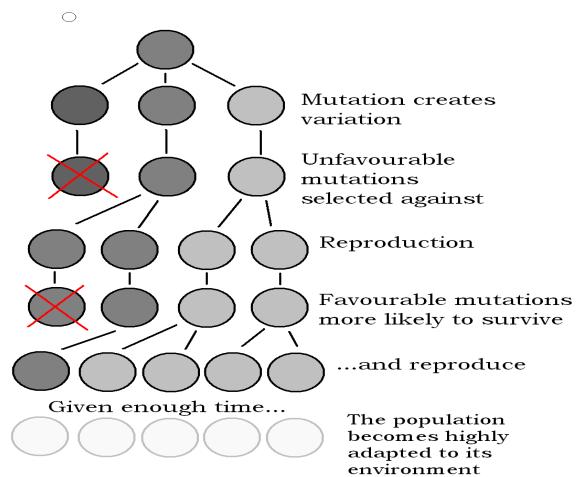


Evolución Molecular



Romina V. Sepúlveda

Jonathan Canan

Daniel Aguayo

2017

1. Proteínas y evolución

La función de las moléculas biológicas están directamente ligadas a sus propiedades fisicoquímicas y como estas son moduladas por el solvente, estructura, interacciones, relaciones y por como estas evolucionaron.

La forma que adoptan las moléculas biológicas están directamente ligadas a las propiedades del agua, por ejemplo, los fosfolípidos adaptan su estructura para alejar sus zonas hidrofóbicas (cadenas aciles) del agua, formando estructuras más complejas que han permitido la compartimentalización celular. Por su parte, las estructuras de las proteínas también están definidas por la interacción de sus componentes –los aminoácidos– con el solvente, ya que residuos hidrofóbicos tienden a empaquetarse para minimizar su interacción con el agua (ver cuadro Entropía y el plegamiento proteico), mientras que los grupos polares son estabilizados por sus propias interacciones y con la formación de enlaces de hidrógeno intra e intermoleculares.

Las estructuras que adoptan las biomoléculas también cumplen un rol importante y como veremos más adelante muchas veces son la clave para entender su función. Las proteínas adoptan patrones simétricos que finalmente se condensan adoptando estructuras tridimensionales más complejas. Entre ellas cuentan las α -helices y las laminas- β , las que a su vez condensan (o compactan) entre ellas favoreciendo la formación de enlaces de hidrógeno en el backbone y maximizando las interacciones entre residuos polares y cargados que han perdido su interacción con las moléculas de agua, debido a las restricciones impuestas por la compactación del núcleo hidrofóbico.



Entropía y el plegamiento proteico. Desde un punto de vista simple, una proteína desplegada tiene una alta entropía configuracional, asociada a la alta cantidad de configuraciones teóricas que esta puede adoptar, pero también una alta entalpía ya que su estructura está estabilizada por un bajo número interacciones. Por otro lado una proteína plegada tiene una entropía considerablemente menor, pero con una alta entalpía. A partir de la ecuación $\Delta G = \Delta H - T\Delta S$, es posible inferir que el plegamiento proteico está dominado por un "juego" entre la entropía, la entalpía y la temperatura, sin embargo, esta explicación del plegamiento proteico omite algo fundamental, el proceso ocurre en presencia de un solvente particular, el agua. Cada vez que un dominio hidrofóbico está expuesto a solvente, este interrumpe la red de enlaces de hidrógeno y constriñe las configuraciones que las moléculas de agua adyacentes pueden adoptar. De esta forma, su compactación en el llamado núcleo hidrofóbico, permite que las moléculas de agua aumenten su entropía con respecto al estado desplegado. Aun más, la red de enlaces de hidrógeno de los residuos polares y el esqueleto de la proteína (backbone) en agua es máxima tanto para el estado plegado como desplegado, por lo cual la diferencia de entalpía entre ambos estados es nula ($\Delta H = 0$) y el cambio está casi en su totalidad guiado por efectos entrópicos.

Las estructuras que adoptan las biomoléculas están directamente ligadas con su función, siendo ambas seleccionadas durante la evolución. El mejor ejemplo de la relación entre estructura, función y evolución es a través del estudio de las enzimas. Las enzimas catalizan reacciones químicas a través de mecanismos complejos que les confieren una alta eficiencia con respecto a las reacciones químicas regulares. Habitualmente las reacciones enzimáticas ocurren en una región cerca de la superficie proteica, denominado “**sitio activo**” (Figura 1), una región pequeña comparada con el tamaño de una proteína. Cab destacar que no tan solo el sitio activo es necesario para que ocurra la reacción enzimática, sino que todo el resto de la proteína. Esto habré una pregunta interesante ¿Cómo llegaron las proteínas a ser una máquina tan sofisticada? La respuesta es que fueron desarrolladas a través de un proceso temporal, en los cuales cambios en la secuencia aminoacídica modificaron las propiedades fisicoquímicas de la cadena polipeptídica, lo cual influye directamente en la estructura que estas pueden adoptar y, por ende, la función química que estas catalizan. A este proceso se le denomina “evolución enzimática” y para entenderlo es necesario conocer tanto como evolucionan la

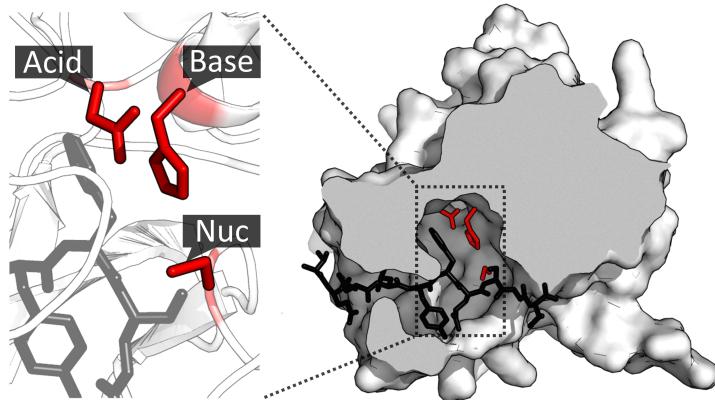


Figura 1 Sitio activo y triada catalítica de la enzima Serine Protreasa TEV (PDB: 1lvm). Los residuos aspartato (Acido), histidina (base) y serina (nucleofilo) están destacados en rojo, mientras que el sustrato se encuentra en negro. Adaptado de Shafee (2014).

secuencia aminoacídica como la estructura proteica.

Ancestro común

A pesar que el ADN es quien se ve afectado por mutaciones (ver cuadro Evolución y el ADN), estas pueden o no tener un efecto sobre la secuencia aminoacídica, lo cual no necesariamente afecta la función proteica; ya que esta última depende de un bajo número de residuos críticos entre los que destacan lo que conforman el sitio activo, el sitio de unión a sustrato o que contribuyen a la estabilidad de la estructura tridimensional. Debido a su importancia, estos residuos presentan una baja variabilidad ya que cumplen un rol relevante para mantener la función, a diferencias de otros en los que su cambio por otro residuo – de igual o distinta propiedades físicoquímicas – no produce un efecto significativa en el arreglo tridimensional de la cadena polipeptídica y en la función que esta desarrolla.

El hecho que las mutaciones pueden ocurrir sin perturbar en gran medida las propiedades físicoquímicas que definen la estructura secundaria y terciaria de una cadena proteína, tienen como consecuencia que las estructuras evolucionan a un ritmo diferente de las secuencias de las que derivan, lo cual se traduce en que las estructuras están más conservadas que las secuencias.

Hoy en día, gracias a los métodos experimentales y teóricos disponibles, es posible observar estos procesos evolutivos a nivel molecular, la llamada “evolución molecular”. En este ámbito, el concepto de **evolución convergente** implica que no hay una relación ancestral, solo una convergencia en un arreglo estructural estable, por el contrario, **evolución divergente** implica que no hay una relación ancestral, solo una convergencia en un arreglo estructural estable.

