

Universidad de La Habana
Facultad de Matemática y Computación



Simulador de sistema inmune para predecir la respuesta de lactantes a vacunas antineumocócicas conjugadas

Autora: Amanda Cordero Lezcano

**Tutores: MSc. Celia T. González González
MSc. Wilfredo Morales Lezca**

Asesora: DrC. Dagmar García Rivera

Trabajo de Diploma
presentado en opción al título de
Licenciado en Ciencia de la Computación



Junio 2025

“Todos los modelos son incorrectos, pero algunos son útiles”
George E. Pelham Box

Opinión de los tutores

Título de la tesis: Simulador de sistema inmune para predecir la respuesta de lactantes a vacunas antineumocócicas conjugadas

Autora: Amanda Cordero Lezcano

El trabajo de diploma presentado por la estudiante Amanda Cordero Lezcano constituye una contribución significativa al campo de la inmunología computacional y al desarrollo de vacunas, al abordar un problema de alto impacto para la salud pública: la predicción de la respuesta inmune en lactantes mediante modelos computacionales.

Este proyecto se enmarca en la colaboración estratégica entre la Facultad de Matemática y Computación de la Universidad de La Habana y el Instituto Finlay de Vacunas (IFV), institución con más de tres décadas de experiencia en el diseño de vacunas innovadoras. Bajo el compromiso social de fortalecer los Programas Ampliados de Inmunización, y aprovechando las lecciones del Proyecto Soberana, esta alianza académico-científica impulsa el uso de herramientas computacionales en el desarrollo de nuevos candidatos vacunales, articulando la excelencia investigativa del IFV con el rigor metodológico de la modelación matemática.

La investigación se desarrolla en un contexto prioritario para Cuba: la necesidad de contar con modelos predictivos eficientes que permitan evaluar formulaciones vacunales como el candidato Quimi-Vio7® antes de emprender ensayos clínicos a gran escala.

Amanda demostró una capacidad excepcional para integrar conocimientos avanzados en inmunología, matemáticas y programación —muchos de ellos fuera del currículo tradicional de la carrera—, destacando por su rigor científico, creatividad metodológica y autonomía intelectual. Diseñó un simulador que modela procesos clave del sistema inmune, como la activación de células B, la dinámica de los centros germinales y la generación de memoria inmunológica.

Entre sus principales aportes técnicos, sobresale la implementación de la selección clonal mediante distribución de Boltzmann, así como el uso de la

métrica de Chamfer para la validación cuantitativa del modelo.

La herramienta desarrollada en Python integra bibliotecas científicas como NumPy y SciPy, y fue aplicada sobre datos clínicos reales del ensayo PCV7-TT del IFV. El modelo fue validado con resultados experimentales, logrando ajustes prometedores para serotipos como 18C (Chamfer = 2.773) y 5 (3.241). La autora identificó limitaciones (como la menor precisión para el serotipo 14) y propuso mejoras como la inclusión de mecanismos de inmunidad innata, lo cual evidencia una visión crítica y prospectiva.

Durante todo el proceso, Amanda mantuvo una comunicación fluida y proactiva con el equipo de tutoría. Mostró una notable perseverancia frente a las limitaciones de datos y recursos, así como una actitud investigativa sostenida que le permitió apropiarse de contenidos complejos como la inmunología en lactantes, a través del autoaprendizaje y el análisis profundo de la bibliografía especializada.

El trabajo representa un aporte metodológico original en la intersección entre computación y biomedicina, y sienta las bases para una línea de investigación sostenible en biocomputación aplicada a la salud pública cubana.

Amanda ha demostrado ser una investigadora emergente con gran potencial, capaz de combinar creatividad técnica, rigor analítico y compromiso social. En un tiempo limitado y bajo condiciones retadoras, logró alcanzar con solvencia los objetivos propuestos.

Como toda obra humana, su tesis es perfectible —lo cual es natural en una primera versión—, pero como tutores estamos plenamente satisfechos con su desempeño. Consideramos que su trabajo merece la calificación de Excelente, y confiamos en que continuará realizando aportes valiosos al desarrollo de herramientas computacionales en colaboración con instituciones como el IFV.

MSc. Celia T. González González
 MSc. Wilfredo Morales Lezca
 Facultad de Matemática y Computación
 Universidad de la Habana

Resumen

Las enfermedades causadas por *Streptococcus pneumoniae* representan un importante problema de salud pública, especialmente en lactantes. Este trabajo presenta el desarrollo de un simulador computacional del sistema inmune infantil para predecir la efectividad de la vacuna antineumocócica conjugada Quimi-Vio7®. El modelo implementado representa los componentes clave del sistema inmune, incluyendo células B, antígenos y la dinámica de los centros germinales. Se emplearon datos reales de ensayos clínicos con la vacuna VCN7-T para validar el modelo, utilizando la distancia de Chamfer como métrica de similitud entre los resultados simulados y los datos experimentales. Este enfoque computacional ofrece una herramienta para acelerar el desarrollo y evaluación de vacunas, reduciendo la dependencia de ensayos clínicos extensivos.

Abstract

Diseases caused by *Streptococcus pneumoniae* represent a significant public health problem, particularly in infants. This work presents the development of a computational simulator of the infant immune system to predict the effectiveness of the pneumococcal conjugate vaccine Quimi-Vio7®. The implemented model represents key immune system components, including B cells, antigens, and germinal center dynamics. Real clinical trial data with the VCN7-T vaccine were used to validate the model, employing the Chamfer distance as a similarity metric between simulated results and experimental data. This computational approach offers a tool to accelerate vaccine development and evaluation, reducing dependence on extensive clinical trials.

Índice general

Introducción	9
1. Preliminares	12
1.1. Perspectiva Inmunológica	12
1.1.1. Características clave	12
1.1.2. Inmunología básica	13
1.1.3. Vacunas antineumocócicas conjugadas	14
1.1.4. Examen ELISA	14
1.2. Perspectiva Matemático-Computacional	15
1.2.1. Metaheurísticas	15
1.2.2. Métricas para comparar conjuntos (nubes) de puntos .	17
1.2.3. Distribución de Boltzmann	20
1.2.4. Acercamiento a la concurrencia	20
2. Estado del Arte	22
2.1. Enfoques Computacionales	22
2.1.1. Modelos de Transmisión Dinámica	22
2.1.2. Enfoques de Inmunología de Sistemas	23
2.2. Aplicaciones de los Modelos en Estrategias de Vacunación . .	24
2.2.1. Predicción del Impacto Poblacional	24
2.2.2. Diseño de Inmunógenos	24
2.3. Respuestas Inmunológicas Específicas a Vacunas	24
2.3.1. Modelación de Respuestas Inmunes Innatas	25
2.3.2. Caracterización de Respuestas Inmunes Adaptativas .	25
3. Metodología	26
3.1. Parámetros	26
3.2. Estructura	27
3.3. Modelado de componentes biológicos	27

3.4. Centros Germinales	28
3.5. Sistema Inmune	32
3.6. Salida del simulador	33
4. Resultados	34
4.1. Datos para la Validación del Modelo	34
4.1.1. Limpieza de Datos	34
4.2. Entrenamiento	38
4.2.1. Exploración del espacio de parámetros	38
4.2.2. Función Objetivo	39
4.3. Validación	39
Conclusiones	42
Recomendaciones	43
Referencias	44
Anexos	47

Introducción

Las enfermedades causadas por *Streptococcus pneumoniae* (neumococo) constituyen un grave problema de salud pública a nivel mundial, siendo responsables de una significativa morbilidad y mortalidad, especialmente en las edades extremas de la vida, como niños y adultos mayores [1]. Este patógeno, grampositivo, encapsulado y con gran diversidad de serotipos, es el agente etiológico de infecciones como neumonía, meningitis y otitis media, y se le atribuye alrededor de un millón de defunciones anuales a nivel global [2]. En Cuba, la vigilancia epidemiológica y el desarrollo de vacunas han sido esenciales para reducir la incidencia de estas enfermedades, aunque la variabilidad genética del neumococo y la aparición de cepas resistentes siguen representando desafíos importantes [3].

El desarrollo de vacunas antineumocócicas conjugadas ha sido un avance crucial en la prevención de estas infecciones, pero su diseño y evaluación requieren largos periodos de investigación y ensayos clínicos costosos, lo que retrasa la introducción de nuevas formulaciones y limita su impacto inmediato en la salud pública [4].

Diversos estudios han explorado los mecanismos inmunológicos involucrados en la defensa contra el neumococo, destacando el papel de la inmunidad innata y adaptativa. El reconocimiento inicial del patógeno es mediado por receptores de reconocimiento de patrón (PRRs), como los receptores tipo Toll y NOD, que desencadenan la activación de factores de transcripción y la producción de citoquinas inflamatorias, esenciales para la activación de la respuesta inmune específica. Sin embargo, la gran diversidad de serotipos, junto con los mecanismos de evasión bacteriana, hace que la predicción de la efectividad de las vacunas mediante métodos tradicionales sea un proceso lento y costoso [5].

La hipótesis que guía esta investigación es que un simulador computacional del sistema inmune de un lactante puede predecir de manera confiable la efectividad de la vacuna Quimi-Vio7®, permitiendo optimizar el proceso de desarrollo y evaluación de vacunas antineumocócicas conjugadas en Cuba.

Los modelos matemáticos permiten representar y predecir situaciones epidémicas, estimar escenarios futuros y evaluar la efectividad de distintas estrategias de vacunación, facilitando la toma de decisiones en salud pública [6]. Gracias a estos modelos, es posible simular la respuesta inmune a diferentes formulaciones vacunales, optimizar el diseño y desarrollo de nuevas vacunas. Además, la simulación computacional permite abordar la complejidad de las interacciones entre células y moléculas del sistema inmune, superando las limitaciones de los modelos tradicionales basados únicamente en ecuaciones diferenciales. Herramientas como los autómatas celulares, los sistemas basados en agentes y la inteligencia artificial están revolucionando el desarrollo vacunal, permitiendo la creación de vacunas más potentes y estables, así como la adaptación ágil a nuevas variantes y contextos epidemiológicos cambiantes [7]. En consecuencia, la integración de la matemática y la computación en la investigación vacunal no solo contribuye al avance científico, sino que también agiliza y optimiza la respuesta frente a los desafíos actuales y futuros en la prevención de enfermedades inmunoprevenibles.

