protein folds classification. In the first layer, OET-KNN (Optimized Evidence-Theoretic K Nearest Neighbors) is used as the component classifier to find the most probable K-folds of the query protein. In the second layer, we use support vector machine (SVM) to build the multi-class classifier just on the K-folds, generated in the first layer, rather than on all the 27 folds. For multi-feature combination, ensemble strategy based on voting is selected to give the final classification result. The standard percentage accuracy of our method at ~63% is achieved on the independent testing dataset, where most of the proteins have <25% sequence identity with those in the training dataset. The experimental evaluation based on a widely used benchmark dataset has shown that our approach outperforms the competing methods, implying our approach might become a useful vehicle in the literature.

Chapter 5: *Improving PSI-BLAST's Fold Recognition Performance through Combining Consensus Sequences and Support Vector Machine*, by Ren-Xiang Yan, Jing Liu, and Yi-Min Tao

Profile-profile alignment may be the most sensitive and useful computational resource for identifying remote homologies and recognizing protein folds. However, profile-profile alignment is usually much more complex and slower than sequence-sequence or profile-sequence alignment. The profile or PSSM (position-specific scoring matrix) can be used to represent the mutational variability at each sequence

position of a protein by using a vector of amino acid substitution frequencies and it is a much richer encoding of a protein sequence. Consensus sequence, which can be considered as a simplified profile, was used to improve sequence alignment accuracy in the early time. Recently, several studies were carried out to improve PSI-BLAST's fold recognition performance by using consensus sequence information. There are several ways to compute a consensus sequence. Based on these considerations, we propose a method that combines the information of different types of consensus sequences with the assistance of support vector machine learning in this chapter. Benchmark results suggest that our method can further improve PSI-BLAST's fold recognition performance.

Chapter 6: *A New Approach for DNA Sequence Similarity Analysis Based on Triplets of Nucleic Acid Bases*, by Dan Wei, Qingshan Jiang, and Sheng Li

Similarity analysis of DNA sequences is a fundamental research area in Bioinformatics. The characteristic distribution of L-tuple, which is the tuple of length L, reflects the valuable information contained in a biological sequence and thus may be used in DNA sequence similarity analysis. However, similarity analysis based on characteristic distribution of L-tuple is not effective for the comparison of highly conservative sequences. In this paper, a new similarity measurement approach based on Triplets of Nucleic Acid Bases (TNAB) is introduced for DNA sequence similarity analysis. The new approach characterizes both the content feature and position feature of a DNA sequence using the frequency and position of occurrence of TNAB in the sequence. The experimental results show that the approach based on TNAB is effective for analysing DNA sequence similarity.

Chapter 7: *MicroRNA Precursor Prediction Using SVM with RNA Pairing Continuity Feature*, by Huan Yang, Yan Wang, Trupti Joshi, Dong Xu, Shoupeng Yu, and Yanchun Liang

MicroRNAs (miRNAs) are endogenous single-stranded non-coding RNAs of ~22 nucleotides in length and they act as post-transcriptional regulators in bacteria, animals and plants. Almost all current methods for computational prediction of miRNAs use hairpin structure and minimum of free energy as characteristics to identify putative pre-miRNAs from a pool of candidates. We discovered a new effective feature named "basic-n-units" (BNU) to distinguish pre-miRNAs from pseudo ones. This feature describes pairing continuity of RNA secondary structure. Simulation results show that a classification method, called Triplet-SVM-classifier, achieved an accuracy of 97.24% when this BNU feature was used. This is a 3% increase caused solely by adding this new feature. We anticipate that this BNU feature may increase the accuracy for most classification methods.