“La evolución selecciona la estructura proteica, NO la secuencia aminoácidica, ya que la estructura es la que determina la función”

De esta forma, es posible encontrar proteínas que no tienen una secuencia similar o función relacionada, pero que si comparten una estructura tridimensional similar. Esta diferencia en la velocidad evolutiva entre las secuencias y la estructura hacen que esta última sea un mejor marcador, ya que el estudio de sus cambios permiten reconocer mejor las relaciones entre ancestros relacionados y determinar si la diferenciación entre las proteínas de interés provienen procesos de evolución divergente y convergente.

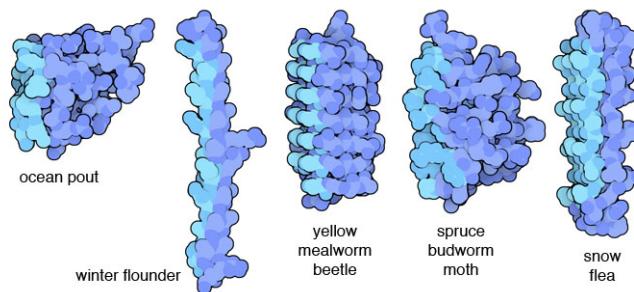


Figura 2Proteínas anticongelantes (AFP). InterPro:IPR003460,SUPERFAMILY 1ezg,Pfam PF02420SCOP 1ezg, Pfam structures.

Evolución Convergente

En general, divergencia y convergencia describen procesos evolutivos por los cuales los organismos se adaptaron a sus ambientes. Es difícil encontrar casos representativo de episodios de evolución convergente. En general hablamos de evolución convergente cuando organismos no cercanamente relacionados entre si han desarrollado características similares para adaptarse a un ambiente similar. Cabe destacar que para cada problema hay infinitas soluciones posibles, sin embargo estas emergen una y otra vez de manera independiente, aunque no siempre a través de un mecanismo similar. Para exemplificar observe lo siguiente, los peces que habitan en la Antártica han evolucionado para sobrevivir en ambientes de baja temperatura, ello utilizan glicoproteínas que les permiten evitar el crecimiento de cristales de agua y disminuir el punto de congelación de sus fluidos corporales.¹. Por su parte, al otro lado de la tierra, los peces del Ártico también tienen proteínas que actúan de manera similar. Es evidente que estos peces tienen relaciones evolutivas distantes, lo cual se traduce que los genes de estas proteínas “anticongelantes” provienen de genes distintos y proteínas de estructura diferente (Figura 2). Esta evidencia sugiere episodios de “evolución molecular” independientes, que resultaron en funciones similares ya que ambos organismos Antárticos y Árticos evolucionaron en nichos ambientales de similares características.

Evolución divergente

Las Serine proteasas están presentes virtualmente en todos los organismos, cumpliendo funciones intra e extracelulares. En general se describen como dos familias: “Trypsin-like” y “subtilisin-like” (traducido como parecidas a tripsina o a subtilisina). Se describen como Serine proteasas por dos razones principales: hidrolizan proteínas y tienen un residuo de Serina en el sitio activo, esencial para realizar la catálisis. Esta serina es significativamente más reactiva que otras serinas de la proteína. En mamíferos, las Serine proteasas conforman una familia de proteínas con diversas funciones, pero de estructura similar. Ejemplos de ella son:^{*2}:

- Chymotrypsin
- Trypsin
- Elastase
- Plasmin
- Thrombin
- Acrosomal protease
- Complement C1
- Keratinase, Collagenase, Fibrinolysin, Cocco-nase, etc.

¹Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod PNAS 1997 94 (8) 3817-3822

²se han conservado sus nombres en inglés

En este práctico Ud. deberá resolver distintas actividades y preguntas, las cuales aparecerán a lo largo del texto. Para diferenciar las actividades de las preguntas estas aparecerán como una letra destacada con un circulo rojo, similares a las que aparecen a continuación y que Ud. debe constatar.

A Traduzca los nombres de las serina proteasas descritas anteriormente.

B Describa su función de forma general.

Como fue mencionado anteriormente, las serine proteasas tienen una estructura tridimensional similar³, sin embargo, tienen diferencias en su superficie – reflejo de sus diferentes secuencias – para mejorar la interacción con los sustratos involucrados en las diferentes actividades fisiológicas en que participan. A pesar de estas diferencias **todas ellas comparten un mismo mecanismo catalítico**. Las diferencias entre sus secuencias y la similitud entre sus estructuras indican sus relación evolutiva. La alta similitud entre chymotrypsin, trypsin y elastase indican que estas proteínas evolucionaron a partir de eventos de duplicación génica a partir de una Serine proteasa ancestral, a partir de la cual sus funciones fueron definidas a partir de un proceso de **evolución divergente**. Por otro lado, Subtilisina (aislada de *Bacillus subtilis* y de secuencia y estructura diferente a la chymotrypsina) los grupos del sitio activo son similar al de chymotripsina, sin embargo su posición tridimensional no lo es. Chymotripsina y Subtilisina son ejemplos claros de un proceso de evolución convergente (ver figura 1).

C Explique con sus palabras los conceptos de evolución convergente y divergente

D Investigue si hemoglobina, glutation peroxidasa y cytochrome c son ejemplos de evolución convergente o divergente.

Cuando se investigan las relaciones evolutivas entre diferentes organismo, es luego importante observar tanto los genes como las proteínas que los diferencian.

En este práctico utilizaremos diferentes herramientas diseñadas para estudiar la evolución de los mecanismos enzimáticos a través de la información genética disponible, como la conservación evolutiva, en conjunto con la información estructural de las proteínas que estos genes codifican, lo cual es una herramienta útil al momento de entender los aspectos evolutivos que llevan a la función proteica y como esta puede ser modulada.



Evolución y el ADN. El ADN está siempre sujeto a mecanismos de reparación que permiten la mantención del código genético, sin embargo, los sistemas de reparación están sujetos a fallas. Si consideramos que el ser humano tiene una tasa de error de $1 \cdot 10^{-8}$ mutaciones por nucleotido por generación y que el genoma humano tiene alrededor de $3 \cdot 10^9$ bases, entonces cada persona tiene en promedio alrededor de 70 mutaciones en su genoma (mutaciones por generación). Esto es una diferencia promedio de 0.1 % en los genoma de dos personas cualquiera (1 de cada 1000 bases).

Homología y similitud

El concepto de homología – presencia de un ancestro evolutivo común – se encuentra mencionado en múltiples ocasiones en este texto, siendo central para el análisis computacional de proteínas y

³lo cual aun no hemos definido

secuencias nucleotídicas. Sin embargo, la relación entre homología y similitud no es tan clara y atención debe prestarse a ella. En términos analíticos, decimos que existe homología cuando el grado de similitud entre dos elementos, ya sea secuencias o estructuras, excede el valor esperado para elementos de la misma naturaleza (secuencias aleatorias de nucleótidos, aminoácidos de un mismo largo) elegidos al azar. Cuando existe un exceso de similitud mayor al esperado, la explicación más simple (parsimoniosa) es que ambas secuencias no fueron generadas de forma independiente, sino que provienen de un ancestro en común. Este ancestro en común permite explicar el exceso de similitud, ya que otras explicaciones requieren que las estructuras proteicas resultante de la expresión de estos genes evolucionen de manera independiente.

Sin embargo, secuencias homólogas no siempre significan similitud de secuencias. Los antecedentes disponibles muestran que existen alineamientos de secuencias que no son significantes, pero que al mirar las estructuras proteicas utilizando alineamientos estructurales están presentan una similitud estadísticamente significativa, ya sea entre ellas o hacia una secuencia intermedia. Thus, when a similarity search finds a statistically significant match, we can confidently infer that the two sequences are homologous; but if no statistically significant match is found in a database, we cannot be certain that no homologs are present.

- E** Discuta el concepto de “homología” bajo el contexto de evolución convergente y divergente.

