

## 736 - Bioinformática

|                            |   |
|----------------------------|---|
| Carácter del curso         | Electivo/Optativo para las carreras de grado de Facultad de Química según corresponda |
| Semestre en que se dicta   | Par   |
| Número de créditos         | 7   |
| Carga horaria semanal (hs) | 4 horas por semana  |
| Previaturas                | Microbiología General, Bioquímica, Biología Molecular ó formación equivalente         |
| Cupo                       | 20  |

### Estructura Responsable:

Centro de Bioinformática Estructural-DETEMA

### Docentes Responsables:

Prof. Margot Paulino

### Docentes Colaboradores

Dr. Federico Iribarne  
M. Sc. Jorge Cantero  
M. Sc. Andres Camilo Ballesteros

### Objetivos:

La Bioinformática es una disciplina que integra métodos matemáticos, estadísticos y de ciencias de la computación, para analizar datos biológicos, biofísicos y bioquímicos. La Bioinformática intenta simplificar y organizar el manejo de grandes volúmenes de datos biológicos, mediante el desarrollo de estrategias de análisis que permitan extraer conocimiento de los mismos.

Este curso introductorio a la Bioinformática tiene como objeto hacer una presentación general de los conceptos más importantes dentro de la disciplina, sin profundizar demasiado en las consideraciones técnicas, que serán diferidas para otros cursos avanzados. Por ello, en este curso se hará énfasis en el empleo de los portales Web de Bioinformática, una manera económica y adecuada para acceder a los datos de la comunidad científica. Dichos portales no solamente hacen posible el acceso a los datos sino también su organización, permitiendo búsquedas e incluso análisis sofisticados a

través de los que se generan y prueban hipótesis sobre el significado de tales datos y se encuentran patrones de comportamiento aún no descritos. Su contenido se ha organizado en dos módulos, abarcando los dos principales desarrollos y aplicaciones de esta área científica. Por un lado, en el MODULO I: la genómica y el desarrollo de algoritmos de la llamada Bioinformática “lineal” (que usa la secuencia aminoacídica o nucleotídica como insumo y que desarrolla aplicaciones en genómica, transcryptómica y otras). Por otro lado, en el MODULO II la “ Bioinformática estructural” (que usa la información tridimensional de las estructuras a través de las coordenadas espaciales de los átomos de biomoléculas y ligandos).

### Contenido:

#### **Programa Teórico**

#### **MODULO I**

##### **1) Introducción (2 hrs)**

*Presentación de la especialización en Bioinformática. Presentación del curso. ¿Qué es la Bioinformática? Ejemplos simples de aplicación ¿Quién puede dedicarse a la Bioinformática? Algoritmos clásicos y bioinformáticos. Definición del problema en Bioinformática. Informática en Biología. Tipos informáticos en Bioinformática Genoma, proteoma y metaboloma. Sinopsis sobre los temas a tratar en el curso.*

##### **2) Repaso de Biología Molecular (2 hrs)**

*La célula como máquina que procesa materia, energía e información. Constructores de sistemas moleculares: DNA, RNA y proteínas. Estructura y contenido informático. El código genético. Estructura jerárquica del DNA. Cromosomas, genes, exones, intrones, regiones regulatorias, señales. Elementos repetidos. Replicación. Los genes codifican proteínas. Distribución aproximada de genes. Transcripción y postranscripción. mRNA y snRNA. Procesamiento y migración del mRNA maduro. Traducción. tRNA, importancia de la estructura. Proteínas. Estructura de proteínas y protein folding. Genomas de organelos.*

##### **3) Métodos de secuenciación (2hrs).**

*Antecedentes. Secuenciación por Maxxam y Gilbert. Método Sanger clásico. Avances en el método Sanger. Metodos de secuenciación masiva (NGS). Secuenciación en tiempo real de una base. Tecnología de nanoporos. Método ion semiconductor. Secuenciación por síntesis. Pirosecuenciación. Método por hibridación.*

|               |            |      |
|---------------|------------|------|
| Fecha         | MA-SGC-2-3 | V.01 |
| Página 2 de 7 |            |      |

