BHARATI VIDYAPEETH
(DEEMED TO BE UNIVERSITY), PUNE, INDIA
PhD Entrance Test – 2026
SECTION-II: Bioinformatics - 35 Marks

The syllabus for the entrance test for Ph.D. in Bioinformatics is based on the syllabus for the Masters in Bioinformatics course conducted by Bharati Vidyapeeth (Deemed to be University). The Entrance examination will be based on the following topics.

Unit 1: Computational Environment and Reproducible Bioinformatics

- Linux/Unix operating systems
- Shell scripting basics
- Environment management using:
  - Conda
  - Docker
- Computational notebooks:
  - Jupyter Notebook
  - RStudio
- Version control using:
  - Git
  - GitHub
- High Performance Computing (HPC)
- Workflow automation:
  - Snakemake
  - Nextflow
- FAIR data principles

Unit 2: Biological Databases and Sequence Analysis

- Biological databases:
  - NCBI
  - UniProt
  - Ensembl
  - Protein Data Bank
- Pairwise and multiple sequence alignment
- Scoring matrices:
  - PAM
  - BLOSUM
- Dynamic programming algorithms:
  - Needleman–Wunsch
  - Smith–Waterman
- Hidden Markov Models (HMMs)
- BLAST and PSI-BLAST
- Motif and domain analysis
- Comparative genomics and pan-genomics
- Molecular phylogenetics
- Evolutionary models and phylogenetic tree construction

Unit 3: Genomics, Transcriptomics, and Functional Annotation

- Genome organization and assembly
- Structural and functional genome annotation
- Gene prediction algorithms
- Functional annotation using:

    - GO
    - KEGG
    - Reactome
    - InterPro
- Next Generation Sequencing (NGS)
- Sequencing platforms:
    - Illumina
    - PacBio
    - Oxford Nanopore
- RNA-seq analysis
- Whole Genome Sequencing (WGS)
- Variant calling and SNP analysis
- Epigenomics and methylation analysis
- Single-cell transcriptomics
- Metagenomics and microbiome informatics
- Multi-omics integration

Unit 4: Proteomics, Metabolomics, and Systems Biology

- Protein identification and characterization
- Mass spectrometry basics
- Protein-protein interaction networks
- Metabolomics and pathway analysis
- Gene Regulatory Networks (GRN)
- Systems biology approaches
- Network biology and graph theory
- Biological network analysis tools:
    - Cytoscape
    - STRING
    - MetaboAnalyst

Unit 5: Programming Fundamentals for Bioinformatics

- Python programming
- R programming
- Data structures:
    - Strings
    - Lists
    - Dictionaries
    - Sets
- File handling and exception handling
- Object-oriented programming basics
- Regular expressions
- Parsing biological file formats:
    - FASTA
    - FASTQ
    - GFF/GTF
    - BED
    - VCF
- Data visualization using:
    - Matplotlib
    - ggplot2
    - Plotly

Unit 6: Algorithms and Computational Biology

- Algorithm design principles
- Big-O notation
- Searching and sorting algorithms
- Dynamic programming
- Graph algorithms
- String matching algorithms
- K-mer based analysis
- Burrows-Wheeler Transform (BWT)
- De Bruijn graphs
- Parallel computing concepts

Unit 7: Bioinformatics Toolkits and Database Systems

- Biopython and BioPerl
- NumPy and Pandas
- SciPy and Bioconductor
- SQL databases:
    - MySQL
    - PostgreSQL
- NoSQL databases:
    - MongoDB
- APIs and web services
- Web applications using:
    - Flask
    - Streamlit
    - Shiny

Unit 8: Structural Bioinformatics and Protein Modeling

- Protein structure hierarchy
- Ramachandran plots
- Protein Data Bank (PDB) formats
- Homology modeling
- Threading and ab initio prediction
- Structure validation:
    - PROCHECK
    - MolProbity
- Molecular visualization:
    - PyMOL
    - UCSF Chimera
    - VMD
- Protein folding and dynamics
- AI-based structure prediction using:
    - AlphaFold

