

Red neuronal artificial para la interpretación de resistencias a la terapia antirretroviral

P. Chausa, E. J. Gómez, C. Cáceres, M.E. Hernando, F. Del Pozo

¹ Grupo de Bioingeniería y Telemedicina, Universidad Politécnica de Madrid, Madrid, España, {pchausa,egomez,ccaceres,elena,fpozo@gbt.tfo.upm.es

Resumen

La aparición de la terapia antirretroviral supuso un punto de inflexión en el tratamiento de las personas infectadas por el VIH. La selección de los fármacos es vital para poder controlar la enfermedad y asegurar la supervivencia de los pacientes. Son muchos los factores que hay que tener en consideración a la hora de efectuar dicha elección. Especialmente relevante es la posible resistencia del virus a uno o varios fármacos, resistencia generada por el propio tratamiento administrado y que anulará el efecto de las drogas. Esto nos da una idea de la importancia de disponer de técnicas de análisis e interpretación de las resistencias que cada paciente ha desarrollado. Este trabajo describe las diferentes técnicas que están siendo utilizadas para abordar el problema y presenta el desarrollo de un algoritmo para la detección de resistencias, aplicado al análisis de resistencias a un fármaco inhibidor de la proteasa.

1. Introducción

El VIH ataca al sistema inmunitario debilitándolo progresivamente, hasta que se produce una inmunodeficiencia significativa y comienzan a aparecer numerosas infecciones y enfermedades que en conjunto dan lugar al síndrome conocido como SIDA (Síndrome de Inmunodeficiencia Adquirida), una enfermedad hasta ahora incurable, progresiva y mortal. Las personas que padecen este síndrome, comienzan a sufrir enfermedades oportunistas que aprovechan la débil condición del organismo para atacar y desarrollarse **¡Error! No se encuentra el origen de la referencia..**

Los tratamientos antirretrovirales no atacan directamente al VIH. Buscan interferir sus procesos de replicación dentro del organismo. Su acción está dirigida al bloqueo de las enzimas del VIH que catalizan los procesos de reproducción del mismo. La eficacia de un tratamiento antirretroviral depende de muchos factores. Entre los más importantes encontramos las resistencias a los diferentes fármacos que el VIH genera, anulando el efecto de dichos fármacos sobre el organismo. Estas resistencias se desarrollan atendiendo a diversas causas [1]. Unas hacen referencia a la propia naturaleza del virus, como puede ser la elevada tasa de replicación del VIH, la tasa de mutación espontánea en el ciclo típico de replicación del VIH o la ausencia en los retrovirus de un mecanismo para corregir los errores de transcripción. Otras estarían relacionadas con la falta de adherencia o el uso de regímenes de potencia inadecuada. Por último, podríamos

hablar de problemas derivados del huésped, como por ejemplo, una mala absorción o la existencia de procesos metabólicos excepcionales.

El análisis de resistencias se ha erigido como una importante herramienta para los médicos a la hora de seleccionar el tratamiento antirretroviral más adecuado. Es importante conocer qué tipo de test utilizar, cuándo y cómo interpretar los resultados para mejorar el cuidado de los pacientes con VIH. Además, algunos médicos pueden no estar familiarizados con la información molecular medida y presentada por los tests de resistencias realizados en los laboratorios. Existen dos técnicas fundamentales que estudian las resistencias antirretrovirales: pruebas genotípicas (GRT, genotypic resistance testing) y fenotípicas (PRT, phenotypic resistance testing).

En las pruebas genotípicas, el código genético de la muestra del virus es comparado con el código del virus tipo salvaje. El código genético está formado por una larga cadena de moléculas llamadas nucleótidos [3]. Cada grupo de tres nucleótidos, llamado "codón", define a un aminoácido particular que se utiliza para construir un nuevo virus. Las mutaciones son descritas usando una combinación de números y letras, K103N por ejemplo. La primera letra (K) es el código para el aminoácido en el virus tipo salvaje. El número (103) identifica al codón mutante. La segunda letra (N) es el código para el aminoácido que "cambió" en la muestra mutante. En general, las pruebas genotípicas son las más usadas debido a la mayor disponibilidad de estudios que evalúan su utilidad clínica y a su menor coste. El resultado de una prueba genotípica consiste en un listado de diferencias observadas frente a una secuencia de referencia, siendo esas diferencias lo que conocemos como mutaciones. A partir de ellas los especialistas deberían ser capaces de determinar a qué fármacos es resistente el virus y que deberían descartarse a la hora de elegir el tratamiento del paciente. La interpretación de estos tests de resistencias es complicada, por lo que la prueba genotípica se suele acompañar de una descripción del grado de resistencia del virus a los diferentes fármacos.

