

2005
Lozano Aguirre Beltrán, Luis Fernando

00551

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO



INSTITUTO DE BIOTECNOLOGÍA

**“Estabilidad Evolutiva de Estructuras Secundarias
en Proteínas Homólogas”**

T E S I S
QUE PARA OBTENER EL GRADO DE:
MAESTRO EN CIENCIAS BIOQUIMICAS
P R E S E N T A

Biol. Luis Fernando Lozano Aguirre Beltrán

TUTOR
DR. ALEJANDRO GARCIARRUBIO GRANADOS

Cuernavaca, Mor.

2005



m345188



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

La presente tesis se realizó en el Departamento de Ingeniería Celular y Biocatálisis del Instituto de Biotecnología de la Universidad Nacional Autónoma de México bajo la dirección del Dr. Alejandro Garcíarrubio Granados y con el apoyo del Consejo Nacional de Ciencia y Tecnología.

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

NOMBRE: LUIS FERNANDO
LOZANO AGUIRRE BELTRAN

FECHA: 7 JUNIO 2005

FIRMA: HA

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

CONTENIDO

Contenido.....	1
Resumen.....	2
Introducción.....	3
Características principales de las estructuras secundarias.....	4
Estabilidad de las estructuras secundarias.....	5
Predicción de estructuras secundarias.....	7
Antecedentes.....	8
Objetivos.....	10
Métodos.....	10
Alineamientos estructurales.....	10
Comparación de estructuras secundarias.....	11
Análisis de las regiones terminales de las hélices- α y las hojas- β	12
a) Patrones de hidropatía.....	12
b) Conservación de secuencia.....	14
c) Patrones de capping.....	15
Predicción de estructuras secundarias.....	16
Resultados.....	16
Comparación de estructuras secundarias.....	15
Análisis de las regiones terminales de las hélices- α y las hojas- β	20
a) Patrones de hidropatía.....	21
b) Conservación de secuencia.....	25
c) Patrones de capping.....	29
Predicción de estructuras secundarias.....	32
Discusión.....	33
Comparación de estructuras secundarias.....	33
Características de las regiones terminales de las hélices- α y las hojas- β	35
I. Estabilidad de las regiones terminales.....	35
II. Diferencias entre regiones terminales.....	35
III. Patrones de hidropatía en estructuras secundarias.....	37
IV. Predicción y estabilidad de regiones terminales.....	39
Modificación de tamaño de hélices- α y hojas- β	40
Conclusión.....	42
Referencias.....	43

RESUMEN

Las hélices- α y las hojas- β son estructuras ordenadas y repetitivas de la cadena principal de las proteínas, las cuales se mantienen estables debido principalmente a interacciones con residuos de aminoácidos (a.a.) cercanos en la secuencia. Las interacciones cercanas evitan que ambas estructuras secundarias (ESs) se desplieguen por sus regiones amino o carboxilo terminal. Aun así los cambios en la secuencia pueden afectar la estabilidad estructural de las ESs provocando que presenten cambios que afectan a la estructura tridimensional o al plegamiento de la proteína.

En el presente trabajo se investigaron las diferencias entre las hélices- α y las hojas- β en relación con el cambio de tamaño y de secuencia en proteínas homólogas. Posteriormente el análisis se centró en las regiones terminales de estas estructuras. Dichas regiones terminales comprenden los últimos a.a. que forman parte de la ES y los primeros de forman parte del "coil". En estas posiciones se analizó la conservación de secuencia, los patrones hidropáticos (a.a. polares y no polares), la accesibilidad al solvente y en el caso específico de las hélices- α , se analizaron los patrones de "capping" descritos con anterioridad. Para estos análisis se seleccionaron varios alineamientos de secuencia para distintas familias de proteínas, que contienen información estructural codificada. Los alineamientos se obtuvieron de 2 bases de datos, Homstrad y PASS2. Finalmente se realizaron predicciones de ESs mediante el programa PsiPred para analizar el efecto de las características de cada región terminal en el nivel de predicción.