### 4) Archivos y bases de datos (2hrs)

*La Web como soporte para la Bioinformática. Fundamentos de Bases de Datos y de "Information Retrieval". Bases de datos primarias y secundarias. Bases de Datos en la Web. Bases de datos de secuencias de ácidos nucleicos. Bases de datos de genomas. Bases de datos de secuencias de proteínas. Bases de datos de estructuras. Bases de datos especializadas ("boutique"). Bases de datos de expresión y proteómica. Bases de datos de caminos metabólicos. Bases de datos bibliográficas. Portales. ENTREZ. SRS. PIR. ExpASy. ENSEMBL.*

### 5) Técnicas del diseño de algoritmos (2hrs)

*Algoritmo y complejidad. Definiciones. Algoritmos biológicos y computacionales. Ejemplos. Algoritmos correctos e incorrectos. Iteración y recursividad. Algoritmos rápidos y lentos. Notación big-O. Técnicas de diseño: Búsqueda exhaustiva, Ramificación y poda, algoritmos greedy, programación dinámica, dividir y conquistar, algoritmos aleatorios, machine learning.*

### 6) Transcriptómica (2hrs)

*Conceptos generales. Secuenciación masiva y microarrays. Preparación de bibliotecas, secuenciado y análisis de datos. Cuantificación del ARNm en microarrays y en secuenciación masiva. Calidad de los datos. Principios biológicos de la cuantificación. Diseño experimental. Alineamientos: mapeando sobre genomas, sobre transcriptomas, ensamblados de novo. Expresión génica diferencial. Visualización y análisis de datos.*

### 6) Alineamiento de secuencias. Programación dinámica (2 hrs)

*Definición del problema y motivación biológica. Similitud y diferencia. Nomenclatura biológica e informática. Matrices de sustitución. Dayhoff PAM. PAM250. Henikoff BLOSUM. Alineamiento de a pares. Gráfico de puntos. Tipos de alineamiento (local, global, extremos libres). Medidas de similitud de secuencias. Alineamiento global, algoritmo de Needleman-Wunsch y Needleman-Wunsch-Sellers. Alineamiento local, algoritmo de Smith-Waterman. Esquemas de ponderación. Penalidad de ruptura. Penalidad de propagación. Programación dinámica. Alineamiento múltiple. Objetivo del alineamiento múltiple. Secuencia consenso. Generalización de programación dinámica al caso de N dimensiones. Complejidad computacional. Algoritmos de aproximación. MSA. Métricas de scoring, suma de pares.*

### 7) Alineamiento de secuencias. Heurísticas (2 hrs)

*Motivación. Problemas con los algoritmos de programación dinámica. Estrategias para*

|               |            |      |
|---------------|------------|------|
| Fecha         | MA-SGC-2-3 | V.01 |
| Página 3 de 7 |            |      |

acelerar la programación dinámica. Métodos heurísticos. *Algoritmos para alineamientos de a pares y múltiples: FASTA. BLAST. BLAST mejorado. Psi-BLAST y PSSM. CLUSTALO. Significación de un alineamiento. Sensibilidad y selectividad. Valor esperado. Alineamientos al azar y distribución de Gumbel para scores. Tests de significación.*

### **8) Análisis Filogenético (2 hrs)**

*Evolución. Significado. Teorías científicas. Filogenia y evolución. Filogenia molecular. Árboles filogenéticos. Ortólogos, parálogos y xenólogos. Datos usados en la construcción de árboles filogenéticos. Modelos de Jukes-Cantor y Kimura de distancia entre secuencias de DNA. Distancias de a pares. Distancia entre proteínas. Matrices PAM). Árboles como estructuras de datos. Métodos basados en distancias entre especies. Unión de vecinos y UPGMA. Métodos basados en caracteres (bases o aminoácidos). Parsimonia. Algoritmo de Fitch. Algoritmo de Sankoff. Máxima Verosimilitud. Aproximaciones probabilísticas. Evaluación de árboles. Bootstrap.*

