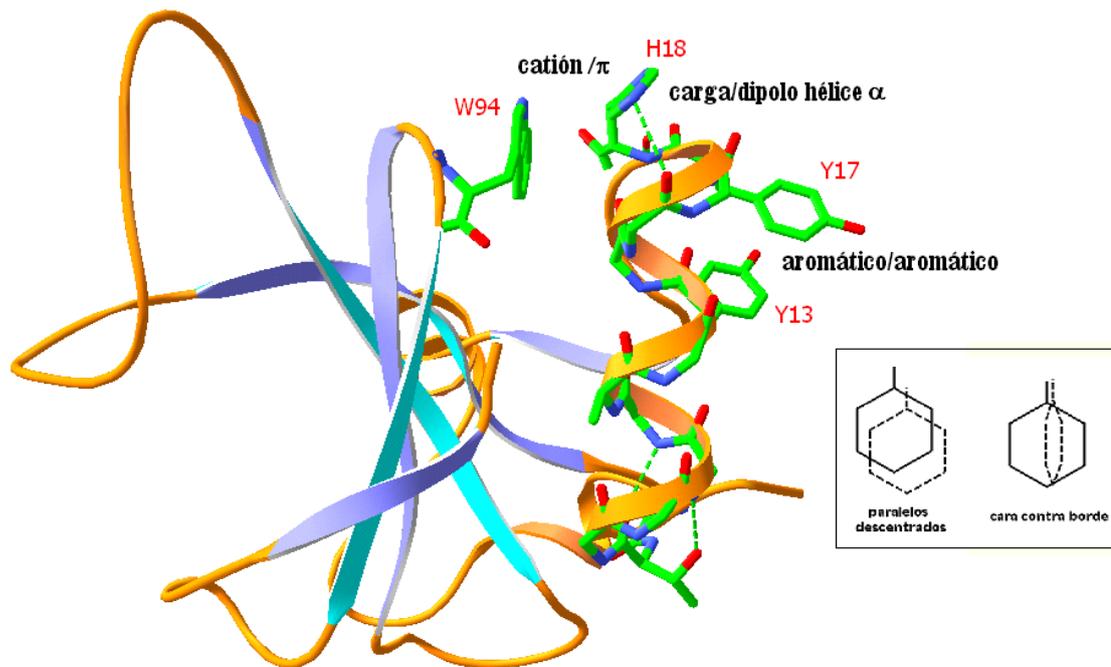


# Prácticas de Estructura de Macromoléculas 2021-2022

## Grado de Biotecnología



Javier Sancho [jsancho@unizar.es](mailto:jsancho@unizar.es)  
Marta Martínez [mmartine@unizar.es](mailto:mmartine@unizar.es)

Departamento de Bioquímica y Biología Molecular y Celular. Facultad de Ciencias &  
Instituto de Biocomputación y Física de Sistemas Complejos (BIFI).  
Universidad de Zaragoza



## **Práctica nº 1: Obtención de coordenadas de macromoléculas del PDB y visualización de su estructura tridimensional.**

### **1. INTRODUCCION**

Las coordenadas XYZ de los átomos de una macromolécula son la información necesaria para representar su estructura. Las coordenadas de las macromoléculas cuya estructura tridimensional se ha resuelto por difracción de Rayos X o resonancia magnética nuclear se encuentran depositadas en el **Protein Data Bank (PDB)**. Al PDB se accede fácilmente por Internet: <http://www.rcsb.org/>. Una vez ahí se introduce el código PDB de la proteína deseada o se realiza una búsqueda por palabras (en inglés).

### **2. VISUALIZACION DE ESTRUCTURAS DE PROTEINAS CON EL PROGRAMA SPDBviewer**

SPDBviewer (<http://us.expasy.org/spdbv/>) es un programa gratuito que permite representar la estructura tridimensional de moléculas utilizando coordenadas con formato pdb. Se puede utilizar, como una aplicación independiente al margen de Internet, abriéndolo y cargando las coordenadas de la molécula, o se puede utilizar como un programa auxiliar del navegador de internet que se activa automáticamente cada vez que hace falta. Para esta práctica y para la práctica 2, descargaremos a nuestro ordenador de trabajo la aplicación (versión 4.1 para MS Windows), la guía de usuario (User Guide.pdf) y el tutorial (Tutorial Material).

Una vez cargada una molécula en **SPDBviewer** (abriendo desde el menú de la aplicación el archivo .pdb correspondiente a la estructura que se desea visualizar) podemos verla de diversas maneras: como varillas, como bolas y varillas, como bolas, sólo los carbonos alfa de cada aminoácido, sólo unas cintas que representan el discurrir del polipéptido, etc.

La molécula se puede rotar, trasladar y ampliar o reducir convenientemente. Se puede ver los puentes de hidrógeno, calcular distancias, los vecinos de un residuo, las cavidades de la molécula, etc. Las posibilidades son muchas y hay que descubrirlas explorando el menú o leyendo el manual.