Quimi-Vio7® es una vacuna cubana desarrollada a partir de polisacáridos capsulares (PsC) de diferentes serotipos de *Streptococcus pneumoniae*, conjugados a una proteína portadora. La misma se encuentra registrada para su uso en población preescolar. Sin embargo el ensayo clínico que demostraría su efectividad en lactantes fue interrumpido por la pandemia de COVID-19. Durante este período se demostró la necesidad y posibilidad de acortar los tiempos de producción de vacunas en el país. El presente trabajo contribuye al establecimiento de una nueva línea de investigación en Cuba orientada hacia la predicción de la efectividad de vacunas basada en serotipos, con especial énfasis en las vacunas neumocócicas conjugadas.

El objetivo general de esta investigación es el desarrollo de un simulador computacional del sistema inmune de un lactante para la predicción de la efectividad de la vacuna Quimi-Vio7®.

Los objetivos específicos de esta investigación son:

- Analizar los principales mecanismos inmunológicos involucrados en la respuesta a la vacuna antineumocócica conjugada.
- Diseñar y programar un modelo computacional que represente la interacción entre el sistema inmune y *Streptococcus pneumoniae*.
- Validar los resultados del modelo utilizando los datos del Estudio en Lactantes Fase I/II para PCV7-TT (Instituto Finlay de Vacunas, 2019)

Este enfoque contribuye tanto al avance teórico en inmunología computacional como a la optimización práctica del proceso de desarrollo de vacunas de este tipo, con potencial para reducir costos y tiempos de investigación y mejorar la toma de decisiones en salud pública.

La tesis se estructura en cuatro capítulos:

1. **Preliminares:** Se sientan las bases biológicas y matemático - computacionales para el desarrollo de la investigación.
2. **Estado del Arte:** Analiza los antecedentes y actualidad en cuanto a la modelación de sistema inmune y su interacción con vacunas.
3. **Metodología:** Describe el diseño y desarrollo del simulador.
4. **Experimentación y Resultados:** Expone y analiza los resultados obtenidos con el simulador, además de las fuentes de datos utilizadas y los criterios de validación empleados.

Capítulo 1

Preliminares

En este capítulo se sientan las bases para el desarrollo de la investigación.

1.1. Perspectiva Inmunológica

Streptococcus pneumoniae es una bacteria Gram-positiva, encapsulada y con forma lanceolada, que coloniza las vías respiratorias superiores en humanos. Es un patógeno de relevancia clínica global, asociado tanto a enfermedades invasivas (como meningitis, neumonía bacteriémica y sepsis) como a infecciones no invasivas (otitis media aguda y sinusitis) [1].

1.1.1. Características clave

- **Variabilidad serotípica:** Existen más de 90 serotipos¹ diferenciados por la composición de su cápsula polisacárida, aunque apenas 12 de ellos son responsables de más del 80% de las infecciones neumocócicas invasoras [8]
- **Poblaciones vulnerables:** Lactantes, adultos mayores (mayores de 65 años) e individuos inmunocomprometidos presentan mayor riesgo de infección grave.

Importancia de la vacunación

La diversidad de serotipos y la emergencia de cepas resistentes a antibióticos subrayan la importancia de estrategias preventivas como la vacunación.

¹Microorganismo infeccioso clasificado según los antígenos presentes en su superficie celular.

Las vacunas conjugadas han demostrado reducir significativamente la incidencia de enfermedad neumocócica al inducir inmunidad específica frente a los serotipos más prevalentes [9].

1.1.2. Inmunología básica

El sistema inmunitario humano responde a *Streptococcus pneumoniae* mediante dos grandes componentes [10]:

- **Inmunidad innata:** Primera línea de defensa, compuesta por células como neutrófilos, macrófagos y células dendríticas, así como factores humorales como el sistema del complemento.
- **Inmunidad adaptativa:** Proporciona especificidad y memoria inmunológica, mediada por linfocitos T ($CD4^+$, $CD8^+$) y B.

Respuesta inmune a *Streptococcus pneumoniae*: linfocitos B, memoria inmunológica y anticuerpos

Los linfocitos B son fundamentales en la defensa frente a *Streptococcus pneumoniae*, especialmente tras la vacunación:

- Producen anticuerpos específicos que neutralizan al patógeno.
- Experimentan maduración de afinidad y cambio de isotipo (por ejemplo, de IgM a IgG).
- Generan células de memoria que permiten una respuesta más rápida y eficaz ante exposiciones futuras.

Características inmunológicas en lactantes

- **Anticuerpos maternos:**
 - Transferencia placentaria de IgG (especialmente IgG1) mediante el receptor FcRn.
 - Protección adicional a través de IgA secretora en la leche materna.
 - Posible *blunting* (interferencia de anticuerpos maternos con la respuesta vacunal), aunque su impacto clínico es limitado.[11]
- **Desarrollo inmune:** El sistema inmunitario neonatal es funcional pero inmaduro, dependiendo inicialmente de la inmunidad materna.

Definiciones clave

- **Inmunogenicidad:** Capacidad de una vacuna para inducir una respuesta inmunitaria medible (por ejemplo, producción de anticuerpos).
- **Efectividad clínica:** Protección real conferida contra la enfermedad en condiciones del mundo real.
- **Conjugación proteica:** Las vacunas conjugadas unen polisacáridos capsulares a proteínas transportadoras para potenciar la respuesta inmunitaria, especialmente en niños pequeños.

1.1.3. Vacunas antineumocócicas conjugadas

Las vacunas antineumocócicas conjugadas (PCV por sus siglas en inglés) emplean polisacáridos capsulares conjugados a proteínas para inducir una respuesta inmunitaria robusta y duradera y generar memoria inmunológica mediante la activación de linfocitos T y B.

Cobertura y correlatos de protección

Las formulaciones actuales de PCV cubren los serotipos más prevalentes y virulentos. Los correlatos de protección² pueden variar según el serotipo y la vacuna específica, y son utilizados como criterios en ensayos clínicos y en la evaluación de programas de vacunación [9].

1.1.4. Examen ELISA

El examen ELISA (Ensayo de Inmunoabsorción Ligado a Enzima) es una técnica inmunoenzimática ampliamente utilizada para la detección y cuantificación específica de antígenos o anticuerpos en muestras biológicas. El principio fundamental del ELISA se basa en la unión específica entre un anticuerpo y un antígeno, donde uno de estos componentes está inmovilizado en una superficie sólida, generalmente una placa de microtitulación.

La detección se realiza mediante un anticuerpo conjugado con una enzima, que al reaccionar con un sustrato específico produce un cambio de color proporcional a la cantidad de analito presente en la muestra. Este cambio es medido espectrofotométricamente, permitiendo cuantificar la concentración del antígeno o anticuerpo.

Existen diferentes formatos de ELISA, entre ellos:

²Mediciones de parámetros inmunitarios que permiten predecir el grado de protección contra la infección o enfermedad inducida por un patógeno.

- **ELISA directo:** el anticuerpo marcado con enzima se une directamente al antígeno inmovilizado.
- **ELISA indirecto:** utiliza un anticuerpo primario para detectar el antígeno y un anticuerpo secundario marcado con enzima para amplificar la señal.
- **ELISA sándwich:** un anticuerpo captura el antígeno, y otro anticuerpo marcado con enzima detecta el antígeno capturado, aumentando la sensibilidad y especificidad.
- **ELISA competitivo:** el antígeno en la muestra compite con un antígeno marcado por la unión a un anticuerpo específico.

Esta técnica es fundamental en diagnóstico clínico, investigación biomédica y control de calidad, debido a su alta sensibilidad, especificidad y rapidez [12].

1.2. Perspectiva Matemático-Computacional

En esta sección se analizan los conceptos teóricos sobre los que se fundamenta esta investigación. Temas como: Metaheurísticas, métricas de similitud y concurrencia serán tratados en las siguientes subsecciones.

1.2.1. Metaheurísticas

Las metaheurísticas son estrategias de alto nivel diseñadas para encontrar soluciones aproximadas a problemas de optimización complejos, especialmente aquellos donde los métodos exactos son computacionalmente inviables. Estas técnicas, inspiradas frecuentemente en procesos naturales o comportamientos sociales, exploran espacios de soluciones de manera inteligente para escapar de óptimos locales y aproximarse al óptimo global [13].

A continuación se presentan algunas de las metaheurísticas más conocidas:

Recocido Simulado (Simulated Annealing)

Inspirado en el proceso metalúrgico de recocido, donde un material se calienta y luego se enfría gradualmente para reducir sus defectos [14]. Este método:

- Comienza con una solución inicial s_0 y temperatura alta T_0

- En cada iteración, genera una solución vecina s'
- Acepta soluciones mejores ($f(s') < f(s)$) siempre
- Acepta soluciones peores con probabilidad $P = e^{-\Delta f/T}$ donde $\Delta f = f(s') - f(s)$
- Reduce temperatura gradualmente según esquema $T_{k+1} = \alpha T_k$ ($\alpha \in (0, 1)$)

La curva de enfriamiento (disminución de T) controla el balance exploración-explotación:

$$T(k) = \frac{T_0}{\ln(1+k)} \quad (\text{esquema logarítmico}) \quad (1.1)$$

Algoritmos Genéticos (Genetic Algorithms)

Basados en la evolución biológica mediante selección natural [15].