Actividades a Realizar

La disciplina denominada “enzyme evolution” involucra la aplicación de conceptos bioinformáticos en el área de la enzimología para la describir el proceso de evolución de la estructura y función. En esta actividad práctica Ud.

- Investigará los distintos niveles de la comisión de enzimas, desde el punto de vista de las estructuras cristalinas que hay en cada una de ellas.
- En la base de datos UNIPROT, investigará relevante información sobre proteínas y enzimas.
- Utilizará PROSITE, Pfam, InterPro para encontrar proteínas homólogas a partir de secuencias aminoacídicas.
- Utilizará alineamientos múltiples para encontrar proteínas homólogas a partir de perfiles obtenidos de secuencias aminoacídicas.
- Utilizará los datos e información recopilada para plantear un experimento relacionado con la evolución de enzimas.
- Utilizará VMD y sus herramientas para relacionar la estructura y función con los conceptos de evolución convergente y divergente.

Comprenda y realice todas las instrucciones (destacadas con números arábigos), conteste las distintas preguntas (destacadas con letras en mayúsculas), y entregue un informe escrito con los resultados y discusión de los temas abordados. A lo largo del texto se entregará material complementario en forma de “cuadros” independientes, los que incluyen información y preguntas de distintos aspectos como importancia biológica, parámetros a profundizar y atajos para VMD.

Programas Requeridos

Los siguientes programas serán requeridos en este práctico:

- **VMD⁴** (para todas las plataformas)

⁴<http://www.ks.uiuc.edu/Research/vmd/>

- NAMD⁵

- Jalview⁶

- **Programa de graficación matemática:** Será necesario utilizar programas para graficar las salidas de VMD y NAMD. VMD tiene un programa básico de graficación incorporado. Algunos ejemplos de otros programas son:

- Unix/Linux: xmgrace ⁷
- Windows: Excel ⁸ (Pago)
- Mac/Multiples Platforms: Mathematica ⁹ (Pago); gnuplot ¹⁰(Descarga gratuita)

Los archivos de este práctico se encuentran en el directorio **practico-evolucion**.

2. Nomenclatura de Enzimas

Las enzimas se clasifican de acuerdo a la nomenclatura asignada por el Nomenclature Committee de la International Union of Biochemistry (descrita por primera vez en 1961, mejorada por última vez en 1992). Bajo esta nomenclatura, las enzimas **NO** son clasificadas por su mecanismo, secuencia aminoácídica (ie. homología) o por la estructura que estas adoptan, **sino a través de la reacción química que estas catalizan**. La llamada *enzyme commission* clasifica las enzimas en 6 grupos o clases principales, de acuerdo a la reacción que estas catalizan:

- | | |
|-------------------------|--------------------|
| ■ EC 1. Oxidoreductasas | ■ EC 4. Liasas |
| ■ EC 2. Transferasas | ■ EC 5. Isomerasas |
| ■ EC 3. Hidrolasas | ■ EC 6. Ligasas |

A A partir de los datos de la página de la Enzyme comission¹¹ describa de manera general las clases en las cuales se agrupan las enzimas y las reacciones que estas catalizan.

La enzyme commission asigna a cada enzima un código recomendado (Enzyme classification, EC) que contiene 4 secciones o números a,b,c y d; “a” es la clase, “b” la subclase y “c” es la sub-subclase. Mientras “b” y “c” describen la reacción, “d” es utilizada para distinguir enzimas que tienen una función similar sobre el sustrato de la reacción.

Por ejemplo, las “Serine proteasas”(también llamadas serine endopeptidasas) son enzimas capaces de romper los enlaces peptídicos de proteínas. Su nombre se debe a que en todas ellas, un residuo conservado de Serina sirve como nucleófilo en el sitio activo. Además, todas las serine proteasas contienen tres residuos en su sitio activo: Serina, histidina y aspartato. Habitualmente, para identificar los residuos de una proteína se utiliza la numeración de los residuos de un miembro característico de una familia; en el caso de las serine proteasas es habitual utilizar el esquema de numeración de la Chymotripsina, por ejemplo la Chymotrypsin-C Humana¹² (EC:3.4.21.2), donde

⁵<http://www.ks.uiuc.edu/Research/namd/>

⁶<http://www.>

⁷<http://plasma-gate.weizmann.ac.il/Grace/>

⁸<http://office.microsoft.com>

⁹<http://www.wolfram.com/>

¹⁰<http://www.gnuplot.info/>

¹¹<http://www.chem.qmul.ac.uk/iupac/jcbn/>

¹²<http://www.uniprot.org/uniprot/Q99895>

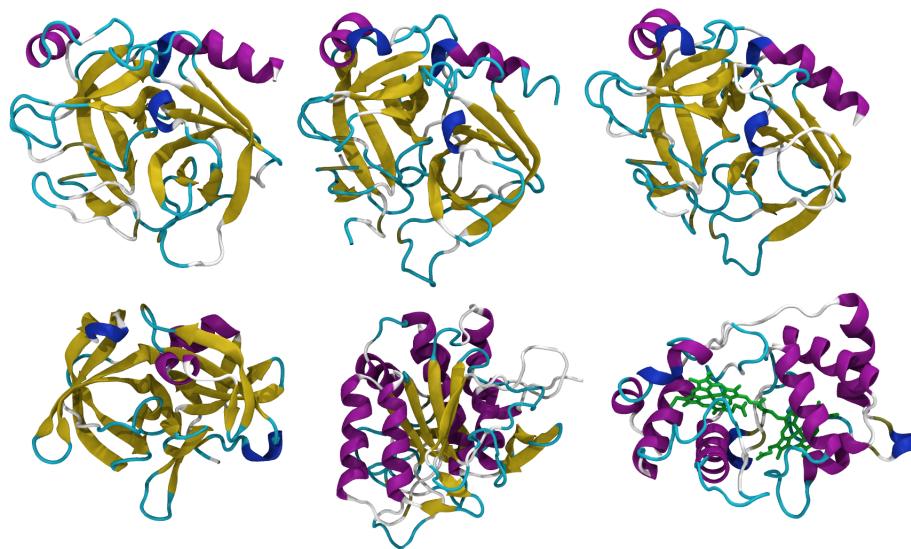


Figura 3 Estructura de cuatro Serine Protreasas y dos proteínas no relacionadas. (A) Tripsina Bovina (pdb 3OTJ) (B) Chymotripsina bovina (pdb 1YPH) (C) Tripsina de *S. griseous* (pdb 1SGT) (D) Proteasa A *S. griseous* (E) Subtilisina (pdb 1SBT) (F) Citocromo c4 (pdb 1ETP).

los residuos que componen la llamada “**triada catalítica**” son His⁷⁴, Asp¹²¹ y Ser²¹⁶,(figura 1)¹³. Estos residuos, también llamados “residuos catalíticos”, se encuentran alejados en la secuencia aminoacídica (cuadro 1), pero cercanos espacialmente en la conformación tridimensional que adopta esta proteína (Figura 1).

■

Serine Proteasas. Cerca de un tercio de las proteínas conocidas pertenecen a la familia de las Serine proteasas. Entre ellas las tripsinas son las más predominantes, con funciones que participan en la digestión, coagulación, fibrinolisis, desarrollo, fertilización, apoptosis e inmunidad, entre otras. ^a. Por otro lado, cabe destacar que los residuos que componen la triada catalítica son “altamente” superponibles entre diferentes serine proteasas. ¿Qué se puede inferir de esto?

^aDi Cera, Serine Proteases. Life. 2009 May; 61(5): 510–515.

Las serines proteasas pertenecen a la subsubclase EC 3.4.21. El cuadro 2 señala 5 de los 121 miembros entre los descritos para esta subsubclase de enzimas según la enzyme commission.

