ESCUELA DE PREGRADO – FACULTAD DE CIENCIAS – UNIVERSIDAD DE CHILE

"UN ROL PARA LA DINÁMICA MOLECULAR EN EL MODELADO DE MACROMOLECULAS BIOLÓGICAS"

Seminario de Título entregado a la Universidad de Chile en cumplimiento parcial de los requisitos para optar al Título de Ingeniero en Biotecnología Molecular.

JOSÉ ANTONIO GÁRATE CHATEAU

Dr. Ricardo Cabrera Paucar
Director Seminario de Título

Dr. Tomás Pérez-Acle
Co-Director Seminario de Título

Comisión de Evaluación Seminario de Título

Dr. Osvaldo Alvarez Araya
Presidente Comisión

Dr. José Jaime Arbildua Sepúlveda
Corrector

Santiago de Chile, Noviembre de 2007

José Antonio Gárate Chateau

Nacido el 29 de Julio de 1983 en Santiago de Chile, en 1989 ingresó al Colegio Calasanz donde a lo largo de trece años completó su educación básica y media. En el año 2002 optó por estudiar a la carrera de Ingeniería en Biotecnología Molecular de la Facultad de Ciencias de la Universidad de Chile. En Septiembre del 2005 ingresó al Centro de Genómica y Bioinformática (actual Centro de Bioinformática, CBUC) de la Facultad de Ciencias Biológicas de la Pontificia Universidad Católica de Chile, dónde hasta la fecha se ha desempeñado activamente en las diversas actividades de investigación del centro, incluyendo el desarrollo del presente seminario de título. En noviembre del año 2006 egresó de su carrera obteniendo el grado académico de Licenciado en Ingeniería en Biotecnología Molecular. En noviembre del presente año José Antonio fue admitido para realizar sus estudios doctorales en el aérea de la Nanomedicina en el School of Chemical and Bioprocess Engineering of the University College Dublin (UCD), Irlanda.

Entre sus intereses académicos destacan la biofísica dando énfasis en el área de la simulación molecular siendo de su particular interés la técnica de la Dinámica Molecular. donde además de su seminario de título, ha desarrollado diversos trabajos que incluyen el modelamiento y simulación de proteínas de transmembrana, mutantes de enzimas y ensambles de surfactantes no-iónicos en solvente orgánico.

José Antonio se define como un colocolino empedernido, amante de la buena cerveza, la ciencia ficción, el rock y la informática en especial todo lo que tenga relación con los proyectos open source.

FACULTAD DE CIENCIAS
UNIVERSIDAD DE CHILE

"Un rol para la Dinámica Molecular en el modelado de macromoléculas biológicas"

Seminario de título
Entregado a la
Universidad de Chile
en cumplimiento parcial de los requisitos
para optar al título de:
Ingeniero en Biotecnología Molecular

por
José Antonio Gárate Chateau
Noviembre 2007
Santiago-Chile

Director Seminario de Título: Dr Ricardo Cabrera, Ph.D
Co-director Seminario de Título: Dr Tomás Pérez-acle, Ph.D

Agradecimientos

En primer lugar me gustaría darle las gracias a mis padres por darme el apoyo, herramientas y oportunidad de estudiar, lejos es el mejor regalo que me han podido dar. También es oportuno agradecer a mi hermanos y sobrinos, gracias por el apoyo en particular a Pilar, gracias por el ejemplo y constante interés en mi quehaceres, bueno y también por el financiamiento.

Por otro lado debo mencionar a mis tutores, Tomás y Ricardo, gracias por su guía y constante preocupación por mi trabajo esta memoria es en parte de ustedes también. Debo destacar a los "cabros" del CGB (me gusta más el nombre antiguo), gracias por los consejos, enseñanzas y los buenos carretes.

Bueno esta no pudo haberse logrado si antes haber estudiado en la gloriosa Facultad de Ciencias, por lo que en general quiero agradecerle a todos mi compañeros de generación dando énfasis al movimiento Charista y sus integrantes en particular a Richard Kawada. Bueno esta es para la jevita, gracias Susi por todo, te pasaste :).

Finalmente, al lector de esta memoria, gracias por el interés en mi primeros paso en el exigente pero hermoso camino de la ciencia.

ÍNDICE

I Índice de tablas

II Índice de figuras

III Lista de abreviaturas

Å: angstrom, 10^{-10} m

Apertura: gaps

ANOLEA: Evaluación del ambiente atómico no local (Atomic Non-local Environment Assessment)

BD: Base de datos.

BLOSUM: Matriz de substitución de bloques (Blocks Substitution Matrix)

BLAST: Herramienta básica de búsqueda de alineamientos locales (Basic Local Alignment Search Tool)

Cα: Carbonos alfa.

CATH: Clase, arquitectura, topología homología (Class, Architecture, Topology, Homologous superfamiliy)

CHARMM: Química en Harvard mecánica molecular (Chemistry at Harvard Molecular Mechanics)

dE/dt: Derivada de la energía total con respecto al tiempo.

DM: Dinámica molecular.

ECR. : estructura cristalográfica de referencia.

G: Energía libre de Gibbs.

fs: femtosegundos, 10^{-15} s

HA1: Acuaporina humana 1.

Indel: apertura aleatoria, gap.

M: moles/litro.

MVP: Modelado validación y predicción.

NAMD: Dinámica molecular en nanoescala (Nanoscale Molecular Dynamics)

NVT: Número de átomos volumen y temperatura constante.

RMN: Resonancia magnética nuclear.

ns: nanosegundos 10^{-9} s

PBC: Condiciones periódicas de borde (Periodic Boundary Conditions)

Perl: Lenguaje Práctico para la extracción y reporte (Practical extraction and report language)

PDB: Base de Datos de Proteínas (Protein Data Bank)

pdf: Función de Densidad de Probabilidad (Probability Density Function).

PME: Método del acoplamiento de partícula de Ewald (Particle Mesh Ewald)

ps: picosegundos, 10^{-12} s

RMSD: Desviación cuadrada de las coordenadas medias (Root Mean Square Deviation)

RECR: RMSD de carbones alfa con respecto a la estructura cristalográfica de referencia.

SCOP: Clasificación estructural de proteínas (Structural Classification of Proteins).

TCL: Lenguaje de herramientas de comandos (Tool Command Language)

VMD: Dinámica molecular visual (Visual Molecular Dynamics)

IV Abstract

Bioinformatics, syncretism between biology and the technologies of information takes more and more relevance in the biological sciences because of the big increment in the technological capacity to obtain empirical data. One of the topics that bioinformatics studies is the generation of three-dimensional molecular models of proteins that no experimental structure is yet available. In this field one of the most popular methods for the accomplishment of this task is comparative modeling by satisfaction of spatial restrains, implemented in the software MODELLER. The three-dimensional structures derived form this process must be evaluated, thus, there are several tools to accomplish this objective, but they work in an adequate manner in the range of high sequence identity, failing consistently for models obtained in the range of low sequence identity. Previous works of our laboratory in comparative modeling in the range of low sequence identity have required some tool which allows the evaluation of the structural stability of the generated models. Molecular Dynamics, mechanical-statistics technique for the characterization of molecular systems at atomic level, has been systematically used with this purpose. The present seminar has as objective the demonstration of the utility of molecular dynamics as an evaluation tool in the modeling of biological macromolecules, in the range of low sequence identity. For that, a data base of crystallographic structures from the protein data base was generated. For each member of the data base, a templates search using BLAST-P was performed, selecting the templates whose identity was located in the range of 30% and 50%. With these templates, a conjunto of comparatives models was developed using non intervened and intervened multiple sequence alignments to generate errors in the alignments. To evaluate the resulting models structures,

they were evaluated with different tools, and relaxed using a molecular dynamics protocol in explicit solvent and spherical boundary conditions. The models generated with intervened sequence alignments presented higher RMSD (Cα (> 3Å) (instable dynamics) with respect to the reference crystallographic structures, demonstrating the power of the molecular dynamics evaluation in the discrimination between comparatives models of different quality, in the range of low sequence identity.

IV Resumen

La bioinformática, sincretismo científico-técnico entre la biología y las tecnologías de la información, cobra cada vez más relevancia dentro las ciencias biológicas debido al gran aumento en la capacidad tecnológica para la obtención de datos empíricos. Dentro de los tópicos que estudia la bioinformática se encuentra la generación de modelos tridimensionales para la estructura de proteínas que no han sido determinadas experimentalmente. En este ámbito, uno de los métodos más populares para la realización de esta tarea dice relación con el modelado comparativo por satisfacción de restricciones espaciales, implementado en el programa MODELLER. Las estructuras tridimensionales derivadas de este proceso deben ser evaluadas, objetivo para el cual existen diversas herramientas que funcionan de manera adecuada cuando se trabaja en el ámbito de la alta identidad de secuencia, fallando consistentemente para modelos obtenidos a baja identidad de secuencia. Trabajos anteriores de nuestro laboratorio en modelado comparativo en el ámbito de la baja identidad de secuencia, han necesitado de alguna herramienta que permita evaluar la estabilidad estructural de los modelos generados. La dinámica molecular, técnica mecánico-estadística para la caracterización de sistemas moleculares a nivel atómico, ha sido sistemáticamente utilizada con este propósito. La presente memoria tiene como objetivo demostrar la utilidad de la dinámica molecular en el modelado de macromoléculas biológicas, en el ámbito de la baja identidad de secuencia. Para esto se generó una base de datos de estructuras cristalográficas provenientes de la Base de Datos de Proteínas. Para toda la base de datos se realizó búsquedas de patrones usando BLAST-p, siendo seleccionados aquellos cuya identidad de secuencia se ubicase entre 30% y 50%. Usando estos patrones, se generó una

serie de modelos comparativos a partir de alineamientos múltiples no intervenidos, e intervenidos para generar errores en el alineamiento a partir de aperturas aleatorias en los alineamientos (indel o gaps). Los modelos resultantes fueron evaluados por diversas herramientas, siendo sometidos a un protocolo de dinámica molecular en una esfera de solvente, con un número suficiente de moléculas de agua para solvatar el sistema. Los modelos generados a partir de alineamiento intervenidos presentaron altos valores RMSD de Cα (> 3Å) (dinámicas inestables) en relación a las estructuras cristalográficas de referencia, demostrando el poder de evaluación de la dinámica molecular para la discriminación entre modelos moleculares comparativos de distinta calidad, en el ámbito de la baja identidad de secuencia.

V Introducción

•*Bioinformática*

La bioinformática puede ser definida como la aplicación a la biología de las técnicas de la informática, siendo su rol fundamental el manejo, análisis e integración de información biológica [Baldi y col, 2001]. Su origen se remonta a la década de los sesenta con los trabajos desarrollados por la Dra. Margaret Oakley Dayhoff (1925-1983) pionera en el uso de computadores en biología y química. Uno de sus mayores logros fue la realización de la primera base de datos de estructuras proteicas conocida como el atlas de secuencia y estructuras de proteínas [Dayhoff y col, 1965]. La bioinformática cobra cada vez más relevancia dentro de las ciencias biológicas debido al gran aumento en la capacidad tecnológica para la obtención de datos empíricos siendo gatillada por la explosión de la información genómica disponible en bases de datos públicas, resultado de la concreción exitosa del Proyecto Genoma Humano [Venter y col, 2001]. Tal como ocurre en otros ámbitos científicos el objetivo principal de la bioinformática es el modelado, validación y predicción (MVP) de los fenómenos biológicos. Cuando las predicciones son validadas contra los datos empíricos se cierra el círculo virtuoso del conocimiento tan propio del desarrollo de las ciencias aplicadas.

•*Modelado Molecular*

Dentro de los tópicos que estudia la bioinformática, se encuentra la generación de modelos tridimensionales (3D) de macromoléculas biológicas, en particular de proteínas. La estructura de una proteína está definida por la posición en el espacio 3D de todos los átomos que la componen, se puede determinar mediante técnicas empíricas como resonancia magnética nuclear (RMN)[Wüthrich, 2001], cristalografía de rayos x [Drenth, 1999] o modelar a través de las aproximaciones teóricas. Se ha propuesto que la estructura tridimensional de una proteína está determinada por dos distintos principios que operan en escalas de tiempos muy diferentes: las leyes de la física y la teoría de la evolución. Por un lado, de acuerdo a las leyes de la física, una proteína en solución es un sistema de átomos que interactúan a través de una serie de fuerzas tales como interacciones coulómbicas, Van der Walls, enlaces covalentes y de hidrógeno. Estas fuerzas (bajo condiciones apropiadas) definen la estructura proteica, proceso que se alcanza en el intervalo de los milisegundos o segundos luego del fenómeno de traducción. Por otro lado, en millones de años la deriva estructural de proteínas ha resultado en familias de variantes que comparten una función, lo cual requiere la conservación de su estructura y por ende de partes de su secuencia [Fiser y col, 2002]. A partir de estos dos principios, las aproximaciones teóricas se pueden dividir en métodos *de novo* y métodos que se basan en conformaciones conocidas (figura 1). Los métodos *de novo* predicen la estructura 3D solo a partir de la secuencia de aminoácidos, suponen que la conformación del estado nativo de una proteína esta en el mínimo local accesible de energía libre de Gibbs (G) y llevan a cabo una búsqueda a gran escala (utilizando las leyes de la física) del espacio conformacional de las estructuras terciarias que poseen valores particularmente bajos de ΔG [Baker & Sali, 2001]. Los métodos basados en estructuras conocidas, entre lo cuales destacan el *threading* o reconocimiento de plegamiento [Leonard y col, 2004] y el modelado comparativo, dependen de estructuras

3D proteicas que se han determinado empíricamente (datos que son depositados en bases de datos públicas como PDB [Sussman y col 1998]) dejando el proceso de modelado a manos de las restricciones espaciales detectadas en al menos una estructura relacionada [Fiser y col, 2002].

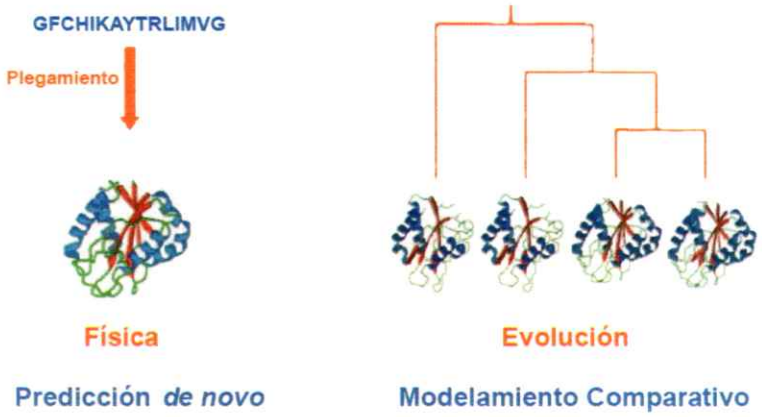


Figura 1: Predicción de estructura *de novo* y modelado comparativo de estructuras de proteínas. Las proteínas obedecen dos principios distintos, las leyes de la física y la teoría de la evolución, cada una de las cuales dan origen a una variedad de los métodos de predicción correspondientes, figura modificada de Fiser y col, 2001.

Es importante entender que cualquier método de predicción de estructura 3D de proteínas puede ser visto como una optimización, con respecto a cierta función objetivo. Desde este punto de vista los métodos *de novo* tratan de encontrar la conformación de una secuencia proteica dada las fuerzas entre los átomos, mientras que los métodos basados en una conformación conocida buscan la estructura 3D de una secuencia de aminoácidos primordialmente mediante su relación con una o varias proteínas similares en secuencia y de estructura 3D conocida [Fiser y col, 2002].

Dada la existencia de distintos métodos para la obtención de la estructura de una proteína, trabajos emblemáticos [Baker & Sali, 2001] se han encargado de determinar la valía y el rango de aplicación de los distintos métodos.

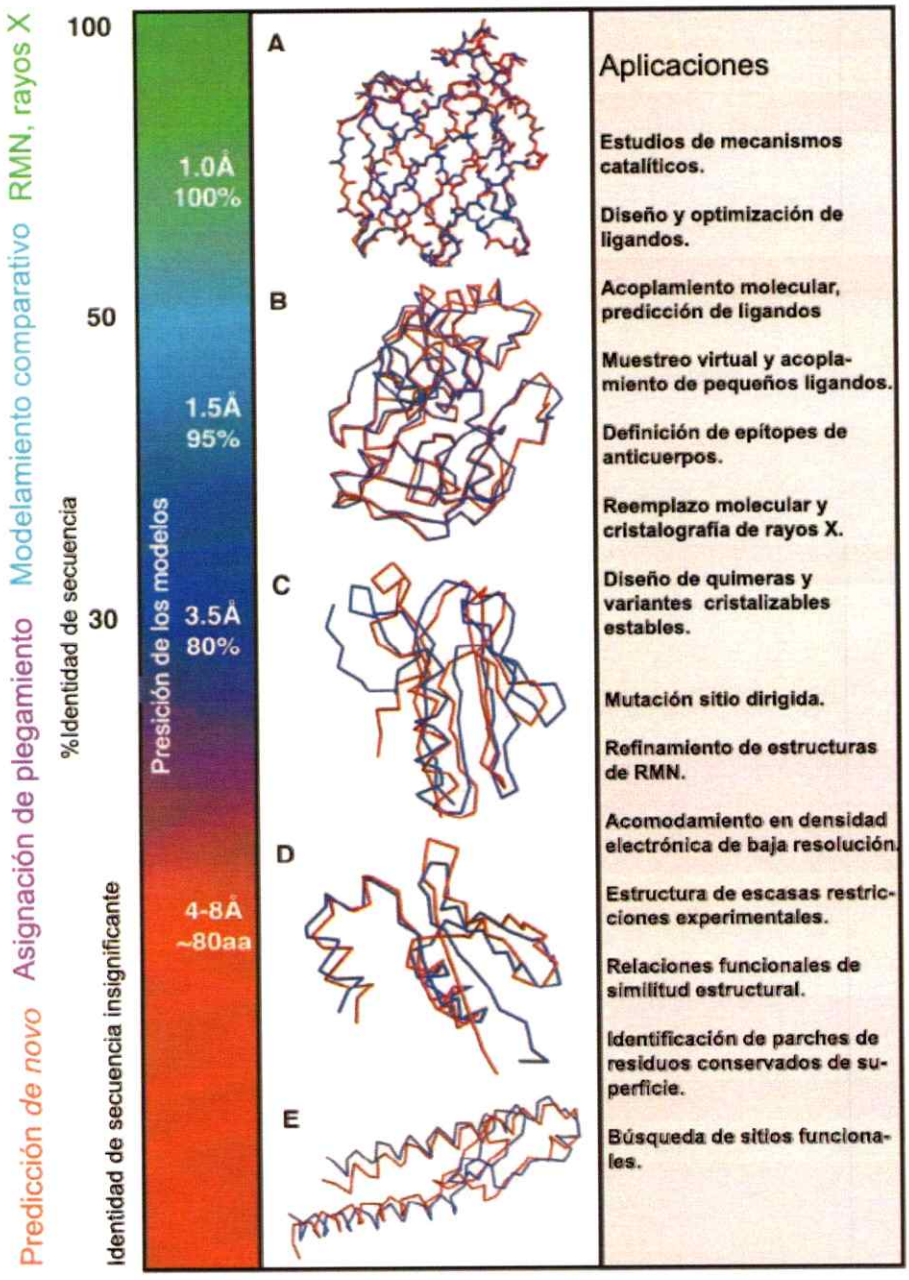


Figura 2: Precisión y aplicación de modelos de estructuras de proteínas. Se muestran los diferentes rangos de aplicabilidad del modelado comparativo, *threading* o asignación de plegamiento y los métodos de predicción estructural *de novo*; Se aprecian gradualmente las exactitudes determinadas para cada tipo de modelo, así como ejemplos de sus aplicaciones. Paneles(A a la C). Ejemplos de modelos comparativos basados en 60% (A), 40% (B) y 30% (C) de identidad de secuencia con sus estructuras patrones. Paneles (D y E). Ejemplos de predicciones estructurales *de novo*. Estructuras predichas están en rojo y las estructuras reales en azul. La precisión de los modelos decrece significativamente de (A) a (E), pero la estructura general es correcta. Figura modificada de [Baker & Sali, 2001].