- Representan soluciones como cromosomas (vectores)
- Operadores clave:
 - **Selección:** Prefiere soluciones con mejor *fitness*
 - **Cruzamiento:** Combina partes de dos soluciones
 - **Mutación:** Modifica aleatoriamente componentes
- Mantienen una población de soluciones diversa
- Reemplazo generacional o estado estable

La aptitud (fitness) de una solución s_i se calcula como:

$$fitness(s_i) = \frac{1}{1 + f(s_i)} \quad (1.2)$$

Optimización por Colonia de Hormigas (Ant Colony Optimization)

La Optimización por Colonia de Hormigas es una metaheurística bioinspirada que emula el comportamiento colectivo de las hormigas reales en su búsqueda de caminos óptimos hacia fuentes de alimento. Cuando las hormigas exploran su entorno, depositan sustancias químicas llamadas feromonas que sirven como señalización para otras hormigas [16]. Este mecanismo de comunicación indirecta (estigmergia) permite a la colonia encontrar rutas eficientes mediante:

- Retroalimentación positiva: Caminos más cortos acumulan más feromonas al ser transitados más frecuentemente
- Evaporación dinámica: Las feromonas se evaporan gradualmente, evitando convergencia prematura
- Exploración estocástica: Decisiones basadas en probabilidades que balancean explotación y exploración

El proceso sigue estos pasos iterativos:

1. *Inicialización*: Feromonas uniformes en todos los caminos ($\tau_{ij} = \tau_0$)
2. *Construcción de soluciones*: Cada hormiga k construye una solución completa mediante selección probabilística:

$$p_{ij}^k = \frac{[\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{l \in N_i^k} [\tau_{il}]^\alpha \cdot [\eta_{il}]^\beta}$$

donde η_{ij} es información heurística (ej. $1/d_{ij}$)

3. *Actualización de feromonas*: Combinación de evaporación y depósito:

$$\tau_{ij} \leftarrow (1 - \rho) \cdot \tau_{ij} + \sum_{k=1}^m \Delta \tau_{ij}^k$$

$$\Delta \tau_{ij}^k = \begin{cases} Q/L_k & \text{si la hormiga } k \text{ usó } (i, j) \\ 0 & \text{otro caso} \end{cases}$$

4. *Reforzamiento elitista*: Las mejores soluciones históricas reciben depósito adicional

1.2.2. Métricas para comparar conjuntos (nubes) de puntos

Para evaluar la similitud entre dos nubes de puntos, se utilizan diversas métricas según el contexto de análisis. A continuación se presentan los métodos más relevantes:

Distancia Euclidiana Promedio

Mide la proximidad espacial absoluta entre puntos correspondientes:

$$D_{\text{avg}} = \frac{1}{n} \sum_{i=1}^n \sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2} \quad (1.3)$$

Características:

- Interpretación física directa (unidades espaciales)
- Sensible a traslaciones globales
- Cálculo computacionalmente eficiente $O(n)$

Similitud del Coseno

Evalúa la orientación relativa de los vectores posición, independiente de su magnitud:

$$S_{\text{cos}} = \frac{\sum_{i=1}^n (x_i x'_i + y_i y'_i)}{\sqrt{\sum_{i=1}^n (x_i^2 + y_i^2)} \cdot \sqrt{\sum_{i=1}^n (x'^2_i + y'^2_i)}} \quad (1.4)$$

Aplicaciones:

- Comparación de formas geométricas
- Análisis donde la posición absoluta es irrelevante
- Sistemas de coordenadas relativas

Métrica Basada en Trabajo de Transformación

Calcula la energía requerida para transformar una nube en otra:

$$S_{\%} = 100 \cdot \left(1 - \frac{W}{W_{\text{máx}}} \right) \quad (1.5)$$

donde W es el trabajo computado y $W_{\text{máx}}$ el trabajo máximo teórico.

Implementación:

1. Calcular la transformación óptima (traslación, rotación)
2. Integrar las diferencias posicionales residuales
3. Normalizar respecto a la peor transformación posible

Métricas de Clustering

Evalúan relaciones intra e inter-conjuntos:

$$D_{\text{intra}}^A = \frac{1}{n(n-1)} \sum_{i \neq j} \|p_i^A - p_j^A\| \quad (1.6)$$

$$D_{\text{inter}}^{\text{mín}} = \min_{i,j} \|p_i^A - p_j^B\| \quad (1.7)$$

$$H_{\text{hausdorff}} = \max \left(\sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|a - b\| \right) \quad (1.8)$$

Interpretación:

- D_{intra} : Compacidad de cada nube
- $D_{\text{inter}}^{\text{mín}}$: Proximidad mínima entre nubes
- $H_{\text{hausdorff}}$: Máxima discrepancia entre fronteras

Distancia Chamfer

Cuantifica la similitud entre dos conjuntos de puntos en el espacio. Evalúa la proximidad bidireccional entre cada punto de un conjunto y su vecino más cercano en el otro conjunto.

El cálculo de la distancia de Chamfer se realiza en dos pasos principales:

1. Para cada punto en el primer conjunto, se encuentra el punto más cercano en el segundo conjunto y se calcula la distancia euclidiana entre ellos. Se obtiene el promedio de estas distancias.
2. Se repite el procedimiento invirtiendo los roles de los conjuntos, es decir, desde el segundo conjunto hacia el primero.

Se define como la suma (o promedio) de estos dos valores, formalmente expresada como:

$$D_{\text{Chamfer}}(S_1, S_2) = \frac{1}{|S_1|} \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2 + \frac{1}{|S_2|} \sum_{y \in S_2} \min_{x \in S_1} \|y - x\|_2$$

donde $\|\cdot\|_2$ denota la distancia euclidiana.

Esta métrica es relevante porque permite evaluar de forma robusta la similitud entre nubes de puntos, incluso cuando los puntos no están emparejados uno a uno.

1.2.3. Distribución de Boltzmann

La **distribución de Boltzmann** es un concepto fundamental en física estadística que describe cómo las partículas en un sistema en equilibrio térmico distribuyen su energía entre los diferentes estados accesibles [17]. Esta distribución establece que la probabilidad de encontrar una partícula en un estado con energía E_i a una temperatura T está dada por:

$$P(E_i) = \frac{1}{Z} g_i e^{-\frac{E_i}{k_B T}} \quad (1.9)$$

donde:

- $P(E_i)$: probabilidad del estado con energía E_i ,
- g_i : degeneración del nivel de energía (número de estados con la misma energía),
- k_B : constante de Boltzmann,
- T : temperatura absoluta,
- Z : función de partición, definida como $Z = \sum_i g_i e^{-\frac{E_i}{k_B T}}$, y que normaliza la distribución para que la suma de todas las probabilidades sea igual a uno.

Esta distribución es válida para sistemas clásicos en equilibrio térmico y es especialmente útil cuando las partículas son distinguibles y las interacciones entre ellas pueden considerarse débiles o promediadas.

1.2.4. Acercamiento a la concurrencia

La programación concurrente es un paradigma esencial para maximizar la eficiencia en sistemas modernos, permitiendo la ejecución aparentemente simultánea de múltiples tareas. Su dominio es crucial para desarrollar aplicaciones escalables y receptivas, aunque introduce complejidades que requieren técnicas especializadas para garantizar la corrección y estabilidad del software. Los contenidos tratados en lo adelante están fundamentados en lo tratado en [18].

Concurrencia vs. Paralelismo

La concurrencia y el paralelismo son conceptos fundamentales en la programación moderna, aunque a menudo se confunden debido a su estrecha relación. La concurrencia se refiere a la capacidad de gestionar múltiples tareas de forma que parezcan avanzar al mismo tiempo, aunque no necesariamente se ejecuten simultáneamente. Esto se logra mediante la planificación eficiente del procesador, donde un único núcleo puede alternar entre diferentes tareas en intervalos de tiempo muy pequeños, creando la ilusión de ejecución simultánea.

La concurrencia permite que varias tareas progresen de manera ordenada incluso cuando compiten por recursos limitados.

Por otro lado, el paralelismo implica la ejecución real y simultánea de múltiples tareas utilizando recursos físicos separados, como varios núcleos de CPU o dispositivos de cómputo dedicados. Su enfoque está orientado a acelerar procesos computacionalmente intensivos dividiendo el trabajo en partes que pueden resolverse en paralelo. Aunque todo paralelismo implica concurrencia, ya que las tareas deben coordinarse, no toda concurrencia requiere paralelismo. Esta distinción es crucial a la hora de diseñar sistemas que prioricen el rendimiento o la interacción con el usuario.

Hilos

Un hilo (*thread*) es la unidad básica de ejecución dentro de un proceso, gestionada directamente por el sistema operativo. Los hilos permiten dividir un programa en componentes que pueden ejecutarse concurrentemente, lo que resulta especialmente útil para tareas que requieren manejo simultáneo de distintas operaciones. Una característica clave de los hilos es que comparten el espacio de memoria y los recursos del proceso padre, lo cual facilita la comunicación entre ellos, pero también introduce desafíos relacionados con la consistencia de datos.

Retos en Programación Concurrente

La programación concurrente presenta una serie de desafíos, siendo uno de los más comunes la condición de carrera (*race condition*). Este fenómeno ocurre cuando dos o más hilos intentan acceder y modificar un recurso compartido simultáneamente, causando resultados inconsistentes o impredecibles. Para evitar este tipo de problemas, se utiliza la exclusión mutua, que garantiza que solo un hilo pueda acceder a un recurso crítico en un momento dado.

Entre las herramientas más utilizadas para implementar exclusión mutua están los semáforos, *mutex* y monitores. Estas estructuras permiten controlar el acceso a secciones críticas del código y asegurar la integridad de los datos compartidos. Sin embargo, su uso incorrecto puede llevar a otro problema conocido como interbloqueo (*deadlock*), que sucede cuando dos o más hilos quedan esperando indefinidamente por recursos que nunca serán liberados porque cada uno posee un recurso que el otro necesita.

Capítulo 2

Estado del Arte

La modelación matemático-computacional ha emergido como una herramienta para comprender y predecir complejas interacciones en el sistema inmunológico humano. Los documentos analizados en esta sección revelan algunas metodologías complementarias que abordan estos procesos desde perspectivas poblacionales e individuales.

2.1. Enfoques Computacionales

En la literatura científica actual se identifican dos metodologías computacionales para modelar la respuesta inmune a las vacunas neumocócicas conjugadas.