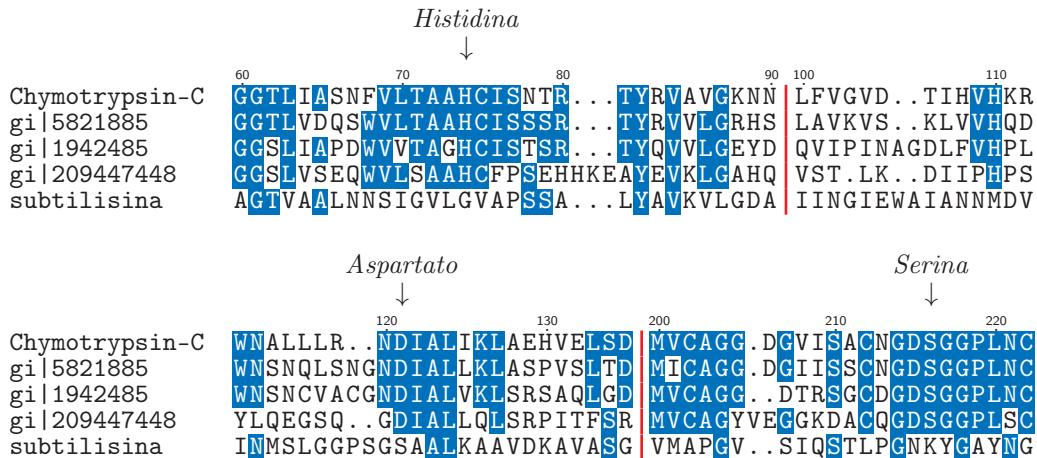
- B** A partir de la clasificación de la enzyme commission, identifique y justifique cada clase, subclase y subsubclase hasta encontrar el número identificador EC de la enzimas **Serine Proteasas** listadas anteriormente en el cuadro 2.

Recuerde que esta jerarquización esta realizada en base a la reacción catalizada por la enzima. De esta forma cuando dos enzimas que pertenecen a una misma subsubclase, el número final ‘EC a.b.c.d’ o identificador esta relacionado con los metabolitos y cofactores involucrados en la reacción.

- C** A partir de la clasificación de la enzyme commission, identifique y diferencie las reacciones de la enzimas de las subsubclases listadas en el cuadro 2.

¹³Thomas, Shafee, (2014). “Evolvability of a viral protease: experimental evolution of catalysis, robustness and specificity”. PhD Thesis. University of Cambridge.

Cuadro 1 Diferentes Serine proteasa y su triada catalítica.



Cuadro 2 Ejemplo de miembros de la subsubclase EC Serine Proteasas

- Trypsin-like
- Chymotrypsin-like
- Thrombin-like
- Elastase-like
- Subtilisin-like

En la tabla 2.1 Ud. encontrará los códigos que permiten acceder a información y a las secuencias de diferentes miembros de estas familias.

La primera clasificación detallada de las estructuras de las proteínas fue realizada por Jane Richardson en 1981 ¹⁴. Actualmente existen múltiples clasificaciones, sin embargo, destacan dos clasificaciones:

- CATH: cuyo acrónimo viene dado por los niveles de organización estructural de las proteínas: Clase, arquitectura, topología y homología.
- SCOP: Su nombre viene de Clasificación estructural de proteínas (en inglés). Que organiza las proteínas en: Clase, Plegamiento (Fold), Superfamilia, Familia y dominio.

Cuadro 3 Miembros de la familia de Serine Proteasas

PDB-Id	cadena	residuos	resolución	largo	PM	Fuente
3OTJ	E	281		223	23324.3	Bos taurus
1YPH	C	482	1.34	131	13934.6	Bos taurus
1SGT	A	223	1.7	223	23076.8	Streptomyces griseus
2SGA	A	181	1.5	181	18016.6	Streptomyces griseus
1SBT	A	275	2.5	275	27552.5	Bacillus amyloliquefaciens

¹⁴<http://kinemage.biochem.duke.edu/teaching/anatax/>

PDB-Id	uniprot Id	Nombre	cath	cathId	pfam
3OTJ	P00760	Cationic trypsin	Trypsin-like SP	2.40.10.10	PF00089
1YPH	P00766	Chymotrypsinogen A	Trypsin-like SP	2.40.10.10	PF00089
1SGT	P00775	Trypsin	Trypsin-like SP	2.40.10.10	PF00089
2SGA	P00776	Streptogrisin-A	Trypsin-like SP	2.40.10.10	PF00089
1SBT	P00782	Subtilisin BPN'		3.40.50.200	PF00082

2.1. UNIPROT

Provee anotación detallada de secuencias biológicas, incluyendo: estructura, función, clasificación en familias de proteínas, dominios estructurales, sitios catalíticos, cofactores, modificaciones postraduccionales, vías metabólicas, asociación a enfermedades. También provee links a otros recursos de interés, y es muy poco redundante.

La información de secuencias deriva de TrEMBL, una base de datos de secuencias traducidas de ácidos nucleicos.

La anotación de cada entrada es cuidadosamente curada por expertos que obtienen información de la literatura científica, y por tanto es de buena calidad.

Vaya Uniprot ¹⁵, y en la pestaña titulada “Retrieve” ingrese los códigos PDB de 5 proteínas. Luego presione el botón “Retrieve” y siga el link que dice: “UniProtKB (5)”. Esto le dará acceso directo a las páginas en UniProt de cada una de las 5 proteínas que estamos estudiando.

- D** Que información puede obtener de UNIPROT.
- E** Describa cada una de sus proteínas.
- F** Anote los códigos UNIPROT, EC, Pfam asociados a cada proteína.
- G** Descargue la secuencia de cada proteína.
- H** Compruebe los datos de la tabla .

3. Uso Básico de VMD

3.1. Inicio

En esta unidad utilizaremos VMD para visualizar y analizar la estructura de una serine proteasa, el cuál servirá en el resto de este práctico. Solo se abordarán aspectos básicos de uso, por lo que se recomienda leer el manual de usuario disponible en ¹⁶.

- 1 Para iniciar VMD escriba vmd en una ventana de terminal a Unix, doble-click en el ícono de la aplicación de VMD en la carpeta Applications en Mac OS X, o haga click en el menu Inicio → Programas → VMD en Windows. En Unix escriba vmd en una ventana de terminal.



Webpdb. VMD puede descargar estructuras desde en formato pdb desde el Protein Data Bank. Si hay conexión a internet disponible escriba el código de cuatro letras del cristal seleccionado, por ejemplo 1STP.

¹⁵<http://www.uniprot.org/>

¹⁶<http://www.ks.uiuc.edu/Research/vmd/current/ug/>

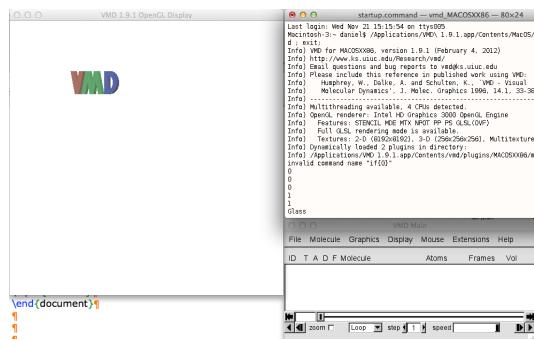


Figura 4 Ventanas abiertas al ejecutar VMD.

3.2. Cargando el sistema

Para iniciar se deben cargar las coordenadas de los distintos átomos del sistema, las cuales se pueden obtener desde distintas fuentes como el *protein data bank* en el caso de proteínas o directamente con VMD. El archivo *3OTJ.pdb* contiene las coordenadas atómicas de la cadena E de la proteína Serine Protease de origen bovino.

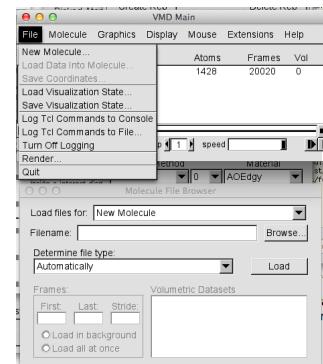


Figura 5 Cargando una molécula.

