

Proceedings

**Sixth IEEE Symposium on
BioInformatics and BioEngineering
BIBE 2006**

**16-18 October 2006
Arlington, Virginia**

Co-Sponsored by
IEEE Computer Society
Biological and Artificial Intelligence Society

In cooperation with
ITRI at Wright State University



Los Alamitos, California
Washington • Tokyo

Copyright © 2006 by The Institute of Electrical and Electronics Engineers, Inc.

All rights reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries may photocopy beyond the limits of US copyright law, for private use of patrons, those articles in this volume that carry a code at the bottom of the first page, provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Other copying, reprint, or republication requests should be addressed to: IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, P.O. Box 133, Piscataway, NJ 08855-1331.

The papers in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and, in the interests of timely dissemination, are published as presented and without change. Their inclusion in this publication does not necessarily constitute endorsement by the editors, the IEEE Computer Society, or the Institute of Electrical and Electronics Engineers, Inc.

IEEE Computer Society Order Number P2727

ISBN 0-7695-2727-2

ISBN-13 978-0-7695-2727-7

Library of Congress 2006933280

Additional copies may be ordered from:

IEEE Computer Society
Customer Service Center
10662 Los Vaqueros Circle
P.O. Box 3014
Los Alamitos, CA 90720-1314
Tel: + 1 800 272 6657
Fax: + 1 714 821 4641
<http://computer.org/cspress>
csbooks@computer.org

IEEE Service Center
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
Tel: + 1 732 981 0060
Fax: + 1 732 981 9667
[http://shop.ieee.org/store/
customer-service@ieee.org](http://shop.ieee.org/store/customer-service@ieee.org)

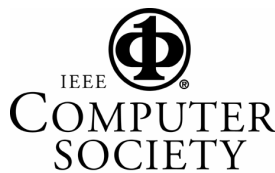
IEEE Computer Society
Asia/Pacific Office
Watanabe Bldg., 1-4-2
Minami-Aoyama
Minato-ku, Tokyo 107-0062
JAPAN
Tel: + 81 3 3408 3118
Fax: + 81 3 3408 3553
tokyo.ofc@computer.org

Individual paper REPRINTS may be ordered at: reprints@computer.org

Editorial production by Silvia Ceballos

Cover art production by Joseph Daigle/Studio Productions

Printed in the United States of America by Commercial Communications, Inc.



IEEE Computer Society

Conference Publishing Services

<http://www.computer.org/proceedings/>

**Sixth IEEE Symposium on
BioInformatics and BioEngineering
BIBE 2006**

TABLE OF CONTENTS

Message from the Chairs ix
BIBE-06 Program Committees x

Session 1 : Protein Structure, Function, and Classification

IUP : Intrinsically Unstructured Protein Predictor - A Software Tool for Analyzing
Polypeptide Sequences 3
Jack Yang and Mary Yang

A Language Modeling Text Mining Approach to the Annotation of Protein
Community 12
Xiaodan Zhang, Daniel Wu, Xiaobua Zhou, and Xiaobua Hu

FIK Model: Novel Efficient Granular Computing Model for Protein Sequence Motifs
and Structure Information Discovery 20
Bernard Chen, Phang C. Tai, Robert Harrison, and Yi Pan

On Integrating Peptide Sequence Analysis and Relational Distance-Based Indexing 27
Weijia Xu, Rui Mao, Shu Wang, and Daniel Miranker

Assessing Protein Function Using a Combination of Supervised and Unsupervised Learning 35
Jack Yang, Mary Yang, and Okan K. Ersoy

Session 2 : Pattern Discovery and Data Mining

Partition Frequency Distance Based Filter Method for Finding Approximate
Repetitions in DNA Sequences 45
*Di Wang, Guoren Wang, Baichen Chen, Qingquan Wu, Changyong Yu,
Yi Zhao, and Ge Yu*

Finding Patterns in Biological Sequences by Longest Common Subsequences
and Shortest Common Supersequences 53
Kang Ning, Hoong Kee Ng, and Hon Wai Leong

