Syllabus and Suggested Curriculum for Candidacy Exam in CS IBAM (Informatics in Biology and Medicine)

**Suggested Curriculum:**

**Core Material (all on list)**

CS 284A Representations and Algorithms for Molecular Biology  
CS 284B Probabilistic Modeling of Biological Data  
CS 284C Computational Systems Biology

**Artificial Intelligence/ Machine Learning/Data Management/Networks (three from list)**

CS 215 Data Mining  
CS 271 Introduction to Artificial Intelligence  
CS 273A Machine Learning  
CS 273B Kernel-Based Learning  
CS 274A Probabilistic Learning: Theory and Algorithms  
CS 274B Learning in Graphical Models  
CS 276 Network-Based Reasoning/Belief Networks  
CS 222 Principles of Data Management  
CS 288A Biological Networks

**Algorithms (one from list)**

CS 260 Fundamentals of the Design and Analysis of Algorithms  
CS 263 Analysis of Algorithms  
CS 265 Graph Algorithms  
CS 266 Computational Geometry

**Probability and Statistics (one from list)**

CS 206 Principles of Scientific Computing  
Stat 200A Intermediate Probability and Statistical Theory  
Stat 210 Statistical Methods I: Linear Models  
Stat 225 Bayesian Statistical Analysis  
Stat 230 Statistical Computing Methods
Reading for Candidacy Exam

Core Reading

*Bioinformatics: Sequence and Genome Analysis*, by David Mount
Cold Spring Harbor Laboratory Press

*Bioinformatics: the Machine Learning Approach*, by P. Baldi and S. Brunak
MIT Press (Second Edition)

Additional Reading

*Biological Sequence Analysis* by R. Durbin, S. Eddy, A. Krogh, G. Mitchison

*Biochemistry: A Short Course* by Harry R. Matthews, Richard Freedland, and Roger L. Miesfeld.

*Introduction to Protein Structure* by Carl Branden and John Tooze.

*Molecular Modeling: Principles and Applications* by A.R. Leach.

*Introduction to Computational Molecular Biology* by Joao Setubal and Joao Meidanis.

*Computational Molecular Biology: An Algorithmic Approach* by Pavel A. Pevzner

*Handbook of Graphs and Networks: From the Genome to the Internet* by Bornholdt and Schuster (Editors), Wiley, 2003.

*Computational Modeling of Genetic and Biochemical Networks* by Bower and Bolouri (Editors), 2001.

*DNA Microarrays and Gene Expression*, by P. Baldi and G.. W. Hatfield

*An Introduction to Chemoinformatics*, by A. R. Leach and V. J. Gillet
Topics for Candidacy Exam

Overview:

Basic understanding of molecular biology: atoms, nucleotides, DNA, RNA, amino acids, genes, chromosomes, peptides, proteins, enzymatic reactions, metabolic networks, signaling networks, regulatory networks, cellular organization

Basic understanding of main databases: GenBank, Swissprot, PDB, Medline, etc.

Good understanding of modern AI, statistical machine learning, data mining, and scientific modeling methods and algorithms, Bayesian methods, evolutionary algorithms, reinforcement learning, neural networks, graphical models, kernel methods

Excellent understanding of computational analysis and prediction of the structure, function, interactions, and evolution of DNA, RNA, proteins, molecules, and processes.

Specific Topics:

Sequence modeling, similarity, and alignment: graphical models, basic Markov models; global and local dynamic programming alignment; BLAST, sequence database searching

Models of evolution: phylogenetic tree reconstruction, parsimony, maximum likelihood, neighbor joining; multiple alignments, HMMs, profiles

Sequence patterns: pattern matching, pattern induction, signal peptides, k-mer analysis, Gibbs sampling, expectation-maximization (EM); gene finding, regulatory regions, intron/exon calling, splice site prediction

Protein structure prediction: homology modeling, fold recognition, threading, ab-initio; secondary structure prediction: various methods including neural networks

Atomic Force Fields: functional form, parameterization, molecular mechanics

Rational drug design: docking algorithms; rational drug discovery: small molecules, QSAR, combinatorial methods, basic chemoinformatics

Microarray data, differential analysis, clustering

Modeling regulatory, signaling, and metabolic reactions and networks: reaction rates, stoichiometric constants, mass action, stochastic modeling and master equation, ordinary and partial differential equation models

Network representation of biological data; network models, properties, analysis; and complexity; graph clustering; graph alignment heuristics (e.g. PathBLAST).