Chapter 8: *Four Long-Chain Acyl-Coenzyme A Synthetase Genes That Might be Involved in the Biosynthesis of Lipids in Brassica Napus*, by Fuge Zhu, Xiaoli Tan, Juan Li, Mingyu Wei, and Lili Yu

Long chain acyl-coenzyme A synthetases (LACSs) activate free fatty acid to acyl-CoA thioesters, and play important roles in the biosynthesis and degradation of lipids. In this study, four cDNAs (Complementary DNA) encode long chain fatty acyl-CoA synthetase activity has been found in Brassica napus. Sequence analysis indicated that the four LACSs possessed typical molecular characteristics of LACS. Compared with low oil content varieties seed, the four genes are strongly expressed in high oil content varieties seeds at 35 days after pollination (DAP). The expression pattern suggested that the four LACSs might be involved in the biosynthesis of lipids and oil accumulation in rapeseed.

Section 2: Biological Networks and Systems Biology

Chapter 9: *Topological Analysis of Axon Guidance Network for Homo Sapiens*, by Xuning Chen and Weiping Zhu

Axonal outgrowth is usually guided by a variety of guidance factors, such as netrins, ephrins, slits and semaphorins, and is one of the critical steps for the proper formation of neural networks. However, how the signal molecules function and why some of these play more important roles than others in guiding the axonal directional outgrowth has not been fully understood. In this study, we try to solve the problem by using the complex network analysis method. The signal molecules and interactions are treated as the nodes and edges to construct the axon guidance network model for *Homo sapiens*. The data of the model are taken from the KEGG database, and an analysis workbench named Integrative Visual Analysis Tool for Biological Networks and Pathways (VisANT) is employed to analyze the topological properties, including the degree distribution and the top co-expressed genes of the axon guidance network. This study has just opened a window into understanding the mechanism of axon guidance.

Chapter 10: *Evaluation of Coupled Nuclear and Cytoplasmic p53 Dynamics*, by Tingzhe Sun, Meihong Cai, Jun Cui, and Pingping Shen

The tumor suppressor protein p53 predominantly serves as a sequence specific transcription factor that may be activated upon exposure to diverse stimuli. One potent death inducer, p53-upregulated mediator of apoptosis (PUMA), is transcriptionally induced by p53. Once released into the cytoplasm, PUMA can lead to the activation of Bcl-2 apoptotic network. The cytoplasmic proapoptotic roles of p53 have recently been discovered, and these findings have placed p53 into the chemical interaction network with Bcl-2 family members. PUMA can also relieve p53 from the sequestration of antiapoptotic members. Released p53 further enters the nucleus and induces PUMA expression. We proposed that this positive feedback loop could lead to bistability. Further sensitivity analysis suggested that the system which covers the interactions between p53 and BCL-2 family members is considerably sensitive to p53 production rate. Meanwhile, downstream network components are much more affected by certain parameters than upstream effectors. Therefore, this newly discovered positive feedback loop might play critical roles in apoptotic network.

Chapter 11: *Agonist Fluctuation Maintained Calcium Signaling in a Mesoscopic System*, by Lin Ji and Haiyan Wang

Signals in transduction cascades are widely exposed to stochastic influences. In this work, we investigate the effects of agonist release noises on calcium signaling. Besides the usually considered “amplitude noise”, the case of “frequency noise” is also discussed. Simulation results show that the transduction cascades may amplify these noises when its intensity is bigger than certain critical value. The amplified noise show constructive effect to maintain the calcium signaling in critical signal-free cases. Moreover, the signal is more sensitive to the “frequency noise” than to the “amplitude noise”. This suggests frequency fluctuations in signaling cascades may have greater influence than the amplitude ones, which is an important finding for signal transduction in complex pathways. Since biological systems are inherently stochastic, this work demonstrates how the calcium system takes advantage of the environmental fluctuations to maintain signaling, and therefore provide effective, sensitive signal communication.