### **3. EJERCICIOS**

- ⇒ **1fn** ¿Qué proteína es? ¿A qué tipo estructural pertenece? ¿Cuál es el aminoácido C-terminal de la hélice N-terminal?
- ⇒ **1flv** ¿Qué proteína es? ¿A qué tipo estructural pertenece? ¿Qué aminoácidos aromáticos están en contacto con el grupo prostético FMN?
- ⇒ **1hti** ¿Qué proteína es? ¿A qué tipo estructural pertenece? ¿Cuántas metioninas contiene?
- ⇒ **1aqb** ¿Qué proteína es? ¿A qué tipo estructural pertenece? ¿Cuál es su cofactor? ¿Qué aminoácidos polares se encuentran en contacto con el cofactor?
- ⇒ ¿Cuántos mutantes de la lisozima del fago T4 se han cristalizado?
- ⇒ Visualizar algún complejo proteína/DNA
- ⇒ Visualizar alguna molécula de DNA
- ⇒ Visualizar algún carbohidrato
- ⇒ Comparar la estructura de la ribonucleasa barnasa resuelta por difracción de rayos X (**1a2p**) con la determinada por RMN (**1bnr**).

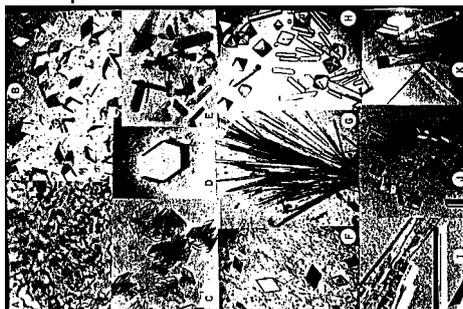
## Práctica nº 2: Resolución estructural de una proteína mediante cristalografía de rayos X: cristalización y refinamiento.

### 2.1.- Cristalización de lisozima

#### 1. INTRODUCCION

La característica principal de un cristal es su estructura interna periódica y ordenada en tres dimensiones. El hecho de que un sistema renuncie a sus grados de libertad parece una contradicción desde el punto de vista entrópico y, sin embargo, la cristalización de moléculas es un fenómeno espontáneo y reversible cuyos parámetros cinéticos y termodinámicos dependen de las propiedades químicas y físicas del solvente y del soluto involucrados. La cristalización se produce cuando, bajo ciertas condiciones (de supersaturación), el sistema se dirige hacia el estado de equilibrio entre dos fases, una soluble y una sólida. A pesar de que las moléculas individuales fijan su posición, disminuyendo así la entropía del sistema, establecen al mismo tiempo nuevas interacciones reduciendo la energía libre total del sistema. Las moléculas, por tanto, cristalizan minimizando su energía libre al disponerse periódicamente en unidades repetitivas y simétricas en el estado sólido, maximizando las interacciones intermoleculares favorables.

La solvatación es un factor vital para la cristalización de macromoléculas. En soluciones extremadamente concentradas, donde no hay agua suficiente para mantener la hidratación, las moléculas pueden o bien agregar como un precipitado amorfo o bien cristalizar (figura 1). La estrategia para inducir la cristalización consiste en llevar al sistema hacia el estado de mínima solubilidad lentamente y lograr así un grado limitado de supersaturación. Para lograr que las moléculas tengan oportunidad de formar el mayor número de interacciones favorables con los vecinos se modifican las propiedades del solvente con agentes precipitantes o se alteran algunas propiedades físicas como la temperatura.



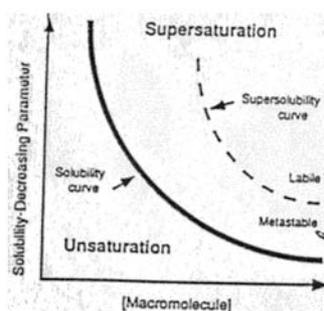
**Figura 1. Algunos cristales de proteínas crecidos mediante una variedad de técnicas y usando diferentes agentes precipitantes.** (a) catalasa de hígado, (b) fructosa 1,6-difosfatasa, (c) proteína de unión a cortisol, (d) concanavalina, (e) catalasa de hígado, (f) proteína de función desconocida de piña, (g) factor de elongación Tu, (h) tRNA fenilalanina, (i) proteína de desenrollamiento de DNA, (j) glicerol 3-fosfato deshidrogenasa, (k) canavalina.

#### 2. FUNDAMENTOS FÍSICOS DE LA CRISTALIZACIÓN

La determinación de la estructura cristalográfica empieza con el crecimiento de un cristal adecuado. Para ello pueden aplicarse distintos métodos que se diferencian por las condiciones y solutos que emplean. Para decidir el método mejor para cristalizar una proteína determinada debemos entender los fundamentos de la cristalización.

La formación de un cristal de proteína macroscópico que contiene hasta  $10^{15}$  moléculas, empieza con la aparición de agregados de proteína mediante contactos intermoleculares (*agregación*). Estos agregados prenucleares eventualmente alcanzan el tamaño nuclear crítico. Una vez formado un núcleo estable (*nucleación*), el crecimiento prosigue por la adición de moléculas a la red cristalina. Tanto la nucleación del cristal como el *crecimiento* ocurren en soluciones supersaturadas donde la concentración de proteína es mayor que su valor de solubilidad en equilibrio. La región del *parámetro de solución* aconsejable para la cristalización viene recogida en el diagrama de fase (figura 2).