Mining the Database of Transcription Binding Sites 61
Wei Peng, Tao Li, and Giri Narasimhan

Extending Pattern Branching to Handle Challenging Instances 65
Jaime Davila and Sanguthevar Rajasekaran

Discovering DNA Motifs with Nucleotide Dependency 70
Francis Chin and Henry Leung

Session 3: Microarrays

Exploratory Tools for Follow-Up Studies to Microarray Experiments.....	81
<i>Kausubik Sinha, Ruoming Jin, Gagan Agrawal, and Helen Piontekivska</i>	
Mining Positive and Negative Co-regulation Patterns from Microarray Data.....	86
<i>Yubai Zhao, Guoren Wang, Ying Yin, and Ge Yu</i>	
Building a Classifier for Integrated Microarray Datasets through Two-Stage Approach	94
<i>Youngmi Yoon, Jongchan Lee, and Sanghyun Park</i>	
IntClust: A Software Package for Clustering Replicated Microarray Data.....	103
<i>Wei Peng and Tao Li</i>	
Quick Hierarchical Biclustering on Microarray Gene Expression Data	110
<i>Liping Ji, Kenneth Mock, and Kian-Lee Tan</i>	

Session 4: Sequence Alignment and Database Search

An Improved Algorithm for the Regular Expression Constrained Multiple Sequence Alignment Problem	121
<i>Abdullah Arslan and Dan He</i>	
Large Grain Size Stochastic Optimization Alignment.....	127
<i>Perry Ridge, Hyrum Carroll, Dan Sneddon, Mark Clement, and Quinn Snell</i>	
Sequence Homology Search Based on Database Indexing Using the Profile Hidden Markov Model	135
<i>Sakti Pramanik, Qiang Xue, and James Cole</i>	
A Tool for Supporting Integration across Multiple Flat-File Datasets	141
<i>Xuan Zhang and Gagan Agrawal</i>	
Simplicity in RNA Secondary Structure Alignment: Towards Biologically Plausible Alignments	149
<i>Rimon Mikhael, Guobui Lin, and Eleni Stroulia</i>	

Session 5: Phylogenies, Trees, and Visualization

Improving Progressive Alignment for Phylogeny Reconstruction Using Parsimonious Guide-Trees.....	159
<i>Usman Roshan, Dennis Livesay, and Satish Chikkagoudar</i>	
Learning the Tree of Phenotypes Using Genomic Data and VISDA.....	165
<i>Yuanjian Feng, Zuyi Wang, Yitan Zhu, Jianhua Xuan, David Miller, Robert Clarke, Eric Hoffman, and Yue Wang</i>	
Visualization of DNA/RNA Structure Using Temporal CGRs	171
<i>Margaret Dunham, Donya Quick, Yubang Wang, Monnie McGee, and Jim Waddle</i>	
Isolation and Visualization of Codon Usage Biases	179
<i>Douglas Raiford, Dan Krane, Travis Doom, and Michael Raymer</i>	

Finding Correlations in Functionally Equivalent Proteins by Integrating Automated and Visual Data Exploration.....	183
<i>Daniel A. Keim, Daniela Oelke, Royal Truman, and Klaus Neubauss</i>	

Evidence of Multiple Maximum Likelihood Points for a Phylogenetic Tree	193
<i>B-B Zhou, M. Tarawneh, D. Chu, P. Wang, C. Wang, A. Zomaya, and R. Brent</i>	

Session 6: DNA & RNA Sequence and Structure

A New Kernel Method for RNA Classification	201
<i>Xiaoming Wu, Jason T.L. Wang, and Katherine G. Herbert</i>	

RADAR: An Interactive Web-Based Toolkit for RNA Data Analysis and Research.....	209
<i>Mugdha Khaladkar, Vivian Bellofatto, Jason T. L. Wang, Bin Tian, and Kaizhong Zhang</i>	