### **9) Predicción génica (2 hrs)**

*Motivación. Estructura de genes en procariontas y eucariotas. Splicing. Clasificación de métodos de búsqueda génica. Métodos ab initio: basados en señales y basados en contenido. Métodos basados en homología. Métodos mixtos. Encontrando genes en procariontas. ORFs. Lectura en 6 marcos de referencia. Frecuencia de codones en exones. Hexámeros. Encontrando genes en eucariotas. HMM y redes neurales. Algoritmos. Detección de las regiones de promoción. TATA box. Matriz de peso posicional. Función de scoring.*

## **MÓDULO II**

### **10) Estructura tridimensional de proteínas (2hrs)**

*Estructura de proteínas. Estructura secundaria. Estructura del backbone. Conformación de cadenas laterales. Mapa de Ramachandran. Estructura terciaria y estructura cuaternaria. Clasificación estructural de proteínas: SCOP y CATH. Superposición de estructuras. Alineamiento estructural. Modelado por homología. Predicción de novo de estructuras tridimensionales de proteínas: métodos de primeros principios, redes neurales, threading, algoritmos genéticos, etc.*

### **12) Diseño de compuestos bioactivos y Medicina Química (2 hrs)**

*Generalidades sobre metodologías LBDD (Ligand Based Drug Design). Construcción de Bases de Datos de Ligandos. Definición del concepto de Farmacóforo y métodos de determinación del Farmacóforo.*

|               |            |      |
|---------------|------------|------|
| Fecha         | MA-SGC-2-3 | V.01 |
| Página 4 de 7 |            |      |

**13) QSAR(Quantitative Structure-Activity Relationship)-2D y QSAR-3D. (2 hrs).**

*Calculo de Descriptores fisicoquímicos. Modelos basados en fingerprints. Modelos obtenidos utilizando algoritmos genéticos.*

**14) Metodologías basadas en la estructura tridimensional de proteínas SBDD (Structure Based Drug Design) (2hrs)**

*Diseño de Fármacos basado en la estructura 3D del blanco farmacológico. Métodos de Anclaje Molecular (Docking y Docking reverso (target fishing)).*

**15) Dinámica Molecular de complejos Ligando-Biomolécula (2 hs).** *Algoritmos de dinámica molecular. Efecto solvente explícito. Fuerza iónica. Condiciones periódicas. Evaluación de la interacción ligando-biomolécula en equilibrio y en situación de no equilibrio (potencial de fuerza media).*

**TOTAL HORAS DE TEÓRICO: 30**

**Programa Práctico**

**1) Bases de Datos (2 hrs)**

*Búsqueda de información disponible sobre una secuencia génica utilizando herramientas online sobre bases de datos de secuencias de DNA (GenBank, RefSeq, EMBL, DDBJ), secuencias de proteína (GenPept, SwissProt, TrEMBL, PIR), expresión génica (UniGene, SAGE-Genie), genómicas (Genome Viewer, Ensembl), enzimáticas (Reactome), ontológicas (Gene Ontology) y bibliográficas (PubMed, MedLine). Familiarización con los grandes servidores online de Bioinformática a nivel mundial (NCBI, EBI, SIB, CIB-DDBJ).*

**2) Alineamiento de secuencias. Programación dinámica (2 hrs)**

*Aplicación manual de algoritmos Needleman-Wunsch y Smith-Waterman a problemas simples de alineamiento global y local de a pares. Alineamiento de a pares local y global usando online el programa LALIGN. Alineamiento múltiple con el programa MSA. Generación de matrices de puntos usando online el programa DOTLET.*

**3) Alineamiento de secuencias. Heurísticas (2 hrs)**

*Alineamientos de 2 secuencias sencillas usando online el programa bl2Seq. Búsquedas de homólogos en base de datos usando online los programas BLAST y FASTA. Búsqueda de homólogos remotos usando online el programa PSI-BLAST. Alineamiento múltiple online con el programa CLUSTALO.*

**4) Análisis Filogenético (4 hrs)**

*Aplicación manual de algoritmos UPGMA y Finch a un problema simple. Ejercicios sobre representación de New Hampshire. Análisis filogenético usando online distintos programas para métodos de distancias, máxima parsimonia y máxima verosimilitud*

| Fecha         | MA-SGC-2-3 | V.01 |
|---------------|------------|------|
| Página 5 de 7 |            |      |

aplicados a diferentes problemas evolutivos. Evaluación de árboles usando online el programa *BOOTSTRAP*.