La prueba de resistencia fenotípica es un examen de laboratorio que nos permite saber la cantidad de virus que puede crecer en presencia de los medicamentos antirretrovirales. Se realiza una comparación entre el crecimiento de una muestra de virus del paciente y una

muestra de referencia con la presencia de diferentes concentraciones de fármacos antirretrovirales. Representa una medida más directa de la susceptibilidad a un fármaco. Las pruebas proporcionan ratios que indican la concentración necesaria de un fármaco para conseguir una inhibición del crecimiento en un 50% (IC₅₀).

Una técnica alternativa, más barata y rápida, sería el fenotipo virtual, en la cual se estima el fenotipo real que se derivaría del genotipo del paciente. Se obtiene a partir de bases de datos con miles de muestras de VIH a las cuales se les ha realizado el correspondiente test genotípico y fenotípico. Proporciona información rápida acerca de la sensibilidad del VIH a los antirretrovirales una vez conocido el genotipo, sin que sea absolutamente necesario llevar a cabo una prueba fenotípica.

2. Algoritmos de interpretación de genotipo

El resultado de una prueba genotípica consiste en un listado de mutaciones que el especialista debe interpretar para poder inferir el grado de susceptibilidad del paciente a los diferentes fármacos antirretrovirales. Existen diferentes aproximaciones básicas para convertir una lista de mutaciones en un informe más simple y fácil de interpretar.

2.1. Algoritmos basados en reglas

Son algoritmos desarrollados por diferentes expertos haciendo uso del conocimiento adquirido sobre las resistencias a fármacos debidas a mutaciones del virus. Esta aproximación tiene diversas ventajas como su modularidad, es decir, la posibilidad de añadir nuevas reglas al dominio. Asimismo los algoritmos presentan un conocimiento explícito que permite al experto validar más fácilmente el resultado de la decisión. Como principal inconveniente, los algoritmos basados en reglas no soportan incertidumbres ni razonamientos no monótonos, aquellos en los que la llegada de un nuevo hecho puede modificar la creencia de deducciones anteriores. Algunos de los sistemas basados en reglas han sido elaborados por el Resistance Collaborative Group (RCG) [4]; Stanford HIVdb algorithm [5]; Agencia Nacional Francesa para la investigación en SIDA (ANRS) [6]; Instituto Rega y Retrogram (Virology Networks). Asimismo, otros ejemplos de algoritmos desarrollados por diversas empresas son: GuideLines (Bayer Diagnostics), ViroSeq (Celera Diagnostics) o TruGene (Visible Genetics, Canadá).

Aunque los algoritmos de interpretación de resistencia a fármacos están evolucionando con la publicación de nuevos datos y la incorporación de nuevos agentes antirretrovirales, el software necesario para implementar dichos algoritmos debería permanecer estable para permitir comparaciones entre las interpretaciones proporcionadas por los mismos y las reglas de las que están compuestos. En este sentido, existe una plataforma estándar para el desarrollo de los algoritmos [7]: ASI (Algorithm Specification Interface), en la que se

especifica el algoritmo en formato XML y se transforma después en código ejecutable.

2.2. Algoritmos basados en redes probabilísticas

El razonamiento probabilístico se rige por las reglas de la teoría de la probabilidad representadas en el Teorema de Bayes, teoría que puede utilizarse en procesos de clasificación. Las estimaciones necesarias para realizar la inferencia en la red se pueden realizar por métodos paramétricos o no paramétricos. En la estimación paramétrica suponemos que la función de probabilidad marginal tiene una forma conocida y una serie de parámetros que pueden ser estimados. La estimación no paramétrica sigue tres posibles direcciones: estimación de la función de densidad de probabilidad dado un conjunto de muestras, estimación de las probabilidades a posteriori de cada una de las clases o la transformación del espacio de parámetros de manera que en el nuevo espacio se puedan aplicar técnicas paramétricas.

Los métodos basados en técnicas probabilísticas se han empleado para predecir el fenotipo [8] y la respuesta virológica a una terapia [9] a partir de los datos obtenidos con la prueba genotípica.