Se halló que después de un mismo tiempo de divergencia, las hélices- α cambian más de tamaño que las hojas- β . Relacionado a esto, se encontró que en general las hojas- β mantienen una mayor conservación de secuencia que las hélices- α después de un mismo tiempo de divergencia. En las regiones terminales de las ESs se hallaron claras diferencias entre las hélices- α y las hojas- β . Las hojas- β presentaron una mayor conservación de secuencia y de uso de patrones hidropáticos en sus regiones terminales. Para el caso específico de las hélices- α se hallaron claras diferencias entre sus dos regiones terminales (amino y carboxilo), estos resultados se comparan con trabajos teóricos que analizan la tendencia a desplegarse de las regiones terminales de las hélices- α . Finalmente las predicciones de ES mostraron diferencias en el nivel de predicción en las distintas regiones terminales siendo más marcada en las hélices- α , donde la región amino terminal tiene niveles de predicción más altos que cualquier otra región terminal, aun cuando es la región terminal menos conservada.

Todos estos resultados muestran que las hojas- β presentan una mayor estabilidad evolutiva en sus regiones terminales por estar más conservadas que las hélices- α . Así mismo, la mayor estabilidad de la región carboxilo terminal de las hélices- α se contrapone a su mayor tendencia a desplegarse. El mejor nivel de predicción en la región amino terminal de las hélices- α resulta controversial debido su bajo nivel de conservación. Por otro lado, los resultados sugieren que esta región podría tener una mayor señal evolutiva en ciertas posiciones, lo cual aprovecha el método de predicción de PsiPred.

INTRODUCCIÓN

Las proteínas son moléculas biológicas importantes capaces de desarrollar distintas funciones en los organismos vivos. Se postula que estas funciones se pueden llevar a cabo sólo si las proteínas tienen una estructura tridimensional (3D) específica. Hoy en día se conocen un gran número de estructuras de proteínas que han permitido realizar diversos estudios y generar distintas clasificaciones estructurales. Muchos de estos estudios han analizado la modificación de las estructuras 3D con el fin de entender cómo se ha generado la diversidad estructural de las proteínas ^(1, 2, 3, 4). Desde los primeros análisis resultó evidente que la formación de la estructura 3D depende en gran medida de la secuencia de aminoácidos (a.a.) ⁽⁵⁾. Es por esta razón que la mayoría de los trabajos sobre estructuras de proteínas analizan las preferencias y cambios que presentan las secuencias de a.a., y los relacionan con las modificaciones en las estructuras 3D.

Las clasificaciones estructurales se han realizado en base a las semejanzas de estructura y de secuencia entre proteínas. Clasificaciones como SCOP ⁽⁶⁾ y CATH ⁽⁷⁾ muestran que las proteínas pueden adquirir una gran variedad de tipos de estructuras 3D, llamadas plegamientos. Las relaciones entre proteínas que estas clasificaciones muestran, parten del hecho de que si dos proteínas tienen una alta similitud en la secuencia, también presentarán una alta similitud en la estructura 3D. Así mismo, debido a que la estructura 3D no siempre se ve afectada por las mutaciones, presenta una mayor conservación en comparación con la secuencia de a.a. ⁽⁸⁾. Debido a lo anterior, la evolución estructural de las proteínas se ha estudiado tomando en cuenta la relación entre la secuencia de aa. y la estructura 3D, y mediante comparaciones entre estructuras de proteínas homólogas.

Aun cuando existen muchos trabajos sobre las estructuras 3D de las proteínas, todavía no se sabe de qué forma la secuencia de a.a. tiene la información precisa para que se obtenga una estructura 3D específica y funcional. Para entender mejor este fenómeno se ha estudiado la relación entre la secuencia de a.a. y la estructura 3D de las proteínas. Una característica principal es que las mutaciones no aparecen en cualquier posición de la cadena polipeptídica. Por el contrario, hay posiciones donde los cambios de a.a. se toleran preferentemente ⁽⁹⁾. Lo anterior se debe a que las proteínas presentan regiones en su estructura 3D con diferentes propiedades tanto químicas como físicas. Las posiciones donde aparecen un mayor o menor número de mutaciones están relacionadas con la accesibilidad al solvente y con las diferentes estructuras secundarias (ESs) de las proteínas ^(10, 11). En sí mismas, las ESs pueden encontrarse en el interior o exterior de la proteína, lo cual también provoca que presenten posiciones con distintos niveles de mutaciones. Es por estas razones que la relación secuencia-estructura, no puede ser analizada en su totalidad sin considerar las propiedades de las diferentes ESs.