2.1.1. Modelos de Transmisión Dinámica

Los modelos de transmisión dinámica basados en ecuaciones diferenciales constituyen uno de los enfoques fundamentales para analizar el impacto de las vacunas neumocócicas a nivel poblacional [19]. Estos modelos matemáticos utilizan sistemas de ecuaciones diferenciales para simular la propagación de la enfermedad neumocócica invasiva (IPD por sus siglas en inglés) a través de poblaciones estratificadas¹. El estudio [9] demuestra que estos modelos incorporan múltiples componentes críticos para garantizar su precisión predictiva.

La estratificación por edad es un elemento esencial de estos modelos, dividiendo la población en categorías específicas que reflejan las diferencias en susceptibilidad y transmisión entre diversos grupos etarios. Esta segmentación

¹Clasificadas en grupos o capas diferenciadas.

permite capturar tanto los efectos directos como indirectos de la vacunación en diferentes cohortes poblacionales.

Adicionalmente, estos modelos categorizan los serotipos neumocócicos según su inclusión en diversas formulaciones vacunales, facilitando la evaluación del impacto diferencial de las vacunas en la circulación de serotipos específicos.

2.1.2. Enfoques de Inmunología de Sistemas

En contraste con los modelos poblacionales, los enfoques de inmunología de sistemas se centran en las respuestas inmunológicas individuales a la vacunación [20].

Las soluciones computacionales empleadas en inmunología de sistemas permiten el procesamiento de conjuntos de datos inmunológicos complejos, facilitando la identificación de biomarcadores predictivos² de respuesta vacunal. Esta capacidad tiene implicaciones significativas para la personalización de estrategias de vacunación según perfiles inmunológicos individuales. En [21] los autores presentan DAIMP (Discriminant Analysis via Mixed-Integer Programming), un modelo de aprendizaje automático que predice firmas genéticas asociadas a la respuesta inmune. A través de este enfoque se analizan datos genómicos complejos e identifican conjuntos mínimos de genes predictivos capaces de determinar si un individuo desarrollará inmunidad robusta o débil tras la vacunación.

Avances recientes en el modelado de respuestas inmunes celulares han permitido comprender los mecanismos moleculares subyacentes a la generación de memoria inmunológica. El estudio de Murugan et al. [22] demuestra mediante secuenciación de genes de inmunoglobulina a nivel de célula única, y producción de anticuerpos monoclonales recombinantes, como la selección clonal de precursores de células B *naïve* y de memoria, impulsa respuestas protectoras en infecciones controladas de malaria. El desarrollo de modelos matemáticos que explican cómo la eficiencia de la maduración por afinidad disminuye con la complejidad del antígeno proporciona un marco teórico para entender la dinámica de las respuestas B a vacunas conjugadas.

²Característica biológica medible que puede utilizarse para identificar prospectivamente a los pacientes que tienen mayor probabilidad de beneficiarse de una intervención específica, o por el contrario, presentar efectos adversos.

2.2. Aplicaciones de los Modelos en Estrategias de Vacunación

Los modelos computacionales ofrecen aplicaciones para la planificación e implementación de estrategias de vacunación neumocócica a diferentes niveles.

2.2.1. Predicción del Impacto Poblacional

Los modelos de transmisión dinámica han demostrado capacidad para prever el impacto de programas de vacunación a largo plazo, incluyendo efectos directos en poblaciones vacunadas y efectos indirectos mediante inmunidad de grupo. La capacidad de estos modelos para predecir con precisión los cambios en la incidencia de IPD subraya su utilidad para la planificación sanitaria estratégica.

En [20] se sugiere la posibilidad de integrar, en los modelos estructurados por edad, datos de vigilancia y estimaciones epidemiológicas para predecir específicamente las respuestas a la vacuna PCV en recién nacidos, demostrando reducciones en la transmisión de enfermedad y susceptibilidad. En [23] la autora construye un modelo compartimental para evaluar cómo la vacunación de preescolares con PCV7-TT puede influir indirectamente en la colonización nasofaríngea de lactantes, usando algoritmos como Metropolis-Hastings para estimar parámetros clave como la fuerza de infección. Estas aplicaciones ilustran la versatilidad de estos modelos para evaluar intervenciones en poblaciones vulnerables.

2.2.2. Diseño de Inmunógenos

Los modelos de selección clonal a nivel molecular tienen aplicaciones directas en el diseño racional de vacunas. Murugan et al. [22] sugieren que para antígenos complejos, el diseño de inmunógenos que apunten específicamente a células B que expresan anticuerpos germinales de alta afinidad puede optimizar la respuesta inmune.

2.3. Respuestas Inmunológicas Específicas a Vacunas

El estudio de Obermoser et al. [20] ofrece información sobre la modelación de respuestas inmunológicas específicas desencadenadas por vacunas neumocócicas, tanto a nivel innato como adaptativo.

2.3.1. Modelación de Respuestas Inmunes Innatas

El estudio examinó respuestas inmunes innatas a vacunas neumocócicas dentro de las horas posteriores a la vacunación mediante análisis de firmas transcripcionales. Sin embargo, la información disponible no detalla completamente la integración de estos datos en modelos computacionales.

La modelación de estas respuestas innatas tempranas representa un área prometedora para comprender determinantes precoces de la respuesta vacunal y potencialmente identificar individuos con mayor probabilidad de desarrollar protección eficaz.

2.3.2. Caracterización de Respuestas Inmunes Adaptativas

Las respuestas inmunes adaptativas fueron evaluadas al día 7 post-vacunación mediante medición de respuestas de plasmablastos, estableciendo correlaciones con títulos de anticuerpos posteriores. Esta caracterización temporal de la respuesta adaptativa proporciona un marco para predecir la efectividad vacunal antes de que se establezcan respuestas de anticuerpos completas.

La modelación de estas respuestas adaptativas complementa los modelos poblacionales al proporcionar mecanismos inmunológicos subyacentes que explican las tasas de protección observadas a nivel epidemiológico.

Capítulo 3

Metodología

En este capítulo se describe la metodología y la estructura empleada en la implementación de un simulador computacional del sistema inmune de un lactante para modelar la respuesta a vacunas conjugadas neumocócicas.

3.1. Parámetros

Cuadro 3.1 Parámetros utilizados en la simulación del sistema inmune.

Parámetro	Tipo	Descripción
duration_days	Entero	Duración total de la simulación en días.
vaccination_schedule	Lista de enteros	Días en los que se aplican las vacunas (dosis 1, 2 y refuerzo).
lf_decay	Flotante	Tasa de decaimiento natural de linfocitos no activos.
decay_factor	Flotante	Factor de decaimiento de anticuerpos.
plasma_production_factor	Diccionario (str: float)	Tasa de producción de anticuerpos a partir de células plasmáticas por cada serotipo.
memory_production_factor	Diccionario (str: float)	Tasa de producción de anticuerpos a partir de células de memoria activadas por cada serotipo.

Cuadro 3.2 Parámetros utilizados en la simulación del sistema inmune.

Parámetro	Tipo	Descripción
affinity_threshold_memory	Flotante	Umbral mínimo de afinidad para que una célula B se convierta en célula de memoria.
affinity_threshold_plasma	Flotante	Umbral mínimo de afinidad para que una célula B se diferencie en célula plasmática.
mutation_p	Flotante	Probabilidad de mutación de una célula B.
mutation_rate	Flotante	probabilidad de mutar cada componente del vector de receptores.
mutation_strength	Flotante	Intensidad del cambio en la afinidad tras una mutación celular.
temperature	Flotante	regula la presión selectiva.
THRESHOLD	Flotante	Umbral mínimo de afinidad para que una célula B entre a un centro germinal.

3.2. Estructura

El simulador se estructura en dos capas principales:

- Modelado de componentes biológicos:
 - Antígenos
 - Células B
- Entornos de Interacción:
 - Centros germinales
 - Sistema inmune

3.3. Modelado de componentes biológicos

1. **Antígeno:** Cualquier sustancia que haga que el cuerpo produzca una respuesta inmunitaria contra ella. Conformado por epítopo¹ y una

¹Parte de una molécula que será reconocida por un anticuerpo y a la cual se unirá.

proteína portadora. En el estudio este se modela como un objeto con propiedades: **id** (entero que representa el identificador del antígeno), **serotype** (*string* que representa código del serotipo), **carrier_protein** (*string* que representa la proteína portadora) y **epitope_vector** (vector de *float* que representa el epítipo).

2. **Célula B:** Cualquier linfocito encargado de reconocer y responder a un antígeno específico mediante la producción de anticuerpos. Conformada por su receptor BCR (receptor de célula B), que reconoce y se une a un epítipo particular. En el estudio, esta se modela como un objeto con propiedades: **id** (entero que representa el identificador de la célula), **bcr_vector** (vector de *float* que representa el BCR), **associated_serotype** (*string* que representa código del serotipo asociado al antígeno reconocido) y **affinity** (*float* que representa la afinidad de unión al epítipo).

3.4. Centros Germinales

Los centros germinales constituyen conglomerados de células B que reconocen el mismo antígeno en los cuales ocurren diferentes procesos inmunológicos.

Cálculo de afinidad:

La afinidad entre una célula B y un antígeno es una medida fundamental que cuantifica la fuerza de interacción entre el BCR y el epítipo del antígeno. Este cálculo se basa en la similitud estructural entre sus vectores característicos.

Definimos la afinidad A mediante la transformación exponencial de la distancia euclidiana entre los vectores:

$$A = e^{-d(\mathbf{r}_B, \mathbf{e}_A)} = \exp(-\|\mathbf{r}_B - \mathbf{e}_A\|) \quad (3.1)$$

donde:

- $\mathbf{r}_B \in \mathbb{R}^n$ es el vector característico del receptor de la célula B, que codifica las propiedades moleculares de su sitio de unión.
- $\mathbf{e}_A \in \mathbb{R}^n$ es el vector característico del epítipo antigénico, representando sus características estructurales clave para el reconocimiento inmunológico.