- 1 Elija el menú File → New Molecule... Fig. 5 en la ventana VMD Main. Aparecerá otra ventana Molecule File Browser, la que permite elegir los archivos dentro de los directorios disponibles.
- 2 Use el botón Browse... para localizar el archivo *3OTJ.pdb* en *vmd-tutorial-files*. en el directorio *practico-evolucion*. Note que luego de seleccionar este archivo se encuentra nuevamente en la ventana Molecule File Browser. Para cargar el archivo, presione el botón Load. Recuerde revisar que esta seleccionado New Molecule en la sección Load Files for. Si lo desea, puede cerrar la ventana Molecule File Browser.
- 3 El sistema se visualiza en la ventana OpenGL Display. Utilice el ratón para rotar el sistema, para escalar la representación utilice el tercer botón del ratón o haga click la tecla "S" en el teclado y luego mueva el cursor en la ventana OpenGL Display. Para volver a rotar haga click en la tecla "R" en el teclado. Para trasladar el sistema, utilice la tecla "T".
- 4 Para volver a la visualización inicial presione Display → Reset View.
- 5 En el menú Menu → Graphics elija Representations. Se abrirá la ventana Graphical Representations, en verde se destaca la representación que se está utilizando para mostrar su molécula.
- 6 En la ventana Graphical Representations elija la viñeta Draw Style, esta permite cambiar el estilo de visualización de cada representación creada a través del menú Drawing Method. El método o estilo predeterminado es *Lines*, que muestra en forma de líneas los enlaces entre átomos.

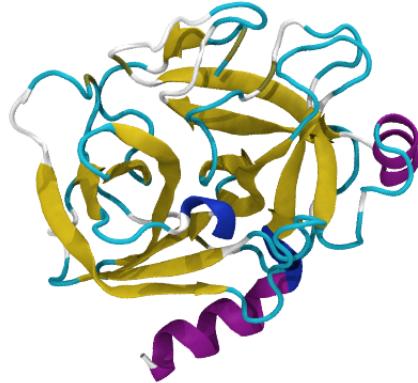
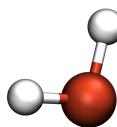


Figura 6 Serine Proteasa Bovina, pdb 3OTJ cadena E

- 7 Cada método de dibujo tiene sus propios parámetros. Por ejemplo, modifique el campo Thickness para aumentar el ancho de la líneas que representan los enlaces atómicos.
- 8 Cambie el método de dibujo a VDW (van der Waals). Cada átomo es ahora representado como una esfera, esto permite visualizar la forma y el volumen de la proteína. Elija ahora el método de dibujo New Cartoon.

El método New Cartoon permite visualizar de forma simplificada la estructura secundaria de su proteína. Las hélices se visualizan en forma de espiral, láminas β en forma de flechas y las otras estructuras en forma de un tubo delgado. Este método es el de mayor uso en representaciones de sistemas proteicos

- 9 En el menú Coloring Method es posible elegir el color de la representación. Diferentes estilos se encuentran disponibles, por ejemplo: Nombre, tipo de átomo, nombre del residuo, estructura, etc. Coloree la representación usando la información de estructura secundaria.



Otros métodos de dibujo. CPK y Licorice son métodos de dibujo que permiten observar estructuras con facilidad. La primera emula los sistemas de bolas de radio variable y tubos utilizados en química para visualizar moléculas, mientras que la segunda representa los átomos como esferas de radio fijo que no puede ser modificado.

10 La entrada de texto “Selected Atoms” de la ventana Graphical Representations permite acotar los átomos que se están visualizando. Reemplace la palabra *all* por *protein* y presione **Apply** (o presione **Enter** en el teclado), lo cual debe realizar cada vez que modifica un campo en la ventana Graphical Representations.

11 En la ventana Graphical Representations elija la viñeta **Selections**. En la sección **Singlewords** se encuentran los diferentes campos por los cuales se pueden realizar selecciones rápidas. Seleccione “beta” en el campo de texto para seleccionar solo los átomos de la proteína que se encuentran formando láminas β .

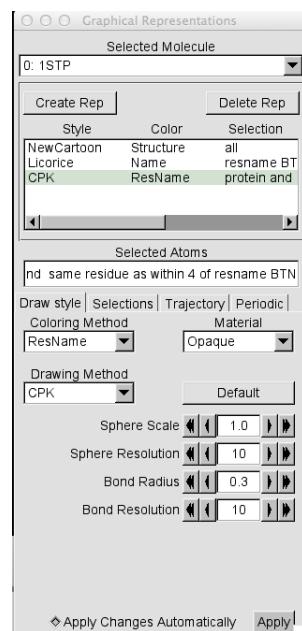


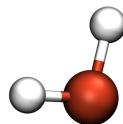
Figura 7Ventana Graphical Representations

12 Para visualizar la proteína sin helices y láminas β escriba en Selected Atoms: (*not helix*) and (*not betasheet*).

Con el ratón se pueden crear y seleccionar diferentes representaciones. Además, estas pueden activar o desactivar haciendo doble-click sobre ellas o ser borradas utilizando el botón **Delete Rep** button, etc.

13 Modifique la selección de átomos para representar toda la proteína, luego desactívela haciendo doble click sobre la representación.

14 Cree una nueva representación, sin borrar la existente, para esto utilice el botón **Create Rep**.



Selecciones geométricas complejas. Distintos tipos de geométricas se pueden lograr utilizando selecciones complejas en relación al espacio cartesiano. Cuál es la selección apropiada para obtener una esfera a partir de las moléculas presentes. Recuerde que una molécula no puede quedar incompleta!

En la viñeta **Selections** en el campo **Keywords** es posible ver distintas palabras claves que sirven para seleccionar átomos. Inspeccione en **Keywords** los nombres de los residuos que componen el sistema. Para este caso, se muestran los nombres de los residuos aminoacídicos de la proteína.

15 Cree una nueva representación, para esto utilice el botón **Create Rep**

16 Al crear una nueva representación se copian los valores de la selección anterior. Para seleccionar un residuo de nombre conocido, por ejemplo el residuo de nombre HIS, utilizaremos la palabra clave “*resname*” seguida del nombre del residuo como aparece en el archivo PDB, para ello escriba “*resname HIS*” en el campo “*Selected Atoms*”. Para una mejor representación cambie el método de dibujo por *licorice* y coloore la representación por nombre de átomo.

En esta nueva representación se ha visualizado los residuos de histidina que están presentes en esta proteína.

Diferentes campos pueden utilizarse para crear selecciones de mayor complejidad. El cuadro 4 muestra algunas selecciones posibles.

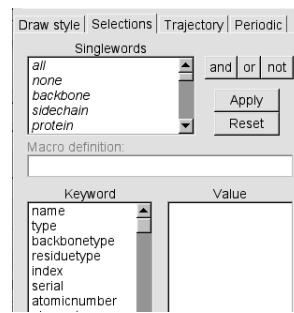


Figura 8 Campos disponibles para seleccionar

Selección	Acción
<code>protein</code>	muestra una proteína si esta presente.
<code>protein and chain E</code>	muestra la cadena E si esta es proteína y si esta presente.
<code>resid 1</code>	el primer residuo
<code>resid 1 2 12</code>	el primer residuo, el segundo y el decimosegundo residuo
<code>x >0</code>	átomos que tienen la coordenada <i>x</i> positiva
<code>same residue as x>0</code>	átomos de residuos que tienen un átomo cuya coordenada <i>x</i> es positiva
<code>name CA and x>0</code>	átomos de nombre CA (carbonos alfa) y coordenada <i>x</i> es positiva
<code>within 5 of resid 1</code>	átomos a 5 Angstrom del residuo 1
<code>backbone</code>	átomos del backbone

Cuadro 4 Ejemplo de selecciones

17 Para observar los residuos de la proteína que se encuentran cercanos al residuo *HIS* se utiliza una selección de mayor complejidad. Cree una nueva representación y escriba la siguiente selección para los átomos que están a 4 Å de distancia del residuo 63: `protein and within 4 of resname HIS and resid 63`. Utilice el método de dibujo CPK y coloore por nombre del residuo.