2.3. Algoritmos basados en Data Mining: Fenotipo virtual

El algoritmo VirtualPhenotype (Virco; Cambridge, Reino Unido y Mechelin, Bélgica) es un algoritmo propietario que se basa en la correlación entre genotipos y fenotipos almacenada en la base de datos de Virco. El procedimiento exacto no ha sido publicado, pero el concepto básico es el siguiente: el algoritmo recibe como entrada una lista de mutaciones o diferencias con una secuencia que por consenso se considera de referencia. El programa la compara con las secuencias que dispone en una base de datos. Se analiza el fenotipo de las secuencias que más se parecen a la de estudio para calcular la media y el rango de los niveles de resistencia. Si dichos valores superan los umbrales correspondientes a cada fármaco se considera que es resistente. Tanto el procedimiento como la base de datos empleada por VirtualPhenotype son propietarios. No se conoce las mutaciones que se emplean para comparar una nueva secuencia con aquellas que ya están presentes en la base de datos. Dos estudios prospectivos, GenPheRex [10] y Realvirfen [11], emplean el algoritmo VirtualPhenotype y Antivirogram, otro algoritmo desarrollado por Virco. En ellos se asocian las reducciones similares de carga viral en plasma después de un cambio de tratamiento.

2.4. Redes Neuronales Artificiales

Las redes neuronales artificiales se han utilizado en el campo del VIH/SIDA como técnicas de análisis para diagnósticos clínicos, análisis de la estructura del VIH o predicción del estado de personas infectadas con VIH [12]. También se han utilizado para predecir las resistencias a determinados fármacos antirretrovirales. Por

ejemplo, existe un estudio en el que se analiza la resistencia del virus al Lopinavir [13]. Otro estudio realizado en la Universidad de Detroit [14] utiliza una arquitectura típica dentro de las redes neuronales, los mapas auto-organizativos (SOFM), para extraer los parámetros significativos que servirán al clasificador diseñado para determinar la resistencia a diferentes inhibidores de la proteasa.

3. Ejemplo de red neuronal para la interpretación de resistencias

3.1. Planteamiento del problema

Como método de acercamiento al problema global, se ha desarrollado un sencillo algoritmo que obtiene el fenotipo virtual, es decir, la resistencia de la secuencia de la proteasa a un fármaco un inhibidor de la proteasa, en concreto el Saquinavir. La secuencia de la proteasa está formada por 99 aminoácidos. De ellos se seleccionan aquellos que están en una posición relevante, es decir, en una posición en la que dependiendo del aminoácido que haya, el virus será resistente o no al fármaco. Las posiciones importantes y las mutaciones que pueden producirse en ellas se han obtenido de investigaciones realizadas en este campo [15]. En el ejemplo del Saquinavir, son las referidas en la Tabla 1.

| Posición | Aminoácido Sustituido | Aminoácido Presente | Mutación |
|----------|-----------------------|---------------------|----------|
| 10 | Leucina (L) | Isoleucina (I) | L10I |
| 10 | Leucina (L) | Arginina (R) | L10R |
| 10 | Leucina (L) | Valina (V) | L10V |
| 48 | Glutamina (G) | Valina (V) | G48V |
| 54 | Isoleucina (I) | Valina (V) | I54V |
| 54 | Isoleucina (I) | Leucina (L) | I54L |
| 71 | Alanina (A) | Valina (V) | A71V |
| 71 | Alanina (A) | Treonina (T) | A71T |
| 73 | Glutamina (G) | Serina (S) | G73S |
| 77 | Valina (V) | Isoleucina (I) | V77I |
| 82 | Valina (V) | Alanina (A) | V82A |
| 84 | Isoleucina (I) | Valina (V) | I84V |
| 90 | Leucina (L) | Metionina (M) | L90M |

Tabla 1. Mutaciones que oponen resistencia al Saquinavir

El trabajo realizado se fundamenta en la base de datos de pacientes del Servicio de Enfermedades Infecciosas del Hospital Clínic de Barcelona en la que se registran, entre otras cosas, las mutaciones obtenidas de diferentes muestras a partir de las pruebas genotípicas y la resistencia que presentan a diversos fármacos. Una vez conocidas cuáles son las 13 mutaciones que nos interesan para el caso concreto del Saquinavir, las 13 que aparecen en la Tabla 1, se han seleccionado 500 muestras útiles de la base de datos, con información sobre dichas

mutaciones y la resistencia de cada muestra al Saquinavir. Los datos utilizados para el entrenamiento y evaluación de la red neuronal (404 y 96 muestras respectivamente) siguen el formato de las muestras de ejemplo que aparecen en la Tabla 2.

