Curriculum Vitae

Yong Kong, Ph.D.

373 Race Hill Road
Madison, CT 06443, USA

1-203-871-2422 (O), 1-203-421-8415 (H), 1-203-481-2075 (Fax)
E-mail: ykong@alum.wustl.edu   yong_kong@yahoo.com

Objective:

Faculty position for research and teaching in the fields of computational biology, computational chemistry, and Bioinformatics.

Education:

Ph.D.	Molecular Biophysics Washington University School of Medicine, St. Louis, 1997 Dissertation: Multiple Electrostatic Methods for Protein Modeling with Reaction Field Treatment Advisor: Jay W. Ponder
M.S.	Neurobiology Shanghai Institute of Physiology, Chinese Academy of Sciences (CAS), Shanghai, 1991 Advisor: Fuchuan Sun
B.S. B.S	Biomedical Engineering Computer Sciences Qinghua University (Tsinghua University), Beijing, 1988 Graduated with highest honor

Employment:

2003-present:	Project Leader in Bioinformatics, CuraGen Corporation
1999-2002:	Senior Research Scientist in Bioinformatics, CuraGen Corporation
1998-1999:	Research Scientist in Bioinformatics, CuraGen Corporation
1997-1998:	Postdoctoral Scientist in Bioinformatics, CuraGen Corporation
1995-1997:	Research Assistant, Washington University School of Medicine
1993-1994:	Research Assistant, University of Illinois at Urbana-Champaign
1991-1992:	Research Associate, Shanghai Institute of Physiology, Chinese Academy of Sciences (CAS)
1988-1991:	Research Assistant, Shanghai Institute of Physiology, Chinese Academy of Sciences (CAS)

Professional Experience:

Statistics
Developed new methods to systematically study distributions of runs, the longest-runs, and other statistics of runs in samples from multi-letter alphabet systems. The methods have direct applications to biological sequence analysis. The manuscript has been submitted to a premier statistics journal (the *Journal of the American Statistical Association* (*JASA*)) and is currently under formal review. To illustrate the applications to computational biology and Bioinformatics, the newly developed methods are applied to the

proteome of bacterium *Mycoplasma genitalium.* A software package of various run-related statistics for general-purpose use with arbitrary precision arithmetic was written in C programming language.

Computational biology and Computational chemistry
Developed the first force field with polarizable multipole electrostatics for flexible biological macromolecules. Developed a polarizable multipole water model. The force field and the water model are parts of TINKER package for molecular mechanics and dynamics of macromolecules. TINKER is available at http://dasher.wustl.edu/.

Developed general and efficient formulas to calculate reaction field due to off-center point multipoles (*J. Chem. Phys.,* 107, 481, 1997).

Developed a new and simple method to evaluate configurations (partition functions) of finite or infinite linear molecules (*J. Phys. Chem. B*, 105, 2001).

Developed a general method to tackle two or three-dimensional ligand binding problems in molecular biology and Ising model in statistical mechanics. Proved a counter-intuitive fact that linear models are in general simpler than the circular models, although the open ends break symmetry in the linear models (*J. Chem. Phys.*, 111, 4790, 1999).

Developed analytical methods to study multivalent binding on ladder lattices and other complex models (*Biophys. Chem.*, 81, 7, 1999; *Biophys. Chem.*, 61, 107, 1996).

Bioinformatics
As a member of the team led by Professor Eugene Myers of UC Berkeley (previously at Celera Genomics), wrote major modules of the whole genome assembly program for 454 Corporation's next generation high-throughput whole genome sequencing technology. 454 Corporation is a majority-owned subsidiary of CuraGen Corporation. The modules include Overlapper, an efficient all-against-all homology search program specially designed for the pyro-sequencing technology.

Developed algorithms and software to map large amounts of transcripts (cDNA and EST) to genome scaffolds. Clustered the transcripts based on their locations on the genome. Determined exon-intron boundaries and transcript orientations based on canonical splice sites. Used the mapping information to investigate the regulatory sequences for the transcripts and gene splice variants.

Developed algorithms and software to cluster and assemble cDNA and EST sequences of whole organisms and specific tissues. As the leader of the group, I developed all the major programs for the cDNA and EST clustering/assembly process: proprietary cDNA sequences and publicly available cDNA and EST data are collected and pre-processed (primer and vector trimming, contamination screening, repeat and low-complexity masking). The processed sequences are clustered and assembled to produce a representation of the transcripts within a whole organism or specific tissue. The algorithms assemble splice variants into separate contigs, while other popular assemblers, like PHRAP, are usually unable to separate splice variant correctly. The assembled contigs and the alignments of their components are stored in relational databases and available to CuraGen internal scientists as well as clients and collaborators through web interfaces. The assembled transcripts come from human, mouse, rat, fly, pig, corn, rabbit, and chicken.

The assembly process and the mapping process were used in the building of the genome-scale protein interaction map of *Drosophila melanogaster* (to appear in the December 5 issue of *Science*).

Developed software to process chromatogram data from electrophoresis sequencing machines.

Developed workflow and user interfaces with PL/SQL and CGI to keep track of information in high-throughput sequencing processes.

Structure biology / X-ray crystallography
Solved the binary complex structure of the large fragment of *Taq* DNA polymerase I and nucleotides
(*Protein Science*, 7, 1116, 1998).

Experimental biochemistry and molecular biology
Cloned and purified photosynthesis proteins from cyanobacteria for physical chemistry studies. (*Biophys. J.*, 66 (2): A227 (38th Annual Biophysics Society Meeting, Abstract, W-AM-C2)).

