

Dongbo Bu

6, Kexueyuan South Road, Zhongguancun
Institute of Computing Technology,
Chinese Academy of Sciences
Beijing, 100190, P. R. China
Telephone: 86-10-6260-1019
Fax: 86-10-6260-1356
E-mail: dbu@ict.ac.cn
WWW: <http://bioinfo.ict.ac.cn/~dbu/>

Employment Record

- Associate Professor, Institute of Computing Technology, Chinese Academy of Sciences, May 2004 – present
- Visiting Scholar, Post-doctoral Fellow, Department of Computer Science, University of Waterloo, Apr. 2006 – July 2008.
- Assistant Professor, Institute of Computing Technology, Chinese Academy of Sciences, Apr. 2001-Apr. 2004.

Education

- Institute of Computing Technology, Chinese Academy of Sciences, Ph.D., 2001 (Advisors: Guojie Li, Thesis title: *Theory of Clustering/Classification and Their Applications in Text Mining.*)
- Institute of Computing Technology, Chinese Academy of Sciences, Master, 1997 (Advisors: Shuo Bai, Thesis title: *SAT Problem: Theory and Algorithms.*)
- Shandong University, Department of Computer Science, B.Sc., 1994.

Teaching

- [CS6012](#) Algorithm design and analysis, Fall 2009.

Research Interests

- Algorithm design and analysis, including SAT problem, information retrieval, and Bioinformatics.

Honors and Awards

- Sony Research Award, 2005.

Research Publications

Articles Published or Accepted in Refereed Journals and Conferences

D. Bu, M. Li, S. C. Li, J. Qian, J. Xu. Finding compact structural motifs. *Theoretical Computer Science*. 410 (30-32) 20, 2009

Gao X, Bu D, Xu J, Li M. Improving consensus contact prediction via server correlation reduction. *BMC Structural Biology*. 2009; 9: 28.

S. C. Li, D. Bu, J. Xu, M. Li. Fragment-HMM: A new approach to protein structure prediction, *Protein Sci*. 2008 17: 1925-1934

S. Sun, C. Yu, Y. Qiao, Y. Lin, G. Dong, C. Liu, J. Zhang, Z. Zhang, J. Cai, H. Zhang, D. Bu. Deriving the probabilities of water loss and ammonia loss for amino acids from tandem mass spectra, *J. Proteome Res.*, 7 (01), 202–208, 2008.

S. C. Li, D. Bu, J. Xu, M. Li. Finding Largest Well-Predicted Subset of Protein Structure Models. *CPM 2008*: 44-55

S. C. Li, D. Bu, X. Gao, J. Xu, M. Li. Designing succinct structural alphabets. *ISMB 2008*: 182-189

Y. Lin, Y. Qiao, S. Sun, C. Yu, G. Dong, D. Bu. A Fragmentation Event Model for Peptide Identification by Mass Spectrometry. *RECOMB 2008*: 154-166

X. Chen, S. He, D. Bu, F. Zhang, Z. Wang, R. Chen, W. Gao. FlexStem: improving predictions of RNA secondary structures with pseudoknots by reducing the search space. *Bioinformatics* 24(18): 1994-2001 (2008)

X Gao, D Bu, SC Li, J Xu, M Li. FragQA: predicting local fragment quality of a sequence-structure alignment. *Genome Informatics*, 2008

J. Qian, S. C. Li, D. Bu, M. Li, J. Xu. Finding Compact Structural Motifs. *CPM 2007*: 142-149

C. Yu, Y. Lin, S. Sun, J. Cai, J. Zhang, Z. Zhang, R. Chen, D. Bu. An Iterative Algorithm to Quantify Factors Influencing peptide Fragmentation during Tandem Mass Spectrometry. *J. Bioinformatics and Computational Biology* 5(2a): 297-311 (2007) 2006

G. Tan, N. Sun, D. Bu. Improving locality of nonserial polyadic dynamic programming. *IPDPS 2006*

S. Sun, Y. Zhao, Y. Jiao, Y. Yin, L. Cai, Y. Zhang, H. Lu, R. Chen, D. Bu. Faster and more accurate global protein function assignment from protein interaction networks using the MFGO algorithm, *FEBS Letters*, 580(7) 2006; 1891-1896

Z. Zhang, S. Sun, X. Zhu, S. Chang, X. Liu, C. Yu, D. Bu, R. Chen. A novel scoring schema for peptide identification by searching protein sequence databases using tandem mass spectrometry data. *BMC Bioinformatics* 7: 222 (2006)