En la ventana OpenGL se observan átomos de distintos colores, sin embargo, es de mayor utilidad seleccionar los residuos que tienen átomos que cumplen una condición.

- 18** Cree representaciones para los siguientes residuos del sitio activo 63 75 77 80 85 107 200, los cuales están descritos un UNIPROT para esta proteína¹⁷. El resultado debe asemejar a la figura 9 en la cual hemos agregado una representación de la superficie de la proteína utilizando el método QuickSurf

<http://www.uniprot.org/uniprot/P00760>.

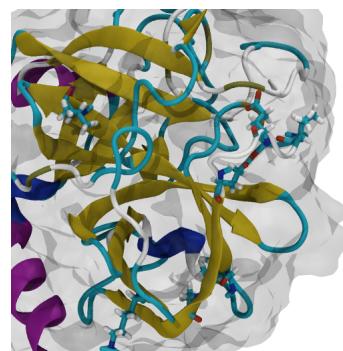
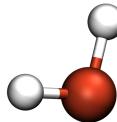


Figura 9Representación Final

- 19** Utilice el ratón para orientar la visualización. Para volver a la visualización inicial presione Display → Reset View.



Enlaces de Hidrógeno. La diferencia de electronegatividad entre los átomos de hidrógeno y oxígeno permite que las moléculas de agua puedan compartir H entre ellas, formando lo que se denomina puente o enlace de hidrógeno. Estos se pueden visualizar utilizando el método Hbond en la ventana de representaciones. Cuales son los parámetros necesarios para que se formen? tiene alguna relación la formación de puentes de hidrógeno con la estructuras observadas?

- 20** Para reconocer un residuo específico es posible utilizar el ratón y la ventana OpenGL para conocer su identificador. Para ello en la ventana VMD Main, elija el menú Mouse → Query y haga click sobre un átomo de interés. Al hacerlo, aparecerá en pantalla información básica del átomo seleccionado. Para obtener mayor información en la ventana VMD Main elija el menú Graphics → Label, que abrirá una ventana con la descripción del átomo seleccionado.

Otra forma de seleccionar un átomo es utilizando el modulo **Sequence Viewer**, el cual se encuentra en la Vineta Extension/Analisis de la pantalla Main

- 21** Utilice Sequence Viewer para verificar que los residuos que selecciono utilizando su número de residuo coinciden con los esperados a partir de la numeración en la secuencia encontrada en UNIPROT.

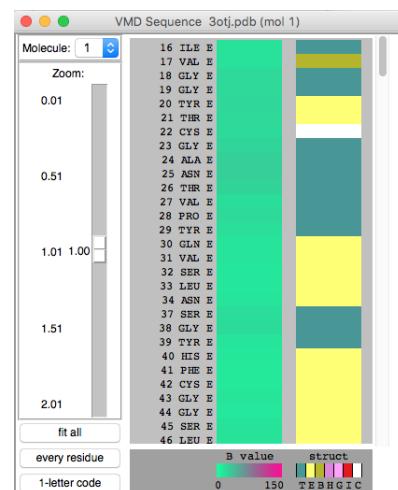


Figura 10Sequence Viewer

VMD es una herramienta util para describir y no hemos explorado todas sus capacidades. Por ejemplo, seleccionando en la ventana VMD Main el menú Mouse → Labels → Bond es posible conocer la distancia entre dos átomos de interés. Al hacerlo, aparecerá en la ventana Label bajo la etiqueta Bonds → value la distancia entre los átomos seleccionados en la configuración actual.

3.3. Guardando su trabajo

La imagen creada puede ser almacenadas en archivos utilizando VMD. Además, las representaciones creadas pueden ser almacenadas utilizando un estado de VMD. El archivo de estado contienen la información necesaria para reiniciar las sesiones de VMD, sin perder el trabajo realizado.

- 1 En la ventana VMD Main, elija el menú File → Save State. Escriba un nombre apropiado para el archivo (i.e, 30TJ.vmd) y guárdelo en el directorio de trabajo.
- 2 cierre y reinicie VMD para cargar el estado guardado

El menú File → Load State permite cargar el estado guardado previamente. Para guardar en imágenes las representaciones realizadas, VMD puede generar un archivo imagen que puede ser utilizado en otros programas.

- 3 Usando lo aprendido, genere una imagen de calidad, que muestre alguna característica de interés del sistema (por ejemplo la interacción entre dos residuos de interés o un metal). Ponga cuidado en la resolución de cada representación.



Operadores lógicos. Habitualmente utilizamos los operadores booleanos "Y", "O" y "NO". El primer es un operador "incluyente", el segundo de "exclusión", mientras que el tercer operador es considerado de "reunión" o suma lógica. Podemos ver la diferencia entre ellos observando la cantidad de átomos de selecciones que las incluyan como: "[protein or x>0]", "[protein and x>0]" y "[protein and not x>0]".

- 4 Cambie el color del fondo seleccionando el menú Graphics → Colors. Elija la categoría Display → Background → 8 white. El fondo debe estar en blanco ahora.
- 5 Elija el menú File → Render.... Se abrirá una ventana llamada File Render Controls.
- 6 Puede utilizar distintos métodos para generar la imagen, seleccione Snapshot o Tachyon.



Troubleshooting. Si el directorio donde guardara los archivos contienen espacios, es necesario que antes de utilizar File Render Controls, seleccionar el directorio de trabajo utilizando la Console Tk. Para esto seleccione el menú Extensions → Tk Console. Use comillas para navegar a la dirección, i.e., cd 'C:\ Documents and Settings'.

- 7 Escriba el nombre del archivo a generar en el campo de texto Filename , i.e., **imagen.bmp**.
- 8 Presione el botón Start Rendering y se generará el archivo con la imagen que aparece en la ventana OpenGL. Esto puede tardar algún tiempo dependiendo de las capacidades del computador utilizado. Al finalizar, el archivo se encuentra en el directorio seleccionado o en el directorio de trabajo.
- 9 Antes de comenzar con la siguiente sección explore VMD libremente. Al terminar, reinicie VMD.



Archivos de Coordenadas y de Conectividad. Los programas de visualización como VMD, infieren los distintos enlaces a partir de las distancias entre átomos, por lo que habitualmente se pueden encontrar átomos con una mayor cantidad de enlaces que los permitidos, ie. carbonos con 5 átomos asociados, moléculas de agua con 4 o más hidrógenos. Para solucionar esto se hace uso de un archivo extra que contiene la conectividad de cada átomo, además de otros parámetros necesarios para realizar una simulación, por ejemplo las cargas atómicas.

3.4. PROSITE

Este programa permite encontrar proteínas homólogas a partir de secuencias a través de la búsqueda en bases de datos utilizando HMMs.

1 Ingrese a Prosite¹⁸.

- A** Describa que información que le entrega esta página

A partir de los códigos UNIPROT de las proteínas elegidas (Pista: Se puede ingresar un máximo de 10 códigos UniProt por búsqueda).

- B** Registre el/los nombre(s) de los dominios y patrones encontrados para cada código UniProt.

- C** ¿Cuántos patrones únicos diferentes identificó entre las secuencias? informe el nombre del patrón, y la secuencia consenso para cada uno de ellos.

3.5. Pfam

Pfam contiene información de grupos (familias) de proteínas homólogas alineadas, sus anotaciones funcionales, y modelos extraídos de cada alineamiento que pueden ser usados para clasificar proteínas en familias.

Ingresé a Pfam:¹⁹ y en la sección titulada “search” busque cada una de las de las proteinas indicadas anteriormente. Para esto necesitará el código UNIPROT.

- D** ¿Que información le entrega esta página?

- E** ¿Como se clasifican las proteinas en ella?

- F** ¿A qué familia(s) corresponde cada secuencia.

