

Daniel Huson, Oliver Kohlbacher, Andrei Lupas,
Kay Nieselt and Andreas Zell (eds.)

German Conference on Bioinformatics

GCB 2006

**19.09.2006-22.09.2006
in Tübingen, Germany**

Gesellschaft für Informatik 2006

Lecture Notes in Informatics (LNI) - Proceedings

Series of the Gesellschaft für Informatik (GI)

Volume P-83

ISBN 978-3-88579-177-5

ISSN 1617-5468

Volume Editors

Prof. Dr. Daniel H. Huson

Zentrum für Bioinformatik (ZBIT), Universität Tübingen, 72076 Tübingen, Germany

Email: huson@informatik.uni-tuebingen.de

Prof. Dr. Oliver Kohlbacher

Zentrum für Bioinformatik (ZBIT), Universität Tübingen, 72076 Tübingen, Germany

Email: kohlbacher@informatik.uni-tuebingen.de

Prof. Dr. Andrei Lupas

Max-Planck-Institut für Entwicklungsbiologie, 72076 Tübingen, Germany

Email: andrei.lupas@tuebingen.mpg.de

Dr. Kay Nieselt

Zentrum für Bioinformatik (ZBIT), Universität Tübingen, 72076 Tübingen, Germany

Email: nieselt@informatik.uni-tuebingen.de

Prof. Dr. Andreas Zell

Zentrum für Bioinformatik (ZBIT), Universität Tübingen, 72076 Tübingen, Germany

Email: zell@informatik.uni-tuebingen.de

Series Editorial Board

Heinrich C. Mayr, Universität Klagenfurt, Austria (Chairman, mayr@ifit.uni-klu.ac.at)

Jörg Becker, Universität Münster, Germany

Ulrich Furbach, Universität Koblenz, Germany

Axel Lehmann, Universität der Bundeswehr München, Germany

Peter Liggesmeyer, TU Kaiserslautern und Fraunhofer IESE, Germany

Ernst W. Mayr, Technische Universität München, Germany

Heinrich Müller, Universität Dortmund, Germany

Heinrich Reiner mann, Hochschule für Verwaltungswissenschaften Speyer, Germany

Karl-Heinz Rödiger, Universität Bremen, Germany

Sigrid Schubert, Universität Siegen, Germany

Dissertations

Dorothea Wagner, Universität Karlsruhe, Germany

Seminars

Reinhard Wilhelm, Universität des Saarlandes, Germany

© Gesellschaft für Informatik, Bonn 2006

printed by Köllen Druck+Verlag GmbH, Bonn

Contents

1	Stefan Canzar and Jan Remy Shape Distributions and Protein Similarity	1
2	Sérgio A. de Carvalho Jr. and Sven Rahmann Microarray Layout as Quadratic Assignment Problem	11
3	Matthias E. Futschik, Gautam Chaurasia, Erich Wanker and Hanspeter Herzel Comparison of Human Protein-Protein Interaction Maps	21
4	Thomas Hamborg and Jürgen Kleffe MPI-ClustDB: A fast String Matching Strategy Utilizing Parallel Computing	33
5	Markus Heinonen, Ari Rantanen, Taneli Mielikäinen, Esa Pitkänen, Juha Kokkonen and Juho Rousu Ab Initio Prediction of Molecular Fragments from Tandem Mass Spectrometry Data	40
6	Samatha Kottha and Michael Schroeder Classifying Permanent and Transient Protein Interactions	54
7	Robert Küffner, Timo Duchrow, Kartin Fundel and Ralf Zimmer Characterization of Protein Interactions	64
8	Stefano Lise and David Jones Invited talk: Docking Protein Domains Using a Contact Map Representation	74
9	Claudio Lottaz, Joern Toedling and Rainer Spang Annotation-based Distance Measures for Patient Subgroup Discovery in Clinical Microarray Studies	75
10	Gene Myers Invited talk: Imaging-Based Systems Biology	92
11	Axel Mosig, Ivo L. Hofacker and Peter F. Stadler Comparative Analysis of Cyclic Sequences: Viroids and other Small Circular RNAs	93
12	Cedric Notredame Invited talk: Combining Sequence Information with T-Coffee	103
13	G. Rättsch, B. Hepp, U. Schulze and C.S. Ong PALMA: Perfect Alignments using Large Margin Algorithms	104