Section 3: Computational Predictions of Drug Properties

Chapter 12: *Theoretical Study on the Antioxidant Activity of Alizarin, Purpurin, and Pseudopurpurin*, by Ruifa Jin, Hongzheng Bao, Yin Bai, and Xiuhua Li

Hydroxyanthraquinone derivatives are a large group of natural polyphenolic compounds found widely in plants. The cytotoxic activities of hydroxyanthraquinone derivatives have been demonstrated using cancer cell lines. The pharmacological effect can be explained by their antioxidant activity and

their inhibition of certain enzymes. There are two main kinds of mechanism, H-atom transfer and one-electron transfer, by which antioxidants can play their role. The structural and electronic properties of hydroxyanthraquinone derivatives, alizarin, purpurin, pseudopurpurin, and their radicals were investigated using density functional theory. It turned out that these three molecules appear to be good candidates for high antioxidant activity species, particularly for pseudopurpurin. Taking this system as an example, we present an efficient method for the investigation of antioxidant activity for such kind of hydroxyanthraquinone derivatives from theoretical point of view. With the current work, we hope to highlight the antioxidant activity of hydroxyanthraquinone derivatives and stimulate the interest for further studies and exploitation in pharmaceutical industry.

Chapter 13: *Human Oral Bioavailability Prediction of Four Kinds of Drugs*, by Aixia Yan, Zhi Wang, Jiaxuan Li, and Meng Meng

In the development of drugs intended for oral use, good drug absorption and appropriate drug delivery are very important. Now the predictions for drug absorption and oral bioavailability follow similar approach: calculate molecular descriptors for molecules and build the prediction models. This approach works well for the prediction of compounds which cross a cell membrane from a region of high concentration to one of low concentration, but it does not work very well for the prediction of oral bioavailability, which represents the percentage of an oral dose which is able to produce a pharmacological activity. The models for bioavailability had limited predictability because there are a variety of pharmacokinetic factors influencing human oral bioavailability. Recent study has shown that good quantitative relationship could be obtained for subsets of drugs, such as those that have similar structure or the same pharmacological activity, or those that exhibit similar absorption and metabolism mechanisms. In this work, using MLR (Multiple Linear Regression) and SVM (Support Vector Machine), quantitative bioavailability prediction models were built for four kinds of drugs, which are Angiotensin Converting Enzyme Inhibitors or Angiotensin, Receptor Antagonists, Calcium Channel Blockers, Sodium and Potassium Channels Blockers and Quinolone Antimicrobial Agents. Explorations into subsets of compounds were performed and reliable prediction models were built for these four kinds of drugs. This work represents an exploration in predicting human oral bioavailability and could be used in other dataset of compounds with the same pharmacological activity.

Chapter 14: *In Silico Prediction of Blood Brain Barrier Permeability: A Support Vector Machine Model*, by Zhi Wang, Aixia Yan, and Jiaxuan Li

The ability of penetration of the blood-brain barrier is an important property for the development of Central Nervous System drugs, which is commonly expressed by logBB ($\log BB = \log(C_{\text{brain}}/C_{\text{blood}})$). In this work, a support vector machine was used to build quantitative models of blood brain barrier permeability. Molecular descriptors for 182 compounds were calculated by ADRIANA.Code and 12 descriptors were selected using the automatic variable selection function in Weka. Based on two common physicochemical descriptors (xlogP and Topological Polar Surface Area (TPSA)) and 10 2D property autocorrelation descriptors on atom pair properties, an SVM regression model was built. The built model was validated by an external test set. The reliable predictions of the test set demonstrate that this model performs well and can be used for estimation of logBB values for drug and drug-like molecules.

Section 4: Medical Signal Processing and Analysis

Chapter 15: *The Study of Transesophageal Oxygen Saturation Monitoring*, by Zhiqiang Zhang, Bo Gao, Guojie Liao, Ling Mu, and Wei Wei

In this chapter, the transesophageal oxygen saturation (SpO₂) monitoring system was proposed based on the early experiments, to provide a new program of SpO₂ acquisition and analysis and avoid the limitation of traditional methods. The PPG (photoplethysmographic) signal of descending aorta and left ventricular was monitored in the experiment. The analysis of the peak-to-peak values, the standard deviation and the position of peaks in signal waveforms showed that in vivo signal was more stable and sensitive; and the physiological information was reflected in the left ventricular PPG waveform. Therefore, it can be concluded that the transesophageal SpO₂ monitoring technology has better guidance in clinical applications.