Como se extrae de la figura 2 las distintas metodologías de modelado molecular presentan distintos niveles de precisión y aplicabilidad, desde la identificación de motivos estructurales hasta el estudio de mecanismos catalíticos. En general el modelado basado en estructuras 3D conocidas consiste de cinco pasos (ver figura 3):

i) Identificar estructuras 3D depositadas en las bases de datos, relacionadas con la secuencia objetivo a modelar.

ii) Seleccionar los patrones estructurales.

iii) Alinear la secuencias primarias con sus patrones.

iv) Construcción de modelo.

v) Evaluación del modelo.

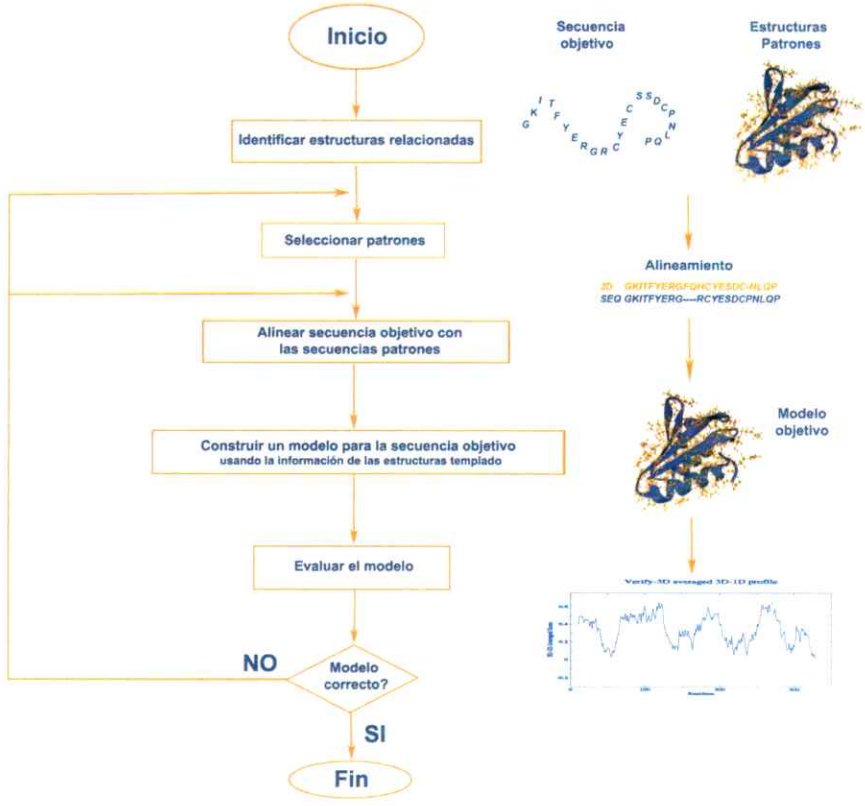


Figura 3: Protocolo de modelado molecular basado en estructuras conocidas, en general tanto para la asignación de plegamiento como para el modelado comparativo se procede de manera similar. Figura modificada de Fiser & Sali, 2003.

Normalmente, tanto para el reconocimiento de plegamiento como para el modelado comparativo los tres primeros pasos son iguales (ver figura 3), siendo el cuarto paso el que define la particularidad del método, elección que va a depender directamente de la identidad de secuencia entre el o los patrones y la proteína a modelar (ver figura 2).El modelado comparativo usa estructuras 3D de proteínas determinadas empíricamente para predecir la conformación de otras proteínas con secuencia similar de aminoácidos. Chotia & Lesk [1986] demostraron, a través del estudio de la relación entre la divergencia de estructura 3D y secuencia primaria en proteínas, que el espacio de secuencias distintas es mayor al espacio de conformaciones existentes (determinadas experimentalmente) , es decir que la estructura 3D es más conservada que la secuencia primaria (ver figura 4). Sin embargo, la diferencia estructural entre la proteína problema (objetivo) y el homólogo con estructura conocida aumenta con la disminución del porcentaje de identidad de secuencia entre ambos para la correcta asignación en el modelo, de los contactos observados en la estructura nativa, una vez que esta ha sido determinada experimentalmente. Este problema se suma a la menor eficacia de los algoritmos de alineamiento para determinar correctamente las posiciones homólogas entre dos secuencias, cuando hay baja identidad de secuencia En resumen, si la similitud entre dos proteínas es detectable a nivel de secuencia y es estadísticamente significativa, probablemente tendrán una estructura similar [Fiser & Sali, 2003, Sali & Blundell, 1993].

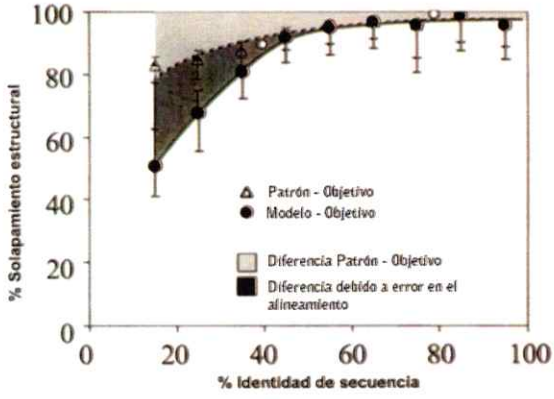


Figura 4: Precisión de modelos comparativos (traslape de Cα) en función de la identidad de secuencia,como se aprecia bajo el umbral del 40% de identidad de secuencia los errores en los alineamientos, las diferencias estructurales entre modelos, proteínas objetivo y patrones tienden a ser preponderantes. Figura modificada de Fiser & Sali , 2003.

Existen variadas metodologías en el modelado comparativo, de las cuales destaca la que utiliza MODELLER [Sali & Blundell, 1993], programa que modela estructuras 3D por satisfacción de restricciones espaciales (figura 5).

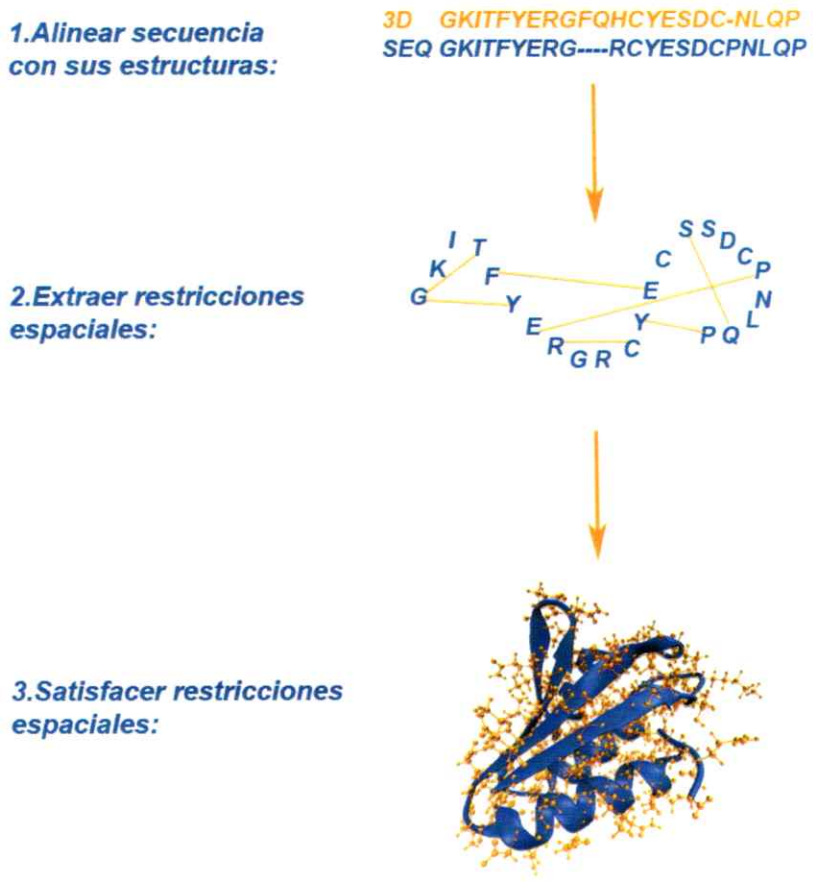


Figura 5: Construcción de un modelo comparativo por el programa Modeller. En primer lugar, restricciones espaciales son extraídas de la estructura patrón, a partir del alineamiento se determinan las posiciones de residuos equivalentes. Las restricciones espaciales son combinadas en una función objetivo, la cual es optimizada hasta que el modelo satisfaga de mejor manera todas estas restricciones.

Estas restricciones espaciales calculadas son expresadas como funciones de densidad de probabilidad (PDF, ver material suplementario 2) para cada característica espacial dado un número de variables que fueron encontradas para la característica estudiada (ej: distancia Cα-Cα), de acuerdo a la figura 6.

$$p(X_1 \leq X < X_2) = \int_{X_1}^{X_2} p(x)\,dx$$

Figura 6: función de densidad de probabilidad(PDF) $p(x)$ para la característica x que es restringida (puede ser de cualquier forma no negativa) y que integra para 1 sobre el rango de todos los valores de x. La probabilidad finita de que un evento $x_1 \leq x < x_2$ ocurra es obtenida por la integración de p.

Existen tres tipos de restricciones espaciales:

• Las que provienen directamente del alineamiento con el o los patrones.

• Derivadas de preferencias estadísticas extraídas de estructuras proteicas de alta resolución y que son expresadas como potenciales estadísticos de fuerza media [Sippl, 1993].

• Las que se obtienen del campo de fuerza de mecánica molecular CHARMM22 [MacKerel y col 1998] (ver material complementario 1) .

El modelo 3D se genera a partir de la optimización (algoritmo de gradiente conjugado) de la PDF molecular (figura 7) que representa la suma ponderada de cada una de las PDF´s obtenidas para cada restricción [Sali & Blundell, 1993].

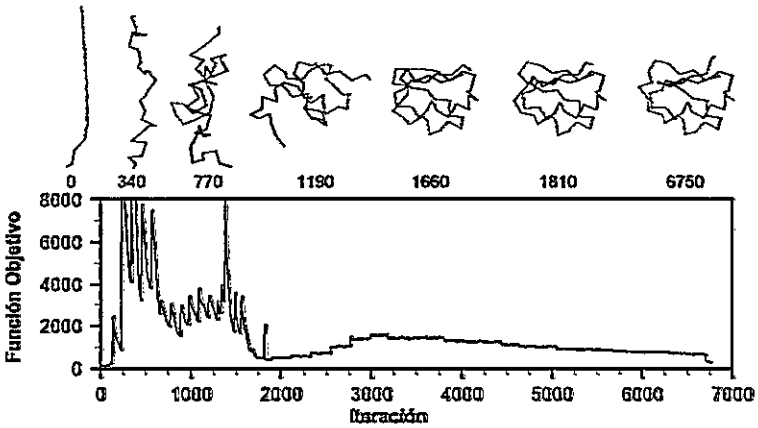


Figura 7: Optimización de la PDF molecular (función objetivo) por el método de gradiente conjugado. Como se observa a medida que aumenta el número de iteraciones el valor de la función objetivo disminuye llegando a un valor mínimo (valor óptimo) que define la estructura de la secuencia a modelar, figura modificada de Fiser & Sali , 2003.

Las estructuras 3D derivadas de un modelado comparativo deben ser evaluadas. Para esto, existen variadas herramientas útiles, destacando :

- ProsaII [Sippl, 1993] y Anolea, [Melo y col, 1997] que se basan en el uso de potenciales estadísticos de fuerza media (mean fields) para analizar las distribuciones de energía en el plegamiento de proteínas.
- Verify 3D que evalúa la compatibilidad de cada residuo con la estructura 3D local adyacente generando puntajes que se comparan con valores promedio (3D-ID scores) [Lüthy y col, 1992].
- Procheck, que evalúa la distribución de la geometría de los residuos en un modelo dado en comparación con distribuciones provenientes de bases de datos [Laskowski y col, 1993].

En forma adicional a las metodologías nombradas, el cálculo del valor de desviación de las coordenadas cartesianas, RMSD (ver material suplementario 2), permite obtener una medida de la diferencia global entre dos estructuras 3D superpuestas. Los métodos para evaluar modelos 3D, en particular modelos comparativos, funcionan de manera adecuada cuando se trabaja en el ámbito de la alta identidad de secuencia (≥40%) [Baker & Sali, 2001], dado que estas herramientas están parametrizadas y validadas frente a modelos generados con patrones en este intervalo de similitud. Diversos estudios han demostrado que la distribución de frecuencias del porcentaje de identidad de secuencia para proteínas secuenciadas y las estructuras 3D determinadas empíricamente (depositadas en el PDB) tiene la forma de valores extremos, en donde la mayoría de la secuencias comparten una identidad de secuencia en el rango del 20 % al 30% [Sanchez y col, 2000] (ver figura 8), siendo éste el límite inferior para el cual , normalmente el modelado comparativo es posible [Fiser & Sali, 2003, Melo y col, 2002].

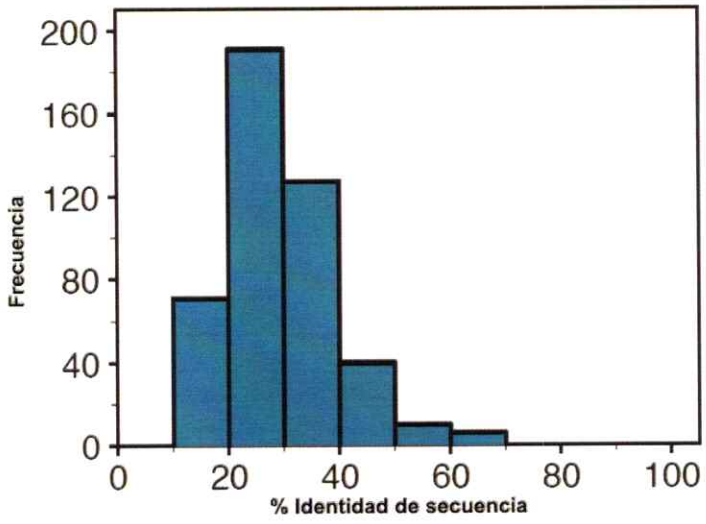


Figura 8: Ejemplo de la distribución del porcentaje de identidad de secuencia entre todas las proteínas secuenciadas para *Mycobacterium genitalium* y las estructuras conocidas, como se aprecia la mayoría de las secuencias comparten una identidad de secuencia entre 20% y 30%, limite inferior aceptable para el modelado comparativo y rango en que la herramientas de evaluación actuales no son lo suficientemente confiables, figura extraída de Sanchez y col 2000.

Trabajos anteriores de nuestro laboratorio demuestran que el modelado comparativo en el rango de la baja identidad de secuencia es posible, bajo condiciones particulares que requieren de un alto grado de conocimiento biológico [Larrondo y col ,2005, Tischler y col ,2005]. En este ámbito límite, las herramientas de evaluación comúnmente usadas no permiten evaluar la estabilidad estructural de los modelos, en este seminario propongo usar Dinámica Molecular con este propósito.

•Dinámica Molecular

La visión antigua de las proteínas como estructuras relativamente rígidas ha sido reemplazada por un modelo dinámico en dónde los movimientos internos y los cambios conformacionales resultantes juegan un rol esencial en su función[Karplus & McCammon, 2002]. La Dinámica Molecular (DM), solución aproximada al clásico problema de los n-cuerpos [Sundman, 1912], se define como una técnica mecánico-estadística que implica la integración, dependiente del tiempo, de las ecuaciones clásicas de movimiento en sistemas moleculares [Beck & Dagget, 2004](ver material complentario 1). Los usos habituales de la DM abarcan: el muestreo del espacio conformacional para determinar y refinar estructuras con datos obtenidos empíricamente; obtención de una descripción del sistema al equilibrio incluyendo propiedades estructurales y motrices; análisis del desarrollo de la dinámica efectiva donde los movimientos y su evolución en el tiempo son el interés principal. [Norberg & Nielson, 2003].

Variadas lineas de investigación han utilizado simulaciones de DM en el marco del modelado molecular y el análisis de la estabilidad estructural de proteínas [Flohil y col, 2002, Fan & Mark, 2004]; se estudió la estabilidad relativa de estructuras de proteínas determinadas por cristalografía de rayos x y RMN mediante simulaciones de DM [Fan & Mark, 2003]; la calidad de estructuras de proteínas de membrana fue evaluada mediante simulaciones de DM [Law y col, 2005] ;Richard J. Law y Mark S. P. Sansom (2004) en el trabajo "Modelado Comparativo y simulaciones de Dinámica Molecular: estudios comparativos de la acuaporina humana 1(HA1)", se evaluó la calidad de estructuras modeladas derivadas de patrones cuya estructura fue obtenida con distintas

técnicas experimentales , de acuerdo a la estabilidad estructural que mostraban en simulaciones de DM [Law & Sansom, 2004].

A pesar de la aplicabilidad de la DM en la evaluación de la calidad de modelos moleculares de proteínas, no se ha desarrollado estudios sistemáticos enfocados en la capacidad de la DM para discriminar entre estructuras 3D de proteínas pequeñas modeladas a partir de alineamientos de distinta calidad. En este sentido las herramientas de evaluación convencionales están sesgadas hacía la comparación de rasgos estructurales específicos del modelo con respecto a la representatividad de dicha configuración en las bases de datos de proteínas (potenciales estadísticos). En cambio, la DM puede evidenciar en que medida un modelo presenta una baja probabilidad de asignación correcta de contactos nativos, dado que esto afectará la estabilidad estructural durante una simulación de DM [Ponder & Case, 2003].

V.2 Hipótesis de trabajo.

"La Dinámica Molecular es una herramienta útil para evaluar la calidad de estructuras 3D obtenidas por modelado comparativo"

V.3 Objetivos

•General

Demostrar la capacidad diferenciadora de la estabilidad estructural de modelos comparativos desarrollados a partir de alineamientos de alta y baja calidad ofrecida por la dinámica molecular.

•Específicos

1) Generar una base de datos de estructuras de proteínas cuyas secuencias sean utilizadas para modelado comparativo y su respectivo conjunto de estructuras patrones.

2) Usando MODELLER desarrollar modelos comparativos a partir de :

2.1) Alineamientos usando ClustalW.

2.2) Alineamientos intervenidos: introducción *ex profeso* de aperturas aleatorias (indel, gaps).

3) Comparar modelos generados con su estructura cristalográfica de referencia utilizando funciones clásicas de evaluación.

4) Aplicar DM a los modelos generados y comparar la estabilidad estructural durante las trayectorias respecto a su estructura cristalográfica de referencia.

VI Materiales y métodos

Se desarrolló una base de datos (BD) de estructuras proteicas provenientes de PDB. Para cada estructura se realizó una búsqueda de patrones, filtrando los resultados de acuerdo a un criterio mínimo establecido. A continuación se hizo alineamientos globales múltiples para cada estructura con sus patrones. Para cada alineamiento se generaron diferentes conjuntos de alineamientos intervenidos. Usando estos conjuntos de alineamientos se generaron modelos comparativos para luego ser comparados contra las estructuras cristalográficas de referencia y evaluados con ProsaII y Anolea. Finalmente determinados modelos fueron templados con herramientas de dinámica molecular, para luego ser sometidos al mismo proceso de evaluación, con la idea de poder determinar si la DM tiene la capacidad de distinguir entre ambos conjuntos de modelos. En la figura 9 se presenta un diagrama de la metodología desarrollada.