| Muestra | L10I | L10R | L10V | ... | I84V | L90M | Res |
|----------|------|------|------|-----|------|------|-----|
| Muestra1 | 0 | 0 | 1 | 0/1 | 0 | 0 | 3 |
| Muestra2 | 0 | 0 | 1 | 0/1 | 1 | 0 | 1 |
| Muestra3 | 0 | 0 | 0 | 0/1 | 0 | 0 | 2 |
| Muestra4 | 0 | 0 | 0 | 0/1 | 0 | 0 | 0 |

Tabla 2. Formato de los datos de entrenamiento y evaluación

Los trece primeros unos y ceros nos indican la presencia o no de la mutación y el valor de resistencia es el que aparece en último lugar y varía entre el 0 y el 3. La codificación de los niveles de resistencia está descrita en la Tabla 3.

| Código Resistencia | Descripción |
|--------------------|-----------------------------|
| 0 | - |
| 1 | Resistencia |
| 2 | Resistencia parcial |
| 3 | No evidencia de resistencia |

Tabla 3. Codificación de los niveles de resistencia

3.2. Diseño de la red

El objetivo es crear una red que clasifique las muestras en función de las mutaciones que presenta. Con ello obtendríamos un sistema que permitiría decidir a un especialista en VIH si recetar o no Saquinavir a un paciente. Esto mismo se podría generalizar y aplicar a diferentes fármacos, de manera que, a partir del resultado de la red, se podría elegir el tratamiento óptimo a administrar a los pacientes.

El clasificador diseñado sigue una estructura de perceptrón multicapa de 3 niveles. La capa de entrada tiene 13 neuronas, la de salida 1 y la capa oculta tiene un número arbitrario de neuronas en cada realización, de manera que se puedan comparar los resultados obtenidos con 10, 7, 5 y 3 neuronas. El diseño, simulación, entrenamiento y evaluación de la red se ha realizado con el Neural Network toolbox presente en la versión 6.1 del Matlab.

EL diseño de la red con Matlab utiliza la función general "network", y no funciones que crean redes predefinidas como perceptrones o redes de Hopfield. Esto proporciona más flexibilidad a la hora de elegir la arquitectura de la red y sus diferentes parámetros. Las funciones de inicialización de los pesos de cada capa, de entrenamiento y de validación han sido las mismas. Esto nos permite realizar una comparación real entre las redes creadas variando el número de neuronas de la capa oculta. Dichas funciones son: 1) Función de inicialización: Método de Nguyen-Widrow; 2) Función de entrenamiento:

Retropropagación de Levenberg-Marquardt; y 3) Función de validación: Minimización del error cuadrático medio

3.3. Resultados obtenidos

La Tabla 4 muestra los resultados en el proceso de entrenamiento, es decir, el número de iteraciones producidas hasta alcanzar la convergencia y el error cuadrático medio obtenido. Después aparece el error cuadrático medio que se obtiene en el proceso de evaluación de la red. Se explica también cuál ha sido el motivo de alcanzar la convergencia en cada caso. Los resultados obtenidos en el proceso de entrenamiento nos llevarían a la conclusión de que la arquitectura de red que ofrece el menor error en el proceso de evaluación es la red con 3 nodos en la capa oculta.

| | Entrenamiento | | Evaluación | Convergencia |
|-----------------|----------------|----------|------------|---|
| | Learning Epoch | MSE | MSE | |
| 10 nodos | 18 | 0.125833 | 0.1840 | Valor mín de gradiente / Máx valor de μ |
| 7 nodos | 72 | 0.125833 | 0.2673 | Valor mín de gradiente |
| 5 nodos | 100 | 0.310973 | 0.2821 | Número máx de iteraciones |
| 3 nodos | 100 | 0.126697 | 0.0998 | Número máx de iteraciones |

Tabla 4. Comparativa de resultados entre las diferentes redes

4. Conclusiones y trabajos futuros

La selección de los fármacos más adecuados en la terapia antirretroviral es vital para el control de la enfermedad y asegurar la supervivencia de los pacientes. Una tarea compleja que requiere gran experiencia del profesional médico, así como el acceso a amplias bases de datos de muestras de VIH analizadas según el test genotípico y fenotípico.