Características Principales de las Estructuras Secundarias

La cadena polipeptídica de las proteínas se divide en 1) cadena principal, formada por los sucesivos enlaces peptídicos (entre los grupos amino y carboxilo), y en 2) las cadenas laterales de los a.a. Entre las ESs más abundantes y estudiadas están las hélices- α y las hojas- β . Estas ESs son conformaciones ordenadas y repetitivas de la cadena

principal, y se hallan en casi todas las proteínas conocidas. Las hélices- α y las hojas- β se encuentran conectadas por regiones de la secuencia conocidas como “coil”, “turn” y “loop”. Ambas estructuras tienen distintas conformaciones espaciales de la cadena polipeptídica, estabilizadas por interacciones cercanas y lejanas. Entre estas interacciones cercanas y lejanas están los puentes de hidrógeno que se generan en cada ES debido a su propia conformación espacial, las fuerzas de Van der Waals, fuerzas electrostáticas, interacciones hidrofóbicas, etc ⁽¹²⁾. Las interacciones son aquellas que involucran a a.a. cercanos en la secuencia, mientras que las interacciones lejanas son todas aquellas que involucran a otros a.a. lejos en la secuencia que llegan a hacer contacto principalmente con las cadenas laterales de los a.a. de la ES ⁽¹³⁾, y se ha propuesto que tienen un papel más importante en estabilizar la estructura 3D ⁽¹⁴⁾.

Se ha encontrado que existe una relación entre las estructuras secundarias y las mutaciones en la secuencia, debido a que los a.a. muestran distintas preferencias por una estructura secundaria en lugar de otra ^(13, 15, 16). Estas preferencias surgen debido a que las cadenas laterales de los a.a. adquieren distintas conformaciones espaciales, que pueden generar impedimentos estéricos con la cadena principal de las ESs. Así mismo, la posición de las hélices- α y hojas- β dentro de la estructura 3D influye en el tipo de a.a. que pueden presentar. Dado que las hojas- β tienden a estar en el interior de la estructura 3D y tienen poco contacto con el solvente, presentan preferencias por a.a. hidrofóbicos. Lo contrario sucede con las hélices- α y las regiones “coil”, las cuales presentan un mayor uso de a.a. hidrofílicos por estar situados en la superficie de las proteínas en contacto con el solvente ⁽¹²⁾. Por otro lado, se sabe que las ESs usan preferentemente distintas secuencias de a.a. polares y no polares, debido también a su posición en la estructura 3D y a su conformación espacial ⁽¹⁷⁾. Todas estas propiedades deben ser consideradas cuando se quieren analizar las diferentes ESs y entender sus semejanzas y diferencias, y el efecto de estas en la evolución estructural de las proteínas.

Estabilidad de las Estructuras Secundarias

Existen trabajos que han analizado las ESs tomando en cuenta su longitud y composición de a.a. ^(5, 18, 19). Así mismo, la relación entre las mutaciones en la secuencia de a.a. y el cambio estructural de las proteínas ha sido estudiada ampliamente. Aun cuando en algunos trabajos se hallaron diferencias de estas características entre las distintas ESs ⁽⁵⁾, no se logró una descripción completa de dichas diferencias, y más aun, algunas características quedaron vagamente exploradas. Por otro lado, el número de familias de proteínas y plegamientos conocidos sigue en aumento por lo que resulta necesario realizar nuevos análisis sobre distintos aspectos de la evolución de las ESs.