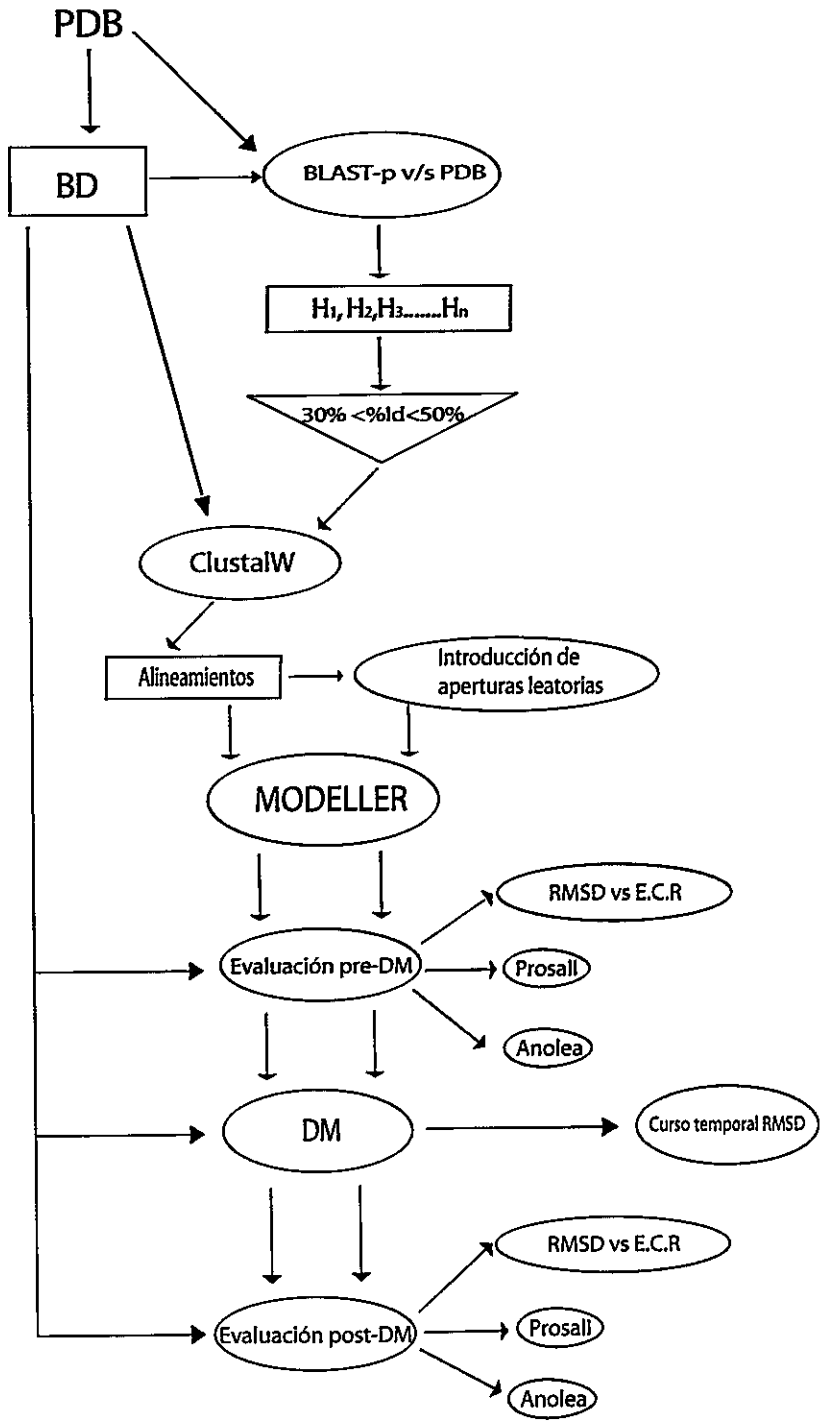


Figura 9: Diagrama de flujo de la metodología empleada.H_1,H_2,H_3 ....H_n , hits de BLAST-P [Altschul y col, 1997]. E.C.R, estructura cristalográfica de referencia.

VI.1 Base de datos y conjunto de prueba

La segmentación de la BD fue de acuerdo a resolución (< 2Å), número de residuos (< 100), estructuras completas y estado oligomérico (monómeros). Las estructuras obtenidas a partir de esta búsqueda fueron utilizadas como entrada para una búsqueda BLAST-p [Altschul y col, 1997] (usando BLOSUM62, [Henikoff & Henikoff, 1992] expectancy :10, word size: 3) contra la base de datos PDB seleccionado aquellos elementos de la BD que obtuviesen al menos un acierto con porcentaje de identidad de secuencia entre 30% y 50%.

VI.2 Modelado comparativo

•*Modelos generados a partir de alineamientos no intervenidos:*

Para cada una de las cinco proteína de la BD se generó modelos de su estructura 3D a partir de sus correspondientes estructuras primarias y los patrones de conformación conocida, utilizando el método de modelado comparativo por satisfacción de restricciones espaciales. Los alineamientos entre patrones y secuencias a modelar fueron realizados con ClustalW [Thompson y col, 1994], para el modelado se uso el programa MODELLER v8.2 con la clase automodel. Para cada proteína de la BD se generó cinco modelos, uno sin refinamiento de lazos (loops) y 4 con refinamientos de lazos .

•*Modelos generados a partir de alineamientos intervenidos con n- iteraciones de error:*

En el modelado comparativo existen varias fuentes de error: selección inapropiada de patrones; alineamientos incorrectos; modelado erróneo de cadenas laterales; modelaje

inadecuado de lazos [Fiser & Sali, 2003]. Dentro de lo errores más comunes en el ámbito de baja identidad de secuencia son los del alineamiento y por lo tanto para la generación de modelos comparativos intervenidos, fue al nivel de los alineamientos en donde se introdujo tales errores. Específicamente, se generó, vía programación en Perl [Wall, 1997], un conjunto de alineamientos intervenidos insertando aperturas aleatorias (ver material suplementario 3 y figura 29). Descartando lo anterior, el protocolo de modelado comparativo fue igual al utilizado en el conjunto de modelos no intervenidos.

VI.3 Simulaciones de dinámica molecular

Las simulaciones fueron desarrolladas con minimizaciones energéticas iniciales (algoritmo de gradiente conjugado),usando agua como solvente explícito, con condiciones esféricas de borde (potencial armónico aplicado sobre moléculas de solvente en los bordes del sistema), grupo termodinámico NVT (número de átomos, volumen y temperatura constante), el número de átomos (que en ningún caso sobrepasó los 30000 átomos) y volumen dependen directamente del sistema estudiado. La temperatura del grupo fue de 310 K y la concentración de iones igual a 0,05 M. El tiempo total de simulación fue de 1ns, el integrador fue de 2fs (timestep), se aplicó el protocolo Shake [Ryckaert y col, 1977] como restricción a los hidrógenos. El campo de fuerza utilizado fue CHARMM27 implementado en el programa de dinámica molecular en nanoescala NAMD (dinámica molecular en nanoescala, *nanoescale molecular dynamics*), [Phillips y col 2005]. Es importante destacar que para las simulaciones exploratorias hubo ciertas variaciones en el protocolo utilizado, como fueron la utilización de condiciones periódicas de borde (cúbicas) y protocolos de relajación y minimización diferentes. Todas las simulaciones fueron ejecutadas en un linux-cluster (Fedora core2) de ocho

procesadores Intel Pentium IV, demorando aproximadamente 24 hrs por cada ns simulado. En la figura 10 se presenta un ejemplo de los sistemas simulados.

Figura 10: Ejemplo de sistema usado para la simulación de dinámica molecular para la estructura cristalográfica 1PCH; condiciones esféricas de borde, solvente (agua) explícito, grupo termodinámico NVT, campo de fuerza CHARMM27, motor de cálculo NAMD, número de átomos 14252. Esfera gris representa las moléculas de agua; Representación de la estructura secundaria de 1PCH: amarillos hojas β, violeta hélices α, calipso y blanco lazos (loops).

VI.4 Evaluación de modelos comparativos y análisis de simulaciones

Los modelos comparativos fueron evaluados antes y después de las simulaciones de dinámica molecular con los evaluadores clásicos ProsaII y Anolea utilizando las opciones por omisión de ambos programas. Los valores del curso temporal de la RMSD (con respecto a la estructura inicial), energías de las simulaciones, divergencia estructural y alineamientos estructurales (STAMP [Russell & Barton, 1992], además de los RECR. antes y después de las DM, fueron calculados con el programa VMD (dinámica molecular visual, *visual molecular dynamics*)[Humphrey y col, 1996] a través de programación en TCL [http://www.tcl.tk/] y Perl. El análisis matemático fue llevado cabo con los programas Grace [http://plasma-gate.weizmann.ac.il/Grace/], Octave [http://www.gnu.org/software/octave/] y Openoffice [http://www.openoffice.org/].

VII Resultados

VII.1 Base de datos y conjunto de prueba

Los resultados de la búsqueda de secuencias patrones BLAST-p se presentan en la figura 11. El análisis de la figura 11 permite destacar dos aspectos fundamentales, la mayor cantidad de los aciertos para los elementos de la base de datos preliminar se encuentran localizados en la región comprendida entre 25% y 50% de identidad de secuencia BLAST-p. PDB es redundante, lo que se ve reflejado por la presencia de un gran número de aciertos cercanos al 100% de identidad de secuencia BLAST-p.

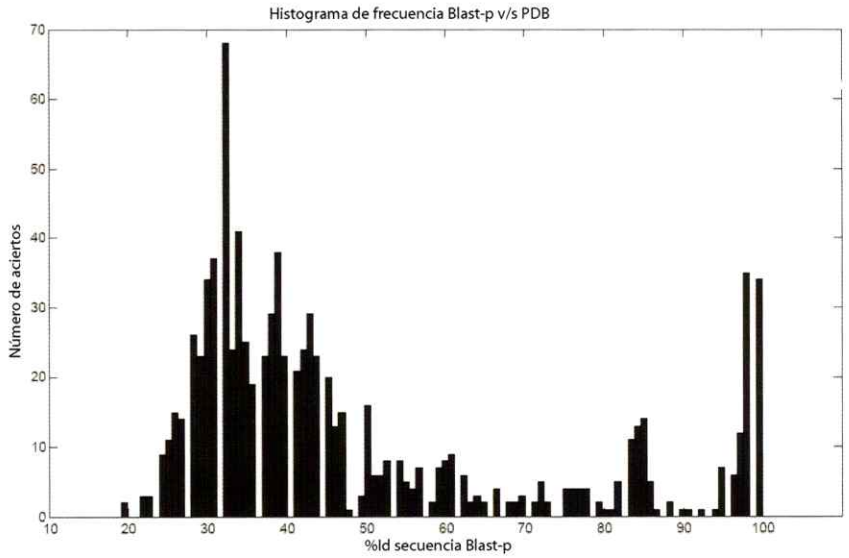


Figura 11: Resultados de las búsqueda de patrones vía Blast-p para la base de datos preliminar expresado en un histograma de frecuencia de hits en función del % id Blast-p. Nótese que la distribución es del tipo de valores extremos (ver figura 8).

Luego de filtrar los resultados de la búsqueda de secuencias patrones (ver métodos), la base de datos quedó compuesta por 5 estructuras cristalográficas y sus correspondientes secuencias patrones (figura 12 y tabla 1).

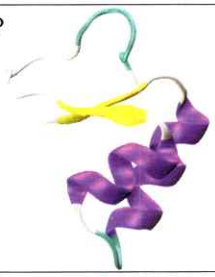
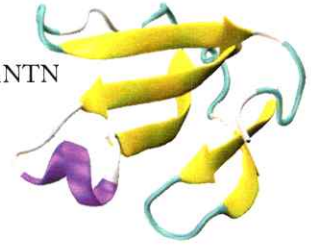
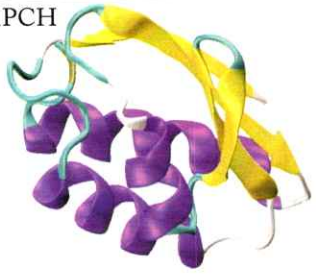
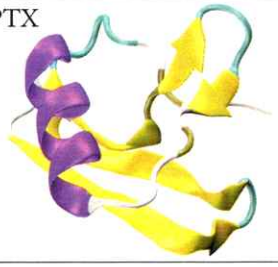
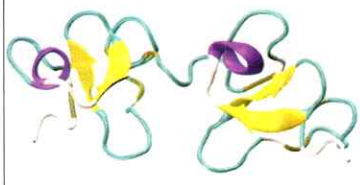
Estructura	Descripción	Datos Experimentales
1BHP 	Purotionina beta, perteneciente a la clase las tioninas, toxinas de plantas que aumentan la permeabilidad de membranas.	Nº residuos: 45 Resolución: 1.7Å Fuente: *Triticum aestivum* CATH: alfa-beta SCOP: pequeñas proteínas
1NTN 	Neurotoxina 1, componente del veneno de la familia de serpientes elapidas. Esta toxina actua sobre la subunidad-α del receptor de acetilcolina nicotínico al nivel postsináptico de la unión neuromuscular.	Nº residuos: 72 Resolución: 1.9Å Fuente: *Naja naja oxiana* CATH: mayormente beta SCOP: pequeñas proteínas
1PCH 	Proteína transportadora de fósforo contenedora de histidina HE. (HPr). Estas proteínas son parte del sistema fosfoenolpiruvato, el sistema de trasporte de azucar fosfotranferasa (PTS).	Nº residuos: 88 Resolución: 1.8Å Fuente: *Micoplasma capricolum* CATH: alfa y beta SCOP: proteínas alfa y beta
1PTX 	Toxina II (AaHII), parte del veneno de escorpiones, se une a canales de Na^+ .	Nº residuos: 64 Resolución: 1.3Å Fuente: *Andronoctus australis* CATH: alfa-beta SCOP: pequeñas proteínas
1UHA 	Tipo de lectina específica para sacaridos que contengan n-acetilglucosamina, y que causan aglutinación celular y sedimentación de glicoproteínas.	Nº residuos: 82 Resolución: 1.5Å Fuente: *Phytolacca americana* CATH: no asignada. SCOP: pequeñas proteínas

Figura 12: Configuración final de la base de datos, las proteínas aparecen esquematizadas de acuerdo a su estructura secundaria, donde amarillo representa sábanas β, violeta hélices α , calipso y blanco distintos tipos de lazos.

Tabla 1: Patrones para modelado comparativo para cada una de las estructuras de la base de datos.

Secuencia a modelar	Patrones	Nombre	Resolución Å	Residuos	% Identidad global
1BHP					
	1CNR	Crambina	1.0	46	37
	1AB1	Crambina	0.8	46	34
	1CRN	Crambina	1.5	46	34
	1EJG	Crambina	0.5	46	33
1NTN*					
	1TGX (A)***	Toxina gama	1.5	60	35
	1F94	Bucandina	0.9	63	35
	1QM7	Proteína quimérica	2.1	61	34
	1KXI (A)***	Cytotoxina	2.1	62	32
	1HOJ(A)***	Cardiotoxina	1.9	60	32
	1CDT(A)***	Cardiotoxina V4III	2.5	60	32
1PCH					
	2HPR	Hpr	2.0	87	45
	1PTF	Hpr	1.6	88	43
	1MU4 (B)***	CRH	1.8	87	41
	1FU0 (A)***	Hpr 46	1.9	87	41
	1MO1 (A)***	Seleno	1.8	87	38
	2JEL	FabHPR	2.5	85	36
	1OPD	Hpr	1.5	85	35
	1CM2	His 15asp Hpr	1.8	85	35
	1CM3	His 15asp Hpr	1.6	85	35
1PTX**					
	1NPI	Neurotoxina TS1	1.1	61	43
1UHA					
	1UKL (A)***	Importina beta	3.0	126	53

* Estructura modelada sin los últimos 6 residuos de carboxilo terminal.
** Estructura modelada sin los últimos 7 residuos del carboxilo terminal.
*** Cadena utilizada como patrón para el modelado.

Tal como se aprecia en la tabla 1 cada elemento de la base de datos, códigos PDB: 1BHP [Stec y col, 1995], 1NTN [Nickitenko y col, 1995], 1PCH [Pieper y col, 1995], 1PTX [Housconjunto y col, 1994] y 1UHA [Fujii y col, 2004], posee una serie de estructuras cristalográficas patrones para el modelado, resultantes de la búsqueda BLAST-p. En general se acepta que mientras mayor sea el número de patrones empleados en un modelado comparativo mejor serán los resultados obtenidos de este modelo [Fiser & Sali, 2003]. Aun cuando este no fue el caso para todos los elementos de la base de datos esto no resultó ser un problema ya que permitió estudiar distintos casos de modelado (con uno o varios patrones).

VII.2 Definición de un protocolo *ad-hoc* de dinámica molecular

Cada ECR de la base de datos fue sometida a un protocolo de DM particular (ver materiales y métodos) con el objetivo de determinar la estabilidad estructural de las proteínas de la base de datos y establecer si las energías de los sistemas estudiados llegaban a valores estables dentro de los tiempos de simulación disponibles de acuerdo a nuestra infraestructura. Para ello se analizó las trayectorias estructurales vía el cálculo del curso temporal de la RECR contra la estructura inicial y la derivada de la energía total con respecto al tiempo. Como se desprende de la figura 13 todas las proteínas de la base de datos son relativamente estables, no sobrepasando los 3Å la RMSD (Cα) en ninguno de los casos. Tal como se aprecia en la figura 13B, en la cual se presenta la variación en el tiempo de la primera derivada de la energía total del sistema con respecto al tiempo para todas las ECR de la base de datos, los valores energéticos para todas las simulaciones tienden a estabilizarse en alrededor de los 6 ps de simulación.

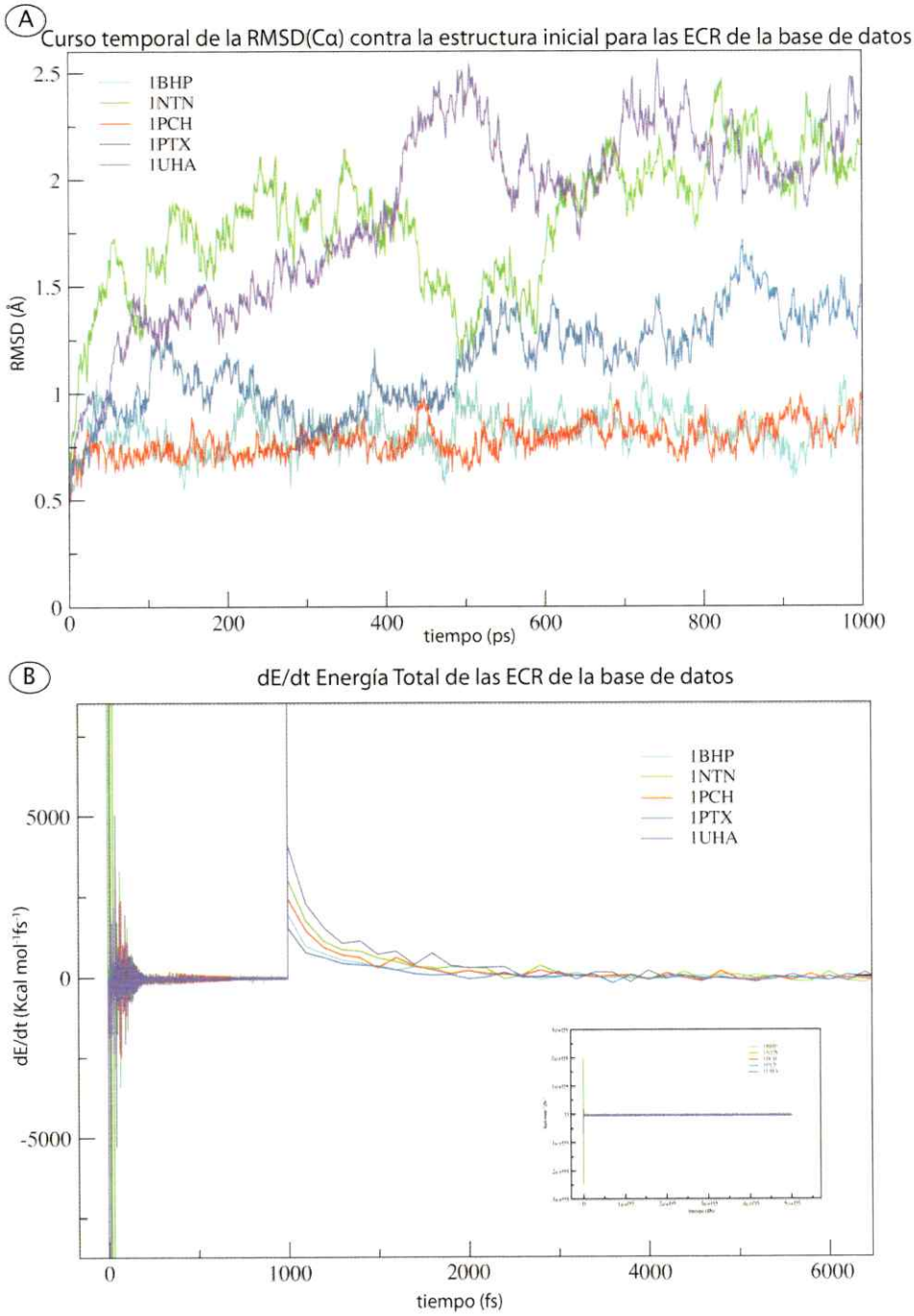


Figura 13: DM para las ECR. Panel A, Curso temporal de la RMSD(Cα) de la ECR de la base de datos. Panel B, dE/dt para los primeros pasos de la energía total a lo largo de la simulación de DM para las ECR de la base de datos todos los cristales de la base de datos (figura esquina inferior derecha, presentan el gráfico de dE/dt completo). __ 1BHP, __1NTN,__1PCH,__1PTX,__1UHA. Nótese los incrementos abruptos de energía al final de la etapa de minimización (paso 1000), reflejo del inicio de la etapa de DM.