Este trabajo ha presentado un estudio de los métodos más utilizados en el análisis e interpretación de las resistencias y ha descrito en detalle la implementación de una sencilla red neuronal artificial que analiza el grado de resistencia del paciente a un fármaco inhibidor de la proteasa. Los resultados preliminares de este trabajo han permitido conocer en profundidad los fundamentos del problema clínicos y las soluciones metodológicas más adecuadas para el desarrollo futuro de esta línea de investigación.

5. Agradecimientos

Los autores expresan su agradecimiento al Servicio de Enfermedades Infecciosas del Hospital Clínic de Barcelona, y en especial al Dr. F. García, Dr. J.M. Miró y y Dr. J.M. Gatell.

Este trabajo ha sido parcialmente financiado por el proyecto del MCYT TIC2002-02129, la Red de "Telemedicina" del MSC y la empresa Glaxo Smith Kline.

Referencias

- [1] Nekane Basabe, Darío Páez, Ricardo Usieto, Henri Paicheler, Jean-Claude Deschamps. El desafío social del SIDA. Centro de estudios sociales aplicados. 1997.
- [2] Jeffrey P. Nadler, Antiretroviral resistance testing and therapeutic drug monitoring. HIV/AIDS Primary Care Guide, Florida/Caribbean AIDS Education and Training Center. 2005.
- [3] InfoRed SIDA: Análisis de resistencia del VIH. Julio 2002. <http://www.aegis.com/pubs/nmap/1999/414e-resistance.html> (Consultado Jun-2005)
- [4] DeGruttola V, Dix L, D'Aquila R, et al. The relation between baseline HIV drug resistance and response to antiretroviral therapy: re-analysis of retrospective and prospective studies using a standardized data analysis plan. *Antivir Ther.* 2000; 5: 41-48.
- [5] Shafer RW, Jung DR, Betts BJ. Human immunodeficiency virus type 1 reverse transcriptase and protease mutation search engine for queries. *Nat Med.* 2000; 6: 1290-1292.
- [6] Meynard JL, Vray M, Morand-Joubert L, et al. Phenotypic or genotypic resistance testing for choosing antiretroviral therapy after treatment failure: a randomized trial. *AIDS.* 2002; 16: 727-736.
- [7] Betts BJ, Shafer RW. Algorithm specification interface for human immunodeficiency virus type 1 genotypic interpretation. *J Clin Microbiol.* 2003; 41:2792-2794.
- [8] Sevin AD, DeGruttola V, Nijhuis M, et al. Methods for investigation of the relationship between drug-susceptibility phenotype and human immunodeficiency virus type 1 genotype with applications to AIDS clinical trials group 333. *J Infect Dis.* 2000; 182: 59-67.
- [9] Quigg M, Frost SD, McDonagh S, et al. Association of antiretroviral resistance genotypes with response to therapy-comparison of three models. *Antiviral Ther.* 2002; 7: 151-157.
- [10] Mazzota F, Lo Caputo S, Torti C, et al. Real versus virtual phenotype to guide treatment in Heavily pretreated patients: 48-week-follow-up of the Genotipo-Fenotipo di Resistenza (GenPheRex) trial. *J Acquir Immune Defic Syndr.* 2003;32:268-280.
- [11] Pérez-Elías MJ, García-Arata I, Muñoz V, et al. Phenotype or virtual phenotype for choosing antiretroviral therapy after failure: a prospective, randomized study. *Antiviral Ther.* 2003;8:577-584.
- [12] Chang W, Lee, Jung-A Park: Assessment o HIV/AIDS-related health performance using artificial neural network. Elsevier. August 2000.
- [13] Wnag D, Larder B. Enhanced prediction of Lopinavir resistance from genotype by use of artificial neural networks. *J Infect Dis.* 2003; 188: 653-660.
- [14] Sorin Draghici and R. Brian Potter. Predicting HIV drug resistance with neural networks. *Bioinformatics.* Vol. 19 no. 1 2003. 98-107.
- [15] Special Contribution-Drug Resistance Mutations. Volume 11 Issue 6. International AIDS Society-USA. October