BIOINFORMATICS AND COMPUTER SCIENCE—SOMETHING FOR EVERYONE

L. S. Heath and L. T. Watson
Departments of Computer Science and Mathematics
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0106 USA

http://www.cs.vt.edu/~heath/

Virginia Tech
1872

The scientific problems in bioinformatics.

- **Analyzing Raw Experimental Data**
  The raw signals from most lab instruments are analog ones that vary with time. These are often to be interpreted as discrete values — a base in a DNA sequence or the expression/non-expression of some trait. More and more the quantity of data to be interpreted requires that computation be used to perform the interpretation.

- **Assembling DNA Fragments**

- **Physical Mapping of DNA**

- **Identifying Genes**

- **Annotating Gene Function**

- **Deriving Phylogenies**

- **Genome Rearrangements**

- **Protein Folding**
  Predicting protein structure from amino acid sequence. Complex problem involving physics and chemistry. Both continuous and discrete approaches have been tried, though a combination is likely to be the ultimate winner.
    - ◇ Molecular dynamics
    - ◇ Quantum chemistry
    - ◇ Global optimization
- **Determining Protein Function**
- **Relating Multiple Genes to Complex Traits**
  Carole Cramer (tobacco), Saghai Maroof (soybeans), Ina Hoeschele (farm animals).
- **Dealing With Complex Experimental Results**
  For example, expression information from DNA microarray technology. Image processing, expert systems, recommender systems.
- **Modeling Cellular Functions and Trait Expression**
  PSEs, simulation, numerical analysis, PDEs, ODEs, nonlinear equations, parameter estimation, optimization.

- **Sequence Matching**

- **Data Visualization**
  Dave Bevan (molecular docking), CAVE.

- **Data Organization, Compression, and Searching**
  Data integrity/accuracy is a problem. Old suspect information tends to stay in the data base. Original data needs to be preserved so it can be reanalyzed with new algorithms. Knowledge representation.

- **BLAST and Similar Sequence Analysis Tools**
  A typical biologist may put unwarranted faith in the results returned by these heuristics, because it is too difficult to understand meaning and reliability.

- **Data Mining**
  Make biologically relevant inferences or conjectures from sequence and other experimental data.

- **Statistical Genetics**
  Identify genes in organisms (e.g., cows, humans) from expressed traits. Robust data analysis.

- **Image Processing**
  Much of the data acquisition is visual and subjective. Perfect application area for image processing and computer vision.

- **Drug Discovery and Design**
  Pharmaceutical companies have enormous data bases that need to be mined; extreme need for new discrete and continuous algorithms. Mathematical formulation of drug design goals.

- **Evolutionary Algorithms**
  Genetic algorithms as a technology for solving bioinformatics problems.

- **Model Higher Level Cell, Organ, System Functions**
  ODEs, bifurcation, parameter estimation, nonlinear PDEs, reaction diffusion equations, Navier-Stokes equations, fluid-solid interaction models.

  ◇ Cell models: ODEs, nonlinear PDEs, reaction diffusion equations, optimization, systems of nonlinear equations.

  ◇ Heart models: Navier-Stokes equations, fluid-solid interaction.

- **Getting Accurate Answers in a Timely Fashion**
  High performance and parallel computing, robotics.