En la figura 13B puede observarse incrementos abruptos en la energía calculada al inicio de la simulación Con el objetivo de determinar la estabilidad energética con énfasis en la primeras etapas de la DM, se estudió 3 distintos protocolos de dinámica molecular para el sistema 1UHA, dado que este es el que presenta mayor inestabilidad estructural (RMSD):

•Protocolo 1: Condiciones esféricas de borde, minimización y relajación en un solo paso con todo el sistema completo.

•Protocolo 2: Condiciones periódicas de borde, PME [Darden y col, 1999], minimización y relajación del solvente más iones manteniendo la proteína fija, para luego minimizar y relajar el sistema completo con la proteína libre.

•Protocolo 3: Condiciones esféricas de borde, minimización y minimización y relajación del solvente más iones manteniendo la proteína fija, para luego minimizar y relajar el sistema completo con la proteína libre.

Como se aprecia en la figura 14, para los tres tipos de protocolos de DM se observan incrementos abruptos en la energía al inicio de la simulación, sin embargo a los pocos picosegundos transcurridos la energía tiende a estabilizarse, lo cual se ve reflejado en el valor 0 de la derivada de la energía total con respecto al tiempo. Para los protocolos 2 y 3 se observan dos incrementos abruptos en la energía que pueden explicarse por las características de estos protocolos, ya que incorporan minimización energética y relajación por partes. Primero las moléculas de agua del sistema, manteniendo la estructura protéica fija, y luego todo el sistema completo, es decir, proteína más solvente. La aparición de dos incrementos abruptos de energía (ver figura 14 B y C) son reflejo del inicio de cada simulación de DM.

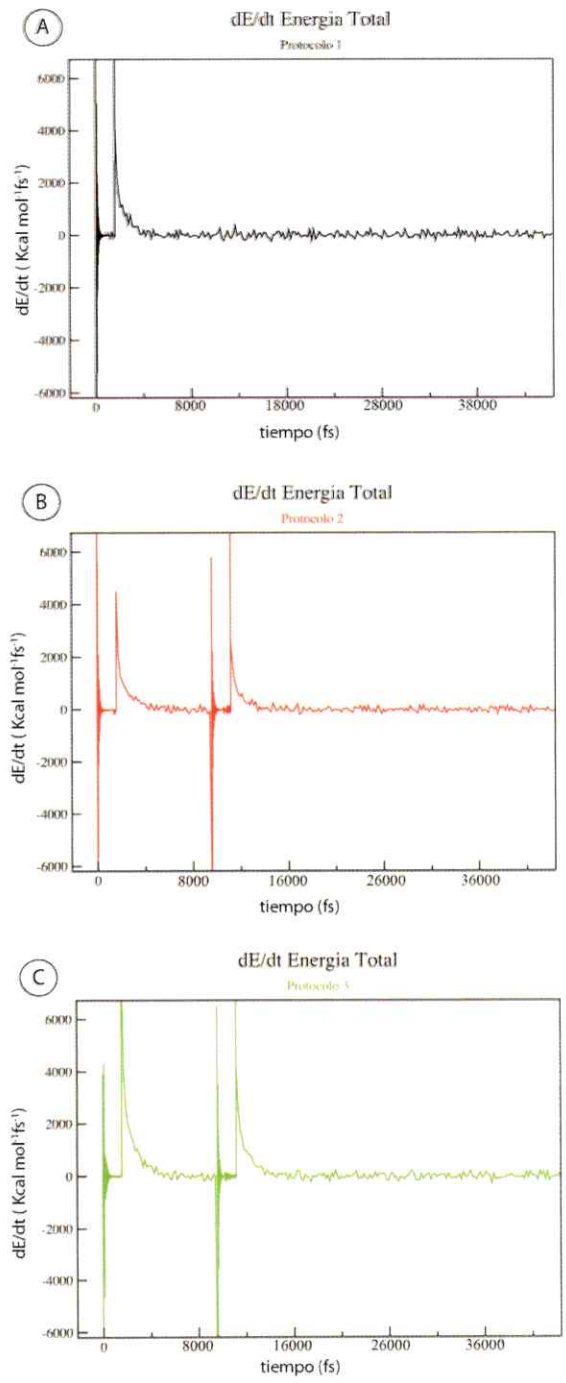


Figura 14: Derivada de la energía total con respecto al tiempo (dE/dt) para distintos protocolos de DM utilizados en el sistema 1UHA.: Panel A protocolo 1, condiciones esféricas de borde, minimización y relajación en un solo paso con todo el sistema completo. Panel B, protocolo 2 condiciones periódicas de borde, PME, minimización y relajación del solvente más iones manteniendo la proteína fija, para luego minimizar y relajar el sistema completo con la proteína libre; Panel C protocolo 3, condiciones esféricas de borde, minimización y relajación del solvente más iones manteniendo la proteína fija, para luego minimizar y relajar el sistema completo con la proteína libre.

Para complementar los análisis energéticos se estudió la estabilidad estructural para cada tipo de protocolo, a través del cálculo del curso temporal de la RMSD(Cα) respecto a la estructura inicial para cada una de estas simulaciones. En la figura 15 se observa que las estabilidades estructurales para los distintos protocolos de DM utilizados resultan ser similares no sobrepasando la RMSD(Cα) en ninguno de los casos los 2.5 Å.

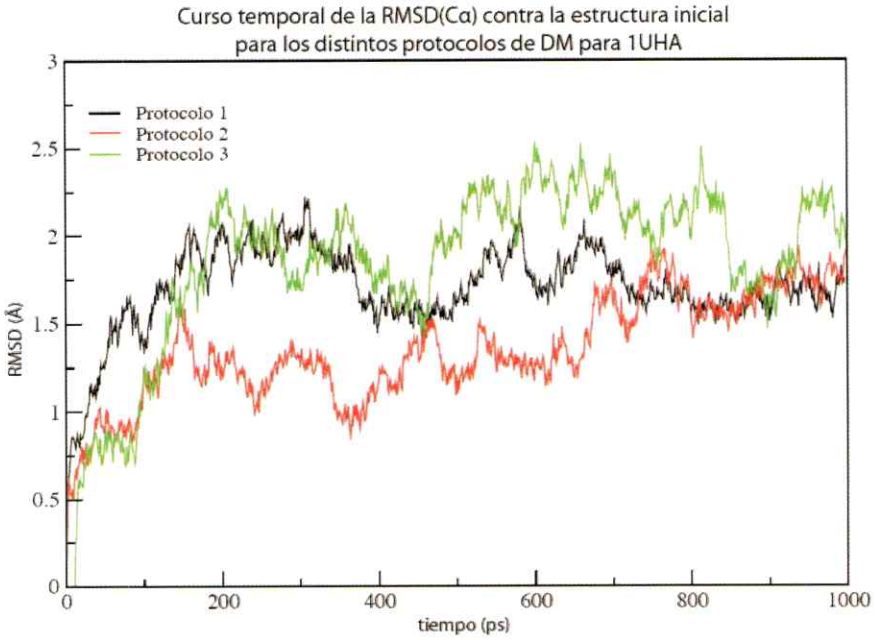


Figura 15: Curso temporal de la RMSD(Cα) para los tres protocolos de DM aplicados sobre las estructura 1UHA. __Protocolo 1: condiciones esféricas de borde, minimización y relajación en un solo paso con todo el sistema completo; __Protocolo 2: condiciones periódicas de borde, PME, minimización y relajación del solvente más iones manteniendo la proteína fija, para luego minimizar y relajar el sistema completo con la proteína libre; __Protocolo 3: condiciones esféricas de borde, minimización y relajación del solvente más iones manteniendo la proteína fija, para luego minimizar y relajar el sistema completo con la proteína libre.

Debido al comportamiento similar en la variación energética y el curso temporal de la RMSD(Cα) de la estructura 3D de la proteína 1UHA en los tres protocolos de DM y al menor costo computacional del protocolo 1, se decidió ocupar éste para todas las simulaciones posteriores.

VII.3 Modelado comparativo

A partir de los alineamientos entre las estructuras primarias de cada una de las 5 proteínas de la base de datos y de sus respectivos patrones, se generó para cada ECR, un un conjunto de 5 estructuras 3D (modelos comparativos),una estructura sin refinamiento de lazos y cuatro con diferentes refinamientos de lazos (un total de 15 modelos comparativos). Por otro lado, se intervino los alineamientos de cada ECR con sus respectivos patrones agregando aperturas aleatorias (gaps). Este proceso se iteró siete veces generando 7 alineamientos de estructuras primarias cada vez más alejados de los originales. Para cada uno de los alineamientos resultantes se generó 5 estructuras 3D por modelado comparativo, una estructura sin refinamiento de lazos y cuatro con diferentes refinamientos de lazos (un total de 175 modelos comparativos). Cada modelo resultante fue evaluado con ProsaII, Anolea, y RECR.

A partir de los datos desplegados en la figura 16 se desprende que, a medida que aumentan las aperturas aleatorias en los alineamientos, el porcentaje de identidad de secuencia con respecto a la secuencia a modelar disminuye en la mayoría de los casos. Al mismo tiempo, a medida que aumenta el número de iteraciones de intervención en los alineamientos, las evaluaciones de ProsaII y Anolea para las estructuras empeoran consistentemente en la mayoría de los casos, al igual que los RECR (medida directa de la calidad de los modelos).

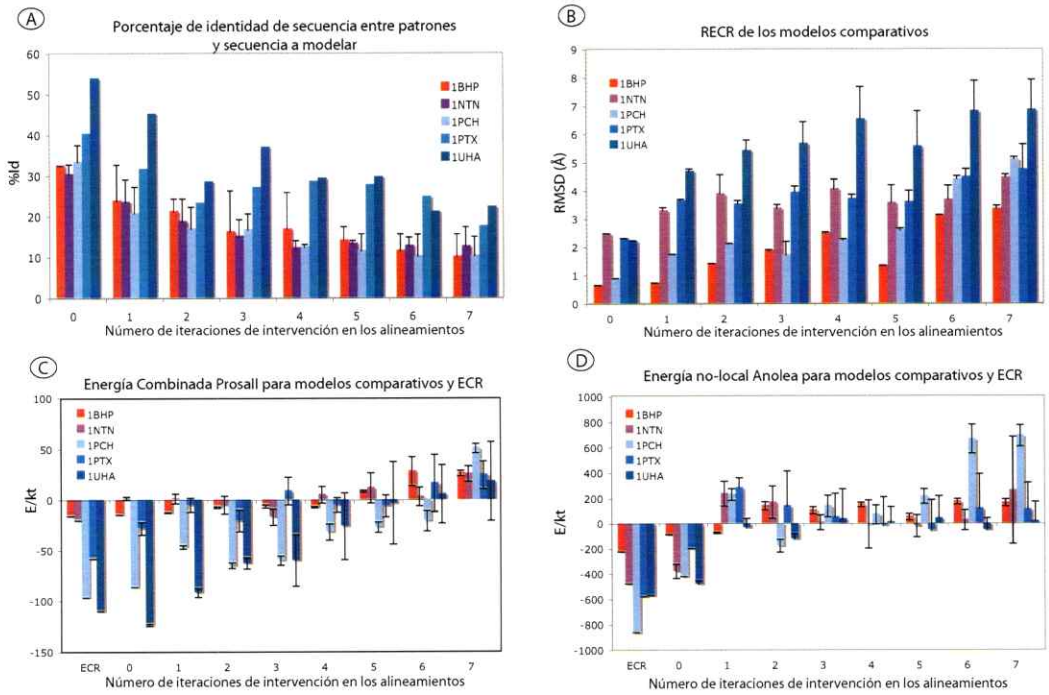


Figura 16: Análisis de los modelos generados en función de la iteración en la intervención de los alineamientos. A) Porcentaje de identidad de secuencia entre patrones y secuencia a modelar; B) RECR; C) Energía combinada calculada por Prosa II D) Energía no-local calculada por Anolea. 1BHP, 1NTN, 1PCH, 1PTX, 1UHA Las columnas representan el promedio y error típico de cada cantidad calculado sobre las 5 estructuras 1 sin + 4 con lazos optimizados.

VII.4 Simulaciones de dinámica molecular

Para cada estructura de la base de datos se simularon dos conjuntos de modelos, los generados a partir de alineamientos no intervenidos y de alineamientos intervenidos. El número de iteraciones de error para la evaluación por DM fue escogido de acuerdo a los valores de RECR a partir del análisis de la figura 16B, se eligieron aquellos modelos en donde existiese una diferencia apreciable con respecto a los modelos generados a partir de alineamientos no intervenidos. De esta forma para cada elemento de la base de datos se incorporaron: 6 aperturas aleatoria para 1BHP y 1PCH; 7 aperturas aleatorias para 1NTN y 1PTX; 3 aperturas aleatorias para 1UHA. En las figuras 17-21 se presentan los los cursos temporales de la RMSD (Cα) respecto de la estructura inicial para cada conjunto de modelos.

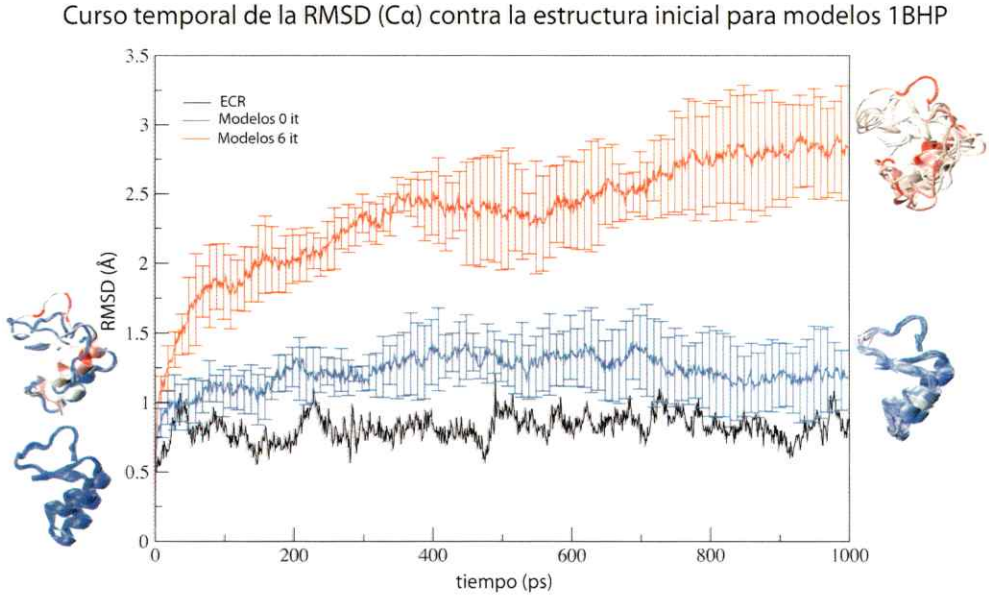


Figura 17: Curso temporal de la RMSD (Cα) contra la estructura inicial para la ECR y ambos conjuntos de modelos para 1BHP. A la izquierda y a la derecha se presentan alineamientos estructurales (representación de estructura secundaria) de los modelos y la ECR antes y después de las simulaciones de DM coloreados de acuerdo a identidad estructural (escala de azul, mayor identidad estructural, a rojo, baja identidad estructural. __ECR, __Modelos 0 iteración, __Modelos intervenidos con 6 iteraciones. Cada punto representa el promedio y error típico sobre las 5 estructuras 1 sin + 4 con lazos optimizados.

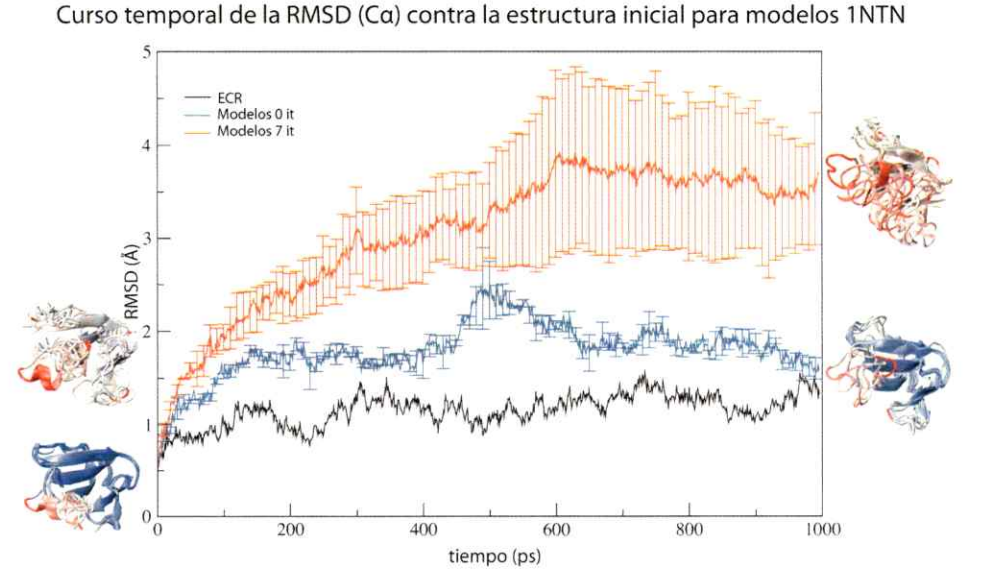


Figura 18: Curso temporal de la RMSD (Cα) contra la estructura inicial para la ECR y ambos conjuntos de modelos para 1NTN. A la izquierda y a la derecha se presentan alineamientos estructurales (representación de estructura secundaria) de los modelos y la ECR antes y después de las simulaciones de DM coloreados de acuerdo a identidad estructural (escala de azul, mayor identidad estructural, a rojo, baja identidad estructural. __ECR, __Modelos 0 iteración, __Modelos intervenidos con 7 iteraciones. Cada punto representa el promedio y error típico sobre las 5 estructuras 1 sin + 4 con lazos optimizados.

