# CURRICULUM VITAE
# Jun (Luke) Huan

Department of Computer Science
University of North Carolina
322 Sitterson Hall
Chapel Hill, NC 27599-3175

Telephone: +1-919-962-1843
Fax: +1-919-962-1799
E-mail: huan@cs.unc.edu
http://www.cs.unc.edu/~huan

## Research Interests:

**Data Mining & Knowledge Discovery**: pattern discovery in biological/complex/high dimensional data, pattern visualization, classification, clustering, and density estimation in high dimensional space.

**Bioinformatics & Computational Biology**: *Structural Bioinformatics & Proteomics*: functional motifs inference, protein-protein interaction prediction, RNA structural classification, pharmacophore discovery, and structure motif discovery in protein evolution. *Functional Genomics*: structure-based functional annotation. *Systems Biology*: genome wide metabolic pathways annotation/reconstruction and comparative study of signal transduction pathways.

## Education:

**Ph.D.** University of North Carolina at Chapel Hill – Expected August, 2006
Major: Computer Science, Advisors: Dr. Jan Prins & Dr. Wei Wang,
Dissertation title: "*Graph Database Mining in Structure Bioinformatics*",
Certificate: UNC Bioinformatics Graduate Training Program, Advisor: Alexander Tropsha.

**M.S.** Oklahoma State University – August 2000, Major: Computer Science.

**B.S.** Peking University – May 1997, Major: Biochemistry & Molecular Biology.

## Honors & Awards:

- Alumni Fellow – University of North Carolina, 2005
- Scholar for Tomorrow – University of North Carolina, 2001
- Honorable Mention in Predoctor Fellowship – Howard Hughes Medical Institution,1998
- Excellent Student Scholarship (First Class) – Peking University, 1993, entered with the national entrance examination waived
- Top 4 – National Physics Olympiad, China, 1992
- First Class Honor – State Mathematics Olympiad, China, 1992
- Three Travel Awards – RECOMB'04, PSB'04, ISMB'05

## Research Experience:

| | |
|---|---|
| 08/2001 – Present | Research Assistant, Computer Science Department, University of North Carolina, *Research in data mining, bioinformatics, and high performance computing.* |
| 05/2004 – 12/2004 | Intern, GlaxoSmithKline, Research Triangle Park, NC, *Developed predictive models to infer hormone receptor genes from the human genome.* |
| 01/2000 – 08/2001 | Software Designer, Nortel Networks, Richardson, TX, *Implemented/sustained the mobile relocation protocol for the 3nd generation wireless communication system UMTS.* |

05/1999 – 08/1999 Intern, Mathematics and Computer Science Division, Argonne National Laboratory, Supervisor: Dr. Ross Overbeek,
*Designed an I/O efficient algorithm to discover recurring q-grams from the human genome.*

08/1997 – 08/1998 Research Assistant, Department of Molecular Genetics, Medical School, University of Illinois, Chicago,
*Experimentally identified novel genes that interact with tumor suppressor gene BRCA2 using yeast two-hybrid systems.*

08/1996 – 05/1997 Research Assistant, Institute of Developmental Biology, China,
*Measured the activities of calcium oscillations related genes in an embryo stem (ES) cell line.*

## Teaching Experience:

08/2004 – 12/2004 Instructor, Computer Science Department, University of North Carolina,
*Lectured a course on introduction to programming.*

08/1999 – 12/1999 Teaching Assistant, Computer Science Department, Oklahoma State University,
*Directed discussion sessions and graded assignments of a course on compiler writing.*

## Conference Presentations

- "Mining Protein Family Specific Residue Packing Patterns From Protein Structure Graphs", *International Conferences on Research in Computational Molecular Biology* (**RECOMB)**, San Diego, California, April, 2004.
- "Classification of Protein Structural Families Based on Coherent Subgraph Mining", *Pacific Symposium of Biocomputing* **(PSB)***,* Big Island, Hawaii, January, 2004

## Invited Talks:

- "Discovering Patterns in Families of Protein Structures", School of Computer Science, University of Waterloo, April, 2006
- "Discovering Patterns in Families of Protein Structures", Department of Electrical Engineering & Computer Science, University of Kansas, March, 2006
- "Discovering Patterns in Families of Protein Structures", Department of Electronic Engineering and Computer Science, Case Western Reserve University, February, 2006
- "Discovering Patterns in Families of Protein Structures", Department of Computer Science & Engineering, University of South Florida, February, 2006
- "Mining Family-Specific Residue-Packing Patterns from Protein Structure Graphs", *Algorithm Seminar*, Duke University, November, 2004.