- $d(\mathbf{r}_B, \mathbf{e}_A) = \|\mathbf{r}_B - \mathbf{e}_A\|$ es la distancia euclidiana en \mathbb{R}^n , calculada como:

$$d = \sqrt{\sum_{i=1}^n (r_{B,i} - e_{A,i})^2} \quad (3.2)$$

- $A \in (0, 1]$ es la medida de afinidad resultante.

Selección de células B mediante distribución de Boltzmann:

La selección de las células B que sobreviven en cada generación se realiza mediante un algoritmo estocástico basado en la distribución de Boltzmann, que asigna probabilidades de supervivencia proporcionales a la afinidad de cada célula mientras mantiene diversidad poblacional. Este proceso consta de tres etapas:

1. **Cálculo de probabilidades:** Para cada célula B con afinidad A_i , se calcula una probabilidad de supervivencia:

$$P_i = \frac{e^{A_i/T}}{\sum_{j=1}^N e^{A_j/T}}$$

donde:

- T es el parámetro de temperatura que regula la presión selectiva
 - N es el tamaño total de la población
 - $A_i \in [0, 1]$ es la afinidad de la célula i
2. **Normalización numérica:** Para evitar inestabilidades computacionales, se aplica una transformación estable:

$$A_i^{\text{norm}} = A_i - \max_j (A_j)$$

garantizando valores negativos o cero sin alterar las proporciones relativas.

3. **Muestreo estocástico:** Se seleccionan K células sin reemplazo según:

$$\text{Seleccionados} \sim \text{Multinomial}(K; P_1, P_2, \dots, P_N)$$

donde $K = \min(\text{max_survivors}, N)$ modela restricciones de recursos.

Interpretación de parámetros:

- **Temperatura (T):** Regula el balance diversidad-optimización:
 - $T \rightarrow 0^+$: Selección elitista (solo células con máxima afinidad)
 - $T \rightarrow \infty$: Selección aleatoria uniforme (máxima diversidad)
 - $T \approx 0.5$: Balance biológico típico en simulaciones
- **max_survivors:** Modela restricciones biológicas:
 - Espacio en nichos linfoides
 - Disponibilidad de factores de crecimiento
 - Capacidad de presentación antigénica

Diferenciación:

Según su afinidad, las células B pueden diferenciarse en células plasmáticas o de memoria [24].

En la simulación se establecen umbrales en la afinidad para esta clasificación.

Hipermutación somática de células B:

El proceso de mutación simula la hipermutación somática que ocurre en los centros germinales durante la respuesta inmune adaptativa. Cada componente del vector receptor \mathbf{r}_B de una célula B tiene una probabilidad p_m de sufrir una modificación estocástica, modelada como:

$$r'_i = \text{clip}(r_i + \delta_i, 0, 1) \quad \text{donde} \quad \delta_i \sim \begin{cases} \mathcal{N}(0, \sigma^2) & \text{con probabilidad } p_m \\ 0 & \text{en caso contrario} \end{cases} \quad (3.3)$$

Parámetros clave:

- *Tasa de mutación (p_m):* Probabilidad de que un componente individual mute (valores típicos: 0.05-0.15). Biológicamente, refleja la actividad de la enzima AID (citidina desaminasa) en linfocitos B.
- *Intensidad de mutación (σ):* Desviación estándar de la distribución gaussiana que determina la magnitud de las alteraciones (valores típicos: 0.01-0.1). Modela el impacto estructural de las mutaciones puntuales.

Procesamiento post-mutación:

- *Normalización de valores:* Los componentes se limitan al intervalo $[0, 1]$ mediante:

$$\text{clip}(x, a, b) = \max(a, \min(x, b))$$

garantizando vectores biológicamente plausibles.

- *Reseteo de afinidad:* La afinidad de la célula mutada se inicializa a cero, requiriendo recálculo contra antígenos.
- *Preservación de linaje:* Se mantienen el identificador único y serotipo para rastreo evolutivo.

Base biológica: Este modelo captura tres propiedades esenciales de la hipermutación somática:

1. *Localidad:* Mutaciones puntuales que modifican ligeramente el paratopo
2. *Independencia:* Cada residuo muta de forma autónoma
3. *No direccionalidad:* Los cambios son estocásticos, no guiados por el antígeno

Las células B ingresan al centro germinal tras calcular su afinidad inicial con el antígeno. Dentro del mismo, el proceso de selección identifica las células candidatas para abandonarlo, las cuales se diferencian en células plasmáticas o de memoria antes de reintegrarse al sistema inmunológico. Las células no seleccionadas enfrentan dos destinos posibles: muerte por apoptosis (pérdida energética), o supervivencia con división y mutación. Para este último grupo, se aplica hipermutación somática a sus receptores y se recalcula su afinidad antigénica, reiniciando el ciclo, como se muestra en la figura 3.1. Este proceso iterativo continúa hasta la desaparición de todas las células del centro germinal.

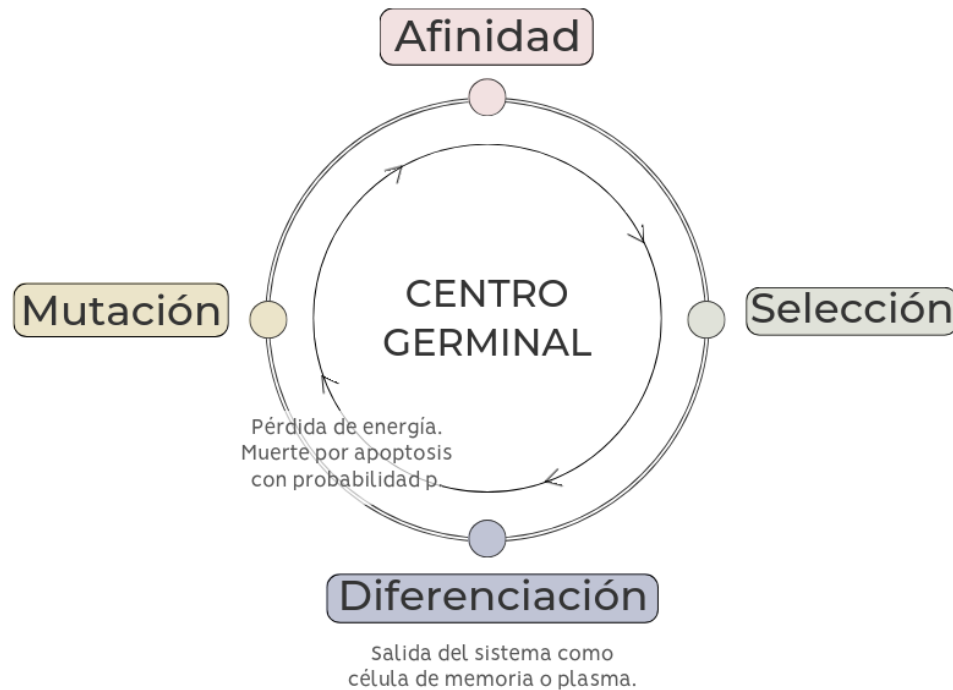


Figura 3.1: Ciclo que ocurre dentro del Centro Germinal

3.5. Sistema Inmune

El sistema inmune consta de un grupo de células B *naïve* que se dividen y mutan o mueren. Ante la presencia de antígeno estas células pueden pasar a formar parte de centros germinales.

Proceso de llenado de los centros germinales

Las células B actúan como recursos que pueden ser reclutados por los antígenos (según su afinidad por el mismo) para formar parte de un centro germinal. En el modelo, se crea un hilo (*thread*) por cada antígeno para simular de forma concurrente el proceso de selección de células B en los centros germinales. Este enfoque permite que todos los antígenos tengan igual prioridad al acceder a dichos recursos. Sin embargo, surge el problema de que dos antígenos podrían intentar reclutar la misma célula B simultáneamente. Para evitar esta condición de carrera, se implementa un mecanismo de exclusión mutua (*mutex*) que garantiza que solo un antígeno pueda analizar una célula B en un momento dado.

Al finalizar cada ciclo, los centros germinales liberan al sistema inmunológico células B diferenciadas en células plasmáticas y células de memoria. La concentración de anticuerpos generada por estas sirve como indicador cuantitativo de la respuesta inmunológica del individuo. Este proceso se muestra en la figura 3.2.

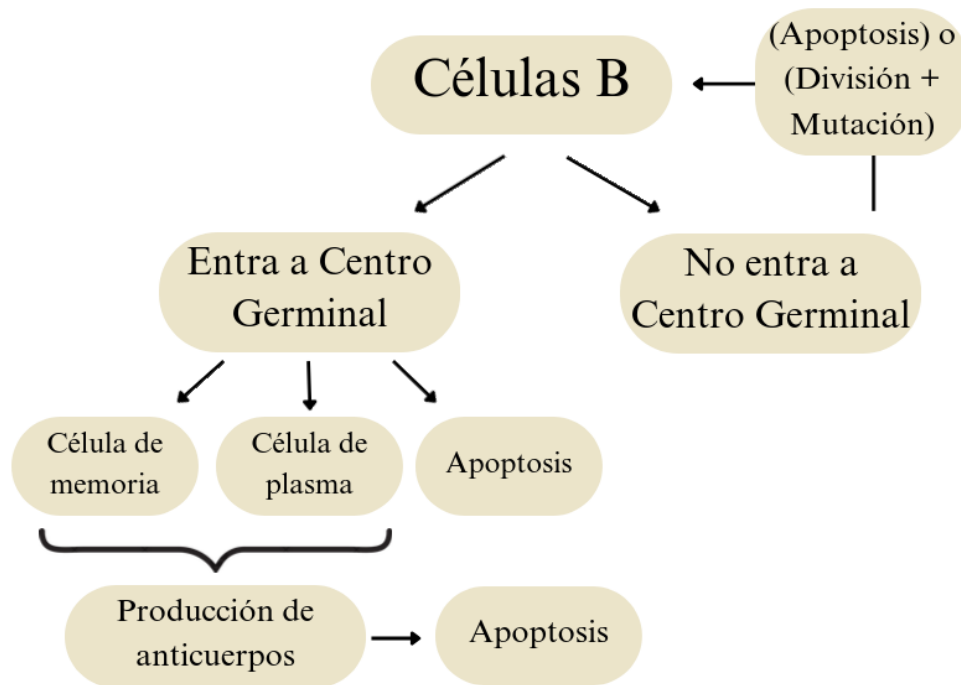


Figura 3.2: Células B dentro del sistema inmune

3.6. Salida del simulador

La salida del simulador consiste en una serie de tiempo que registra diariamente el nivel de anticuerpos específicos para cada serotipo presentes en el organismo durante toda la simulación. Este nivel se interpreta como una medida cuantitativa del resultado que se obtendría al realizar una prueba ELISA. Cada día de la simulación, se almacena el valor del nivel de anticuerpos específicos para el serotipo, calculado como la suma ponderada de las contribuciones individuales de las células productoras de anticuerpos (células de plasma y memoria).