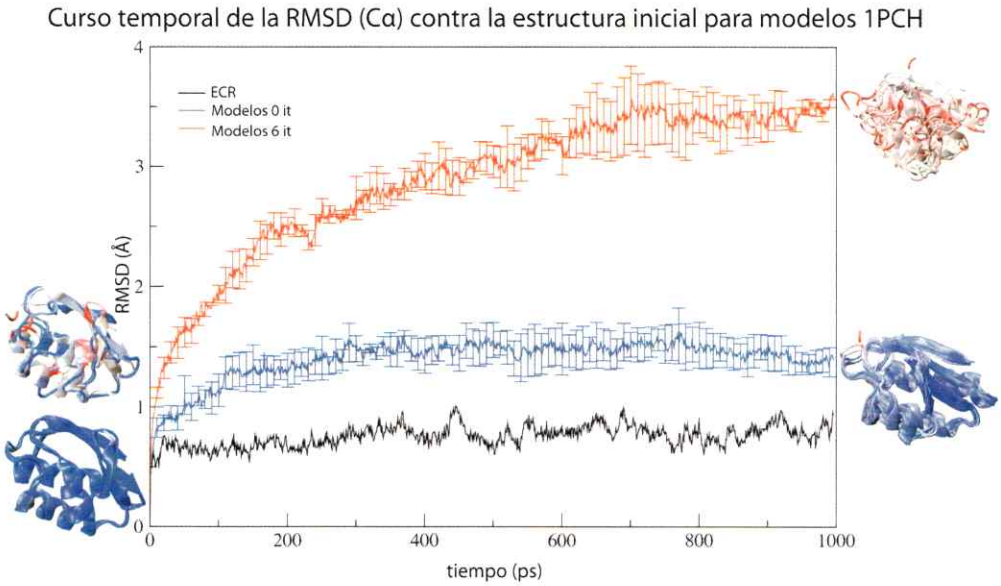


Figura 19: Curso temporal de la RMSD (Cα) contra la estructura inicial para la ECR y ambos conjuntos de modelos para 1BHP. A la izquierda y a la derecha se presentan alineamientos estructurales (representación de estructura secundaria) de los modelos y la ECR antes y después de las simulaciones de DM coloreados de acuerdo a identidad estructural (escala de azul, mayor identidad estructural, a rojo, baja identidad estructural. __ECR, __Modelos 0 iteración, __Modelos intervenidos con 6 iteraciones. Cada punto representa el promedio y error típico sobre las 5 estructuras 1 sin + 4 con lazos optimizados.

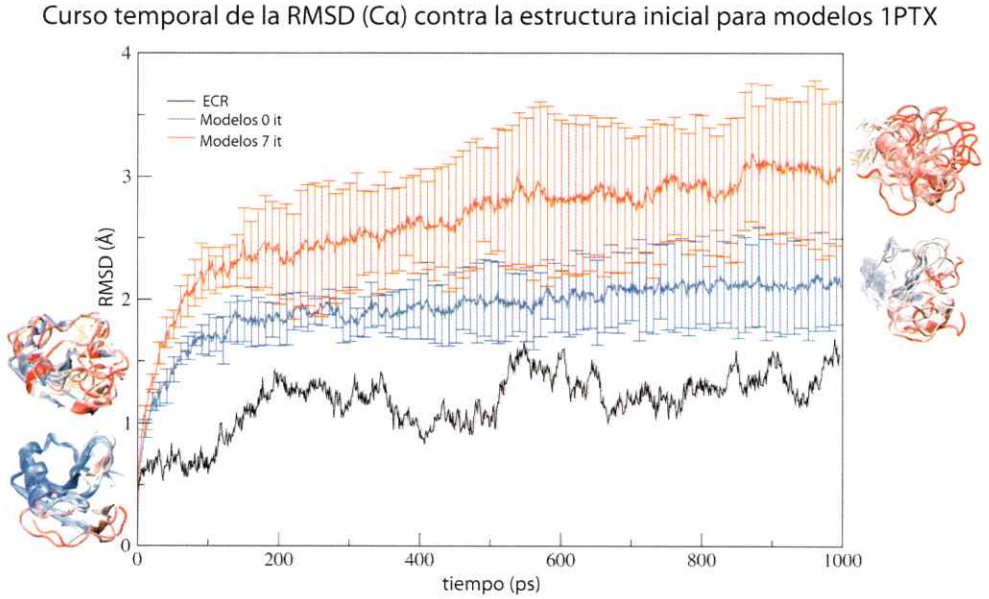


Figura 20: Curso temporal de la RMSD (Cα) contra la estructura inicial para la ECR y ambos conjuntos de modelos para 1PTX. A la izquierda y a la derecha se presentan alineamientos estructurales (representación de estructura secundaria) de los modelos y la ECR antes y después de las simulaciones de DM coloreados de acuerdo a identidad estructural (escala de azul, mayor identidad estructural, a rojo, baja identidad estructural. __ECR, __Modelos 0 iteración, __Modelos intervenidos con 7 iteraciones. Cada punto representa el promedio y error típico sobre las 5 estructuras 1 sin + 4 con lazos optimizados.

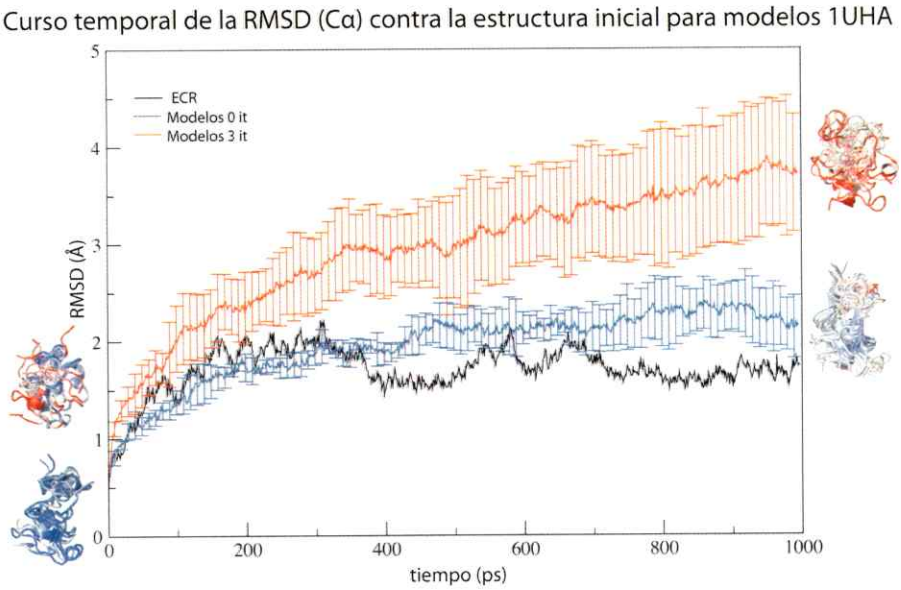


Figura 20: Curso temporal de la RMSD (Cα) contra la estructura inicial para la ECR y ambos conjuntos de modelos para 1UHA. A la izquierda y a la derecha se presentan alineamientos estructurales (representación de estructura secundaria) de los modelos y la ECR antes y después de las simulaciones de DM coloreados de acuerdo a identidad estructural (escala de azul, mayor identidad estructural, a rojo, baja identidad estructural. __ECR, __Modelos 0 iteración, __Modelos intervenidos con 3 iteraciones. Cada punto representa el promedio y error típico sobre las 5 estructuras 1 sin + 4 con lazos optimizados.

Es importante notar, observando las figuras 17-21, que sin importar el sistema simulado, el curso temporal de la RMSD(Cα) contra la estructura inicial para cada uno de los elementos de la base de datos y sus respectivos modelos presentan comportamientos similares. Es decir, en todos los casos estudiados existe una clara diferencia en la estabilidad estructural entre las estructuras cristalográficas de referencia, los modelos de alineamientos no intervenidos e intervenidos, siendo los cristales de referencia más estables que los modelos no intervenidos lo que a su vez son más estables que los modelos intervenidos.

VII.5 Análisis de modelos comparativos post dinámica molecular

Con el objetivo de estudiar los efectos de las simulaciones de dinámica molecular sobre las evaluaciones de los modelos comparativos después de la DM se evaluó la energía de cada conformación usando ProsaII, Anolea y se calculó la RECR. Los resultados de estas evaluaciones se presentan , en la figura 22.

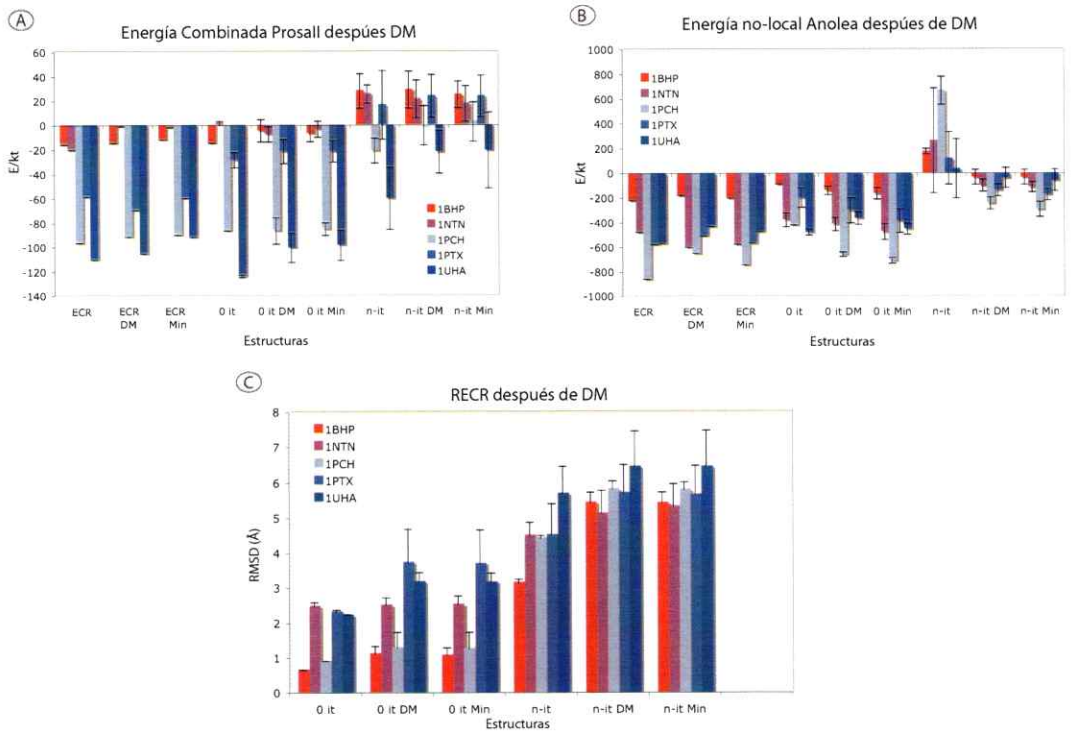


Figura 22: Análisis de dinámica molecular para para la ECR, los modelos no intervenidos y los modelos con iteraciones de intervención en los alineamientos antes y después de las simulaciones de DM.
Panel A, energía combinada calculada por Prosa II.
Panel B, energía no-local calculada por Anolea.
Panel C, RECR.
1BHP, 1NTN, 1PCH, 1PTX, 1UHA.
DM: estructura salida de dinámica molecular.
Min: estructura minimizada energéticamente luego de simulación de dinámica molecular.
0 it : modelo generado sin iteraciones de intervención en el alineamiento.
n-it : modelo intervenido con n-iteraciones de intervención en el alineamiento. 6 iteraciones de intervención para 1BHP y 1PCH, 7 iteraciones de intervención para 1NTN y 1PTX, 3 intervenciones para 1UHA.
Las columnas representan el promedio y error típico de cada cantidad calculado sobre las 5 estructuras 1 sin + 4 con lazos optimizados.

En la figura 22 se aprecia que las simulaciones de DM tienen los siguientes efectos dependiendo del análisis realizado: para los modelos no intervenidos las RECR aumentan levemente y los valores de pseudoenergías conformacionales calculadas por ProsaII y Anolea no varían luego de las DM existiendo un pequeño incremento en las barras de error para ambos análisis. Los modelos intervenidos presentan comportamientos similares para las evaluaciones ProsaII (aunque para el sistema 1NTN hay claras diferencias luego de la DM) y las RECR, no obstante, para estos modelos los valores de pseudoenergías conformacionales calculadas por Anolea cambian drásticamente luego de las simulaciones de DM pasando de valores positivos a negativos (evaluación favorable). Cabe destacar que las estructuras de los cristales de referencia no sufren cambios considerables en sus evaluaciones ProsaII y Anolea después de las DM.

VII.6 Divergencia estructural

Los alineamientos estructurales previos y posteriores a las simulaciones de dinámica molecular para cada unos de los conjuntos de modelos y las estructuras cristalográficas (ver figuras 17 a la 21) indican que las estructuras simuladas tienden a divergir estructuralmente a medida que son sometidas a simulaciones de DM. Con el objetivo de estudiar este fenómeno se analizó la evolución del RMSD(Cα) de cada modelos contra la ECR a lo largo de toda la trayectoria (cada paso contra cada paso). Los resultados de este análisis se presentan en la figura 23. Como se observa en la figura 23 existen claras diferencias para ambos conjuntos de modelos, sin embargo el momento exacto en dónde ocurre la divergencia estructural no se logra apreciar lo cual puede ser debido a la resolución temporal del análisis.

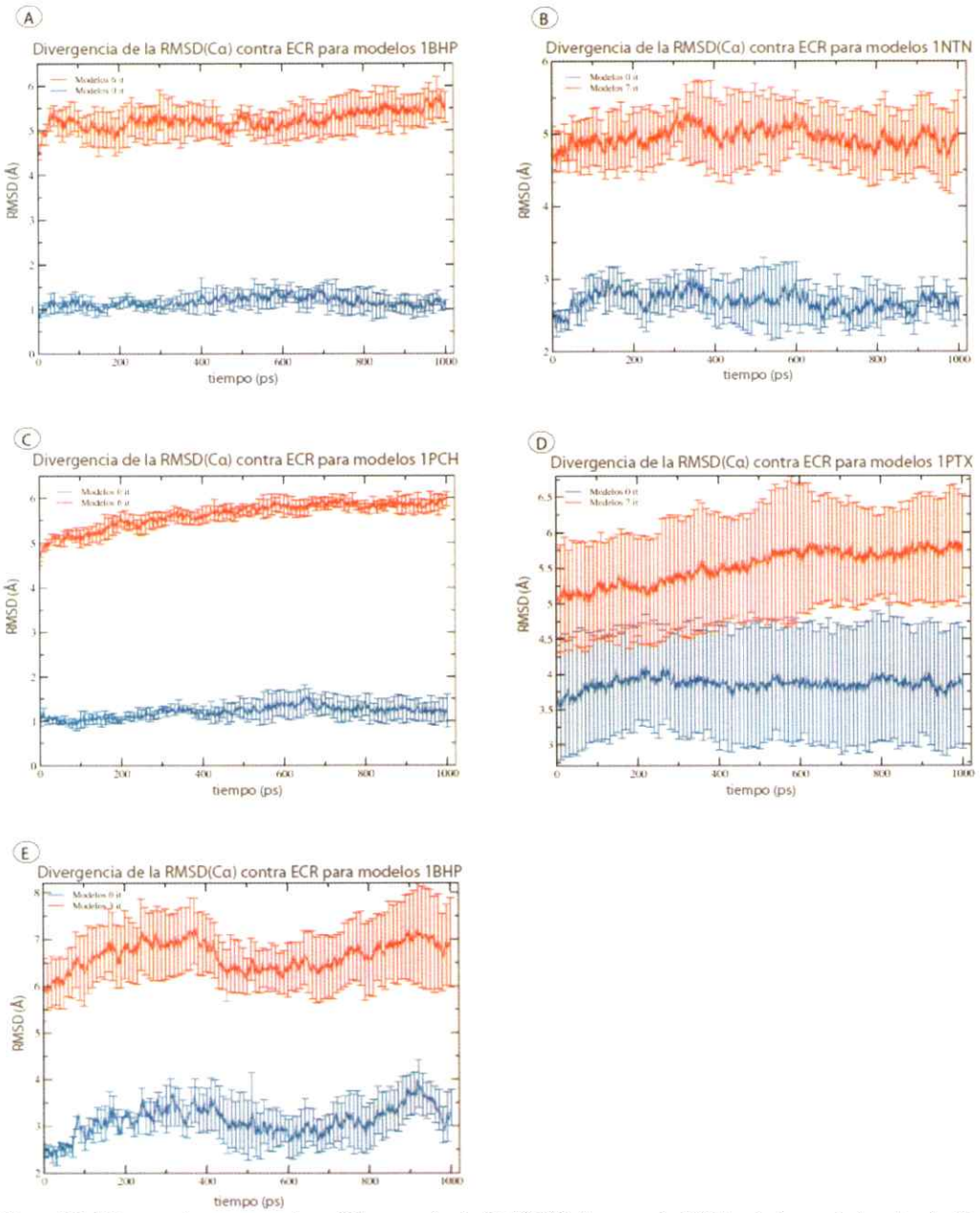


Figura 23: Divergencia estructural medida a través de RMSD(Cα) contra la ECR a lo largo de las simulación de dinámica molecular para cada uno de los conjuntos de modelos de los elementos de la base de datos:
Panel A, análisis para modelos 1BHP, __ Modelos 0 iteración de intervención en el alineamiento , __ Modelos con 6 iteraciones de intervención en el alineamiento;
Panel B, análisis para modelos 1NTN, __Modelos 0 iteración de intervención en el alineamiento, __Modelos con 7 iteraciones de intervención en el alineamiento;
Panel C, análisis para modelos 1PCH, __Modelos 0 iteración de intervención en el alineamiento, __Modelos con 6 iteraciones de intervención en el alineamiento ;
Panel D, análisis para modelos 1PTX, __ Modelos 0 iteración de intervención en el alineamiento, __Modelos con 7 iteraciones de intervención en el alineamiento;
Panel E, análisis para modelos 1UHA, __ Modelos 0 iteración de intervención en el alineamiento __Modelos con 7 iteraciones de intervención en el alineamiento;
Cada punto representa el promedio y error típico sobre las 5 estructuras 1 sin + 4 con lazos optimizados.

VII.7 Modelado de la ubicuitina.

Con el propósito de validar la metodología en un protocolo clásico de modelado molecular, se generó una serie de modelos comparativos a partir de patrones con distinta identidad de secuencia con respecto a la secuencia a modelar (desde 26% hasta 57%) para la estructura cristalográfica de la ubicuitina, proteína de 76 residuos (código pdb 1UBQ) [Vijay-Kumar y col, 1987] para luego ser sometidos a simulaciones de dinámica molecular (ver sección VI.3). La estabilidad estructural de los distintos modelos fue comparada con las evaluaciones de ProsaII y Anolea, resultados que se presentan en la figura 24.