La supersaturación es una función de la concentración de la macromolécula y de diversos parámetros que afectan su solubilidad. En general, se logra a altas concentraciones de macromoléculas y a valores crecientes de parámetros de solución reductores de la solubilidad macromolecular. Los factores que pueden influir en la solubilidad proteica son por ejemplo, la inclusión de aditivos tales como alcoholes, polímeros hidrofílicos y detergentes que puedan disminuir la solubilidad proteica. Estos son *agentes precipitantes* ya que influyen en la solubilidad, pudiendo causar la precipitación macromolecular. Otro factor que influye sobre la solubilidad proteica es la presencia de sal, que varía generalmente como una curva asimétrica en forma de campana indicando que la solubilidad decrece tanto a altas como a bajas concentraciones de sal. Parámetros de la solución tales como pH o temperatura también pueden afectar dramáticamente a la solubilidad macromolecular.



**Figura 2. Diagrama de fase.** La curva de solubilidad (sólida) divide el espacio de fase en regiones que promueven el proceso de cristalización (soluciones supersaturadas) de aquellas donde se disolverá el cristal (soluciones insaturadas). La curva de supersolubilidad (líneas entrecortadas) divide además la región supersaturada en condiciones de más alta supersaturación donde compiten la nucleación y crecimiento (fase lábil) y niveles menores donde sólo crecerán cristales (fase metastable).

Las condiciones de supersaturación óptimas para la nucleación y para el crecimiento del cristal son diferentes. Esto se muestra en el diagrama de fase donde la región de supersaturación es además dividida en regiones de más alta supersaturación (región lábil) donde tanto el crecimiento como la nucleación ocurren y la disminución de supersaturación (fase metastable) donde sólo se alcanza el crecimiento.

### 3. MÉTODOS DE CRISTALIZACIÓN

#### **Métodos de difusión de vapor.**

Las condiciones dentro de una solución que contiene proteína son manipuladas mediante la difusión a través del aire (figura 3). Como con los otros métodos de cristalización, el objetivo inicial es formar núcleos. En este método, la nucleación ocurre cuando se aumenta la concentración de proteína por medio de la reducción del volumen de la solución. Esto ocurre porque el vapor de agua de la gota que contiene la

proteína se deshidrata buscando el equilibrio con la solución de reserva más higroscópica. Como muestra la figura 3, una vez que se han alcanzado las condiciones de nucleación, la solución permanece altamente supersaturada, de modo que puede producirse simultáneamente la nucleación y el crecimiento de cristales. Para evitar que los valores de supersaturación excedan los requeridos para la nucleación, se han diseñado métodos para alcanzar muy lentamente el equilibrio. A pesar que dentro de un marco teórico pareciera que las condiciones no son óptimas para cristalizar una proteína, la difusión de vapor por gota colgante (“hanging drop”) y gota posada (“sitting drop”) son amplia y muy exitosamente usados para el crecimiento de cristales de proteína. El hecho que la mayoría de gotas de difusión de vapor no estén llenas de cristales microscópicos sugiere que el alto grado de supersaturación que motiva nucleación no es mantenido a lo largo del experimento de cristalización. Varios factores, incluso la disminución de concentración de macromolécula debido a la formación de núcleos o precipitados, la incorporación en cristales crecientes o la desnaturalización macromolecular en la interfase solución-aire podría disminuir la concentración proteica a valores menos propicios para la nucleación y más adecuados para el crecimiento.



**Figura 3. Diagrama esquemático de la cristalización por difusión de vapor.** En este método de cristalización, las soluciones de proteína que contienen precipitante insaturado (indicado en el diagrama de fase por círculos abiertos) son suspendidos sobre un reservorio. El equilibrio, por medio de vapor de la gota y el reservorio, causa que la solución de la proteína alcance un nivel de supersaturación donde ocurren la nucleación y el crecimiento inicial (círculos sólidos). Como se ve en el texto, es probable que cambios en la concentración de la gota disminuyan la supersaturación a lo largo del tiempo del experimento (círculos sólido de izquierda).

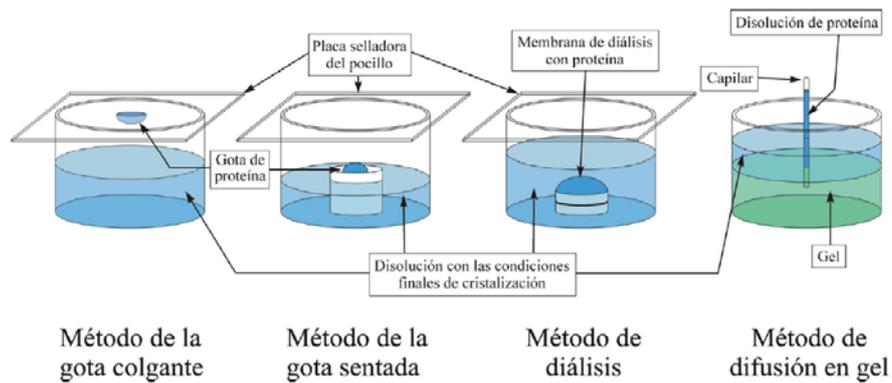
### Otros métodos de cristalización (consultar)

-Methods in Enzymology Volume 276. Macromolecular crystallograpy. Part A. Overview of protein crystallization methods. Patricia C. Weber. 13-22. Academic Press.