Capítulo 4

Resultados

En este capítulo se detalla el proceso de validación del modelo propuesto.

4.1. Datos para la Validación del Modelo

Los datos utilizados en la investigación provienen del Estudio en Lactantes Fase I/II para PCV7-TT (Instituto Finlay de Vacunas, 2019), cuyo principal objetivo era caracterizar la seguridad del candidato vacunal VCN7-T y evaluar la no inferioridad de VCN7-T con respecto a la vacuna Prevnar13® administrados en esquemas concomitante con las vacunas Va-Mengoc- BC® y Heberpenta® incluyendo lactantes de 2 a 3 meses de edad.

Se cuenta con datos de 240 sujetos divididos en dos grupos:

1. Vacunados con VCN7-T
2. Vacunados con Prevnar13®

Para la validación del modelo nos centraremos en el grupo de sujetos vacunados con VCN7-T

4.1.1. Limpieza de Datos

Antes de trabajar con los datos fue realizado un proceso de limpieza de los mismos.

Análisis de Datos Faltantes

Las Figuras 4.1 y 4.2 presentan el análisis de valores faltantes (nulos o vacíos) en las columnas de los conjuntos de datos para las vacunas VCN7-T (Figura 4.1) y Prevnar13 (Figura 4.2).

Descripción de las gráficas:

- **Eje Y:** Nombre de las columnas/variables del estudio
- **Eje X:** Cantidad de valores faltantes (expresada en conteo absoluto)

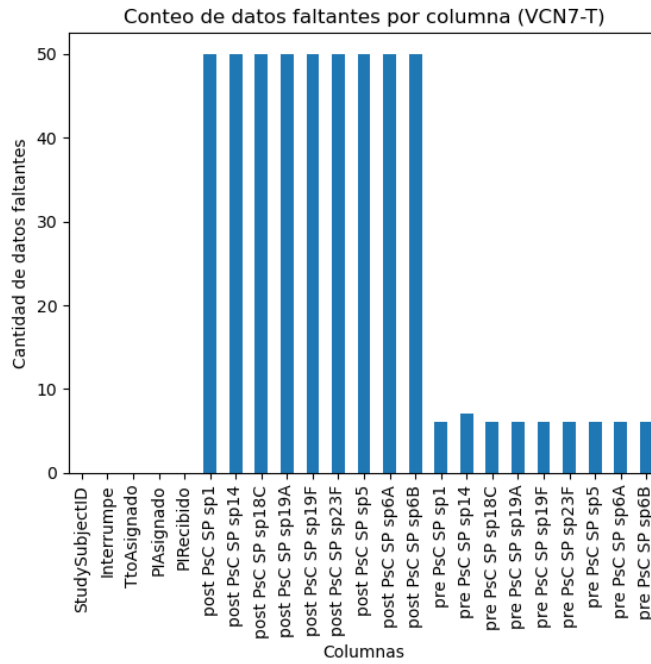


Figura 4.1: Distribución de valores faltantes por columna (VCN7-T)

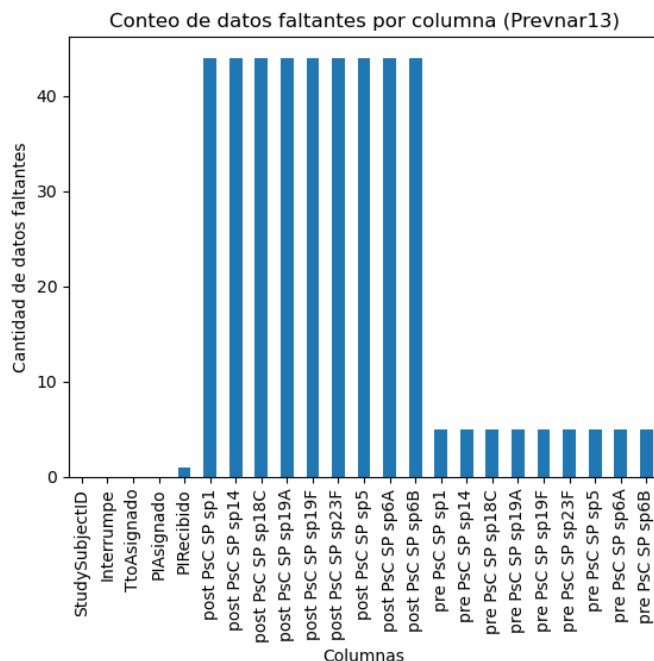


Figura 4.2: Distribución de valores faltantes por columna (Pprevnar13)

La nomenclatura utilizada en los nombres de columna sigue este patrón:

- **pre PsC SP [serotipo]**: Concentración de anticuerpos antes de la aplicación de la dosis de refuerzo (después de la dosis primaria) para un serotipo neumocócico específico
- **post PsC SP [serotipo]**: Concentración de anticuerpos después de la aplicación de la dosis de refuerzo para un serotipo neumocócico específico
- **[serotipo]**: Identificador del serotipo (e.g., sp1, sp14, sp19F, etc.)

Los altos valores en las columnas **post PsC SP [serotipo]** se consideran **no aleatorios** (MNAR - Missing Not At Random) y están directamente relacionados con el diseño del estudio, que incluyó dos esquemas de vacunación:

- **Esquema (2p + 1)**: Dos dosis primarias + refuerzo (mediciones post-refuerzo disponibles)

- **Esquema (3p + 0):** Tres dosis primarias sin refuerzo (mediciones post-refuerzo ausentes)

En la investigación se asume que los sujetos con valores nulos en las columnas `post` `PsC` `SP` [`serotipo`] fueron vacunados siguiendo un esquema (3p + 0), sin embargo para la validación del modelo se trabajó en los sujetos vacunados siguiendo el esquema (2p + 1).

Valores inconsistentes

Durante el análisis de los datos, se detectaron valores atípicos con magnitudes aparentemente inconsistentes. Una inspección minuciosa reveló que dichos valores correspondían a cifras significativas expresadas con distinta notación decimal; en particular, mientras la mayoría de los datos estaban representados con tres cifras significativas correctamente posicionadas, algunos carecían del punto decimal, lo que implicaba una escala incorrecta. Para homogeneizar la representación, se aplicó una corrección que consiste en dividir por 1000 los valores afectados. Esta normalización facilitó una presentación más coherente y uniforme de los datos.

En las gráficas 4.3, 4.4, 4.5 y 4.6 (Gráficas de valores pre-refuerzo contra valores post-refuerzo) se muestra una comparación del estado de los datos antes y después de la corrección (las gráficas para el resto de los serotipos se muestran en 4.3).

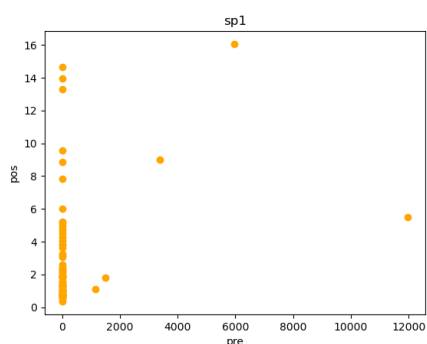


Figura 4.3: Serotipo 1 antes de la corrección

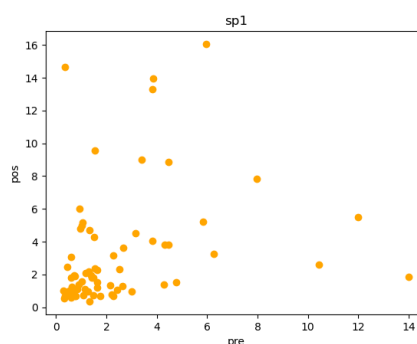


Figura 4.4: Serotipo 1 después de la corrección

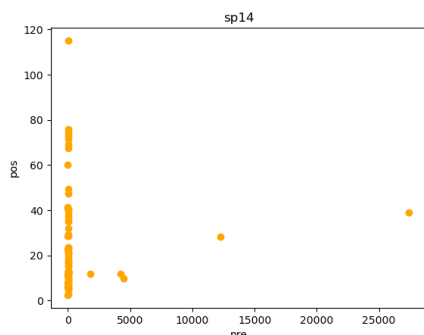


Figura 4.5: Serotipo 14 antes de la corrección

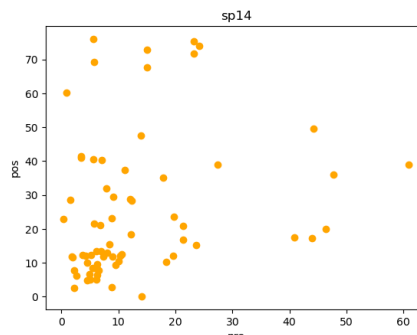


Figura 4.6: Serotipo 14 después de la corrección

En las gráficas pre-procesamiento se observan datos en el orden de las decenas de miles en el eje de las X los cuales pueden ser considerados como *outliers* por un análisis superficial. Sin embargo después de aplicada la corrección antes expuesta estos puntos responden más adecuadamente a la naturaleza de los datos.

4.2. Entrenamiento

Para el entrenamiento del modelo se tomó el 70% de los datos. Este proceso tiene como objetivo estimar los parámetros `plasma_production_factor` y `memory_production_factor`.