- G** informe el nombre de la familia, y al menos una porción del logo que define cada una de las familias

3.6. InterPro

Vaya a InterPro²⁰ y busque nuevamente las secuencias indicadas.

- H** Para cada proteína, vaya a la sección “Detailed signature matches”. Registre los patrones/familias identificadas por InterPro para cada una de las secuencias. (PISTA: Si pasa el mouse sobre cada fila con información en esta sección, se expandirá una ventana con el nombre completo del motivo/familia y la base de datos de donde proviene.
- I** La información obtenida para Pfam y Prosite desde InterPro, ¿corresponde con aquella obtenida por usted desde esas bases de datos directamente en los puntos anteriores?
- J** ¿Qué otras bases de datos, además de Pfam y Prosite, están contenidas en InterPro y aportan información para las proteínas que usted investigó? Nombre las bases de datos.

¹⁸<http://prosite.expasy.org/>

¹⁹<http://pfam.sanger.ac.uk/>

²⁰www.ebi.ac.uk/interpro

3.7. Alineamientos múltiples

Son generalmente alineamientos de 3 o mas secuencias biológicas, con el fin de inferir las relaciones evolutivas en base a los patrones de secuencias y motivos conservados que existan entre ellas. En la página web de EMBL-EBI²¹ podemos encontrar varios servicios que permiten realizar alineamiento multiple de secuencias, entre los cuales están Clustal, MAFFT, MUSCLE, T-coffe, entre otros.

- 1 Vaya a Uniprot²², y en la pestaña titulada “Retrieve” ingrese los códigos Uniprot de las 5 proteínas. Luego presione el botón “Retrieve” y baje el archivo de extensión fasta. Puede omitir este paso si ya tiene las secuencias de sus proteínas.

Para el primer alineamiento múltiple utilizaremos un algoritmo progresivo.

- 2 En la misma página donde se encuentra dentro de UniProt, presione la pestaña titulada “Align”. Ingrese las 5 secuencias que bajo en formato FASTA y luego presione “Align”.

Aquí usted está ejecutando el programa Clustal de alineamiento múltiple progresivo.

- K** ¿Qué tan largo es el alineamiento (pista: cada bloque alineado consta de 60 residuos/gaps)? Describa el alineamiento obtenido.

- 3 Baje el alineamiento en formato FASTA.

Ahora utilizaremos un algoritmo iterativo de alineamiento múltiple.

- 4 Vaya a Genome²³. Cambie el “Output Format” a CLUSTAL. Luego ingrese las 5 secuencias que bajo en formato FASTA desde UniProt y ejecute el alineamiento.

- L** ¿Qué tan largo es el alineamiento (pista: cada bloque alineado consta de 60 residuos/gaps)? Describa el alineamiento obtenido.

- M** A continuación utilizaremos algoritmo basado en bloques de alineamiento múltiple. Vaya a BiBiServ²⁴ submission.html. Ingrese las 5 secuencias que bajo en formato FASTA y presione “Submit”. Una vez que el resultado del cálculo esté listo, siga el link para ver “original dialign output”.

- N** ¿Qué tan largo es el alineamiento (pista: cada bloque alineado consta de 50 residuos/gaps)?

Para comparar los alineamientos podemos utilizar diferentes métricas, por ejemplo, podemos utilizar como criterio que el mejor alineamiento es el que posee una menor cantidad de gaps.

- Ñ** Ordene los alineamientos de peor a mejor. Por ejemplo, basado en bloques <iterativo <progresivo.

Los alineamientos sirven para múltiples propósitos

- 5 Identifique los patrones de dominios identificados en la Parte I.

- 6 ¿Qué cree usted que se podría hacer para mejorar los alineamientos obtenidos?

²¹<https://www.ebi.ac.uk>

²²<http://www.uniprot.org/>

²³<http://www.genome.jp/tools/prrn/>

²⁴<http://bibiserv.techfak.uni-bielefeld.de/dialign/>

3.8. Búsqueda en bases de datos usando PSI-BLAST y HMM, y búsqueda en Pfam.

- 1 Ingrese nuevamente a Uniprot²⁵, y en la pestaña titulada “Retrieve” ingrese los códigos Uniprot de las 5 proteínas. Luego presione el botón “Retrieve” y siga el link que dice: “UniProtKB (5)”. Esto le dará acceso directo a las páginas en UniProt de cada una de las 5 proteínas que estamos estudiando. Vaya a cada una de ellas e investigue los códigos Pfam asociados a cada una. En cada página siga el/los link(s) a Pfam y reporte el número total de secuencias asociadas a cada familia.

Ejecutaremos el programa PSI-BLAST para crear un perfil para cada una de las proteínas estudiadas, y encontrar el mayor número posible de secuencias en la base de datos no redundante.

Para cada una de las 5 proteínas haga lo siguiente.

- 2 Ingrese a la página de BLAST²⁶ y seleccione blastp.

Ingresar el código UniProt de la proteína de interés. Donde dice “Program Selection” seleccione como “Algorithm” PSI-BLAST. Extienda la sección que dice “Algorithm parameters” y en “Max target sequences” seleccione 20000, y en “Expect threshold” escriba 0.00001. Finalmente en la sección que hace referencia a PSI-BLAST al final de la página, en “PSI-BLAST Threshold” escriba 0.00001. Verifique que como “Database” este utilizando “Non-redundant protein sequences (nr)”.

- O** Describa cada uno de estos parámetros.

- 3 Una vez ingresados los parámetros, presione el botón BLAST. Una vez que aparezcan los resultados, vaya a la sección que dice “Sequences producing significant alignments ...”. Presione el botón que dice: Select All. ¿Cuántas secuencias se obtuvieron a partir de la primera búsqueda? Vaya a la sección que dice “Run PSI-Blast iteration 2” y presione el botón “Go”. ¿Cuántas secuencias se obtuvieron a partir de la segunda iteración? Repita una tercera, cuarta y quinta iteración, siempre anotando el número de secuencias obtenidas. (Nota: No todas las proteínas le van a permitir hacer 5 iteraciones.) ¿Cómo se compara el número de secuencias obtenidas usando PSI-BLAST, a aquéllas reportadas para las respectivas familias de cada proteína en Pfam?

c. Búsqueda en bases de datos utilizando HMMs. Ahora realizaremos una búsqueda más compleja. Usando el mejor alineamiento múltiple encontrado en la pregunta 1 (pista: bájelo en formato FASTA), vaya a la página de hmmer²⁷ y ejecute una búsqueda utilizando HMMs, manteniendo todos los parámetros predeterminados. (Note que está haciendo una búsqueda contra la base de datos NR: non redundant, la misma que utilizó en la pregunta anterior). ¿Cuántas secuencias se obtuvieron a partir de la búsqueda? ¿Cuántas secuencias hubiera esperado usted obtener de la búsqueda (pista: compare con los resultados obtenidos en 2a y 2b)?

4. Evolución y estructura

Históricamente, las proteínas han sido clasificadas usando la similitud de secuencia. Las familias de proteínas son combinadas en superfamilias basados en su actividad catalítica, motivos de secuencias u otras cualidades conservadas. Dado lo anterior, se infiere que proteínas dentro de la misma superfamilia provienen de un ancestro en común, aún cuando puedan tener actividades enzimáticas diferentes o desconocidas.

²⁵<http://www.uniprot.org/>

²⁶<http://www.ncbi.nlm.nih.gov/BLAST/>

²⁷<http://hmmer.janelia.org/search/hmmsearch>

La evolución divergente puede ser trazada a través de similaridades de secuencias y/o plegamiento. Las familias divergentes típicamente pertenecen a la misma superfamilia, comparten motivos estructurales y funcionales, y exhiben identidad de secuencia detectable.

En cambio la evolución convergente implica la aparición de propiedades estructurales o funcionales similares dentro de familias de proteínas que no divergieron de un ancestro en común. Típicamente, estas familias pertenecen a distintas superfamilias y no exhiben identidad de secuencia.