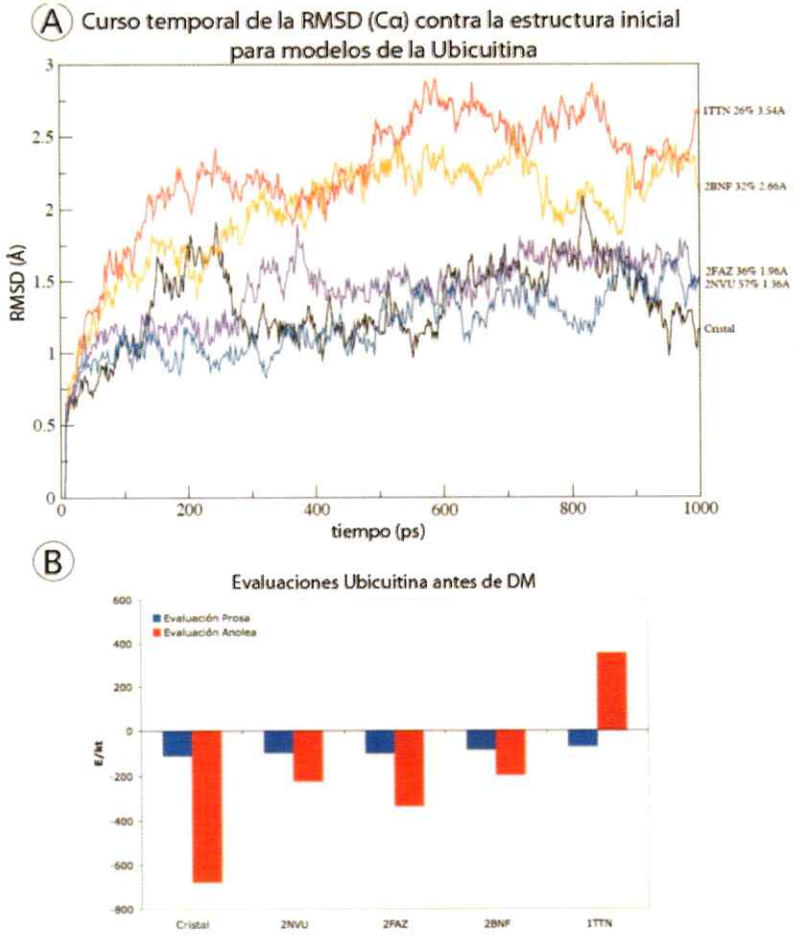


Figura 24: Evaluación para modelos de la ubicuitina provenientes de distintos patrones:
Panel A, RMSD (Cα) de trayectoria para la estructura cristalográfica y los ■modelos de la ubicuitina;
cristal ubicuitina (1UBQ) .
■ modelo de 1.36Å de RMSD (Cα) v/s 1UBQ, desarrollado a partir del patrón 2NVU (57% id).
■modelo de 1.96Å de RMSD (Cα) v/s 1UBQ, desarrollado a partir del patrón 2FAZ (36% id).
■modelo de 2.66Å de RMSD (Cα) v/s 1UBQ, desarrollado a partir del patrón 2BNF (32% id).
■modelo de 3.54Å de RMSD (Cα) v/s 1UBQ, desarrollado a partir del patrón 1TTN (26% id).
Panel B: Evaluación de ProsaII y Anolea para la estructura cristalográfica y los modelos de la ubicuitina;
cristal: estructura cristalográfica de la ubicuitina 1UBQ.
2NVU: modelo de 1.36Å de RMSD (Cα) v/s 1UBQ, desarrollado a partir del patrón 2NVU (57% id).
2FAZ:modelo de 1.96Å de RMSD (Cα) v/s 1UBQ, desarrollado a partir del patrón 2FAZ (36% id).
2BNF modelo de 2.66Å de RMSD (Cα) v/s 1UBQ, desarrollado a partir del patrón 2BNF (32% id).
1TTN modelo de 3.54Å de RMSD (Cα) v/s 1UBQ, desarrollado a partir del patrón 1TTN (26% id).
■ evaluación ProsaII.
■ evaluación Anolea.

Como se aprecia en la figura 24A, la estabilidad estructural media por el curso temporal de la RMSD (Cα) contra la estructura inicial a través de simulaciones de DM permite diferenciar modelos comparativos desarrollados a partir de patrones con distinto porcentaje de identidad de secuencia. Estos resultados son consistentes con las evaluaciones ProsaII y Anolea (figura 24B), sin embargo ProsaII presenta valores similares de pseudoenergía conformacionales sin importar el modelo evaluado. Anolea logra una mejor resolución entre los modelos pero, en particular para este caso, falla en la evaluación del modelo generado a partir del patrón 2FAZ, ya que aunque este modelo presenta RMSD (Cα) contra la estructura cristalográfica de la ubicuitina mayor al modelo generado a partir del patrón 2NVU, su pseudoenergía conformacional es menor (evaluación favorable).

VIII Discusión

VIII.1 Base de datos y definición de un protocolo *ad-hoc* de dinámica molecular

De acuerdo a los resultados de la figura 13 todas las estructuras de la base de datos fueron sometidas a un protocolo de DM presentando simulaciones relativamente estables. Con el propósito de estimar si el protocolo de DM utilizado es apropiado y no genera artefactos en el sistema se probó dos protocolos alternativos de DM para la estructura 1UHA. Tal como se aprecia en la figura 14, el análisis de la variación de la energía total del sistema con respecto al tiempo (dE/dt) demostró que los incrementos abruptos de la energía son un reflejo del inicio de la DM ya que en esta etapa se sube bruscamente la temperatura del sistema. Un protocolo en dónde se subiese la temperatura en forma gradual no debería presentar estos saltos abruptos de energía.De este modo, se concluye que la metodología de DM utilizada resulta adecuada y las simulaciones de las estructuras cristalográficas son estables, lo cual permitió definirlas como comportamiento de referencia en los estudios posteriores.

VIII.2 Modelado comparativo

•*Modelos provenientes de alineamientos no intevenidos :*

El valor de RECR (diferencia estructural) de los modelos no intervenidos fue diferente (figura 16), dependiendo de variados factores. En el caso de este estudio, la identidad de secuencia no fue una variable fundamental en la calidad de los modelos [Sali & Blundell, 1993] debido a que, intencionalmente, se trabajó para todos los modelos en el intervalo de baja identidad de secuencia. No obstante los modelos

obtenidos para 1BHP y 1PCH (ver figura 16) presentaron RECR del orden de 0.7Å, considerados como modelos de alta precisión [Baker & Sali, 2001](ver figura 2). Para 1PTX y 1NTN el proceso de modelado requirió quitar los residuos del extremo carboxilo terminal de ambas estructuras, siendo el caso extremo 1UHA para la cual fue necesario utilizar un patrón de mayor identidad de secuencia (53%, ver tabla 1) ya que con patrones de menor identidad de secuencia fue imposible obtener modelos de calidad razonable bajo el valor de RMSD. Si se analiza el elemento común entre estas estructuras se percibe que poseen lazos (loops) de una extensión apreciable, relativo a su tamaño, siendo estas las regiones más complejas de modelar [Fiser y col, 2000]. Es importante hacer la diferencia entre la definición de lazo estructural, como la ausencia de una estructura secundaria en particular y la definición de lazo de MODELLER la que tiene relación con la ausencia de representación estructural por parte de los patrones, es decir las zonas de aperturas (gaps) en el alineamiento. Estructuralmente, 1BHP y 1PCH presentan pocas regiones de lazos, además de haber sido modeladas a partir de estructuras cristalográficas de alta resolución (especialmente 1BHP) obteniendo modelos de bajos RECR. Aún cuando la identidad de secuencia es la variable fundamental en el modelado comparativo, cuando se trabaja en el ámbito de la baja identidad de secuencia, la resolución comienza a ser un factor fundamental en la calidad y precisión de los modelos [Fiser & Sali, 2003]. Existe una proporcionalidad entre la estabilidad estructural de los cristales estudiados y la calidad de los modelos generados para estas proteínas en términos de RECR. Los lazos estructurales son la zonas más complejas de modelar y al mismo tiempo las regiones más móviles en el contexto de una simulación de DM [Post y col, 1989], conduciendo a la idea que que mientras más inestable sea la dinámica molecular de una estructurada modelada, su modelado es menos confiable.

Si bien las evaluaciones de las ECR presentan valores de pseudoenergías conformacionales menores a los modelos, lo cual es esperable debido a que estos evaluadores están parametrizados a partir de estructuras cristalográficas [Sippl, 1993, Melo y col, 1997], estas son comparables y presentan valores negativos. En conjunto con los cálculos de RECR, los cuales no sobrepasaron los 3Å, y de acuerdo con la literatura [Baker & Sali, 2001] los modelos desarrollados fueron catalogados como adecuados.

•*Modelos provenientes de alineamientos intervenidos con n- iteraciones de inserción de indel:*

En la figura 29 se observa el efecto de la inserción de aperturas aleatorias (gaps) en los alineamientos, en dónde la progresión de esta hace que las posiciones más conservadas en los alineamientos presenten menor definición. Este mismo comportamiento se observa cuando se alinean secuencias que presentan cada vez menor porcentaje de identidad de secuencia y por lo tanto se hace menos probable que una posición en el alineamiento entre el patrón y la secuencia objetivo, efectivamente corresponda a residuos homólogos. A nivel de plegamiento, esto equivale al uso de un patrón que, si bien presenta el plegamiento arquetípico de la superfamilia, no refleja con precisión la estructura del grupo filogenético más cercano.

El error asociado al desarrollo de estos modelos viene dado por la inserción aleatoria de aperturas (gaps) en los alineamientos. Con el incremento en la iteraciones de error se observan ciertos fenómenos que ameritan un análisis más detallado. A diferencia de lo que ocurre con los modelos no intervenidos, donde la resolución de los patrones y la presencia de lazos en su estructura juegan un papel fundamental, para estos sistemas la identidad de secuencia empieza a ser un factor fundamental, como se observa en la figura 16A. El incremento en el error acarrea una disminución progresiva en la similitud de

caracteres, con ciertas excepciones que pueden ser explicadas por el carácter estocástico en la inserción de las aperturas, ya que que ciertas aperturas (gaps) caen en zonas que pueden no tener efecto significativo sobre el alineamiento (ej: zona de gaps) o incluso mejorarlos (ver figuras 16 y 28). Es importante destacar que, mientras más patrones se hayan usado en el alineamiento, se observa una atenuación del error introducido ya que el efecto individual de una particular apertura en cada secuencia tiene menor impacto.

Para el programa MODELLER la definición de lazo no está asociada a la ausencia de una estructura secundaria definida, sino que a la falta de representación estructural de una región de la secuencia a modelar por parte del patrón, es decir, la zonas de aperturasl (gaps) en el alineamiento[Sali & Blundell, 1993]. Es importante destacar que las iteraciones de éstas tienen un efecto gravitante sobre la calidad de los modelos generados, siendo la norma general el empeoramiento de las evaluaciones y aumento del RECR a medida que se insertan aperturas al azar en los alineamientos. Tal como se aprecia en la figura 16 la evaluación Anolea resulta más sensible al error aplicado que ProsaII. Es interesante notar el incremento progresivo en las barras de error a medida que aumenta el número de iteraciones de intervención en los alineamientos en forma independiente del evaluador utilizado. El análisis del protocolo de modelado, que incluye el método de refinamiento de lazos, permite interpretar el comportamiento de ProsaII y Anolea. El error asociado a estos modelos viene dado por la disminución en la representatividad de los patrones sobre la secuencia a modelar, es decir las aperturas cubren mayores zonas del alineamiento, lo cual MODELLER considera como lazos. Esto implica que el modelado de lazos cobra cada vez más relevancia por lo que cada modelo va a tener una conformación de lazos particular y no necesariamente equivalente, resultando en la dispersión en los valores de las pseudoenergías conformacionales calculadas por ProsaII y Anolea que explican este incremento progresivo en la barras de error (desviación estándar).

VIII.3 Simulaciones de dinámica molecular para modelos comparativos

Es importante notar la diferencia entre estabilidad estructural y termodinámica. El uso común del término estabilidad, implica que un objeto, sistema o situación permanece invariable por un período de tiempo considerable. De este modo, se define la estabilidad termodinámica como la entalpía o energía potencial de un sistema con respecto a un estado de referencia, por ejemplo para una reacción exotérmica se dice que los productos son más estables termodinámicamente que los reactantes. En cambio se entiende por estabilidad estructural a la conservación en el tiempo, de la conformación de un determinado sistema (en este caso proteínas) con respecto a sus condiciones iniciales.

•*Efectos de las simulaciones de DM sobre los modelos provenientes de alineamientos no intervenidos:*

Las simulaciones de DM para los modelos desarrollados a partir de alineamientos no intervenidos (0 iteraciones), para todas las estructuras cristalográficas estudiadas, presentan cursos temporales de la RMSD (Cα) contra la estructura inicial que no sobrepasa el límite canónico de 3Å (ver figuras 17 a la 21). Estos resultados son un reflejo del campo de fuerza, función que está parametrizada a partir de estructuras determinadas experimentalmente [Ponder & Case, 2003]. Los modelos al ser una aproximación de la estructuras cristalográficas tienden a mostrar un comportamiento similar pero no necesariamente idéntico, entendiendo que mientras mejor sea la calidad del modelo su comportamiento debe asemejarse más al de la ECR.

•*Efectos de las simulaciones de DM sobre los modelos provenientes de alineamientos intervenidos con n-iteraciones de inserción de indel:*

Para el caso de los modelos generados con alineamientos intervenidos con inserciones de aperturas aleatorias (número de iteraciones característico para cada sistema estudiado), las simulaciones de DM muestran una tendencia general a presentar inestabilidad estructural existiendo una mayor dispersión en los valores observados (ver figuras 17 a la 21). Hay una clara diferencia en la estabilidad estructural entre estos modelos y los generados sin intervención en los alineamientos, en particular para las estructuras que fueron catalogadas como modelos de alta calidad, es decir RECR ≤ 1Å(1BHP y 1PCH) en los que esta diferencia es aún más importante (ver figuras 17 y 19). A medida que la calidad de los modelos generados no intervenidos disminuye, RECR ≈ 2.5 Å, la diferencia entre las curvas (figuras 17 a la 21) disminuye de manera significativa.

•*Efectos de la Dinámica molecular sobre la evaluación de los modelos*

Tal como se aprecia en la figura 22, las evaluaciones de las ECR prácticamente no varían luego de las simulaciones de DM, no importando el evaluador utilizado (ProsaII o Anolea). Este comportamiento es similar al observado para los modelos generados sin iteraciones de intervención en los alineamientos, dónde no existen mayores diferencias antes y después de las dinámicas moleculares salvo un leve aumento en la dispersión de los datos, reflejo de la divergencia estructural de las trayectorias. Por el contrario la estructuras construidas a partir de alineamientos con inserciones de aperturas aleatorias, presentan un comportamiento particular en sus evaluaciones con Anolea luego de las simulaciones de DM. El valor de la pseudoenergía conformacional decae

dramáticamente, pasando de valores positivos a valores negativos, siendo que las ECR y los modelos no intervenidos no sufren cambios en sus evaluaciones Anolea post-DM. Estos resultados propondrían un efecto de refinamiento de los modelos por parte de la DM (si se entiende que la parametrización de Anolea viene de la derivación de potenciales estadísticos de estructuras cristalinas). Sin embargo estos resultados se contradicen con los valores de RECR presentados en la figura 22C en donde se observa que los valores aumentan consistentemente luego de las simulaciones de DM, implicando que los modelos no están siendo refinados ya que divergen estructuralmente de las ECR. Según sus autores, esta contradicción es un caso particular cuando se trabaja con proteínas de menos de 100 residuos[Melo y col, 1997], no obstante, las ECR y los modelos generados sin iteración de inserción de aperturas aleatorias son evaluados adecuadamente antes y después de las DM, lo que el argumento anterior no explica. Estos resultados de RECR son confirmados por ProsaII, el cual no presenta variaciones considerables en la evaluación de los modelos con iteración de intervención en los alineamientos luego de las simulaciones de DM, aunque sí un aumento en la dispersión de los datos.

•*Divergencia Estructural*

El cambio en los valores de la RECR de los modelos luego de ser sometidos a simulaciones de DM reflejan la independencia de las trayectorias (ver figura 22 C). Los alineamientos estructurales antes y después de la simulaciones son otra evidencia de la divergencia estructural (ver figuras 17 a las 21) que experimentan estos modelos al ser sometidos a simulaciones de DM, en las cuales se perturban las condiciones iniciales del sistema, entendiéndose por esto a la diferencia que existe entre la ECR, condición

normal, y lo modelos, condición inicial perturbada [Braxenthaler y col 1997]. Mientras más similares sean los modelos, los valores de RECR a lo largo de la trayectoria, van a ser menores que para los modelos desarrollados con intervención en lo alineamientos. Sería interesante realizar un análisis cuantitativo de estos sistemas para poder determinar si se comportan de manera caótica, en particular el cálculo de exponentes y tiempos de Lyapunov [Braxenthaler y col 1997] pueden ser buenas aproximaciones para realizar este estudio.

•*Modelado de la ubicuitina*

Los resultados expresados en la figura 24, corroboran la utilidad de la DM dentro del ámbito de modelado comparativo, en particular la estabilidad estructural medida lo largo de la trayectoria fue capaz de diferenciar modelos comparativos de la ubicuitina provenientes de distintos patrones, lo cuales se diferenciaban en el porcentaje de identidad de secuencia con respecto a la secuencia a modelar. Al mismo tiempo las evaluaciones ProsaII y Anolea no resultan adecuadas; ProsaII presenta una muy baja sensibilidad lo cual no permite diferenciar de forma clara los distintos modelos; Anolea presenta una excesiva sensibilidad, sin embargo no es robusto y falla en ciertas evaluaciones (en el caso estudiado, para el modelo proveniente del patrón 2FAZ).

IX Conclusiones

I. Con respecto al protocolo de DM utilizado: Los protocolos estudiados permiten afirmar que lo realmente fundamental es minimizar y relajar el sistema por tiempos adecuados para llegar a valores estables de energía.

II. Acerca del modelado de proteínas en el ámbito de la baja identidad de secuencia: En particular para las estructuras 1BHP y 1PCH se generaron modelos de altísima precisión, inclusive modelos obtenidos a partir de alineamientos intervenidos (1 a 5 iteraciones de error) no presentaron RECR superiores a los 3Å, lo cual demuestra que es posible utilizar protocolos de modelado comparativo bajo la barrera del 30% de identidad se secuencia.

III. Acerca de la estabilidad estructural de los modelos a lo largo de las simulaciones de DM : Los modelos no intervenidos presentaron estabilidad estructural, similar a las ECR. Por el contrario los modelos generados a partir de la intervención de los alineamientos con aperturas aleatorias presentan inestabilidad estructural siendo en su mayoría discriminables de los modelos generados sin intervención en los alineamientos de y de las ECR.)

IV. Acerca de la calidad de los modelos y la DM: La estabilidad estructural relativa en una simulación de DM, (medida a través del curso de la RMSD contra la estructura inicial) puede ser considerada como una medida de la calidad de un modelo siempre y cuando las condiciones del sistema bajo las cuales se realiza la simulación sean las adecuadas (nativas a la proteína estudiada).

V. Acerca de los evaluadores utilizados: ProsaII presenta menos sensibilidad a la iteración de intervención en los alineamientos (relativo a Anolea) pero es más robusto, por el contrario Anolea presenta una sensibilidad mucho mayor a la inserción de aperturas aleatorias en los alineamientos, pero es menos robusto presentando un comportamiento erratico frente a los efectos de la dinámica molecular.

VI. Acerca del modelado de la ubicuitina: Estos experimentos corroboran los resultados anteriores, demostrando la utilidad de la DM en el ámbito del modelado comparativo, ya que las evaluaciones ProsaII y Anolea no fueron ni lo consistentes ni lo resolutivas necesarias.

VII. Acerca de la utilidad de la DM como evaluador: Operacionalmente, los evaluadores clásicos, son menos costosos que implementar una simulación de DM y se desempeñan bastante bien cuando se modela en el ámbito de la alta identidad de secuencia. En particular para este trabajo las predicciones de ProsaII y Anolea se correlacionan con las evaluación dinámica, es decir que modelos catalogados como correctos, en su mayoría, presentan trayectorias estructuralmente estables y viceversa. No obstante ambos evaluadores presentan comportamientos erráticos y no proporcionan la confiabilidad necesaria para poder discriminar entre estructuras correctas al nivel de modelos generados en el ámbito de la baja de identidad de secuencia.

VIII. Conclusión final:

Durante el transcurso de esta memoria se ha demostrado que la estabilidad estructural calculada a partir de trayectorias de Dinámica Molecular permite diferenciar modelos comparativos desarrollados a partir de alineamientos de alta y baja calidad. Esta metodología resulta relevante en el ámbito del modelado comparativo de proteínas pequeñas en el intervalo de la baja identidad de secuencia dado que las herramientas actuales de evaluación requieren de un tipo de evaluación independiente en este nivel de modelado comparativo.

X Material complementario

X.1 Dinámica molecular

*Texto extraído y modificado de los artículos "*Molecular Dynamics*" [Chipot, 2000] y "*A molecular dynamics primer*" [Ercolessi, 1997].