4.2.1. Exploración del espacio de parámetros

Para explorar el espacio de parámetros durante el proceso de entrenamiento se utilizó la metaheurística de Recocido Simulado (*Simulated Annealing*). Este método permitió realizar una búsqueda estocástica evitando quedar atrapado en óptimos locales, gracias a su capacidad para aceptar temporalmente soluciones peores con una probabilidad controlada por un parámetro denominado temperatura.

El algoritmo comienza con una temperatura inicial 1.0 que permite explorar el espacio de soluciones, aceptando cambios incluso si empeoran la función objetivo. A medida que avanza el proceso, la temperatura decrece siguiendo un protocolo de enfriamiento, reduciendo gradualmente la probabilidad de aceptar soluciones peores.

Soluciones vecinas:

Estas se generan seleccionando aleatoriamente uno de los parámetros a estimar y añadiéndole un ajuste δ .

Criterios de parada:

- **Enfriamiento del metal:** la temperatura es menor que $1e-3$.
- **Máximo de iteraciones:** se realizan más de 100 iteraciones.

4.2.2. Función Objetivo

La métrica utilizada como función objetivo durante el entrenamiento del modelo es la *distancia de Chamfer* ($D_{\text{Chamfer}}(S_1, S_2)$, donde S_1 y S_2 son los conjuntos de puntos simulados y reales), que es adecuada para comparar nubes de puntos, ya que evalúa la proximidad bidireccional entre cada punto de un conjunto y su vecino más cercano en el otro conjunto.

Minimizar la distancia de Chamfer durante el entrenamiento guía al modelo a generar salidas que se ajusten a los datos.

4.3. Validación

Para el proceso de validación se tomó el 30% restante de los datos. Se generó una conjunto de la misma cantidad de puntos con los parámetros obtenidos en el proceso de entrenamiento y se calculó la distancia Chamfer para estos conjuntos. Los resultados obtenidos se muestran en la tabla 4.1:

Cuadro 4.1 Distancia de Chamfer por Serotipo

Serotipo	Distancia Chamfer
1	3.422
14	26.348
18C	2.773
19F	4.443
23F	4.803
5	3.241
6B	8.539

Aunque los resultados visuales y cuantitativos no son óptimos, aportan información valiosa a la investigación. El modelo presenta un menor desempeño en serotipos, como el 14 y el 6B, cuyos datos experimentales exhiben

una mayor dispersión (nube de puntos más amplia), lo que sugiere que su precisión podría mejorar con un tamaño de muestra más grande.

En las gráficas, desde la 4.7 hasta la 4.13, se observa que, si bien la salida del modelo no captura por completo la variabilidad de los datos, logra aproximarse al centroide de la distribución. Esta aproximación es más precisa en los serotipos donde la dispersión de los datos es menor.

El

**Gráficas de valores pre-refuerzo (post vacunación primaria)
contra valores post-refuerzo**

- Puntos naranjas: datos de validación.
- Puntos azules: salida del modelo.
- Punto violeta: centroide de los datos de validación.

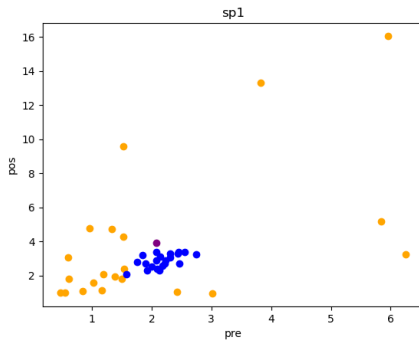


Figura 4.7: Serotipo 1

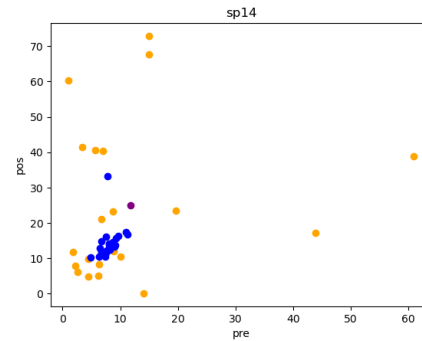


Figura 4.8: Serotipo 14

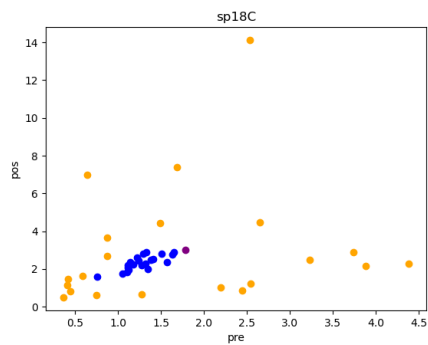


Figura 4.9: Serotipo 18C

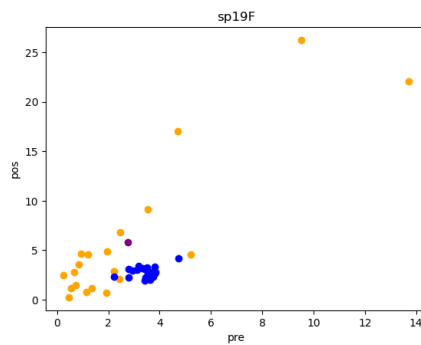


Figura 4.10: Serotipo 19F

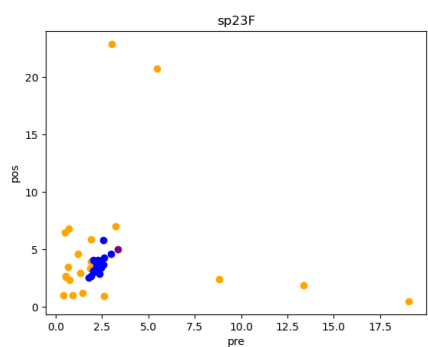


Figura 4.11: Serotipo 23F

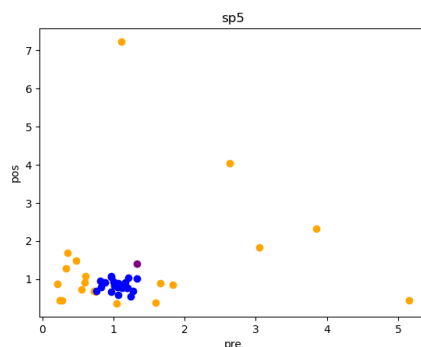


Figura 4.12: Serotipo 5

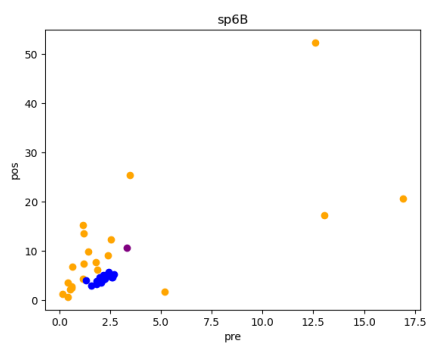


Figura 4.13: Serotipo 6B

Conclusiones

- El presente trabajo, así como otros desarrollados actualmente por el Grupo de Investigación de Biomatemática de la Facultad de Matemática y Computación de conjunto con el Instituto Finlay de Vacunas, abren una nueva línea de investigación en Cuba en la predicción de la efectividad de vacunas dadas por serotipos y vacunas neumocócicas conjugadas en particular.
- Se desarrolló un simulador del sistema inmune de lactantes que modela la respuesta a vacunas antineumocócicas conjugadas, integrando mecanismos inmunológicos clave como la activación de células B, la formación de centros germinales y la generación de memoria inmunológica.
- La validación del modelo contra datos clínicos reales demostró su capacidad para predecir patrones de respuesta inmune, aunque con variabilidad entre serotipos. Los mejores ajustes se obtuvieron para los serotipos 18C (distancia Chamfer = 2.773) y 5 (3.241), mientras que el serotipo 14 presentó mayor discrepancia (26.348).
- Este trabajo contribuye al campo de la inmunología computacional al proporcionar una herramienta para evaluar candidatos vacunales *in silico*, con potencial para reducir costos y tiempos en el desarrollo de vacunas.
- Como limitación, se identificó que el modelo tiene menor precisión para ciertos serotipos (especialmente el 14), lo que sugiere la necesidad de incorporar mecanismos inmunológicos adicionales en futuras versiones del simulador.

Recomendaciones

- Dar continuidad al desarrollo del simulador utilizando los resultados una vez realizado el ensayo clínico para el candidato vacunal futuro, QuimiVio-11 (que incluye 4 nuevos serotipos). Esto permitirá fortalecer la confiabilidad predictiva del modelo mediante una validación empírica más amplia.
- Implementar técnicas de optimización para abordar posibles casos de subajuste, especialmente en serotipos con mayores distancias de Chamfer (ej. serotipo 14).
- Añadir complejidad al modelo mejorando la fidelidad biológica del modelo integrando elementos clave de la inmunidad innata como receptores de reconocimiento de patrones o dinámica de células fagocíticas (neutrófilos, macrófagos).
- Adaptar el simulador para modelar respuestas inmunes en diferentes poblaciones, especialmente para el adulto mayor, mediante el desarrollo de conjuntos de parámetros específicos por edad y la consideración de inmunosenescencia en población geriátrica.
- Tener en cuenta para la realización de próximos ensayos clínicos la posibilidad de realizar más de un examen de sangre a los pacientes en diferente cantidad de días post-vacunación.