Chapter 16: *Digital Auscultation System of Traditional Chinese Medicine and Its Signals Acquisition-Analysis Methods*, by Fanpeng Zhou, Jianjun Yan, Yiqin Wang, Fufeng Li, Chunming Xia, Rui Guo, and Haixia Yan

Digital auscultation of Traditional Chinese Medicine (TCM) is a relatively new technology which has been developed for several years. This system makes diagnoses by analyzing sound signals of patients using signal processing and pattern recognition. The paper discusses TCM auscultation in both traditional and current digital auscultation methods. First, this article discusses demerits of traditional TCM auscultation methods. It is through these demerits that a conclusion is drawn that digital auscultation of TCM is indispensable. Then this article makes an introduction to voice analysis methods from linear and nonlinear analysis aspects to pattern recognition methods in common use. Finally this article establishes a new TCM digital auscultation system based on wavelet analysis and Back-propagation neural network (BPNN).

Chapter 17: *Pulse Wave Analysis of Traditional Chinese Medicine Based on Hemodynamics Principles*, by Rui Guo, Yiqin Wang, Haixia Yan, Fufeng Li, Jianjun Yan, and Zhaoxia Xu

From the perspective of hemodynamics principles, the pressure pulse wave marked in the radial artery is the comprehensive result of pulse wave propagation and reflection in the arterial conduit. The most common pulse charts (also called pulse wave) obtained by Traditional Chinese Medicine (TCM) pulse-taking technique, if quantified and standardized, may become a universal and valuable diagnostic tool. The methods of feature extraction of TCM pulse charts currently involve time-domain analysis, frequency-domain analysis and time-frequency joint analysis. The feature parameters extracted by these methods have no definite clinical significance. Therefore, these feature parameters cannot essentially differentiate different types of TCM pulse. In this chapter, the harmonic analysis method was applied to analyze the common TCM pulse charts (plain pulse, wiry pulse, slippery pulse). Velocity and reflectivity coefficients of pulse were calculated. We found that wave velocities and reflection coefficients of different TCM pulse have different distributions. Furthermore, we studied the clinical significance of velocities and reflection coefficients. The result suggests that wave velocity and reflection coefficient are the feature parameters of TCM pulse with physiological and pathological significance, which can be used to interpret formation of Chinese medicine pulse. Our study reveals the mechanism of TCM pulse formation and promotes non-invasive TCM pulse diagnostic method.

Section 5: Computational Biology

Chapter 18: *A New Mechanical Algorithm for Calculating the Amplitude Equation of the Reaction-Diffusion Systems*, by Houye Liu and Weiming Wang

Amplitude equation may be used to study pattern formation. In this chapter, we establish a new mechanical algorithm AE_Hopf for calculating the amplitude equation near Hopf bifurcation based

on the method of normal form approach in Maple. The normal form approach needs a large number of variables and intricate calculations. As a result, deriving the amplitude equation from diffusion-reaction is a difficult task. Making use of our mechanical algorithm, we derived the amplitude equations from several biology and physics models. The results indicate that the algorithm is easy to apply and effective. This algorithm may be useful for learning the dynamics of pattern formation of reaction-diffusion systems in future studies.