-Preparation and analysis of protein crystals. 1989. Alexander McPherson. 82-159. Krieger Publishin Company.

-How to grow protein crystals:

[http://www-structure.llnl.gov/crystal\\_lab/cystalmake.html](http://www-structure.llnl.gov/crystal_lab/cystalmake.html)



### CRISTALIZACIÓN DE LISOZIMA

En esta práctica intentaremos reproducir condiciones de cristalización previamente establecidas de la lisozima de huevo. Para ellos utilizaremos la técnica de difusión de vapor en gota colgante (hanging drop). Realizaremos un estudio breve de la influencia de la concentración de la proteína y del precipitante en la aparición y tamaño de cristales. Para ello, variaremos ambos parámetros.

Las condiciones de cristalización serán las siguientes. Lisozima a 30, 40 y 50 mg/ml, NaCl 1.0 M a 1.2 M, y buffer Acetato sódico (AcNa) 0,1 M pH4,5. Se probarán condiciones de acuerdo al esquema.

LISOZIMA	NaCl 1.0 M 0.1 M AcNa pH 4.6	NaCl 1.1 M 0.1 M AcNa pH 4	NaCl 1.2 M 0.1 M AcNa pH 4.6
50 mg/ml			
40 mg/ml			
30 mg/ml			

**Paso 1.** Hacer los cálculos respectivos para preparar las soluciones madre a partir de stocks de NaCl (2,5 M), buffer acetato sódico (3 M). El stock de lisozima (50 mg/ml) se diluirá con buffer AcNa 0,1 M pH 4,6 hasta la concentración deseada.

**Paso 2.** Con ayuda de una jeringa, se colocará grasa de animal alrededor de cada pocillo. Para evitar confusiones, marcar cada pocillo con la respectiva concentración de enzima y solución madre empleada.

**Paso 3.** Se pondrán 500 ul de solución madre en cada reservorio. Luego, se prepararán gotas en cristales cubreobjetos (previamente silinizados) mezclando 3 ul de solución proteica con 3 ul de solución madre a la concentración indicada. Con mucho cuidado, voltear el cubreobjetos y colocar sobre el reservorio, presionando ligeramente para que se adhiera bien a la superficie.

**Paso 4.** Tapar y mantener las placas bajo condiciones de temperatura estable y en una superficie libre de movimiento. Esperar de 5 a 12 horas.

**Paso 5.** Se observará cada uno de los pocillos con lupa de 40 aumentos y con lampara de luz fría.

**Paso 6.** Indicar:

- ⇒ ¿Qué condiciones son óptimas para el crecimiento de cristales de lisozima de huevo?.
- ⇒ ¿Cómo es un cristal de lisozima?.
- ⇒ ¿Qué ocurre en las demás condiciones probadas?.

Nota: la lisozima de huevo corresponde a la enzima liofilizada, de la compañía Sigma (código EC 3.2.1.17) contiene 47000 unidades/mg sólido y 95% de pureza.

## 2.2.- Ajuste de una secuencia de aminoácidos a un mapa de densidad electrónica.

### 1. INTRODUCCION

La difracción de un haz de rayos X por un cristal de proteína proporciona reflexiones de distinta intensidad y fase a partir de las cuales se puede calcular un mapa tridimensional de la densidad electrónica de la molécula proteica que ha difractado el haz. El cristalógrafo debe establecer un modelo tridimensional de la proteína que explique la distribución electrónica observada. En el proceso de generación del modelo (lo que al final vamos a llamar la estructura de la proteína) es frecuente realizar ajustes 'manuales' en el ordenador para que los aminoácidos de la proteína encajen bien en la densidad electrónica. Estos ajustes se realizan utilizando un programa capaz de representar simultáneamente el modelo de la proteína y el mapa de densidad electrónica y capaz también de generar conformaciones alternativas de la proteína para conseguir un mejor encaje en la densidad electrónica. En esta práctica vamos a comparar una porción de un mapa de densidad electrónica con un modelo de la proteína y encontraremos una región en la que el aminoácido que se ha introducido en el modelo no encaja con la densidad electrónica del mapa. Nuestro objetivo será encontrar el aminoácido que sí encaja bien para corregir el error del modelo.