Referencias

- [1] World Health Organization. Pneumococcal disease. 2019. (Citado en las páginas 9 y 12).
- [2] KL O'Brien, LJ Wolfson, JP Watt, et al. Burden of disease caused by streptococcus pneumoniae in children younger than 5 years: global estimates. *The Lancet*, 2009. (Citado en la página 9).
- [3] D. González et al. Vigilancia epidemiológica de enfermedades neumocócicas en cuba. *Revista Cubana de Salud Pública*, 2017. (Citado en la página 9).
- [4] S. A. Plotkin, W. A. Orenstein, P. A. Offit, and K. M. Edwards. *Plotkin's Vaccines*. Elsevier, 7th edition, 2017. (Citado en la página 9).
- [5] C. A. Janeway, P. Travers, M. Walport, and M. J. Shlomchik. *Immunobiology: The Immune System in Health and Disease*. Garland Science, 5th edition, 2002. (Citado en la página 9).
- [6] Revistas UC3M. El papel de los modelos matemáticos en la vacunación, 2022. Consultado el 15 de mayo de 2025. (Citado en la página 10).
- [7] Deutsche Welle. Inteligencia artificial produce vacunas superpotentes, 2023. Consultado el 15 de mayo de 2025. (Citado en la página 10).
- [8] VALERIA PREADO J. Conceptos microbiológicos de Streptococcus pneumoniae: BASIC MICROBIOLOGICAL ASPECTS. *Revista chilena de infectología*, 2001. (Citado en la página 12).
- [9] S. Snedecor. Op388 17-year disease reduction predicted by a transmission dynamic model after pneumococcal conjugate vaccine introduction in the united states. *International Journal of Technology Assessment in Health Care*, 2020. (Citado en las páginas 13, 14 y 22).

- [10] Asociación Española de Pediatría. *Manual de inmunizaciones en línea de la AEP*, 2022. (Citado en la página 13).
- [11] Asociación Española de Pediatría (AEP). Manual de inmunizaciones: aspectos inmunológicos de la vacunación maternal. <https://vacunasaep.org/profesionales/noticias/manual-de-inmunizaciones-inmunologia-vacunacion-maternal>, 2023. Consultado el 14 de junio de 2025. (Citado en la página 13).
- [12] R. M. Lequin. Enzyme immunoassay (eia)/enzyme-linked immunosorbent assay (elisa). *Clinical Chemistry*, 2005. (Citado en la página 15).
- [13] El-Ghazali Talbi. *Metaheuristics: From Design to Implementation*. John Wiley & Sons, 2009. (Citado en la página 15).
- [14] Scott Kirkpatrick, C. Daniel Gelatt, and Mario P. Vecchi. Optimization by simulated annealing. *Science*, 1983. (Citado en la página 15).
- [15] John H. Holland. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 2nd edition, 1992. (Citado en la página 16).
- [16] Marco Dorigo, Vittorio Maniezzo, and Alberto Colomi. Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 1996. (Citado en la página 16).
- [17] Frederick Reif. *Fundamentals of Statistical and Thermal Physics*. McGraw-Hill, New York, 1965. (Citado en la página 20).
- [18] Andrew S. Tanenbaum. *Modern Operating Systems*. Prentice Hall, 3rd edition edition, 2008. (Citado en la página 20).
- [19] Yoonjoo Choi, Ulrike Damm, Desmond Curran, Robert J. Pitman, Maarten J. Postma, Philippe Beutels, and Mark Jit. Dynamic transmission models of pneumococcal conjugate vaccination: A review of key assumptions and vaccine effects. *Vaccine*, 39, 2021. (Citado en la página 22).
- [20] G. Obermoser, S. R. Presnell, K. Domico, H. Xu, Y. Wang, E. Anguiano, L. Thompson-Snipes, et al. Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines. *Immunity*, 2013. (Citado en las páginas 23 y 24).

- [21] Eva K. Lee, Helder I. Nakaya, Fan Yuan, Troy D. Querec, Greg Burel, Ferdinand H. Pietz, Bernard A. Benecke, and Bali Pulendran. Machine learning for predicting vaccine immunogenicity. *Interfaces*, 2016. (Citado en la página 23).
- [22] Rajagopal Murugan, Lisa Buchauer, Gianna Triller, Cornelia Kreschel, Giulia Costa, Gemma Pidelaserra Marti, Katharina Imkeller, Christian E. Busse, Sumana Chakravarty, B. Kim Lee Sim, Stephen L. Hoffman, Elena A. Levashina, Peter G. Kremsner, Benjamin Mordmüller, Thomas Höfer, and Hedda Wardemann. Clonal selection drives protective memory b cell responses in controlled human malaria infection. *Science Immunology*, 2018. (Citado en las páginas 23 y 24).
- [23] Elianys García-Pola Córdoves. MCMC en modelos dinámicos: Aplicación a la evaluación de impacto del candidato vacunal PCV7-TT. Master's thesis, Universidad de La Habana, La Habana, Cuba, 2022. (Citado en la página 24).
- [24] Kenneth Murphy and Casey Weaver. *Janeway's Immunobiology*. Garland Science, New York, 9th edition, 2017. Chapter 9: The Humoral Immune Response — Differentiation of B Cells into Plasma Cells and Memory B Cells. (Citado en la página 30).

Anexos

Comparación del estado de los datos antes y después de la corrección

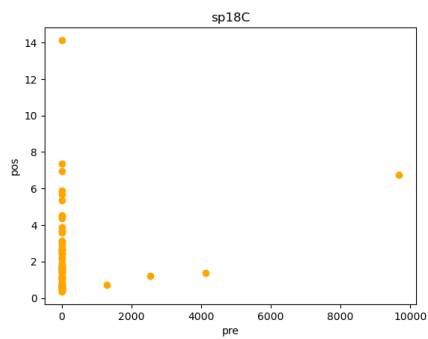


Figura 4.14: Serotipo 18C antes de la corrección

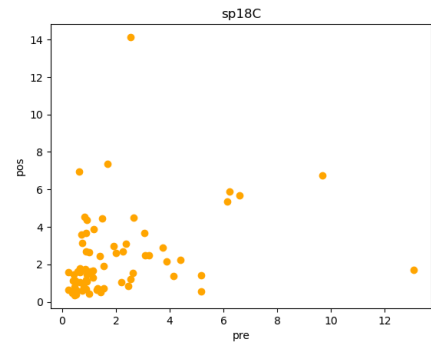


Figura 4.15: Serotipo 18C después de la corrección

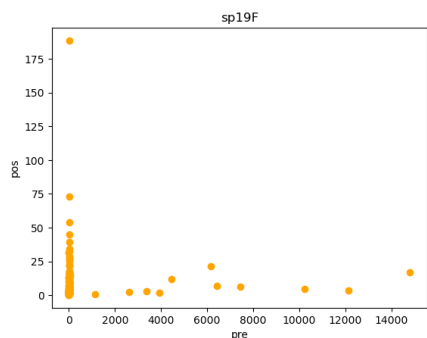


Figura 4.16: Serotipo 19F antes de la corrección

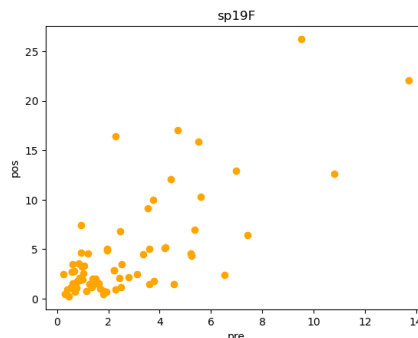


Figura 4.17: Serotipo 19F después de la corrección

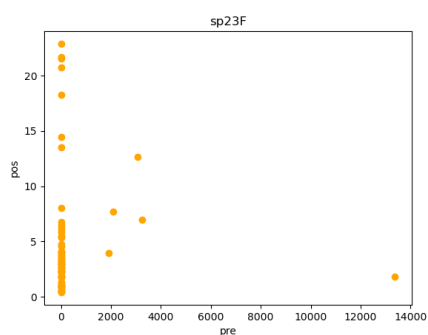


Figura 4.18: Serotipo 23F antes de la corrección

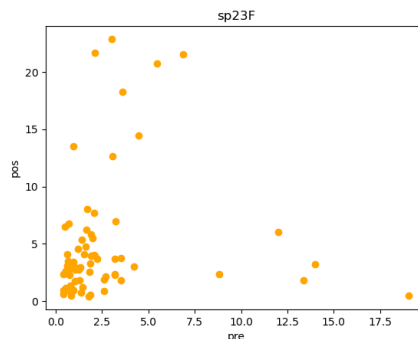


Figura 4.19: Serotipo 23F después de la corrección

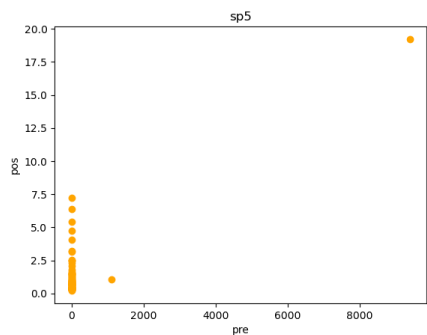


Figura 4.20: Serotipo 5 antes de la corrección

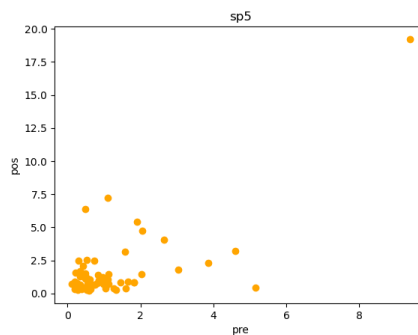


Figura 4.21: Serotipo 5 después de la corrección

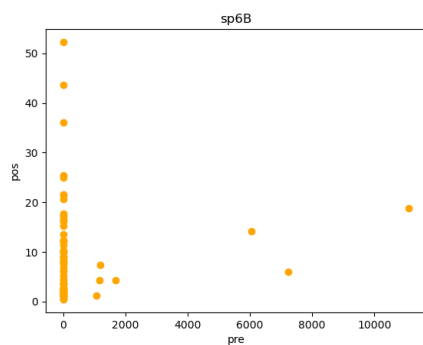


Figura 4.22: Serotipo 6B antes de la corrección

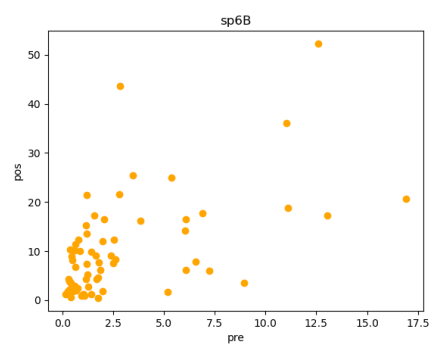


Figura 4.23: Serotipo 6B después de la corrección