La Dinámica Molecular es una técnica mecánico-estadística para el análisis de sistemas moleculares a nivel atómico. Está basada en mecánica estadística y física clásica. La DM ha sido utilizada en diversas áreas como ciencias de los materiales, climatología y biociencias para sistemas de lípidos, ácidos nucleicos y proteínas. La DM es la integración dependiente del tiempo de las ecuaciones clásicas de movimiento(figura 25). Estas ecuaciones, hasta para los sistemas más simples, son de tal complejidad que la integración debe hacerse con métodos numéricos sobre un gran número de discretos pasos temporales, en vez realizarse de manera analítica en una forma continua. La magnitud del paso de tiempo (timestep) está en el rango de los fs, ya que esta metodología supone que por cada paso de tiempo las coordenadas del sistema están fijas. Para cada paso de tiempo estas coordenadas "fijas" son usadas para calcular la Energía Potencial del Sistema y su primera derivada (con respecto a la posición de un determinado átomo del sistema), la fuerza, utilizando un campo de fuerza de Mecánica Molecular.

$$a_n = \frac{f_n}{m} \quad (1)$$

$$v_{n+1} = v_n + a_n \Delta t \quad (2)$$

$$x_{n+1} = x_n + v_n \Delta t + \frac{1}{2} a_n \Delta t^2 \quad (3)$$

Figura 25: 1 segunda ley de Newton, 2 y 3 ecuaciones clásicas de movimiento para la velocidad y posición respectivamente.
n : tiempo n
a_n : aceleración de la partícula en n.
f_n : fuerza actuando sobre partícula n.
v_n : velocidad de la partícula en n.
x_n : posición de la partícula en n.
Δt: paso de tiempo (timestep)

Una larga serie de estos cálculos permite generar una trayectoria a través del espacio fase, espacio de dimensión 6N (dónde N es el número total de átomos del sistema), definido por los tres vectores espaciales de la posición de los átomos y sus velocidades. Normalmente los análisis post-simulaciones se relacionan con el subespacio de las posiciones atómicas (coordenadas) del espacio fase.

Un elemento importante dentro de la DM es el Grupo Termodinámico utilizado, operacionalmente son las condiciones macroscópicas del sistema que uno mantiene constante a lo largo de la simulación, es importante destacar que para cada variable macroscópica existen algoritmos específicos que permiten mantenerlas en valores constantes. Un grupo termodinámico NVT implica que para el sistema estudiado se mantiene el número de átomos, volumen y temperatura constante, existen variados grupos que pueden ser utilizados de acuerdo a lo que uno desee estudiar como los son grupos del tipo NPT (número de átomos, presión y temperatura constante), NVE (número de átomos, volumen y energía total constante), NPE (número de átomos, presión y energía total constante) entre otros. En mecánica estadística el grupo termodinámico viene definido como el conjunto de microestados que representan un mismo macroestado termodinámico, por lo que una trayectoria obtenida por una simulación de DM provee este conjunto de configuraciones (microestados). En el límite de simulaciones muy largas, se puede esperar que el espacio fase fuese completamente muestreado y en ese límite el promedio aritmético de alguna cantidad física va a llevar a la obtención de propiedades termodinámicas. En la práctica la duración de una simulación es finita, por lo que hay que tener cuidado para estimar cuando el muestreo es suficiente. De este modo las simulaciones de DM pueden ser usadas para medir propiedades termodinámicas de un determinado sistema.

¿Que se debería hacer en los bordes del sistema simulado? Una posibilidad es no hacer nada en especial: el sistema simplemente termina, y los átomos cercanos al borde o frontera tendrán menos vecinos que los átomos del interior. En otras palabras, se generaran fenómenos de superficie en el sistema. Un ejemplo de este tipo de condiciones de frontera es la condición esférica de borde, en la cual las moléculas que están en los límites del sistema son sometidas a un potencial armónico manteniendo constante el volumen del sistema. A menos que se quiera simular solo un conjunto de átomos, esta situación no es representativa de una condición real y por lo tanto los efectos de superficie deben ser eliminados. En la práctica se usa las condiciones periódicas de borde, que consisten en una replicación hacia el infinito del grupo finito de partículas confinadas en una caja, normalmente paralelepípeda, en las tres direcciones del espacio cartesiano. El punto clave es que cada partícula *i* del sistema no solo interactúa con otra partícula *j* de la caja sino que también con sus imágenes en las cajas cercanas, implicando que las interacciones pueden ir a través de los límites de las cajas, por lo que virtualmente se han eliminado los efectos de superficie del sistema.

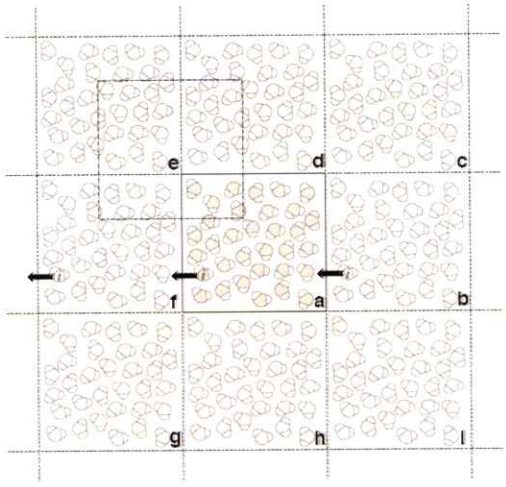


Figura 26: imagen bidimensional de condiciones periódicas de borde(pbc) figura extraída de Chipot , 2000.

La naturaleza pseudo-infinita del sistema implica la necesidad de aproximaciones para las interacciones moleculares. En particular la aproximación de la "imagen mínima" supone que cada partícula *i* de la caja central interactúe con la imagen más cercana de todas las partículas *j*. Además la introducción de una esfera de corte (mejor conocida como función de *cut-off*), se utiliza para ignorar interacciones con átomos que estén más lejos que una distancia arbitraria, igual o menor a la mitad de la dimensión de la caja periódica.

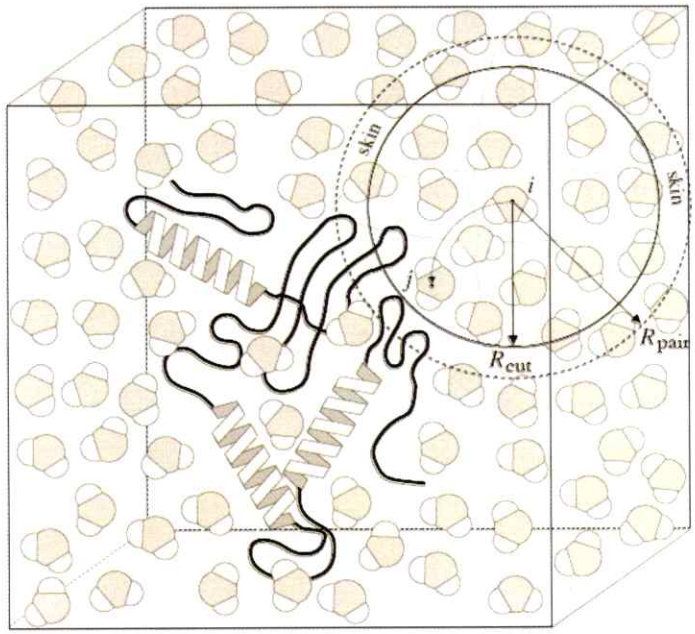


Figura 27: Esfera de corte (función de *cut-off*) para las interacciones de largo alcance, figura extraída de Chipot 2000.

Claramente la validez de estas aproximaciones está condicionada al alcanze de las interacciones moleculares consideradas. En particular para las interacciones electrostáticas de largo alcance las funcione de *cut-off* no son adecuadas, ya que este tipo de interacciones no pueden ser truncadas de una manera esférica. Para superar esta dificultad se utilizan métodos adaptados, que están basados sobre sumas de red (Ewald,

Kornfeld o Ladd), que consisten en una evaluación de las interacciones de una partícula con todas las otras contenidas en la caja central, así como en todas las cajas fantasmas replicadas. Adoptar este tipo de enfoque, sin embargo, aumenta considerablemente el costo computacional, pero es indispensable para una descripción rigurosa y correcta de las interacciones de largo alcance.

El motor de un programa de DM es el algoritmo de integración numérica del tiempo, requerido para integrar las ecuaciones de movimiento de las partículas que interactúan en un determinado sistema y seguir sus trayectorias. Los algoritmos de integración numérica están basados en métodos diferenciales finitos, donde el tiempo es dividido en intervalos pequeños, siendo el paso de tiempo (Δt) la distancia entre puntos consecutivos de la grilla. Sabiendo las posiciones y sus derivadas en el tiempo en un tiempo t (los detalles exactos dependen del tipo de algoritmo), el esquema de integración permite obtener los mismo valores a un tiempo $t + \Delta t$. Iterando este proceso, la evolución en el tiempo del sistema puede ser seguida por largos tiempos. Obviamente estos esquemas son aproximados por lo que hay errores asociados con ellos. En particular, se puede distinguir entre dos tipos de errores:

- Errores de truncado: relacionados con la precisión de método diferencial finito. Estos métodos están usualmente basados en expansiones de Taylor truncadas en algún término, por lo que estos errores no dependen de la implementación: son intrínsecos del algoritmo.

- Errores de redondeo: errores relacionados a la implementación particular de un algoritmo.Por ejemplo el número finito de dígitos usados en la aritmética computacional.

Ambos errores pueden ser reducidos disminuyendo el Δt. Para grandes Δt , los errores de truncado predominan, pero disminuyen rápidamente con la reducción en la magnitud del Δt. Los errores de redondeo decrecen más lentamente con la disminución del Δt y dominan en el límite de los Δt pequeños.

En dinámica molecular las fuerzas son derivadas de una función de energía potencial (V o mejor conocida como campo de fuerza) la cual depende las coordenadas de las partículas. Entonces el problemas de modelar un determinado material viene dado por la obtención de esta función de energía potencial para ese material. ¿Para sistemas moleculares puede existir tal función, como uno puede usar las leyes de Newton para mover átomos, sabiendo que para ese nivel los átomos obedecen a leyes cuánticas ? El comportamiento de moléculas y átomos está controlado por la leyes de la mecánica cuántica más que de la mecánica clásica, por los que lo electrones juegan un papel fundamental en determinar las propiedades de enlace del sistema. Lo anterior implica que inequívocamente habrá que usar con la ecuación de Schrödinger para determinar el valor de la función de onda, lo cual en la práctica es imposible de implementar por lo que aproximaciones han sido desarrolladas. En 1923 Born y Oppenheimer se dieron cuenta que los núcleos son muchos más pesados que lo electrones y que se mueven en escalas de tiempo de dos órdenes de magnitud mayores, por lo que se puede disociar el movimiento de núcleos y electrones. Además considerando que en la mayoría de los casos la longitud de onda termal de de Broglie es mucho más corta que las distancias intermoleculares típicas, los efectos cuánticos pueden ser descartados con seguridad. En definitiva la aproximación mecánica describe a los átomos como esferas (desligándose de los electrones) con carga, en donde los enlaces son simulados como un potencial armónico (en otras palabras un resorte).

Un campo de fuerza (*V*, CF) de mecánica molecular, se define como un operador matemático que analíticamente describe la energía potencial del sistema en términos de las geometrías de los centros atómicos (hamiltoniano del sistema). Los parámetro del CF (distancia ideal de enlace, constante de fuerza del enlace,entre otros) son derivados empíricamente de cálculos *ab initio* de mecánica cuántica, espectroscopia y cristalografía. En la figura 28 se observa la forma funcional de un campo de fuerza con los gráficos idealizados para cada ecuación que lo compone y sus aproximaciones mecánicas idealizadas. (figura 28).

1 2

$$\vec{F}_i = -\nabla V_{\vec{r}_i}(\vec{r}_1, ..., \vec{r}_n)$$

Enlaces

$$V = \sum_i K_{b,i}(b_i - b_{0,i})^2$$

Ángulos de enlace

$$+ \sum_i K_{\theta,i}(\theta_i - \theta_{0,i})^2$$

Torsiones Angulares

$$+ \sum_i K_{\varphi,i}\{1 - \cos[n_i(\varphi_i - \varphi_{0,i})]\}$$

Van der Waals

$$+ \sum_{\text{pairs } i,j} \left[\varepsilon_{ij} \left(\frac{r_{0,ij}}{r_{ij}} \right)^{12} - 2\varepsilon_{ij} \left(\frac{r_{0,ij}}{r_{ij}} \right)^{6} \right]$$

Coulomb

$$+ \frac{1}{4\pi\epsilon_0\epsilon} \sum \frac{q_i q_j}{r_{ij}}$$

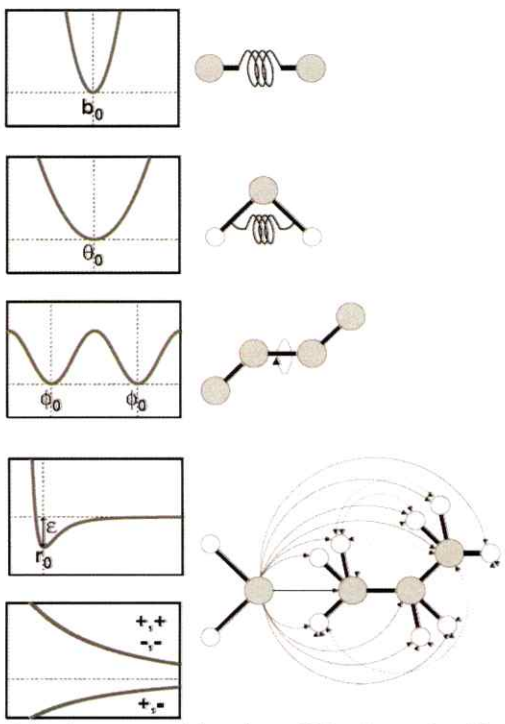


Figura 28: Forma general de una campo de fuerza; columna 1 funciones matemáticas de un CF; columna 2 gráficos idealizados para las funciones de un CF y la aproximación de mecánica clásica de los sistemas simulados.
F_i: fuerza que actúa sobre el átomo *i*; V_{ri}: campo de fuerza o función de energía potencial; r_i: posición en el espacio cartesiano del átomo i; k_b: constante de fuerza para el enlace covalente; b_0: longitud de mínima energía para el enlace covalente; b_i: longitud del enlace covalente; k_θ: constante de fuerza para el ángulo de enlace; θ_0: ángulo de enlace de mínima energía; θ_i: ángulo de enlace; k_φ:constante de barrera de rotación; φ_0: ángulo dihedro de mínima energía; *n*: período o fase del rotor; φ: ángulo dihedro; ε_{ij}: mínimo de energía entre dos átomos para la interacción de Van der Waals; $r_{0,ij}$: distancia de mínima energía entre los átomos i.j para la interacción de Van der Waals; $r_{,ij}$: distancia entre los átomos i.j; ε_0: permisividad del espacio libre; *q*: carga parcial atómica.
Figura extraída y modificada de Chipot, 2000 y Beck & Dagget, 2004.

La energía del enlace covalente es tratada como un oscilador armónico con un mínimo de energía en b_0 y una constante de fuerza k_b. Los ángulos de enlace son tratados de manera similar con un ángulo ideal θ_0 y una constante de fuerza k_θ. El tercer término, usado para torsiones angulares dihedras o fuera del plano, es representado por un coseno con n períodos con una energía mínima en φ_0 y una constante de barrera de rotación $k\varphi$. La interacción de la energía de Van der Waals del par atómico i y j es tratado con un función Lennard-Jones 12/6. Las interacciones electrostáticas del par atómico i y j con cargas parciales q_i y q_j respectivamente separados por una distancia $r_{i,j}$ es expresado con un potencial de Coulomb.

En resumen la dinámica molecular es una herramienta que permite explorar a nivel atómico el espacio fase de sistemas moleculares, permitiendo el cálculo de propiedades termodinámicas. A diferencia de otras técnicas de simulación como lo son montecarlo o la mecánica molecular también se pueden obtener propiedades que dependen directamente de la evolución del tiempo como lo son cálculos de radio de giro y propiedades de correlación espacio-tiempo. A pesar del poder de la DM hay que tener claros sus limitaciones, al utilizar una aproximación de mecánica clásica, para todo sistema en donde los efectos cuánticos sean preponderantes la utilización de herramientas clásicas de DM no va a ser la más adecuada un ejemplo de esto es la simulación de sistemas lo suficientemente pequeños (y livianos) como hidrógeno helio o neón.

X.2 Ecuaciones

•RMSD

Es una sigla en inglés que significa desviación cuadrada de las coordenadas medias (*Root Mean Square Deviation)*. Es usualmente usado en geometría 3D de moléculas para comparar dos conformaciones de un determinado conjuntos de puntos, típicamente átomos. En otras palabras dada una lista de átomos pareados, da la medida de la distancia entre estos puntos.

Normalmente una superposición rígida que minimiza el RMSD es realizada , y este mínimo es el valor calculado. Para dos conjuntos de n puntos v y w el RMSD se define como:

$$RMSD(v,w) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \parallel v_i - w_i \parallel^2}$$

$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n} (v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2}$$

Un valor de RMSD es expresado en medidas de distancia. La unidad que normalmente se usa en biología estructural es el angstrom, que equivale a 10^{-10} m.

•Algoritmo de integración

En dinámica molecular, el algoritmo de integración comúnmente usado es el llamado algoritmo de Verlet. La idea básica es escribir dos expansiones de Taylor para las posiciones *x(t)* (3 valores de *x* cada átomo) una hacia adelante y otra hacia atrás en el tiempo. Llamando *v* las velocidades, *a* a las aceleraciones, *b* a las derivadas de tercer orden de *x* con respecto a *t* y *O* error de truncado se tiene que:

$$x(t+\Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \frac{1}{6}b(t)\Delta t^3 + O(\Delta t^4)$$
$$x(t-\Delta t) = x(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 - \frac{1}{6}b(t)\Delta t^3 + O(\Delta t^4)$$

Sumando estas dos expresiones da:

$$x(t+\Delta t) = 2x(t) - x(t-\Delta t) + a(t)\Delta t^2 + O(\Delta t^4)$$

Esta es la forma básica del algoritmo de Verlet. Como se están integrando las ecuaciones de Newton, *a(t)* es la fuerza dividida por la masa *m*, y la fuerza *f* es a su vez función de la posición *x(t)*:

$$a(t) = -\frac{1}{m}\left(\frac{\partial V}{\partial x}\right)$$

Como se puede ver, el error de truncado del algoritmo *O*, cuando el sistema evoluciona por cada Δt, está en el orden de Δt^4 incluso si la derivadas de tercer orden no aparecen de forma explícita.

•Función de densidad de probabilidad (PDF)

En matemáticas, una función de densidad de probabilidad (pdf) es una función que representa una distribución de probabilidades en términos de integrales. Formalmente una distribución de probabilidad posee una densidad f si f es una función integrable no negativa de $\mathbb{R} \rightarrow \mathbb{R}$ tal que la probabilidad del intervalo *[a,b]* está dada por:

$$\int_a^b f(x)dx$$

para dos valores cualesquiera *a,b*. Esto implica que la integral total para f debe ser 1. Inversamente cualquier función integrable no negativa con valor integral total a 1, es la densidad de probabilidad de una distribución de probabilidades. Una pdf puede entenderse como una versión suavizada de un histograma: si se muestrean empíricamente suficientes valores de una variable aleatoria continua, generando un histograma, entonces este histograma se va a parecer a la densidad de probabilidad de la variable aleatoria. En términos simples una pdf es cualquier función *f(x)* que describe la densidad de probabilidad en términos de una variable de entrada *x* de la siguiente manera:

f(x) ≥ 0 para todos los valores de *x*.

El área total bajo el gráfico es igual a 1: $\int_{-\infty}^{\infty} f(x)dx = 1$

La probabilidad finita puede se calculada calculando la integral de la función *f(x)* en el intervalo de la variable de entrada *x*. Por ejemplo la probabilidad de la variable X dentro del intervalo [1,2] va a ser: $Prob(1 \leq X \leq 2) = \int_1^2 f(x)dx$

X.3 Lenguajes de programación.

Un lenguaje de programación es una interfaz entre el usuario y la máquina. Existen dos niveles de lenguajes, los llamados lenguajes de máquina (*o asembler*) característicos de cada tipo de máquina (*hardware*) y por lo tanto muy difíciles de interpretar y los llamados lenguajes de programación de alto nivel que son una forma de especificar una secuencia de operaciones para realizar una tarea, independiente del tipo de máquina en el que se esté trabajando, una linea en programación de alto nivel puede realizar operaciones complejas, correspondientes a una infinidad de operaciones a nivel de máquina.