**5) Búsqueda de Genes (4 hrs)**

*Predicción de secuencias codificantes usando online los programas ab initio ORFINDER, GENEID, GENESCAN, FGENESH y NETGENE2. Integración de homología de secuencias usando online los programas GENEWISE y GENOMESCAN. Detección de regiones reguladoras usando online los programas PROMOTER y TSSG.*

**6) Transcriptómica (2hrs)**

*Análisis de datos crudos de RNA-Seq: conversión, control de calidad y trimming de secuencias. Alineamientos. Estimación de los valores de expresión génica. Identificación de genes con expresión diferencial. Análisis de ontología y enriquecimiento.*

**7) Modelado de estructuras 3D de proteínas (4 hrs)**

*Métodos de Homología Molecular, alineamiento, selección y modelación. Modelado Ab Initio de proteína, modelación online de estructuras 3D de secuencias problema.*

**8) Diseño de Compuestos Bioactivos Basado en Ligandos (LBDD) (4 hrs)**

*Construcción de Bases de Datos de Ligandos y cálculo de descriptores. Construcción de farmacóforo de Ligandos. SAR y QSAR.*

**9) Diseño de Compuestos Bioactivos Basado en la Estructura molecular (SBDD) (4 hrs)** *Anclaje Molecular y dinámica molecular de complejos proteína-ligando. Medidas de la energía libre de interacción proteína-ligando.(4hs)*

**10) Dinámica Molecular(4 hrs).** *Dinámica molecular. Efecto solvente y fuerza iónica. Medidas de la interacción ligando-proteínas.*

**TOTAL DE HORAS DE PRÁCTICO: 32**

**CRÉDITOS: (TEÓRICOS (30X2)+PRÁCTICOS (32X1.5))/ 15 = 7**

**Modalidad del Curso:**

|   | Teórico | Practico | Laboratorio | Otros (*) |
|---|---------|----------|-------------|-----------|
| Asistencia Obligatoria                    | No      | Sí       |             |           |
| Modalidad Flexible (carga horaria mínima) | Si      |          | No          |           |

(\*) Especificar (talleres, seminarios, visitas, tareas de campo, pasantías supervisadas, etc.)

### **Régimen de ganancia:**

Se requerirá la asistencia al 75% de las clases prácticas. Un guarismo de asistencia menor significará la pérdida del curso.

Se realizarán 2 evaluaciones, la primera un parcial de contenido teórico (25 puntos) y la segunda una exposición oral de tareas realizadas en el práctico (35 puntos) que suman 60 puntos. La aprobación se logrará obteniendo entre 18 y 30 puntos y la exoneración obteniendo más de 30 puntos.

### **Bibliografía:**

D.E. Krane and M.L. Raymer, *Fundamental Concepts of Bioinformatics*, Pearson Education, 2003.

N. C. Jones and P. A. Pevzner, *An Introduction to Bioinformatics Algorithms*, MIT press, 2004.

P. Compeau, and P. A. Pevzner, *Bioinformatics algorithms: an active learning approach*. Vol.1, Active Learning Publishers, 2015

P. Compeau, and P. A. Pevzner, *Bioinformatics algorithms: an active learning approach*. Vol.2, Active Learning Publishers, 2015

M. Zvelebil and J. O. Baum, *Understanding Bioinformatics*, Garland Science, 2008.

C.A. Orengo, D.T. Jones and J.M.Thornton, *Bioinformatics: Genes, Proteins and Computers*, Roulledge, 2003.

A. M. Lesk, *Introduction to Bioinformatics*, Oxford University Press, 2002.

D. Mount, *Bioinformatics: Sequence and genome analysis*, Cold Spring Harbor Laboratory Press, 2001.

P. A. Pevzner, *Computational Molecular Biology: An Algorithmic Approach*, MIT press, 2000.

P. Baldi and S. Brunak, *Bioinformatics: the machine learning approach (2nd edition)*, MIT press, 2001.

|               |            |      |
|---------------|------------|------|
| Fecha         | MA-SGC-2-3 | V.01 |
| Página 7 de 7 |            |      |