5. Descripción del práctico

El objetivo del práctico de hoy, es determinar si existe **evolución divergente o convergente** en cada uno de los sets otorgado. Para ello, utilizaremos cuatro sets de archivos PDB. **Es su misión averiguar cual es la función de cada grupo.*

- **Glutation:** 1AGS (Cadena A), 1EEM (Cadena A), 1FHE (Cadena A), 1GSE (Cadena A) y 1K0D (Cadena A)
- **G-proteinas:** 1AGP (Cadena A), 1BYU (Cadena A), 1EGA (Cadena A), 1EGA (Cadena A) y 1NF3 (Cadena A)
- **Serine:** 1EP5 (Cadena A), 1HAX (Cadena A), 1LMW (Cadena A), 1OP0 (Cadena A) y 1RTF (Cadena A)
- **Subtilisinas:** 1DUI (Cadena A), 1NDU (Cadena A), 1SBI (Cadena A), 1T1E (Cadena A) y 2ID4 (Cadena A)

Puede descargarlos desde la página de Protein Data Bank, o acceder a la carpeta otorgada junto con este práctico.

6. Alineamiento estructural de estructuras

Para realizar esta parte debemos cargar el grupo de estructuras a evaluar en VMD (Si hay dudas en este paso, recurrir a tutoriales anteriores).

Una vez cargadas, iremos a la selección del menú principal: Extensiones, Analysis -> Multiseq. Si estamos abriendo por primera vez Multiseq, el programa nos pedirá instalar o actualizar bases de datos. Si este es el caso, debemos asignar "Yes" esperar que se carguen todas las librerías.

En el menú de Multiseq (Figura 11) se verán las secuencias de las estructuras cargadas en el menú principal de VMD. Si la estructura tiene varias cadenas se cargarán cada una por separado o bien otro tipo de moléculas (Agua, iones, etc), solo se verán signos de interrogación (?). Estos últimos deben ser descartados. Adicionalmente en la parte File, se pueden agregar mas secuencias provenientes de estructuras o en otros formatos.

A continuación utilizaremos el programa STAMP para alinear nuestras estructuras.

Dirigirse a Tools->Stamp Structural Alignment Tool. La ventana mostrará los parámetros requeridos. Por ahora usaremos el cuadro por defecto.

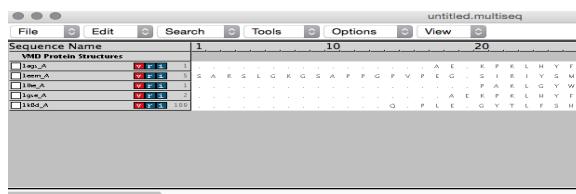
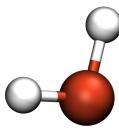


Figura 11 Menú principal de Multiseq



STAMP. El programa STAMP (Structural Alignment of Multiple Proteins) es una herramienta que alinea secuencias de estructuras de proteínas utilizando la estructura tridimensional. El algoritmo de STAMP busca minimizar la distancia entre C β de residuos aplicando rotaciones de cuerpo rígido y traslaciones. STAMP solo puede alinear proteínas que tengan secciones comunes. Información adicional: <http://www.rfcgr.mrc.ac.uk/Registered/Help/stamp/stamp.html>

Una vez aplicado, en la pantalla principal, veremos nuestras proteínas alineadas. En la ventana de Multiseq, veremos que las secuencias se alinearon en relación a las estructuras. En esta etapa puede colorear las estructuras bajo distintos parámetros: Conservación de secuencia, identidad, etc. Buscar esas opciones en la sección View->Coloring (Figure 12)

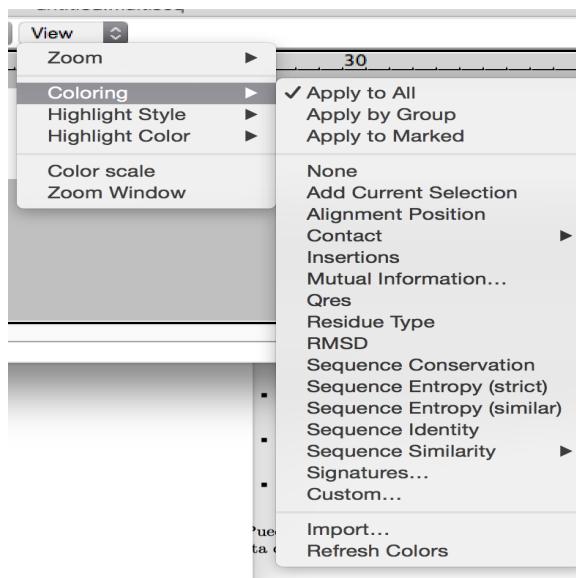


Figura 12 Menú para colorear con respecto a distintos parámetros

De resultado, podremos ver las secuencias y las estructuras coloreadas bajo el parámetro deseado.

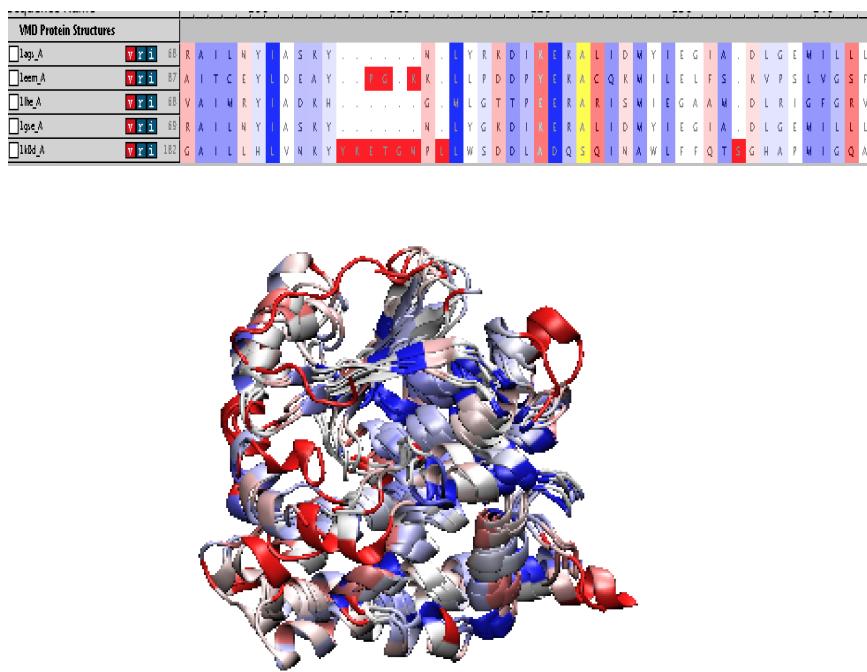


Figura 13 Resultado de secuencias y estructuras coloreadas

6.1. Árbol filogenético

La opción de Arbol Filogenético en el programa Multiseq puede mostrarnos relaciones entre las proteínas consideradas.

Los árboles filogenéticos basados en estructuras se construyen a partir de valores de RMSD o Q value entre las moléculas evaluadas.

Para usar esto debemos

- Alinear las estructuras usando STAMP
- En el programa Multiseq elegir Tools->Phylogenetic Tree. Aquí se puede elegir los parámetros requeridos, entre ellos valor Q_H o RMSD. ¿Qué es el valor Q_H?
- Seleccionar Árbol estructural usando Q_H y presiona OK. (Figura 14)

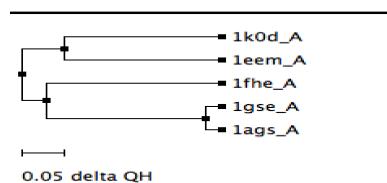
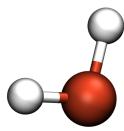


Figura 14 Imagen de ejemplo de árbol estructural



Valor Q_H. El valor Q_H es una métrica que señala homología estructural similar al valor Q. Se compone de dos partes: Qalign (parte alineada) and Qgap (parte con gaps).

Una vez terminada esta parte, podemos repetir la experiencia con los otros set de estructuras.