14 Rob Russell	
Invited talk: Pushing Details into Interaction Networks	114
15 Andreas Schlicker, Carola Huthmacher, Fidel Ramírez, Thomas Lengauer and Mario Albrecht	
Functional Evaluation of Domain-Domain Interactions and Human Protein Interaction Networks	115
16 Mukund Thattai	
Invited talk: Encoding Evolvability: The Hierarchical Language of Polyketide Synthase Protein Interactions	127
17 Detlef Weigel	
Invited Talk: Genomic Variation and Incipient Speciation in <i>Arabidopsis thaliana</i>	128
18 Christof Winter, Thorsten Baust, Bernard Hoflack and Michael Schroeder	
A novel, comprehensive method to detect and predict proteinprotein interactions applied to the study of vesicular trafficking	129

Preface

This volume contains papers presented at the German Conference on Bioinformatics, GCB 2006, held in Tübingen, September 19–22, 2006. This annual international conference provides a forum for the presentation of research results in Bioinformatics and Computational Biology. It is run on behalf of the Fachgruppe “Informatik in den Biowissenschaften (BIOINF)” of the German Society of Computer Science (GI), the AG “Computereinsatz in den Biowissenschaften” of the German Society of Chemical Technique and Biotechnology (DECHEMA), and the Studiengruppe “Bioinformatik” of the German Society for Biological Chemistry and Molecular Biology (GBM).

The conference opened on September 19th, 2006, with the following four tutorials: “Introduction to Phylogenetic Networks” given by Daniel Huson, “Kernel Methods for Predictive Sequence Analysis” given by Gunnar Rätsch and Cheng Soon Ong, “Mining the Biomedical Literature: State of the Art, Challenges and Evaluation Issues” given by Hagit Shatkay, and “Non-coding RNA - No Longer the Dark Matter in a Cellular Universe”, given by Yu Wang.

Six leading scientists were invited to give keynote lectures. Gene Myers (HHMI, Janelia Farms) spoke on “Imaging-Based Systems Biology” and Detlef Weigel (MPI, Tübingen) on “Genome-wide Analysis of Sequence Variation in Arabidopsis”. A thematic focus was placed on Protein-Protein Interactions and this was reflected by the topics of the other invited speakers: David Jones (University College, London) spoke on “Docking protein domains using a contact map representation”, Cedric Notredame (CNRS IGS, Marseille) spoke on “Combining Sequence Information with T-Coffee”, Rob Russell (EMBL, Heidelberg) spoke on “Pushing Details into Interaction Networks”, and Mukund Thattai (NCBS, Bangalore) spoke on “Encoding evolvability: The hierarchical language of polyketide synthase protein interactions”.

The technical program additionally contained seven short papers and 12 long papers, which were refereed and selected from 62 submissions by the program committee. Additionally, over 100 poster abstracts were accepted for presentation at the poster sessions. This volume contains all long papers and abstracts of the invited lectures. The poster abstracts appear in a special abstract book together with the short papers. The conference was concluded with a special session on “Bioinformatics in Germany – State of the Art”, in which the five DFG-funded Bioinformatics Centers at Bielefeld, Leipzig, Munich, Saarbrücken and Tübingen reported on their work.

Thanks to the members of the program committee and their colleagues who gave their time to referee the submissions, and to all that helped locally to organize the meeting. We are also grateful to all presenters and participants, whose contributions and interactions made GCB 2006 a success.

Tübingen, August 2006

Daniel Huson
Oliver Kohlbacher (Local Chair, Program Co-Chair)
Andrei Lupas (Program Co-Chair)
Kay Nieselt
Andreas Zell