La estabilidad de las ESs es uno de los aspectos de interés para el plegamiento de la estructura 3D. La estabilidad se relaciona principalmente con la presencia de factores que fijan la estructura, evitando que se modifique su tamaño o se despliegue totalmente. Entre estos factores se encuentran los puentes de hidrógeno entre la cadena principal, las interacciones iónicas, las fuerzas de Van der Waals, etc. Existe la propuesta de que las ESs se forman primero y sirven como guías para la formación de la estructura 3D ⁽¹⁴⁾. Distintos trabajos apoyan esta idea y resaltan la importancia de estas estructuras en el

correcto plegamiento de la estructura 3D. Otros trabajos han mostrado que el orden y la longitud de las ESs pueden ser un factor importante en la formación de la estructura 3D⁽²⁰⁾. Así mismo se ha encontrado que la desaparición o sustitución por otra estructura depende del tamaño de las ESs⁽²¹⁾ y del tipo de mutaciones que presente su secuencia. Estos trabajos muestran la importancia de la estabilidad de las ESs, evitando que se vean disminuidas en tamaño o sustituidas por otra ES, afectando así a la estructura 3D de la proteína.

Las proteínas homólogas suelen tener una alta similitud en su contenido de ESs. Las principales diferencias se encuentran en el cambio de tamaño de estas estructuras. La estabilidad y crecimiento de las ESs, son aspectos importantes durante el proceso de plegamiento y para mantener una estructura tridimensional funcional. Estas características de las hélices- α y hojas- β , han sido estudiadas mediante dinámicas moleculares^(22, 23, 24). Se han propuesto distintos factores que estabilizan y que promueven el cambio de tamaño de estas estructuras.

Algunos trabajos teóricos que han realizado dinámicas moleculares en péptidos y proteínas, han hallado diferencias en la tendencia a crecer en longitud entre las hélices- α y las hojas- β ⁽²²⁾. Para el caso de las hélices- α se ha encontrado que existen diferencias entre sus dos regiones terminales, amino y carboxilo. Algunos trabajos muestran que la región amino de las hélices- α presenta una mayor estabilidad que la región carboxilo terminal^(22, 25, 26, 27). Esta conclusión se alcanzó mediante dinámicas moleculares, y no se ha hecho un análisis global de este fenómeno con las estructuras de proteínas conocidas. Estos trabajos teóricos y otros experimentales muestran en algunos casos resultados contradictorios. Debido a esto resulta relevante realizar un análisis global de las diferentes ESs.

Predicción de Estructuras Secundarias

La predicción de ESs, a partir de la secuencia de a.a., es un problema de mucha importancia para la biología molecular. Hoy en día el número de secuencias de a.a ha venido en aumento, debido en parte a los proyectos de secuenciación genómica, mientras que la determinación estructural de proteínas sigue siendo un proceso relativamente lento. Conocer la estructura de las proteínas es una necesidad básica para entender su mecanismo de acción y es por esta razón que los métodos para la predicción de ESs y tridimensionales han resultado ser bastante útiles.

Los métodos que han mostrado ser más efectivos son aquellos que basan su predicción en inferencias de carácter evolutivo^(28, 29). Si una secuencia de a.a. es suficientemente similar a otra de estructura conocida, su relación evolutiva garantiza una determinada similitud estructural. Pero lamentablemente no siempre se cuenta con un homólogo cercano de estructura conocida. En la mayoría de los casos hay que emplear técnicas que requieren buscar homólogos lejanos que den más información sobre la variación de la secuencia de a.a.⁽³⁰⁾. Los métodos de predicción resultan ser más eficientes si utilizan información de secuencias relacionadas a la proteína de la cual se quiere predecir su estructura.

correcto plegamiento de la estructura 3D. Otros trabajos han mostrado que el orden y la longitud de las ESs pueden ser un factor importante en la formación de la estructura 3D⁽²⁰⁾. Así mismo se ha encontrado que la desaparición o sustitución por otra estructura depende del tamaño de las ESs⁽²¹⁾ y del tipo de mutaciones que presente su secuencia. Estos trabajos muestran la importancia de la estabilidad de las ESs, evitando que se vean disminuidas en tamaño o sustituidas por otra ES, afectando así a la estructura 3D de la proteína.

Las proteínas homólogas suelen tener una alta similitud en su contenido de ESs. Las principales diferencias se encuentran en el cambio de tamaño de estas estructuras. La estabilidad y crecimiento de las ESs, son aspectos importantes durante el proceso de plegamiento y para mantener una estructura tridimensional funcional. Estas características de las hélices- α y hojas- β , han sido estudiadas mediante dinámicas moleculares^(22, 23, 24). Se han propuesto distintos factores que estabilizan y que promueven el cambio de tamaño de estas estructuras.