G. Tan, L. Xu, D. Bu, S. Feng, N. Sun. Improvement of Performance of MegaBlast Algorithm for DNA Sequence Alignment. *J. Comput. Sci. Technol.* 21(6): 973-978 (2006)

C. Yu, Y. Lin, S. Sun, J. Cai, J. Zhang, Z. Zhang, R. Chen, D. Bu. An Iterative Algorithm to Quantify the Factors Influencing Peptide Fragmentation for MS/MS Spectrum, *CSB 2006*

Z. Zhang, H. Sun, Y. Zhang, Y. Zhao, B. Bai, S. Sun, H. Lu, D. Bu, L. Ling, R. Chen. Genome-wide analysis of

- mammalian DNA segment fusion/fission. *J. Theoretical Bio.* 240(2), 2006
- G. Tan, L. Xu, Y. Jiao, S. Feng, D. Bu, N. Sun. An Optimized Algorithm of High Spatial-temporal Efficiency for MegaBlast. *ICPADS (2) 2005*: 704-708
- C. Liu, B. Bai, G. Skogerbo, L. Cai, W. Deng, Y. Zhang, D. Bu, Y. Zhao, R. Chen. NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Research* 33(Database-Issue): 2005; 112-115
- J. Yu, J. Wang, ..., D. Bu, ..., J. Wang, G. Wong, H. Yang. The Genomes of *Oryza sativa*: A History of Duplications. *PLoS Biol* 3(2): 2005
- H. Lu, Y. Zhao, J. Zhang, Y. Wang, W. Li, X. Zhu, S. Sun, J. Xu, L. Ling, L. Cai, D. Bu, R. Chen. Date of origin of the SARS coronavirus strains. *BMC Infectious Diseases* 2004, 4:3
- H. Lu, X. Zhu, H. Liu, G. Skogerbø, J. Zhang, Y. Zhang, L. Cai, Y. Zhao, S. Sun, J. Xu, D. Bu, R. Chen. The interactome as a tree--an attempt to visualize the protein-protein interaction network in yeast. *Nucleic Acids Res.*, September 8, 2004; 32(16): 4804 – 4811
- H. Lin, Z. Zhang, Q. Zhang, D. Bu, M. Li. A Note on the Single Genotype Resolution Problem. *J. of Comp. Sci. and Tech.*, Vol.19, No.2, March 2004 254-258
- J. Zhang, D. Bu, J. Xu, S. He, W. Gao, M. Li. LSA: More Accurate DNA Fragment Assembly. RECOMB 2003 Satellite meeting at Stanford
- D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, N. Zhang, G. Li, R. Chen. Topological structure analysis of the protein-protein interaction network in budding yeast *Nucleic Acids Research*, 2003. Vol.31, No.9 2443-2450
- D. Bu, S. Bai, G. Li. The Duplex Strategy of Term Weighting in Text Clustering. *J. of Software*. 13(11), 2002: 2083-2089.
- D. Bu, S. Bai, G. Li. Principle of Granularity in Clustering and Classification. *J. Computer*. 25(8), 2002: 810-816.
- D. Bu, H. Xu, S. Bai. An optimization algorithm based on description complexity. *J. of Computer*. 25(8), 2002: 878-882
- B. Wang, H. Xu, Z. Yang, Y. Liu, X. Cheng, D. Bu, S. Bai. TREC-10 Experiments at CAS-ICT: Filtering, Web and QA. TREC 2001
- J. Pang, D. Bu, S. Bai. Implementation of Text Categorization System Based on VSM. *Application Research of Computers*, 2001.
- D. Bu, S. Bai. Modeling the Phase Transition Phenomena of Satisfiability Problem, *Proceeding of International Conference for Young Computer Scientist*, 1999.
- D. Bu, S. Bai. Analysis of the phase transition phenomena of 2-3-SAT problem. *J. of Software*. 9(11), 1998: 828-832.

Book Chapter

- [1] Consensus Approaches to Protein Structure Prediction. Chapter 5 in the book *Machine Learning in Bioinformatics*, John Wiley & Sons, 2008

Recent Software Developments

- FALCON, an *Ab Initio* method for protein structure prediction. Ranked 3rd in the FR-H category in CASP8
- PI, a database-searching package for tandem mass spectra. (See <http://bioinfo.ict.ac.cn/MSMS/>)