## Professional Activities:

Referee for KAIS, TKDE, TKDE Special Issue on Bioinformatics (2005).

Referee for EDBT'06, ICDE'05/06, ICDM'04/05, SIGKDD'04/05, SIGMOD'05, VLDB'05.

Department Committee Member for the Infrastructure Committee and the Colloquium Committee.

## Grant Proposal Writing:

I have participated in preparing the following grant proposals:

| Agency | Title | Total Award |
|---|---|---|
| NSF | Identifying Spatial Motifs for Classification of Protein Structure and Function | $140,000.00 |
| NSF | Mining Salient Localized Patterns in Complex Data | $421,000.00 |

| | | |
|---|---|---|
| Microsoft | A Comprehensive Protein Database Indexed by Spatial Motifs | $50,000.00 |
| Microsoft | New Faculty Fellowship Award | $200,000.00 |

## Publications:

*Refereed Journal Paper & Book Chapters*

1. **J. Huan**, W. Wang, and J. Prins, Protein Local Structure Comparison: Methods and Future Directions, invited to *Advances in Computers* by Chau-Wen Tseng (eds.), Elsevier, 2006.
2. **J. Huan,** W. Wang, and J. Prins, Graph Mining and Protein Local Structure Comparison, invited to *Knowledge Discovery in Bioinformatics: Techniques, Methods, and Applications* by Y. Pan and T. Hu (eds.), John Wiley and Sons, 2006.
3. D. Bandyopadhyay, **J. Huan**, J. Liu, J. Prins, J. Snoeyink, W. Wang, A. Tropsha, Structure-based Function Inference Using Protein Family-specific Fingerprints, invited to *Protein Science* based on our ISMB workshop publication, 2006.
4. **J. Huan,** D. Bandyopadhyay, W. Wang, J. Snoeyink, J. Prins, and A. Tropsha, Comparing Graph Representations of Protein Structure for Mining Family-Specific Residue-Based Packing Motifs, invited to *Journal of Computational Biology (**JCB**)* based on our RECOMB publication, Vol. 12, No. 6: 657-671, 2005.

*Refereed Conference Papers*

5. **J. Huan,** D. Bandyopadhyay, J. Snoeyink, J. Prins, A. Tropsha, W. Wang, Distance-based Identification of Spatial Motifs in Proteins Using Constrained Frequent Subgraph Mining, in *Proceeding of the IEEE Computational Systems Bioinformatics* (***CSB**), 2006*
6. **J. Huan,** W. Wang, J. Prins, J. Yang, SPIN: Mining Maximal Frequent Subgraphs from Graph Databases, in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**ACM SIGKDD**)*, Seattle, Washington, August 2004, pp. 581-586.
7. **J. Huan,** W. Wang, D. Bandyopadhyay, J. Snoeyink, J. Prins, A. Tropsha, Mining Family-specific Residue-Packing Patterns from Protein Structure Graphs, in *Proceeding of the Eighth Annual International Conferences on Research in Computational Molecular Biology (**RECOMB**)*, San Diego, California, March 2004, pp. 308-315.
8. **J. Huan,** W. Wang, A. Washington, J. Prins, R. Shah, A. Tropsha, Accurate Classification of Protein Families based on Coherent Subgraph Mining, in *Proceeding of the Tenth Pacific Symposium on Biocomputing (**PSB**)*, Big Island, Hawaii, January 2004, pp. 411-422.
9. K. Berlin, **J. Huan**, M. Jacob, G. Kochhar, J. Prins, B. Pugh, P. Sadayappan, J. Spacco, C. Tseng, Evaluating the impact of Programming Language Features on the Performance of Parallel Applications on Cluster Architectures, *Language and Compilers for Parallel Computing (**LCPC**)*, Springer-Verlag, 2004.
10. **J. Huan**, W. Wang, J. Prins, Efficient Mining of Frequent Subgraphs in the Presence of Isomorphisms, in *Proceeding of the 3nd IEEE International Conference of Data Mining (**ICDM**)*, Melbourne, Florida, November 2003, pp. 549 -552.