Algunos trabajos teóricos que han realizado dinámicas moleculares en péptidos y proteínas, han hallado diferencias en la tendencia a crecer en longitud entre las hélices- α y las hojas- β ⁽²²⁾. Para el caso de las hélices- α se ha encontrado que existen diferencias entre sus dos regiones terminales, amino y carboxilo. Algunos trabajos muestran que la región amino de las hélices- α presenta una mayor estabilidad que la región carboxilo terminal^(22, 25, 26, 27). Esta conclusión se alcanzó mediante dinámicas moleculares, y no se ha hecho un análisis global de este fenómeno con las estructuras de proteínas conocidas. Estos trabajos teóricos y otros experimentales muestran en algunos casos resultados contradictorios. Debido a esto resulta relevante realizar un análisis global de las diferentes ESs.

Predicción de Estructuras Secundarias

La predicción de ESs, a partir de la secuencia de a.a., es un problema de mucha importancia para la biología molecular. Hoy en día el número de secuencias de a.a ha venido en aumento, debido en parte a los proyectos de secuenciación genómica, mientras que la determinación estructural de proteínas sigue siendo un proceso relativamente lento. Conocer la estructura de las proteínas es una necesidad básica para entender su mecanismo de acción y es por esta razón que los métodos para la predicción de ESs y tridimensionales han resultado ser bastante útiles.

Los métodos que han mostrado ser más efectivos son aquellos que basan su predicción en inferencias de carácter evolutivo^(28, 29). Si una secuencia de a.a. es suficientemente similar a otra de estructura conocida, su relación evolutiva garantiza una determinada similitud estructural. Pero lamentablemente no siempre se cuenta con un homólogo cercano de estructura conocida. En la mayoría de los casos hay que emplear técnicas que requieren buscar homólogos lejanos que den más información sobre la variación de la secuencia de a.a.⁽³⁰⁾. Los métodos de predicción resultan ser más eficientes si utilizan información de secuencias relacionadas a la proteína de la cual se quiere predecir su estructura.

La mayoría de los métodos de predicción logran relativamente altos niveles de predicción en las regiones centrales de las hélices- α y hojas- β (~75%), pero comúnmente fallan en sus regiones terminales ^(28, 31). Esto ha resultado ser un problema para los métodos de predicción de estructuras tridimensionales que emplean predicciones de estructura secundaria.

ANTECEDENTES

Las hélices- α y las hojas- β son los elementos de ES más comunes en las proteínas. Se ha propuesto que la abundancia de estas dos estructuras se debe al balance entre dos fuerzas opuestas: a) *las interacciones locales* de la cadena principal, que favorecen la formación de hélices- α y hojas- β , y b) *las restricciones conformacionales* de las cadenas laterales, que afectan a estas estructuras ⁽²⁰⁾. Las diferentes conformaciones que pueden adquirir las cadenas laterales provocan que los a.a. presenten distintas preferencias por una u otra ES ^(2,3). Lo anterior, se ve reflejado en algunas matrices de sustitución, específicas para hélices- α y hojas- β , que se han hecho con anterioridad ^(10, 11). Así mismo, se sabe que el orden de la secuencia de a.a., clasificados en polares y no polares (características químicas) influye en la formación de las ESs ^(17, 33, 34) y posiblemente en el tamaño que éstas pueden adquirir. En este sentido, se ha demostrado que una secuencia de a.a. típicamente formadores de hélices- α puede formar una hoja- β si los a.a. se ordenan de acuerdo a un patrón hidropático (secuencia de a.a. polares y no polares) característico de hojas- β y viceversa ⁽¹²⁾.

Lo anterior muestra que la secuencia de a.a. y las propiedades de sus respectivas cadenas laterales influyen en la formación y estabilidad de las ESs. Aun así, el contenido y orden de a.a. en la secuencia de una proteína no aporta información suficiente para determinar con precisión sus distintas ESs. Por otro lado, no se tiene claro hasta qué punto la secuencia de a.a. determina la longitud de las ESs al estabilizar sus regiones amino y carboxilo. Así mismo no se sabe qué factores intervienen en la modificación del tamaño de las ESs, ya que se ha observado que estas pueden variar entre proteínas homólogas. Debido a esto, no es posible determinar con total certeza el inicio y término de una ES en la secuencia de a.a. de una proteína, siendo esto uno de los problemas que presentan los programas de predicción de ESs.