### 2. USO DEL PROGRAMA SWISS MODEL PARA ENCAJAR SECUENCIAS EN MAPAS DE DENSIDAD ELECTRONICA

Para obtener los ficheros con los que se va a trabajar:

1. Descargar el paquete que contiene el tutorial de la práctica en formato comprimido (.zip) (<http://us.expasy.org/spdbv/text/tutorial.htm>)
2. Extraer los ficheros contenidos en el archivo zip en una carpeta

#### **Procedimiento:**

1. Abrir el fichero 1LDB\_ala.pdb (que está dentro de la carpeta tutorial): *File* → *Open PDB File*. Este archivo contiene nuestra proteína problema, en la que algunas zonas no tienen asignadas las cadenas laterales y en su lugar hay alaninas.
2. Centrar la proteína (clic derecho) en el residuo ALA63.
3. Abrir el mapa de densidad electrónica: 1LDB.dn6; *File* → *Open electron density map* → *Open DN6 map*. Seleccionar los siguientes parámetros: *display around CA*: 8.000

para las tres casillas y los valores de *contour* de 0.8 (1ª casilla) y de 1.5 (2ª casilla). De este modo solo veremos la parte del mapa a 8 Å del residuo seleccionado.

4. Cargar la secuencia de la proteína (1ldb.txt) desde el menú SwissModel (*SwissModel* → *Load raw sequence to model*). Mostrar la ventana que contiene los archivos abiertos (*Window* → *Layer info*) y hacer invisible la estructura marcada como 1ldb (haciendo clic sobre la “v” que hay debajo de “vis”)

5. Seleccionar (clic izquierdo y arrastrar) los residuos ALA61-ALA65. Desde el panel de control (*control panel*), dejamos a la vista sólo los residuos seleccionados (haciendo clic sobre la palabra “show” mientras los residuos están seleccionados). Comprobar que la vista de las cadenas laterales está activada.

6. Podemos utilizar la vista tridimensional (*Display* → *render in solid 3D*; *Display* → *Use OpenGL rendering*) para ayudarnos a la hora de ver cómo encajan, en el mapa de densidad electrónica, los distintos péptidos que nos va a proponer el programa. También podemos eliminar la cantidad de información en la pantalla, utilizando el modo de eliminación de planos (*Display* → *Slab*).

7. Pedir al programa el conjunto de péptidos que podrían encajar en la zona que vamos a estudiar (*Build* → *Find best fitting peptides...*). Debe aparecer (un poco de paciencia) una lista de pentapéptidos que además indica el número de átomos que quedan dentro de la densidad electrónica y el número de átomos que quedan fuera, así como una puntuación (*score*). Haciendo clic en el péptido de mayor puntuación aparece sustituyendo a la secuencia de alaninas (que habíamos seleccionado) y observamos cómo encaja en el mapa de densidad electrónica. Utilizando los cursores, podemos explorar la correspondencia de las demás opciones. Y buscar aquella que nos parezca que puede ser la solución correcta.

8. Para comprobar si el péptido que hemos elegido es el correcto, abrimos una ventana que contiene el alineamiento de la proteína problema (1ldb\_ala) con la secuencia real (1ldb): *Windows* → *alignent*. Los residuos seleccionados en el panel de control aparecen marcados en gris. Si no hemos encontrado la solución, podemos buscarla en la lista de los péptidos y explorar cómo se disponen dentro del mapa de densidad electrónica.

9. A veces, la solución correcta no encaja bien en el mapa, porque el sistema trabaja creando rotámeros determinados. Podemos realizar un refinamiento manual del residuo, utilizando la herramienta *Torsión* del menú principal. Modificar los ángulos de las laceras laterales para encajar correctamente el residuo y aceptar la conformación desactivando *Torsión*.

### **Ejercicios:**

1. Practicar con la secuencia ALA33-ALA37. Primero, centrar en el residuo ALA35 y seleccionar los residuos ALA33-ALA37. Activar en el panel de control las cadenas laterales y principal (“show” y “side” activados); se puede desactivar la vista del resto de la proteína para limpiar la pantalla de información. En este caso, el péptido correcto aparece entre los primeros puestos de la lista. Comprobarse hemos elegido la solución correcta y refinar las cadenas laterales, si fuera necesario.

2. Practicar con la secuencia ALA38-ALA42. Centrar en el residuo ALA40 y repetir todo el proceso ¿En qué posición está el péptido que corresponde con la secuencia correcta?

3. Practicar con la secuencia ALA43-ALA47 (centrar en el residuo ALA45). El péptido correcto aparece muy alejado de los primeros puestos de la lista. ¿Qué ocurre si repetimos el cálculo con ALA43-ALA46 (centrar en el residuo ALA45)? La solución aparece con la mayor puntuación. Volver a intentarlo con la secuencia ALA43-ALA47, ¿qué ocurre con el residuo 47? Intentar encajar correctamente el GLU47 en el mapa de densidad electrónica (utilizando las herramientas de *Torsión*) Para comprobar cómo ha resultado el refinamiento, cargar el fichero 1LBD.pdb que contiene las coordenadas completas de la proteína.

## Práctica nº 3: Medida de la estabilidad conformacional de una proteína mediante desnaturalización química.

### 1. INTRODUCCION

La conformación plegada, nativa, de las proteínas es la más estable sólo en un determinado intervalo de valores de temperatura, pH, fuerza iónica, etc.