Parsimony Based Approach to Test the Evolving Master Gene Hypothesis for Human <i>Alu</i> Repeats	213
<i>Sridhar Ramachandran, Travis Doom, Michael Raymer, and Dan Krane</i>	

An Algorithm for String Edit Distance Allowing Substring Reversals	220
<i>Abdullah Arslan</i>	

Unsupervised Gene Selection for High Dimensional Data	227
<i>Young Bun Kim and Jean Gao</i>	

Session 7: Clustering, Similarity Metrics, and Near Neighbor Methods

Using Gene Clustering to Identify Discriminatory Genes with Higher Classification Accuracy	235
<i>Zhipeng Cai, Lizhe Xu, Yi Shi, Mohammad Salavatipour, Randy Goebel, and Guohui Lin</i>	

Semi-Supervised Clustering Models for Clinical Risk Assessment.....	243
<i>Yongyang Huo, Francisco Azuaje, Paul McCullagh, and Roy Harper</i>	

Prediction of Protein Function Using Common-Neighbors in Protein-Protein Interaction Networks	251
<i>Chuan Lin, Daxin Jiang, and Aidong Zhang</i>	

Determining Molecular Similarity for Drug Discovery Using the Wavelet Riemannian Metric.....	261
<i>Elinor Velasquez, Emmanuel R. Yera, and Rabul Singh</i>	

Assigning Schema Labels Using Ontology and Hueristics	269
<i>Xuan Zhang, Ruoming Jin, and Gagan Agrawal</i>	

Session 8: Regulatory and Metabolic Networks and Bioengineering I

Optimal Multi-objective Control Method for Discrete Genetic Regulatory Networks	281
<i>Osman Abul, Reda Albajj, and Faruk Polat</i>	
Methods for Random Modularization of Biological Networks	285
<i>Zachary Saul and Vladimir Filkov</i>	
Efficient Modularization of Weighted Protein Interaction Networks Using k -Hop Graph Reduction	289
<i>Young-Rae Cho, Woosung Hwang, and Aidong Zhang</i>	
A Computational Inference Framework for Analyzing Gene Regulation Pathway Using Microarray Data	299
<i>Dong-Guk Shin, John Bluis, Yoo-Ah Kim, Winfried Krueger, Ravi Nori, Nathan Viniconis, Hsin-Wei Wang, Alan Wong, and David Rowe</i>	
The Automation of SIBIOS Workflow Composition	307
<i>Xiang Xiao, Malika Maboni, Zina Ben Miled, and Bhavna Choudhury</i>	

Session 9: Bioengineering II

Using MEDLINE as Standard Corpus for Measuring Semantic Similarity in the Biomedical Domain	315
<i>Hisham Al-Mubaid and Hoa A. Nguyen</i>	
Leveraging Chemical Background Knowledge for the Prediction of Growth Inhibition	319
<i>Lothar Richter, Stefan Hecht, and Stefan Kramer</i>	
A Supervised Learning Approach to Predicting Coronary Heart Disease Complications in Type 2 Diabetes Mellitus Patients	325
<i>Marisol Giardina, Francisco Azuaje, Paul McCullagh, and Roy Harper</i>	
Prediction of Labor for Pregnant Women Using High-Resolution Mass Spectrometry Data	332
<i>J-H Oh, J. Gao, A. Nandi, P. Gurnani, P. B. Greenwood, and K.P. Rosenblatt</i>	
Noninvasive Study of the Human Heart Using Independent Component Analysis	340
<i>Yi Zhu, T-L Chen, W. Zhang, T-P Jung, J-R Duann, S. Makeig, and C-K Cheng</i>	
The Role of Timing in Speech Perception and Speech Production Processes and its Effects on Language Impaired Individuals	348
<i>Anna Esposito and Nikolaos Bourbakis</i>	
Analysis of Invariant Meta-features for Learning and Understanding Disable People's Emotional Behavior Related to Their Health Conditions: A Case Study	357
<i>N. Bourbakis, A. Esposito, and D. Kavrakci</i>	
Author Index	370