*Poster & Demo Papers*

11. **J. Huan**, D. Bandyopadhyay, J. Liu, J. Prins, J. Snoeyink, A. Tropsha, and W. Wang, Rapid Determination of Local Structural Features Common to a Set of Proteins, *Intelligent Systems for Molecular Biology (**ISMB**)* demo*,* 2005.
12. D. Bandyopadhyay, **J. Huan**, J. Liu, J. Prins, J. Snoeyink, A. Tropsha, and W. Wang, Function Inference Using Family-Specific Subgraph Fingerprints Mined from Protein Families, *Intelligent Systems for Molecular Biology (**ISMB**)*, poster, 2005.
13. R. Shah, **J. Huan**, A. Tropsha, W. Wang, Structure Based Identification of Protein Family Signatures for Function Annotation, in *Proceeding of the Ninth Annual International Conferences on Research in Computational Molecular Biology (**RECOMB**)*, poster, Cambridge, MA 2005.
14. **J. Huan**, J. Prins, T. Vision, W. Wang, Reconstruction of Ancestral Gene Order after Segmental Duplication and Gene Loss, *IEEE Computer Science Society Bioinformatics Conference (**CSB**),* poster, Stanford University, August, 2003, pp. 484-485.

*Technical Reports & Thesis*

15. D. Bandyopadhyay**, J. Huan**, J. Liu, J. Prins, J. Snoeyink, A. Tropsha, and W. Wang, Protein Functional Family Identification by Fast Subgraph Isomorphism Using Structure-Based Fingerprints Mined from SCOP Families, *UNC-CS Technical Report TR04-031***,** 2004. Poster presented at *Triangle Biophysics Symposium*, Nov. 2004, Durham, NC.
16. J. Prins, **J. Huan**, B. Pugh, C. Tseng, P. Sadayappan, UPC Implementation of an Unbalanced Tree Search Benchmark, *UNC-CS Technical Report TR03-034*, 2003.
17. **J. Huan**, A Localized Clustering Algorithm and Its Application in DNA Sequence Analysis, *Master's thesis, Computer Science Department, Oklahoma State University*, 2*000.*
18. M. G. D'Souza, **J. Huan**, S. Sutton, M. Romine, and N. Maltsev, PUMA2 -- An Environment for Comparative Analysis of Metabolic Subsystems and Automated Reconstruction of Metabolism of Microbial Consortia and Individual Organisms from Sequence Data, *Argonne National Laboratory Technical Memorandum ANL/MCS-TM-240*, 1999.

*Other Refereed Paper:*

19. Jingmei Liu, Yuan Yuan, **J. Huan**, and Zhiyuan Shen. Inhibition of Breast and Brain Cancer Cell Growth by BCCIP, an Evolutionarily Conserved Nuclear Protein that Interacts with BRCA2, in *Oncogene*, Volume 20, Number 3. pp. 336-345, January 2001.

*Papers under Submission:*

20. **J. Huan**, W. Wang, J. Prins, Mining Geometric Patterns Using Graph-based Techniques.
21. **J. Huan**, W. Wang, J. Prins, Graph Database Mining and its Applications.

## References:

**Dr. Nikolay V. Dokholyan**
Assistant Professor
Department of Biochemistry and Biophysics
School of Medicine
University of North Carolina
Chapel Hill, NC 27599
Email: dokh@med.unc.edu
Tel: 919-843-2513
Fax: 919-966-2852

**Dr. Jack Snoeyink**
Professor
Department of Computer Science
University of North Carolina
Chapel Hill, NC, 27599-3175, USA
Email: snoeyink@cs.unc.edu
Tel: 919-962-1969
Fax: 919-962-1799

**Dr. Wei Wang**
Associate Professor
Department of Computer Science
University of North Carolina
Chapel Hill, NC, 27599-3175, USA
Email: weiwang@cs.unc.edu
Tel: 919-962-1744
Fax: 919-962-1799

**Dr. Jan Prins**
Professor & Department Chair
Department of Computer Science
University of North Carolina
Chapel Hill, NC, 27599-3175, USA
Email: prins@cs.unc.edu
Tel: 919-962-1913
Fax: 919-962-1799

**Dr. Alexander Tropsha**
Professor
School of Pharmacy &
Director
Bioinformatics Graduate Training Program
University of North Carolina
Chapel Hill, NC, 27599-3175, USA
Email: alex_tropsha@unc.edu
Tel: 919-966-2955
Fax: 919-966-0204