Una forma sencilla de determinar la estabilidad conformacional de una proteína es realizar una desnaturalización térmica registrando el valor de alguna propiedad de la proteína en función de la temperatura y ajustando los datos obtenidos a un modelo realista del equilibrio de desnaturalización térmica. Otra forma, análoga, es mediante desnaturalización química, utilizando una sustancia que desestabilice la conformación nativa. Este segundo método, que es extratermodinámico, recibe el nombre de método de extrapolación lineal (LEM). En esta práctica vamos a utilizarlo para determinar la estabilidad de la proteína lisozima empleando cloruro de guanidinio como desnaturalizante.

### 2. REACTIVOS

- Tampón acético/acetato sódico 45 mM, pH 3.0
- Lisozima 10mg/ml en tampón acético/acetato sódico 45 mM, pH 3.0
- Cloruro de guanidinio 7M en tampón acético/acetato sódico 45 mM, pH 3.0

### 3. PROCEDIMIENTO

- Preparar, en viales Eppendorf, 950 µl de cloruro de guanidinio 0.0, 0,5, 1.0, 1.5, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 3.6, 4.0, 4.5, 5.0, 5.5, 6.0 y 6.5 M, por dilución en tampón acético/acetato sódico 45 mM, pH 3.0
- Añadir a cada vial, 50 µl de solución de lisozima
- Incubar a temperatura ambiente durante 20 minutos
- Medir la absorbancia de las disoluciones a 301 nm, empezando por la de menor concentración de cloruro de guanidinio.

### 4. DETERMINACION DE LA $\Delta G$ DE DESPLEGAMIENTO

#### **A: Procedimiento manual:**

- Representar la curva de desnaturalización
- Trazar con regla las líneas base de la absorbancia de los estados plegado y desplegado
- Calcular, en la zona de transición, las constantes de equilibrio a distintas concentraciones de cloruro de guanidinio
- Calcular las correspondientes  $\Delta G$  a esas concentraciones
- Representar las  $\Delta G$  en función de la concentración de cloruro de guanidinio
- Calcular la  $\Delta G$  de la lisozima en ausencia de cloruro de guanidinio, por extrapolación lineal a concentración 0 M de desnaturalizante

#### **B: Ajuste computacional:**

El modelo utilizado para realizar el ajuste manual asume lo siguiente:

- a) La absorbancia observada (A) es la que corresponde a las de los estados nativo ( $A_N$ ) y desplegado ( $A_D$ ) ponderadas por sus fracciones molares:

$$A = A_N X_N + A_D X_D$$

- b) La absorbancia de los estados nativo y desplegado no son siempre igual sino que dependen linealmente de la concentración de agente desnaturante presente en la disolución:

$$A_N = A_{N0} + m_N * D$$

$$A_D = A_{D0} + m_D * D$$

donde  $A_{N0}$  e  $A_{D0}$  son las absorbancias de los estados nativo y desnaturado, respectivamente a concentración 0 M de desnaturante mientras que  $m_N$  y  $m_D$  cuantifican la dependencia lineal de esas absorbancias con la concentración de agente desnaturante.

- c) Cada molécula de proteína sólo puede adoptar dos conformaciones: o bien nativa o bien desplegada relacionadas por la constante de equilibrio de desplegamiento, la cual determina el valor de la diferencia de energía libre estándar:

$$K_{\text{desp}} = X_D/X_N$$

$$\Delta G = -RT \ln K_{\text{desp}}$$

- d) La energía libre estándar depende linealmente de la concentración de agente desnaturante según:

$$\Delta G = \Delta G_w - m * D$$

donde D es la concentración de agente desnaturante, m es una constante de proporcionalidad entre la estabilidad y la concentración de agente desnaturante, y  $\Delta G_w$  es la estabilidad de la proteína en ausencia de desnaturante.

Combinando las ecuaciones anteriores, la dependencia de la absorbancia observada (A) de la concentración de agente desnaturante (D) se puede expresar como:

$$A = \frac{(A_{N0} + m_N D) + (A_{D0} + m_D D)e^{-(\Delta G_w - mD)/RT}}{1 + e^{-(\Delta G_w - mD)/RT}}$$

Se recomienda deducir la ecuación anterior a partir de las asunciones del modelo físico indicado.

Una vez conocida la ecuación, se puede realizar un ajuste no lineal para estimar los valores de los parámetros ajustables  $A_{N0}$ ,  $A_{D0}$ ,  $m_N$ ,  $m_D$  (no interesantes pero necesarios pues son parte del modelo), m (una medida de la cantidad de área expuesta por la proteína al desplegarse, dato de interés para un especialista), y  $\Delta G_w$  (la energía libre de desplegamiento de la proteína en el tampón de trabajo elegido y sin desnaturante: (la magnitud termodinámica que se desea conocer).