Eye movement control during reading
Developed a real-time graphic program to display Chinese texts on monitors in different formats with the display speed controlled by the reader during reading. Developed a real-time graphic data acquisition and analysis system using C and assembly language for eye movement analysis. Studied the effects of eye movement on reading of Chinese and other languages.

Biomedical engineering
Designed and implemented several biomedical engineering instruments, including a real-time acquisition and analysis system for evoked potentials, and a digital ECG simulator, using integrated circuits (ICs) and microprocessors.

Teaching Experience:

January 1996 - May 1996, Washington University, St. Louis, MO.
Teaching Assistant, Fundamentals of Biology I (Biol 296A).
Organized discussions for three classes of students for eight weeks. Designed and graded quizzes.

Honors And Awards:

Fellowship, Washington University School of Medicine, St. Louis, 1994-1996
University Fellowship, University of Illinois at Urbana-Champaign, 1992-1993
The Institute Director Award, Shanghai Institute of Physiology,
                    Chinese Academy of Sciences (CAS), 1991
University Honor Program, Qinghua University, 1984-1988 (the highest honor for undergraduate students)
Third Place, National Mathematical Competition, PR China, 1982

Publications:

Kong, Y. (2003), "Distribution of Runs and Longest Runs: A New Generating Function Approach with Applications to the Proteome of *Mycoplasma genitalium* ", submitted to the *Journal of the American Statistical Association* (*JASA*).

Giot, L., *et. al.* (2003), "A Protein Interaction Map of *Drosophila melanogaster*", *Science*, to appear in the December 5 print edition (cover story). Published online on November 6, 2003.

Kong, Y. (2002), "A Note on the Quantitative Properties of McGhee-von Hippel Model", *Biophysical Chemistry*, 95, 1-6.

Kong, Y. (2001), "A Simple Method for Evaluating Partition Functions of Linear Polymers", *Journal of Physical Chemistry B*, 105, 10111-10114.

Kong, Y. (1999), "General Recurrence Theory of Ligand Binding on Three-Dimensional Lattice", *Journal of Chemical Physics*, 111, 4790-4799.

Kong, Y. (1999), "Ligand Binding on Ladder Lattices", *Biophysical Chemistry*, 81, 7-21.

*Li, Y., *Kong, Y., *Korolev, S. and Waksman, G. (1998), "Crystal Structures of the Klenow Fragment of Thermus Aquaticus DNA Polymerase I Complexed with Deoxyribonucleoside Triphosphates", *Protein Science*, 7, 1116-1123. (*Those authors contributed equally.)

Kong, Y. and Ponder, J. W. (1997), "Calculation of the Reaction Field Due to Off-Center Point Multipoles", *Journal of Chemical Physics*, 107, 481-492.

Di Cera, E. and Kong Y. (1996), "Theory of Multivalent Binding in One and Two-Dimensional Lattices", *Biophysical Chemistry*, 61, 107-124.

Recent Poster Presentations:

Kong, Y., Zhang, Y., Braverman, M, and Leach, M. (2001) "CuraGen's SeqCalling Clustering/Assembly Package". Fifth Annual Conference On Computational Genomics, Baltimore, MD.

Kong, Y., Braverman, M., Walker, E., and Leach, M. (2000) "SeqCalling – A Normalized Process for Generating Protein-Coding Biased Expressed Sequence Databases". 12th International Genome Sequencing and Analysis Conference, Miami Beach, FL.

Szekeres, E.S., Kong, Y., and Leach, M. (2000) "Post-Assembly Analysis Methods of Transcript Sequences for Quality Assessment and SNP Candidate Detection". 12th International Genome Sequencing and Analysis Conference, Miami Beach, FL.

Miscellaneous:

Active contributor to Sloane's "On-Line Encyclopedia of Integer Sequences"
http://www.research.att.com/~njas/sequences/index.html
(Click "word" to search author's name to get entries.)

Kong, Y. (1999) "A summation identity", *American Mathematical Monthly*, 106, 266.
Kong, Y. (1999) "Solution to a problem proposed by D. Knuth - Subtracting square roots repeatedly", *American Mathematical Monthly*, 106, 167.

References:

Dr. Enrico Di Cera   Professor
                                Department of Biochemistry and Molecular Biophysics
                                Washington University School of Medicine
                                Box 8231
                                660 S. Euclid Avenue
                                St. Louis, MO 63110

                                Phone: (314) 362-4185
                                FAX (314) 747-5354
                                Email: enrico@biochem.wustl.edu
                                http://biochem.wustl.edu/~enrico/

Dr. Garland R. Marshall   Professor
                                Department of Biochemistry and Molecular Biophysics and the Center for Computational Biology
                                700 S. Euclid Avenue
                                Campus Box 8036
                                St. Louis, MO 63110

                                Phone: (314) 362-1567

FAX: (314) 747-3330
Email: garland@pcg.wustl.edu
http://www.cmd.wustl.edu/

Dr. Jay W. Ponder   Associate Professor
Department of Biochemistry and Molecular Biophysics
Biochemistry, Box 8231
Washington Univ. School of Medicine
660 South Euclid Avenue
St. Louis, MO 63110

Phone: (314) 362-4195
FAX: (314) 362-7183
Email: ponder@dasher.wustl.edu
http://dasher.wustl.edu/

Dr. Gabriel Waksman   Professor of Structural Moleculal Biology
Head, Joint Research School in Biomolecular Sciences
Director, Institute of Structural Molecular Biology
School of Crystallography
Birkbeck College
University College London
Gower Street, London, WC1E 6BT

Phone: 44-(0)207 631 6833
FAX: 44-(0)207 631 6803
Email: g.waksman@bbk.ac.uk
http://people.cryst.bbk.ac.uk/~ubcg54a