Los primeros trabajos que analizaron la variación estructural de proteínas en relación con la conservación de la secuencia, no hallaron grandes diferencias entre las hélices- α y las hojas- β de proteínas homólogas ^(5, 18, 19). Posteriormente se encontró que entre proteínas lejanamente relacionadas variaban las longitudes de sus ESs, pero no muestran diferencias entre hélices- α y hojas- β ⁽³⁵⁾. Todos estos trabajos, posiblemente no prestaron especial atención al cambio de tamaño de las ESs debido al bajo número de estructuras de proteínas conocidas, lo cual limitaba el número de comparaciones. Por otro lado, los estudios se enfocaron solamente en las regiones conservadas de las estructuras 3D, razón por la cual no se analizaban las modificaciones de tamaño de las ESs. En si, no se hizo ningún análisis detallado sobre la variación de longitud de estas estructuras.

La mayoría de los métodos de predicción logran relativamente altos niveles de predicción en las regiones centrales de las hélices- α y hojas- β (~75%), pero comúnmente fallan en sus regiones terminales ^(28, 31). Esto ha resultado ser un problema para los métodos de predicción de estructuras tridimensionales que emplean predicciones de estructura secundaria.

ANTECEDENTES

Las hélices- α y las hojas- β son los elementos de ES más comunes en las proteínas. Se ha propuesto que la abundancia de estas dos estructuras se debe al balance entre dos fuerzas opuestas: a) *las interacciones locales* de la cadena principal, que favorecen la formación de hélices- α y hojas- β , y b) *las restricciones conformacionales* de las cadenas laterales, que afectan a estas estructuras ⁽²⁰⁾. Las diferentes conformaciones que pueden adquirir las cadenas laterales provocan que los a.a. presenten distintas preferencias por una u otra ES ^(2,3). Lo anterior, se ve reflejado en algunas matrices de sustitución, específicas para hélices- α y hojas- β , que se han hecho con anterioridad ^(10, 11). Así mismo, se sabe que el orden de la secuencia de a.a., clasificados en polares y no polares (características químicas) influye en la formación de las ESs ^(17, 33, 34) y posiblemente en el tamaño que éstas pueden adquirir. En este sentido, se ha demostrado que una secuencia de a.a. típicamente formadores de hélices- α puede formar una hoja- β si los a.a. se ordenan de acuerdo a un patrón hidropático (secuencia de a.a. polares y no polares) característico de hojas- β y viceversa ⁽¹²⁾.

Lo anterior muestra que la secuencia de a.a. y las propiedades de sus respectivas cadenas laterales influyen en la formación y estabilidad de las ESs. Aun así, el contenido y orden de a.a. en la secuencia de una proteína no aporta información suficiente para determinar con precisión sus distintas ESs. Por otro lado, no se tiene claro hasta qué punto la secuencia de a.a. determina la longitud de las ESs al estabilizar sus regiones amino y carboxilo. Así mismo no se sabe qué factores intervienen en la modificación del tamaño de las ESs, ya que se ha observado que estas pueden variar entre proteínas homólogas. Debido a esto, no es posible determinar con total certeza el inicio y término de una ES en la secuencia de a.a. de una proteína, siendo esto uno de los problemas que presentan los programas de predicción de ESs.

Los primeros trabajos que analizaron la variación estructural de proteínas en relación con la conservación de la secuencia, no hallaron grandes diferencias entre las hélices- α y las hojas- β de proteínas homólogas ^(5, 18, 19). Posteriormente se encontró que entre proteínas lejanamente relacionadas variaban las longitudes de sus ESs, pero no muestran diferencias entre hélices- α y hojas- β ⁽³⁵⁾. Todos estos trabajos, posiblemente no prestaron especial atención al cambio de tamaño de las ESs debido al bajo número de estructuras de proteínas conocidas, lo cual limitaba el número de comparaciones. Por otro lado, los estudios se enfocaron solamente en las regiones conservadas de las estructuras 3D, razón por la cual no se analizaban las modificaciones de tamaño de las ESs. En si, no se hizo ningún análisis detallado sobre la variación de longitud de estas estructuras.