Chapter 19: *Pattern Formation Controlled by External Forcing in a Spatial Harvesting Predator-Prey Model*, by Feng Rao

Predator-prey models in ecology serve a variety of purposes, which range from illustrating a scientific concept to representing a complex natural phenomenon. Due to the complexity and variability of the environment, the dynamic behavior obtained from existing predator-prey models often deviates from reality. Many factors remain to be considered, such as external forcing, harvesting and so on. In this chapter, we study a spatial version of the Ivlev-type predator-prey model that includes reaction-diffusion, external periodic forcing, and constant harvesting rate on prey. Using this model, we study how external periodic forcing affects the stability of predator-prey coexistence equilibrium. The results of spatial pattern analysis of the Ivlev-type predator-prey model with zero-flux boundary conditions, based on the Euler method and via numerical simulations in MATLAB, show that the model generates rich dynamics. Our results reveal that modeling by reaction-diffusion equations with external periodic forcing and nonzero constant prey harvesting could be used to make general predictions regarding predator-prey equilibrium, which may be used to guide management practice, and to provide a basis for the development of statistical tools and testable hypotheses.

Chapter 20: *Repetitive Firing and Bursting due to Different Bifurcation Mechanism in Unmyelinated Fibre*, by Junran Zhang, Yongguo Han, Guangcan Xiao, and Sanjue Hu

In neural science, different action potential (AP) firing patterns are typically considered to be dominated by different dynamical mechanisms. Different AP firing patterns in unmyelinated fibres can contribute to pain and sensory information transmission. Experiments in rabbit unmyelinated nerve (axon) show some interesting phenomena, mainly concerned with the AP firing patterns that changed regularly. Investigating the dynamical mechanism of unmyelinated fibre during various kinds of AP firing patterns is useful to understand the neural information processing in axon and pain information transmission. Here, we reproduced these phenomena by constructing a mathematical model, where the discharge of the Hodgkin-Huxley (H-H) neuron under square wave stimulation was studied by simulation. It is shown that square wave can induce bursting firing, especially with long time-course duration. This is different from the popular theory that explained repetitive and bursting firing due to stimulus intensity and instantaneous fluctuation. Through dynamical analysis, we found that the mechanism of the action potential pattern changed according to Hopf bifurcation, a dynamical behavior that emergence and stability of limit cycles of bifurcating from a stable equilibrium. The finding may support the neural information coding hypothesis in unmyelinated axon.

Chapter 21: *The Mathematical Modeling and Computational Simulation for Error-Prone PCR*, by Lixin Luo, Fang Zhu, and Si Deng

Many enzymes have been widely used in industrial production, for they have higher catalytic efficiency and catalytic specificity than the traditional catalysts. Therefore, the performance of enzymes has attracted wide attention. However, due to various factors, enzymes often cannot show their greatest catalytic efficiency and the strongest catalytic ability in industrial production. In order to improve the enzyme activity and specificity, people become increasingly interested in the transformation and

modification of existing enzymes. For the structure modification of proteinase, this chapter introduces a computational method for modelling error-prone PCR. Error-prone PCR is a DNA replication process that intentionally introduces copying errors by imposing mutagenic reaction condition. We then conclude about the mathematical principle of error-prone PCR which may be applied to the quantitative analysis of directed evolution in future studies.

Section 6: Structure and Modeling

Chapter 22: *Molecular Dynamics Simulation of Interlayer Structure and Hydration Properties of Glycine Intercalated Layered Double Hydroxides*, by Guo-Xiang Pan, Feng Cao, Pei-Song Tang, Hai-Feng Chen, Zhe-Ming Ni, Jin-Tian Yang, Li-Geng Wang, and Min-Hong Xu

Interlayer structure, hydrogen-bond, hydration and swelling properties of glycine intercalated layered double hydroxides (LDHs-Gly) were investigated with molecular dynamics (MD) methods. The results show that the interlayer spacing d_c increases as hydration level increases. The computed hydration energies reach the most negative values at low water contents and change rapidly over the range $1 \leq NW \leq 6$, and slowly and gradually approach the potential energy for bulk SPC water at $NW > 6$. But there are no local minima in the energy over the entire hydration range. This result suggests that LDHs-Gly tend to absorb water continuously in water-rich environments and enhance swelling to delaminate the hydroxide layers. The interlayers of LDHs-Gly exhibit complex hydrogen-bond network. With water content increasing, the glycine molecules progressively change their orientation from parallel to the layers to nearly perpendicular. Water molecules firstly form hydrogen-bond with M-OH layers at low water contents. While the hydroxide layers gradually get to saturation state at $Nw > 3$. And then water molecules continuously fill the interlayer to expand interlayer spacing.