El compilador es una herramienta para convertir un programa escrito en lenguaje de alto nivel en una secuencia de instrucciones en lenguaje de máquina requeridos por un computador en particular para completar la tarea requerida.

En la actualidad existen variados lenguajes de programación de alto nivel que son específicos para el tipo de tarea que se quiera realizar, por ejemplo lenguajes como Fortran traducción de fórmula *(formula translation)* o C son ideales para cálculo numérico, Perl lenguaje práctico de extracción y reporte *(practical extraction and report language)* esta enfocado al manejo de texto. Otros lenguajes como Java y Tk están enfocados al desarrollo de programas gráficos multiplataforma.

Para este trabajo se utilizaron dos lenguaje de programación, la automatización de procesos y manejo de textos se programó en Perl; todo el análisis de las simulaciones fue desarrollado con el lenguaje TCL, lenguaje de herramientas de comandos *(tool comand language)* ya que es el código presente en el programa de análisis molecular VMD. Es

importante destacar que ambos lenguajes son interpretados, es decir que no necesitan previa compilación y generación de un binario para su utilización, solo requieren ser ejecutados. A continuación, a modo de ejemplo se presentan algunos programas *(scripts)* desarrollados en este trabajo, el conjunto completo de programas desarrollados se depositaron en el anexo digital 1.

```
set molname ubq

mol new ${molname}.psf
mol addfile ${molname}.pdb

set cen [measure center [atomselect top all] weight mass]
set x1 [lindex $cen 0]
set y1 [lindex $cen 1]
set z1 [lindex $cen 2]
set max 0

foreach atom [[atomselect top all] get index] {
  set pos [lindex [[atomselect top "index $atom"] get {x y z}] 0]
  set x2 [lindex $pos 0]
  set y2 [lindex $pos 1]
  set z2 [lindex $pos 2]
  set dist [expr pow(($x2-$x1)*($x2-$x1) + ($y2-$y1)*($y2-$y1) + ($z2-$z1)*($z2-$z1),0.5)]

    if {$dist > $max} {set max $dist}
  }
set max $max*1.5
set radio $max
mol delete top


package require solvate
solvate ${molname}.psf ${molname}.pdb -t 40 -o del_water

resetpsf
package require psfgen
mol new del_water.psf
mol addfile del_water.pdb
readpsf del_water.psf
coordpdb del_water.pdb

set wat [atomselect top "same residue as {water and ((x-$x1)*(x-$x1) + (y-$y1)*(y-$y1) + (z-$z1)*(z-$z1))<($max*$max)}"]
set del [atomselect top "water and not same residue as {water and ((x-$x1)*(x-$x1) + (y-$y1)*(y-$y1) + (z-$z1)*(z-$z1))<($max*$max)}"]
set seg [$del get segid]
set res [$del get resid]
set name [$del get name]
for {set i 0} {$i < [llength $seg]} {incr i} {
  delatom [lindex $seg $i] [lindex $res $i] [lindex $name $i]
  }
writepsf ${molname}_ws.psf
writepdb ${molname}_ws.pdb

mol delete top

mol new ${molname}_ws.psf
mol addfile ${molname}_ws.pdb
set out [open ${molname}_CMRS.txt w]
puts -nonewline $out "CENTER OF MASS OF SPHERE IS: [measure center [atomselect top all] weight mass]\n"
puts -nonewline $out "RADIUS OF SPHERE IS: $max"
mol delete top
```

Programa 1: Script de TCL que inserta una proteínas en una esfera de agua.

```
#!/usr/bin/perl

if ($ARGV[0] eq ""){
    print "Usage of this script: [/path/to/alineamientos/]\n";
    exit 1;
}

$ARGV[0]=~s/\/$//;
@alineamientos=<$ARGV[0]/*.pir>;

foreach $pir (@alineamientos){
    print "$pir\n";
    open (PIR,"$pir");
    $pir_name=$pir;
    $pir_name=~s/.*\///;
    open (OUT,">$pir_name");
    $flag=1;
    while ($linea=<PIR>){
    if (($linea=~/^\>/)||($linea=~/^s/)){
        print OUT $linea;
    }


    if ($linea=~/^[A-Z\-]/){
        $largo=length($linea);
        $alea=int(rand($largo));
        @sec=split(//,$linea);
        undef $salida;
        for ($i=0; $i<=$#sec; $i++){
        if ($i==$alea){
        $salida="$salida\-";
        $salida="$salida$sec[$i]";
        } else {
            $salida="$salida$sec[$i]";
        }
        }
        print OUT $salida;
    }
    if ($linea=~/^\*/){
        print OUT $linea;
    }
    }
    close OUT;

}
```

Programa 2: Script de Perl que inserta gaps al azar en los alineamientos.

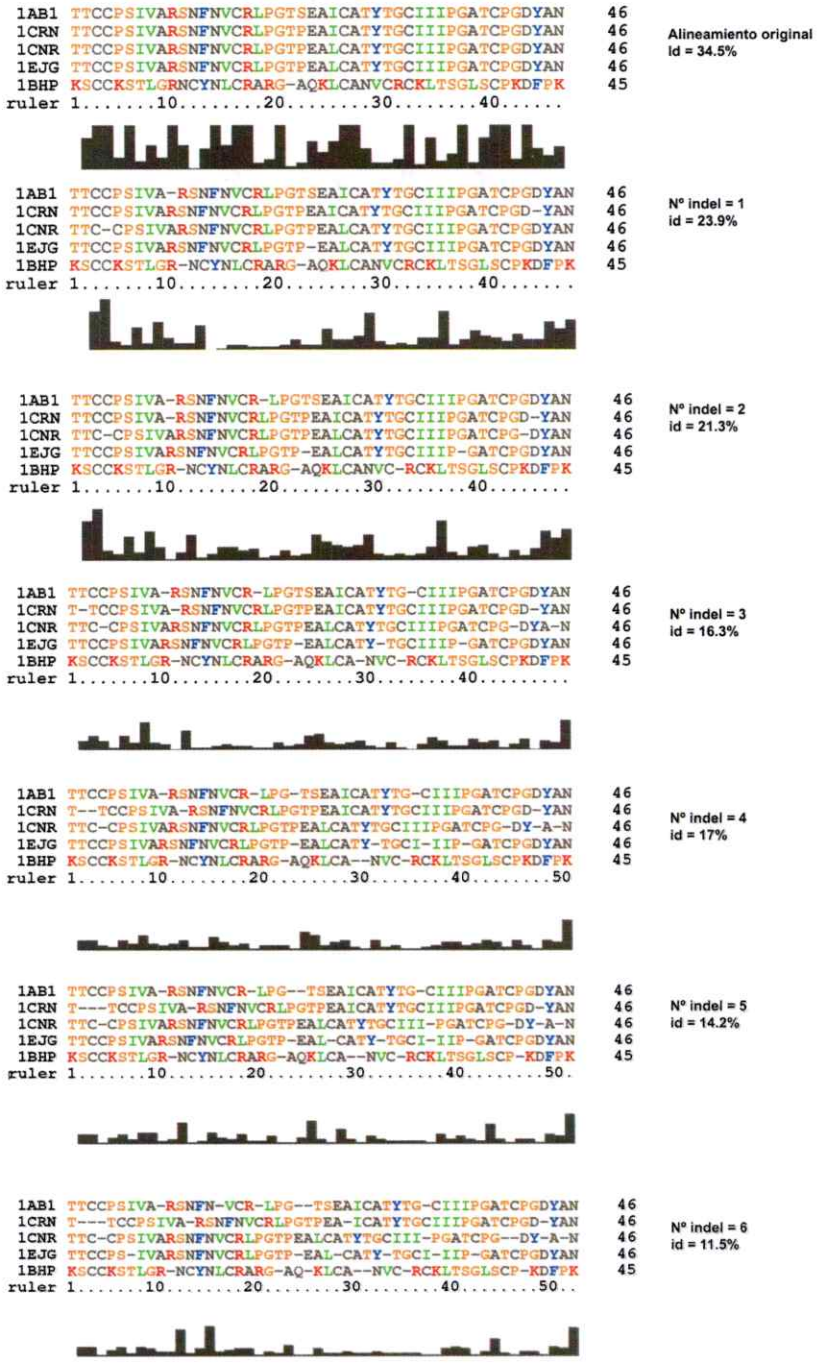


Figura 29: Alineamientos múltiples de 1BHP y sus patrones no intervenidos e intervenidos con inserciones de indel al azar vía programación en Perl. Se destaca que la inserción de indels es tanto para la secuencia objetivo como para las secuencia patrones. Alineamiento original calculado con ClustalW.
Nº indel: número de inserciones al azar de gaps.
Id: porcentaje de identidad de secuencia promedio de cada patrón con respecto a 1BHP.

X.4 Presentación en congreso.

Este trabajo fue presentando en la sección de Paneles I de la *XXIX Reunión Anual de la Sociedad de Bioquímica y Biología Molecular de Chile*, resumen publicado en la revista *Biological Research volumen 39 (suplemento B) página R-77, 2006.*

12. EVALUACIÓN DE MODELOS COMPARATIVOS USANDO SIMULACIONES DE DINÁMICA MOLECULAR. (Assesment of comparative models using molecular dynamics simulations).

Gárate, J. A. & Perez-Acle T.
Centro de Genómica y Bioinformática (CGB), Facultad de Ciencias Biológicas, P. Universidad Católica de Chile.

El modelamiento molecular de proteínas por métodos comparativos requiere de la existencia de herramientas evaluadoras de la calidad de los modelos generados. Múltiples herramientas actualmente disponibles permiten la evaluación de estos modelos, donde la certeza de la evaluación depende directamente de la identidad de secuencia. Datos provenientes de nuestro laboratorio demuestran que modelos generados con baja identidad de secuencia, son estables bajo protocolos de dinámica molecular, a menos que se produzcan errores durante el alineamiento múltiple. Con el objetivo de generar una metodología dinámica de evaluación de modelos comparativos se generó una base de datos (BD) de estructuras cristalográficas provenientes de PDB. Para toda la BD se realizaron búsquedas de patrones usando BLAST-P, siendo seleccionados aquellos cuya identidad de secuencia se ubicase entre 30% y 50%. Usando estos patrones, se generaron una serie de modelos comparativos correctos e incorrectos. Estos últimos fueron producidos generando aperturas aleatorias en los alineamientos. Los modelos resultantes fueron evaluados por diversas herramientas, siendo sometidos a un protocolo de Dinámica Molecular en solvente explícito y condiciones esféricas de borde. Los modelos generados en forma errónea presentaron altos RMSD (> 3Å) en relación al los cristales de referencia, y sus evaluaciones decayeron significativamente respecto de las iniciales, demostrando el poder de la evaluación dinámica.

la concentración local de K^+ (Brelidze y col., 2003 y Haug y col., 2004). Se postula que la región extracelular de h*Slo* modula la concentración local de iones K^+. En este sentido se analizaron las mutaciones D326N, E329Q y D326NE329Q de la región extracelular del canal h*Slo* a través del cálculo de perfil de energía libre de unión (PMF) y de la densidad local de iones K^+. Los resultados del cálculo de PMF y de mediciones electrofisiológicas de conductancia muestran que la mayor contribución electrostática está dada por el residuo D326, mientras que el residuo E329 no tiene ningún efecto en la conductancia ni en las propiedades estructurales y energéticas del canal. De los resultados de la dinámica molecular se puede concluir que el residuo D326 modula la concentración local de iones K^+ en la región extracelular del canal.

Agradecimientos: González-Nilo F. Fondecyt # 1040254. Latorre R. Fondecyt # 1030830

14. CAVIDAD INTRACELULAR DEL CANAL H*SLO*: ANÁLISIS DEL PERFIL DE ENERGÍA LIBRE DE UNIÓN DEL K^+ MEDIANTE SMD. (Intracellular h*Slo* channels vestibule: Analysis of the binding free energy profile of the potassium ion with SMD).

Vidal, M. A.[*,b], Saavedra G.[*,b], Urbina H.[c], González W.[c], González-Nilo F.[c], Carvacho I.[&,§] y Latorre R.[&]
[*]Centro de Bioinformática, Universidad de Talca, 2 Norte 685, Talca, [&]Centro de Estudios Científicos, Av. Arturo 514, Valdivia y [§]Universidad Austral de Chile, Valdivia.

La metodología de Dinámica Molecular Dirigida (Steered Molecular Dynamics, SMD) aplica vectores de fuerzas externas dependientes del tiempo a uno o más átomos en un sistema molecular. Simulaciones con SMD permiten obtener el perfil de energía libre de un ión a lo largo de una coordenada de reacción. El canal de potasio humano

XI Bibliografía

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-402.

Baldi, P and Brunak, S, 2001 Bioinformatics: The Machine Learning Approach, 2nd edition. MIT Press.

Baker, D. and A. Sali (2001). "Protein structure prediction and structural genomics." Science **294**(5540): 93-6.

Beck, D. A. and V. Daggett (2004). "Methods for molecular dynamics simulations of protein folding/unfolding in solution." Methods **34**(1): 112-20.

Braxenthaler, M., R. Unger, et al. (1997). "Chaos in protein dynamics." Proteins **29**(4): 417-25.

Chipot C. (2000). "Molecular Dynamics" Galerne summer school (Le Cap d'Agde), 2002, CNRS summer school (Bordeaux).

Chotia C. and Lesk AM (1986). " The relation between the divergence of sequence and structure in proteins" The EMBO Journal vol.5 no.4 pp.823-826.

Darden T, Perera L, Li L and Pedersen L. (1999). "New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations", Structure 7, R55-R60.

Dayhoff, Eck, Chang, and Sochard, (1965) Atlas of Protein Sequence and Structure, 1965.

Drenth, J (1999). "Principles of Protein X-Ray Crystallography". New York: Springer-Verlag.

Ercolessi F. (1997) "A molecular dynamics primer". Spring College in Computational Physics, ICTP, Trieste Italy.

Fan, H. and A. E. Mark (2003). "Relative stability of protein structures determined by X-ray crystallography or NMR spectroscopy: a molecular dynamics simulation study." Proteins **53**(1): 111-20.

Fan, H. and A. E. Mark (2004). "Refinement of homology-based protein structures by molecular dynamics simulation techniques." Protein Sci **13**(1): 211-20.

Fiser, A., R. K. Do, et al. (2000). "Modeling of loops in protein structures." Protein Sci **9**(9): 1753-73.

Fiser, A., M. Feig, et al. (2002). "Evolution and physics in comparative protein structure modeling Acc Chem Res **35**(6): 413-21.

Fiser, A. and A. Sali (2003). "Modeller: generation and refinement of homology-based protein structure models." Methods Enzymol **374**: 461-91.

Fujii, T., Hayashida, M., Hamasu, M., Ishiguro, M., Hata, Y. (2004). "Structures of two lectins from the roots of pokeweed (Phytolacca americana)". Acta Crystallogr., Sect.D v60 pp.665-673 , 2004

Henikoff, S. and J. G. Henikoff (1992). "Amino acid substitution matrices from protein blocks." Proc Natl Acad Sci U S A **89**(22): 10915-9.

Housconjunto, D., Haberconjuntozer-Rochat, C., Astier, J.P., Fontecilla-Camps, J.C. (1994) "Crystal structure of toxin II from the scorpion Androctonus australis Hector refined at 1.3 A resolution". J.Mol.Biol. v238 pp.88-103 .

Humphrey, W., A. Dalke, et al. (1996). "VMD: visual molecular dynamics." J Mol Graph **14**(1): 33-8, 27-8.

Karplus, M. and J. A. McCammon (2002). "Molecular dynamics simulations of biomolecules." Nat Struct Biol **9**(9): 646-52.

Laskowski R A, MacArthur M W, Moss D S & Thornton J M (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J. Appl. Cryst., 26, 283-291.

Larrondo, L., A. Gonzalez, et al. (2005). "The nop gene from Phanerochaete chrysosporium encodes a peroxidase with novel structural features." Biophys Chem **116**(2): 167-73.

Law, R. J. and M. S. Sansom (2004). "Homology modelling and molecular dynamics simulations: comparative studies of human aquaporin-1." Eur Biophys J **33**(6): 477-89.

Law, R. J., C. Capener, et al. (2005). "Membrane protein structure quality in molecular dynamics simulation." J Mol Graph Model **24**(2): 157-65.

Leonard, N., C. Lambert, et al. (2004). "Modeling of human monoamine oxidase A: from low resolution threading models to accurate comparative models based on crystal structures." Neuro toxicology **25**(1-2): 47-61.

Luthy, R., J. U. Bowie, et al. (1992). "Assessment of protein models with three-dimensional profiles." Nature **356**(6364): 83-5.

Lyapunov, A (1892). "Общая задача об устойчивости движения, A general task about the stability of motion" Ph.D. thesis.

MacKerell, A.D.,Brooks, Brooks C.L., III, Nilsson L., Roux B., Won Y., and Karplus M. (1989). "CHARMM: The Energy Function and Its Parameterization with an Overview of the Program, in The Encyclopedia of Computational Chemistry" 1, 271-277, P. v. R. Schleyer et al.,editors John Wiley & Sons: Chichester.

Melo, F., D. Devos, et al. (1997). "ANOLEA: a www server to assess protein structures." Proc Int Conf Intell Syst Mol Biol **5**: 187-90.

Melo, F., R. Sanchez, et al. (2002). "Statistical potentials for fold assessment." Protein Sci **11**(2): 430-48.

Nickitenko, A.V., Mikhailov, A.M., Betzel, C. (1995). "The Crystal Structure of Neurotoxin-I from Naja Naja Oxiana at 1.9 Angstroms Resolution". To be Published

Norberg, J. and L. Nilsson (2003). "Advances in biomolecular simulations: methodology and recent applications." Q Rev Biophys **36**(3): 257-306.

Ott, Edward (2002). "Chaos in Dynamical Systems. Cambridge". University Press New, York.

Pieper, U., G. Kapadia, et al. (1995). "Structural evidence for the evolutionary divergence of mycoplasma from gram-positive bacteria: the histidine-containing phosphocarrier protein." Structure **3**(8): 781-90.

Phillips, J. C., R. Braun, et al. (2005). "Scalable molecular dynamics with NAMD." J Comput Chem **26**(16): 1781-802.

Ponder JW and Case DA. (2003) "Force fields for protein simulations". Adv. Prot. Chem. 66: 27-85.

Post, C. B., C. M. Dobson, et al. (1989). "A molecular dynamics analysis of protein structural elements." Proteins **5**(4): 337-54.

Russell, R. B. and G. J. Barton (1992). "Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels." Proteins **14**(2): 309-23.

Ryckaert J.P, Ciccotti G, Berendsen HJC (1977) "Numerical integration of the Cartesian equations of motion of a system with constraints: molecular". J. Comput. Phys 23, 327.

Sali, A. and T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints." J Mol Biol **234**(3): 779-815.

Sanchez, R., U. Pieper, et al. (2000). "Protein structure modeling for structural genomics." Nat Struct Biol **7 Suppl**: 986-90.

Stec, B., Rao, U., Teeter, M.M. (1995) "Refinement of Purothionins Reveals Solute Particles Important for Lattice Formation and Toxicity. Part 2: Structure of Beta-Purothionin at 1.7 Angstroms Resolution" Acta Crystallogr.,Sect.D v51 pp.914 , 1995

Sundman, K. E. (1912)." Memoire sur le probleme de trois corps". Acta Mathematica 36 : 105–179.

Sussman, J. L., D. Lin, et al. (1998). "Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules." Acta Crystallogr D Biol Crystallogr **54**(Pt 6 Pt 1): 1078-84.

Tischler, N. D., A. Gonzalez, et al. (2005). "Hantavirus Gc glycoprotein: evidence for a class II fusion protein." J Gen Virol **86**(Pt 11): 2937-47.

Thompson, J. D., D. G. Higgins, et al. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-80.

Vijay-Kumar, S., C. E. Bugg, et al. (1987). "Structure of ubiquitin refined at 1.8 A resolution. " J Mol Biol **194**(3): 531-44.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-51.

Wall, L (1987). www.perl.org

Wuthrich, K. (2001). "The way to NMR structures of proteins." Nat Struct Biol **8**(11): 923-5.