Otros trabajos más recientes han evaluado la variación en la longitud y el orden de las ESs entre proteínas con una identidad (ID) menor de 30%, y encontraron que se podía diferenciar entre los distintos plegamientos de las proteínas ⁽³⁶⁾. También se han hecho matrices de sustitución de ESs que toman en consideración sus longitudes, y se ha visto que la longitud de la ES parece ser el principal factor que determina la sustitución de una hélice- α o una hoja- β por un “coil” ⁽²¹⁾. Los anteriores trabajos muestran que es importante conocer las diferencias en longitud de las ES entre proteínas, así como la estabilidad de sus regiones terminales.

Existen trabajos, en su mayoría teóricos, que han analizado la formación y crecimiento de ESs ^(22, 23, 24). Haciendo uso de simulaciones mediante dinámicas moleculares se halló para el caso de las hélices- α que existen diferencias entre las terminaciones amino y carboxilo ⁽²²⁾. Por un lado se vio que las hélices- α se despliegan más fácilmente en su región carboxilo debido posiblemente a una menor estabilidad de los puentes de hidrógeno en esta región. A diferencia de la región amino, las cadenas laterales de los a.a. cercanos a la terminación carboxilo no forman fácilmente puentes de hidrógeno con los grupos C=O dado que su orientación los aleja ⁽²¹⁾.

Todos estos trabajos muestran diferencias, en algunos casos contradictorias, sobre la estabilidad y crecimiento de las hélices- α y las hojas- β . Debido a lo anterior, se requiere un análisis a profundidad que analice si estas diferencias experimentales y teóricas se conservan en los elementos estructurales de las proteínas conocidas. Así mismo, resulta necesario conocer qué características (conservación de secuencia, patrones de hidropatía y de “capping”, accesibilidad al solvente, etc.) presentan las regiones terminales de estas estructuras que pudieran afectar la modificación de su tamaño. Las diferencias de tamaño entre hélices- α y hojas- β junto con las características de sus regiones terminales pueden ser importantes para las clasificaciones estructurales ⁽³⁶⁾ y para los métodos de predicción de ESs.

OBJETIVOS

Analizar las diferencias en modificación del tamaño y conservación de secuencia entre hélices- α y hojas- β de proteínas homólogas de diversas familias.

Analizar las regiones terminales de hélices- α y hojas- β considerando las siguientes propiedades: conservación de secuencia medida mediante la entropía de Shannon, patrones de hidropatía, patrones de capping, accesibilidad al solvente y polaridad de los residuos en distintas posiciones de las ESs.

Realizar predicciones de ES para analizar la relación entre el nivel de conservación y el nivel de predicción correcta de las regiones terminales.

MÉTODOS

Alineamientos estructurales.

Los alineamientos múltiples se obtuvieron de dos bases de datos, Homstrad ⁽³⁷⁾ y PASS2 ⁽³⁸⁾. Estas bases de datos presentan alineamientos de secuencia con información estructural codificada, es decir, los alineamientos se realizan con información sobre la

Otros trabajos más recientes han evaluado la variación en la longitud y el orden de las ESs entre proteínas con una identidad (ID) menor de 30%, y encontraron que se podía diferenciar entre los distintos plegamientos de las proteínas ⁽³⁶⁾. También se han hecho matrices de sustitución de ESs que toman en consideración sus longitudes, y se ha visto que la longitud de la ES parece ser el principal factor que determina la sustitución de una hélice- α o una hoja- β por un “coil” ⁽²¹⁾. Los anteriores trabajos muestran que es importante conocer las diferencias en longitud de las ES entre proteínas, así como la estabilidad de sus regiones terminales.