Unit 9: Molecular Modeling and Computational Chemistry

- Molecular mechanics
- Force fields:
    - AMBER
    - CHARMM
    - GROMOS
    - MMFF
- Energy minimization

- Solvation models
- Molecular dynamics simulation
- Thermodynamics and free energy
- Quantum mechanics:
    - ab initio methods
    - Density Functional Theory (DFT)
- Molecular docking
- Virtual screening
- Protein-ligand interaction analysis
- GROMACS and NAMD basics

Unit 10: Cheminformatics and AI-Driven Drug Discovery

- Chemical databases
- Molecular descriptors and fingerprints
- RDKit applications
- QSAR and pharmacophore modeling
- Lead optimization
- ADMET prediction
- Rational drug design
- Structure-based and ligand-based drug discovery
- AI/ML in drug discovery
- Generative AI for molecule and peptide design
- Precision medicine and pharmacogenomics

Unit 11: Machine Learning in Bioinformatics

- Supervised and unsupervised learning
- Feature engineering
- Feature selection
- Dimensionality reduction:
    - PCA
    - t-SNE
    - UMAP
- Classification algorithms:
    - SVM
    - Random Forest
    - Decision Trees
    - k-NN
- Clustering:
    - K-means
    - Hierarchical clustering
- Cross-validation and evaluation metrics
- ML applications in:
    - Gene expression analysis
    - Protein classification
    - Disease prediction
    - Biomarker discovery

Unit 12: Artificial Intelligence and Deep Learning

- Fundamentals of Artificial Intelligence
- Neural networks
- Deep learning frameworks:
    - TensorFlow

    - PyTorch
- Convolutional Neural Networks (CNNs)
- Recurrent Neural Networks (RNNs), LSTM, GRU
- Transformers and attention mechanisms
- Large Language Models (LLMs)
- Protein language models
- AI in genomics and proteomics
- AI in drug discovery and disease diagnosis
- Explainable AI and ethical AI

Unit 13: Recent Advances and Emerging Technologies

- Single-cell sequencing
- Spatial transcriptomics
- Multi-omics systems medicine
- Synthetic biology
- CRISPR informatics
- Computational vaccinology
- Immunoinformatics
- AI-guided protein engineering
- Precision medicine
- Microbiome informatics
- Biomedical image analysis
- Quantum computing in biology
- Cancer genomics and rare disease informatics
- Ethical, legal, and social implications (ELSI)

References

1. Haddock, S.H.D. & Dunn, C.W. — *Practical Computing for Biologists*
2. Buffalo, V. — *Bioinformatics Data Skills*
3. Mount, D.W. — *Bioinformatics: Sequence and Genome Analysis*
4. Durbin, R. et al. — *Biological Sequence Analysis*
5. Brown, T.A. — *Genomes*
6. Shendure, J. & Ji, H. — “Next-Generation DNA Sequencing” (*Nature Biotechnology*)
7. Liebler, D.C. — *Introduction to Proteomics*
8. Klipp, E. et al. — *Systems Biology*
9. Cock, P.J.A. et al. — “Biopython: Freely Available Python Tools for Computational Biology and Bioinformatics” (*Bioinformatics*)
10. Mitchell, L. — *Bioinformatics Programming Using Python*
11. Jones, N.C. & Pevzner, P. — *An Introduction to Bioinformatics Algorithms*
12. Gusfield, D. — *Algorithms on Strings, Trees and Sequences*
13. Branden, C. & Tooze, J. — *Introduction to Protein Structure*
14. Leach, A.R. — *Molecular Modelling: Principles and Applications*
15. Bajorath, J. — *Chemoinformatics for Drug Discovery*
16. Larrañaga, P. et al. — “Machine Learning in Bioinformatics” (*Briefings in Bioinformatics*)
17. Goodfellow, I., Bengio, Y., Courville, A. — *Deep Learning*