Chapter 23: *Structure of Hydrogenase in Biohydrogen Production Anaerobic Bacteria*, by Ming Du and Lu Zhang

Hydrogenase plays an important role in the process of biohydrogen production. Hydrogenases have very unique active sites and are classified into three groups according to the metal composition of the active sites: the [Ni-Fe] hydrogenase, [Fe-Fe] hydrogenase, and [Fe-only] hydrogenase. In this paper, the crystal structures and active sites of three kinds of hydrogenases are examined and compared. These enzymes have an unusual structural feature in common. Their similar active site indicates that the catalytic mechanism of hydrogen activation is probably similar. The understanding of the catalytic mechanisms for the three kinds of hydrogenases may help achieve the industrialization process of hydrogen energy production. Moreover, the future research direction about the hydrogenases from auto-aggregative bacteria and the chemical mimic of hydrogenases structure is discussed.

Section 7: Problems and Solutions in Environmental Sciences

Chapter 24: *Seasonal Trade-Off between Water- and Nitrogen-Use Efficiency of Constructive Plants in Desert Riparian Forest in Hyperarid Region of China*, by Shengkui Cao, Qi Feng, Jianhua Si, Yonghong Su, Zongqiang Chang, and Haiyang Xi

Foliar $\delta^{13}C$ values are often used to denote the long-term water use efficiency (WUE) of plants whereas long-term nitrogen use efficiency (NUE) are usually estimated by the ratio of C to N in the leaves. Seasonal variations of $\delta^{13}C$ values, foliar nitrogen concentration and C/N ratios of *Populus euphratica* and *Tamarix ramosissima* grown under five different microhabitats of Ejina desert riparian

oasis of northwestern arid regions in China were studied. The results indicated that *T. ramosissima* had higher $\delta^{13}\text{C}$ value compared with that of *P. euphratica*. The N concentration and C/N ratios of two species were not significantly different. The seasonal pattern of three indexes in two species was different. The $\delta^{13}\text{C}$ values and N concentration decreased during the plant's growth period. However, the change of C/N ratios was increased. Among microhabitats, there were higher $\delta^{13}\text{C}$ values and N concentration as well as lower C/N ratios in the Dune and Gobi habitats. Foliar $\delta^{13}\text{C}$ values significantly and positively correlated with N concentration in *P. euphratica* and *T. ramosissima*, whereas a significantly negative correlation between $\delta^{13}\text{C}$ values and C/N ratios was found for *P. euphratica*. This relation in *T. ramosissima* was weak, but there was a significant quadratic curve relationship between $\delta^{13}\text{C}$ values and C/N ratios, which revealed that there was a trade-off between WUE and NUE for *P. euphratica* and in natural condition, *P. euphratica* could not improve WUE and NUE simultaneously. *T. ramosissima* could simultaneously enhance WUE and NUE. The above characters of WUE and NUE in two plants reflected the different adaptations of desert species to environmental condition.

Chapter 25: *Root Water Uptake Model of Populus Euphratica in Desert Riparian Forest in Extreme Arid Region*, by Yongzheng Tian, Jianhua Si, Qi Feng, and Shengkui Cao

Plant root water uptake is a key way to transfer soil water to the atmosphere. It is an important part of the research on water transforming patterns in the SPAC (Soil-Plant-Air Continuum). So understanding the water absorption patterns of plant root system is a base to recognize the SPAC. Recently there are many studies on the water absorption patterns of plant root system. However, the researched plants are mostly crops and the main researched areas are regions with adequate precipitation. There are only a few studies on the water absorption of natural plants in extreme arid desert regions. This paper studied the root water absorption patterns of *Populus euphratica* and established the corresponding mathematical model based on the data of root density and soil water dynamics in root zone in desert riparian forest in extreme arid region. The finite difference method was used to discretize the soil water movement equation with evaporation boundary conditions. Numerical simulation analysis of soil water movement in root zone of *Populus euphratica* showed that the simulated values were consistent with the measurement values with 92-98% precision. This work provides a theoretical basis for the study of water movement in the SPAC.