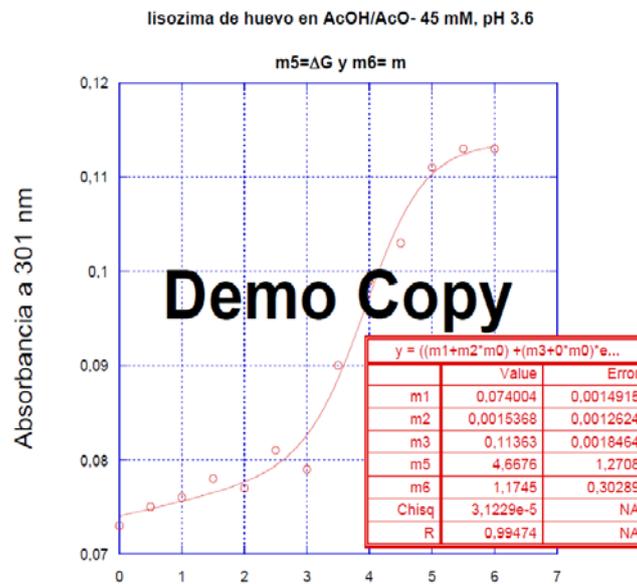
Para poder realizar el ajuste hay que utilizar aplicaciones, raramente gratuitas, que permiten escribir las ecuaciones específicas del modelo que se ha planteado para ajustar los datos experimentales. Utilizaremos la aplicación Kaleidagraph.

- Descargar versión demo de la aplicación KaleidaGraph (<http://www.synergy.com/demowinint.htm>)

- Introducir en MENU CurveFit la ecuación General equivalente a la deducida:

$$((m1+m2*m0)+(m3+m4*m0)*\exp((-m5+m6*m0)/0.56))/(1+\exp((-m5+m6*m0)/0.56));m1=10;m2=0.001;m3=3;m4=0.001;m5=6;m6=3$$

- Ajustar la curva de desnaturalización a esa ecuación para obtener el valor de  $\Delta G$ . Ejemplo:



## Práctica nº 4: Acaba de secuenciar un gen: ¿y ahora qué?

### 1. INTRODUCCION

Tras secuenciar un gen, pueden sentirse deseos irrefrenables de conocer la estructura tridimensional de la proteína correspondiente.

Para lograr semejante propósito, lo habitual es establecer una colaboración con un grupo de cristalografía de proteínas o de resonancia magnética nuclear. En cualquier caso, pueden transcurrir meses, o quizá años, hasta que la estructura quede resuelta.

Mientras tanto, hay distintas herramientas informáticas que nos pueden dar alguna información sobre la **posible** estructura de la proteína.

### 2. RESCRITURA DE SECUENCIAS DE ADN Y DE PROTEINAS

Las secuencias de aminoácidos de una proteína pueden escribirse utilizando código de tres letras (e.g. TyrGlySer, que significa N-tirosina-glicina-serina-C) o de una letra (e.g. YGS, con el mismo significado). Salvo el PDB que sigue consignando las secuencias en código de tres letras, las bases de datos usan código de una letra, que es imprescindible conocer. Existen en internet servidores altruistas que convierten secuencias en código de tres letras a código de una letra, y viceversa. Deberás buscar uno y utilizarlo.

Las secuencias de nucleótidos de un ácido nucleico se escriben en código de una letra. Existen también servidores que traducen de ADN(ARN) a proteína y viceversa. Cuando se traduce de ADN a proteína obtendremos 6 lecturas distintas y sólo una será correcta. Debes entender porqué. Cuando se traduce de proteína a ADN las posibilidades son casi infinitas (¿porqué?) y si el objetivo es diseñar un gen sintético conviene especificar la utilización de una frecuencia de codones característica del organismo en que se desea obtener la proteína.

### 3. HOMOLOGIA DE SECUENCIA

Una herramienta muy útil es la búsqueda de secuencias de proteínas homólogas. Cuando se encuentra una proteína de secuencia altamente similar a la de nuestra proteína problema cuya estructura tridimensional es conocida, la estructura tridimensional de nuestra proteína será parecida a ella.

Hay servidores en Internet que establecen el grado de identidad de una secuencia cualquiera con las secuencias ya registradas en distintas bases de datos. Estos servidores identifican secuencias homólogas y proporcionan información sobre las mismas (código de la proteína, código de su estructura tridimensional si se conoce, etc).

Ejemplos de servidores de este tipo se pueden encontrar en el portal **Expasy**:

[http://www.expasy.ch/proteomics/similarity\\_search\\_alignment](http://www.expasy.ch/proteomics/similarity_search_alignment)

Puede resultar necesario seleccionar una aplicación específica para alinear secuencias de aminoácidos o secuencias de nucleótidos.

### 3. PREDICCIÓN DE ESTRUCTURA TRIDIMENSIONAL

Los más atrevidos pueden hacer predicción completa de la estructura tridimensional en Internet.

#### **Modelado por homología:**

Si existen proteínas con alta identidad de secuencia y estructura conocida, el modelo será razonable (aunque la conformación predicha para las cadenas laterales, especialmente las expuestas al disolvente, es muy difícil que sea la correcta).

(Proteomics en ExPasy) <http://www.expasy.org/proteomics>, Usar SWISS-MODEL

#### **Modelado por threading:**

Si no hay proteínas homólogas de estructura conocida, la probabilidad de que el modelo difiera sustancialmente de la realidad (no solo en los detalles) es más alta.