Chapter 26: *The Fuzzy Integrated Energy Prior-warning Model Based on Entropy Weight*, by Yaqun He, Hua Wei, Weiran Zuo, Xiaobing Wu, Xin Ge, Shan Wu, and Baofeng Wen

Based on the analysis of five factors affecting energy risks, including supply and demand, economy, environment, transport, and disaster, this chapter establishes the prior-warning index system of the energy risk by covering 3 sub-systems and 37 indexes. The three sub-systems are the coal sub-system, the petroleum and natural gas sub-system, and the integrated factors sub-system. Fuzzy synthesis evaluation was applied to confirm the internal estimated index weight of the sub-systems. Moreover, the risk prior-warning model of the energy sub-systems was established by the method. The weights of the three sub-systems were determined through the concept of entropy weight, and the prior-warning indexes of energy risk were applied to evaluate the total energy security of the three sub-systems. Finally, the prior-warning model of energy risk in China was established. The entire situation of energy safety of China was summarized via empirical analysis.

Chapter 27: *A Novel Flowsheet for the Recycling of Valuable Constituents from Waste Printed Circuit Boards*, by Jingfeng He, Yaqun He, Nianxin Zhou, Chenlong Duan, Shuai Wang, and Hongjian Zhang

Waste printed circuit boards (PCBs) contain a number of valuable constituents. It is of great significance to separate precious metals and non-metallic constituents from waste PCBs with appropriate methods for resource recycling and environment protection. A novel flowsheet for the recycling of waste PCBs using

physical beneficiation methods was constructed. Waste PCBs were disassembled into substrates and slots firstly. The substrates were crushed to the size below 1mm through wet impact crushing and separated with a tapered column separation bed. The results indicated that products with integrated separation efficiency of 93.9% and metal recovery ratio of 93.7% were obtained by the primary separation with the water discharge of 5.5 m³/h, feed-rate of 250g/min and inclination angle of 35°. Waste PCBs slots components were crushed to the size of 0.5-5mm through impact crushing and separated with an active pulsing air classifier. The separation results showed that products with integration separation efficiency of 92.4% and metal recovery ratio of 96.2% were obtained with the airflow velocity of 2.90m/s and pulsing frequency of 2.33Hz. Precious metals could be obtained by further separation and purification of the metal components and the non-metal components could be used as refuse derived fuel. The flowsheet has great potential to be applied in the field of waste PCBs treatment and recycling.

Editors:

Limin Angela Liu, Shanghai Jiao Tong University, China

Dong-Qing Wei, Shanghai Jiao Tong University, China

Yixue Li, Shanghai Center for Bioinformation Technology, China

Editorial Assistant:

Huimin Lei, Shanghai Jiao Tong University, China

May, 2010

INTRODUCTION

In the last decade, genomic projects have been generating a large amount of sequence data at an exponential rate. By the year 2009, more than 50 species have been sequenced, whereas the characterization of newly discovered genes is largely lacking. Computational approaches are used to predict function or functional class of many new genes, yet are still considered as preliminary.

Sequence class or k-mer based approaches, in the form of global alignment, local alignment or hidden Markov models, are extensively used to infer the putative functions for new genes. Recent studies show that statistical learning approaches could be an effective alternative and/or supplement, especially in the case of predicting function of domain-related proteins or domainless proteins with different lengths (Cai et al., 2004). Few successful applications have also been reported in predicting protein functional class (Kusner et al., 2002) protein-protein interactions (Bock &