**TASSER online** (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>)

**Apéndice.** Secuencias propuestas para identificar homólogos que insinúen su función y realizar una predicción de estructura tridimensional cuando sea posible.

(1)

ATAGGCGATGAGCCCCTTGAGAACTACCTGGATACTGAGTACTTCGGCACCATCGGCATTGGAACCCCGCTCAG  
GACTTCACTGTCATCTTTGACACCGGCTCCTCCAACCTGTGGGTGCCCTCTGTCTACTGCTCTAGTCTCGCTGCAG  
CGACCACAACCAGTTCAACCCTGATGACTCCTCCACCTTCGAGGCCACCAGCCAGGAGCTGTCCATCACCTATGGC  
ACCGGTAGCATGACAGGCATCCTTGGATACGACACCGTCCAGGTCGGAGGCATCAGCGACCAACCAGATCTTT  
GGCCTGAGCGAGACAGAGCCCGGCTCCTTCTGTACTACGCCCTTCGACGGCATCCTAGGTCTGGCCTACCCCA  
GCATCTCCGCTCCGGGGCCACCCCTGTCTTTGACAACTGTGGGACCAGGGCCTGGTTTCCAAGACCTTCTCTCC  
GTCTACCTGAGCTCCAATGACGATAGCGGCAGCGTGGTGTGTTGGGCGGCATCGATTCTTCTTACTACACAGGAA  
GCCTGAACTGGGTGCCTGTTTCCGTCGAGGGTACTGGCAGATTACCTTGGACAGCATCACCATGGACGGAGAGA  
CCATCGCTTGCAGCGGTGGCTGCCAGGCCATTGTGGACACGGGCACCTCTGCTGACCGGTCCCACAGTGCAA  
TTGCAATCAACATCCAGAGCGACATCGGAGCCAGCGAGAACTCAGATGGCGAGATGGTGATCAGCTGCTCCTCCA  
TCGACAGCCTGCCTGACATCGTCTTACCATCAACGGTGTCCAGTACCCTCTGAGCCCCAGTGCCTACATCCTGCAG  
GACGATGACAGCTGCACCAGCGGCTTCGAGGGCATGGATGTCCCACTCCTCCGGAGAGCTCTGGATCCTGGGT  
GACGTCTTCATCCGCCAGTACTACACCGTCTTTGACAGGGCCAACAACAAGTTCGGCCTGGCTCCCGTGGC

(2)

MILKILNEIASIPSTKQRQAIFEKNKDELLEKLVYRVYTSAGLQWYIKYKPKGIATQSFAMLTISDMLMFIEFLATRKVT  
GNAAFEELTYISDGGKDEVEVMRDLSEGASYSIANKVWGLIPLQPQMLATSYDERGINKQIKFPAQAQLKDDGAFCE  
VRADELDEVRLDRAGNEFLGLDLLKDELIKMTTEARNIHPEGALIDGELVYDEQVKKKPEGLDFAFDAHPEQSKVKDYT  
EVAESRSASNGIANKTLKGTISEKELQCMKFQVYDYVPLLEVYGLPAFLKYDVRWSKLEQLTSGYDCCKLCVILIDNQVV  
NALDEAKVIYNKYIDQGLEGIILKNIDGLWENARSKNLYKFEVIDVDLKVGIYPHRKDPTKAGGFILESECGKIKVNAGS  
GLKDKAGVKTHELDRTKIMENQNYLKGILECDCNGWYKSDGRDYLKFLPLAIRLRDDTKADTFEDVWGDFFELTGL  
L

(3)

CTGAAAGGAGGGCAATCCTTAATGAATCAAGCGAGTTCGTCGGAGNAAGTGTTCCTATGCACTTGCTCAAGGTT  
TGAAAGTGATTGCTGTGTGGTGGTACTCTGAGGAGCGGGAAGCTGGATCGACCATGGATGTTGTNGCTGCC  
AGACCAAGGCTATTGCCATTCGGGTNACAACTGGTCAAATNTTGTATAGCCTATGAACCANTGTGGGCCATTG  
GAACCGNNNGTTCGNTAGCCCAGNCCAAGCTCAAGAAGTNCACGATGAGCTGAGGAATGGCTNGCCAAGANC  
GTGA

(4)

GLY TRP GLU ILE PRO GLU PRO TYR VAL TRP ASP GLU SER PHE ARG VAL PHE TYR GLU GLN LEU ASP GLU  
GLU HIS LYS LYS ILE PHE LYS GLY ILE PHE ASP CYS ILE ARG ASP ASN SER ALA PRO ASN LEU ALA THR LEU  
VAL LYS VAL THR THR ASN HIS PHE THR HIS GLU GLU ALA MET MET ASP ALA ALA LYS TYR SER GLU VAL  
VAL PRO HIS LYS LYS MET HIS LYS ASP PHE LEU GLU LYS ILE GLY GLY LEU SER ALA PRO VAL ASP ALA LYS  
ASN VAL ASP TYR CYS LYS GLU TRP LEU VAL ASN HIS ILE LYS GLY THR ASP PHE LYS TYR LYS GLY LYS LEU