Existen trabajos, en su mayoría teóricos, que han analizado la formación y crecimiento de ESs ^(22, 23, 24). Haciendo uso de simulaciones mediante dinámicas moleculares se halló para el caso de las hélices- α que existen diferencias entre las terminaciones amino y carboxilo ⁽²²⁾. Por un lado se vio que las hélices- α se despliegan más fácilmente en su región carboxilo debido posiblemente a una menor estabilidad de los puentes de hidrógeno en esta región. A diferencia de la región amino, las cadenas laterales de los a.a. cercanos a la terminación carboxilo no forman fácilmente puentes de hidrógeno con los grupos C=O dado que su orientación los aleja ⁽²¹⁾.

Todos estos trabajos muestran diferencias, en algunos casos contradictorias, sobre la estabilidad y crecimiento de las hélices- α y las hojas- β . Debido a lo anterior, se requiere un análisis a profundidad que analice si estas diferencias experimentales y teóricas se conservan en los elementos estructurales de las proteínas conocidas. Así mismo, resulta necesario conocer qué características (conservación de secuencia, patrones de hidropatía y de “capping”, accesibilidad al solvente, etc.) presentan las regiones terminales de estas estructuras que pudieran afectar la modificación de su tamaño. Las diferencias de tamaño entre hélices- α y hojas- β junto con las características de sus regiones terminales pueden ser importantes para las clasificaciones estructurales ⁽³⁶⁾ y para los métodos de predicción de ESs.

OBJETIVOS

Analizar las diferencias en modificación del tamaño y conservación de secuencia entre hélices- α y hojas- β de proteínas homólogas de diversas familias.

Analizar las regiones terminales de hélices- α y hojas- β considerando las siguientes propiedades: conservación de secuencia medida mediante la entropía de Shannon, patrones de hidropatía, patrones de capping, accesibilidad al solvente y polaridad de los residuos en distintas posiciones de las ESs.

Realizar predicciones de ES para analizar la relación entre el nivel de conservación y el nivel de predicción correcta de las regiones terminales.

MÉTODOS

Alineamientos estructurales.

Los alineamientos múltiples se obtuvieron de dos bases de datos, Homstrad ⁽³⁷⁾ y PASS2 ⁽³⁸⁾. Estas bases de datos presentan alineamientos de secuencia con información estructural codificada, es decir, los alineamientos se realizan con información sobre la

estructura secundaria, contacto con el solvente y puentes de hidrógeno que presenta cada a.a. Homstrad presenta alineamientos múltiples de familias de proteínas con una identidad (ID) de secuencia promedio de 30%. Se escogieron 141 alineamientos con un mínimo de 5 proteínas por alineamiento, para un total de 1236 proteínas. PASS2 presenta alineamientos múltiples de superfamilias de proteínas, donde ningún par de proteínas tiene una ID de secuencia mayor al 25%, estas superfamilias se formaron en base a similitud estructural y funcional. Se escogieron 110 alineamientos, con 375 proteínas. Ambas bases de datos están en concordancia con la clasificación de SCOP⁽⁶⁾.

Las proteínas de todos los alineamientos seleccionados se consideran el “Dataset” total (Dt) del presente trabajo. El Dt está conformado por proteínas globulares de todas las clases estructurales y con una gran variedad de funciones. A partir de las proteínas del Dt, se hizo un Dataset no redundante (Dnr). El Dnr se obtuvo usando el programa Cd-Hit⁽³⁹⁾ y se hizo de forma que ningún par de secuencias tuviera una ID >40%. El Dt consta de un total de 1611 proteínas, mientras que el Dnr tiene 433 proteínas.

Comparación de estructuras secundarias.

La relación entre el cambio de secuencia y el cambio de tamaño de las hélices- α y las hojas- β , se analizó usando todos los alineamientos múltiples del Dt. Para cada alineamiento múltiple, se obtuvieron todos los posibles alineamientos sencillos. En cada alineamiento sencillo se determinó la ID del alineamiento y se compararon las estructuras secundarias alineadas para obtener un valor absoluto del cambio de tamaño para hélices- α y hojas- β (Fig.1). El mismo análisis se realizó para las cuatro clases estructurales (All Alpha, All Beta, Alpha+Beta y Alpha/Beta) por separado. Las hélices 3_{10} no se incluyeron dentro de las hélices- α .

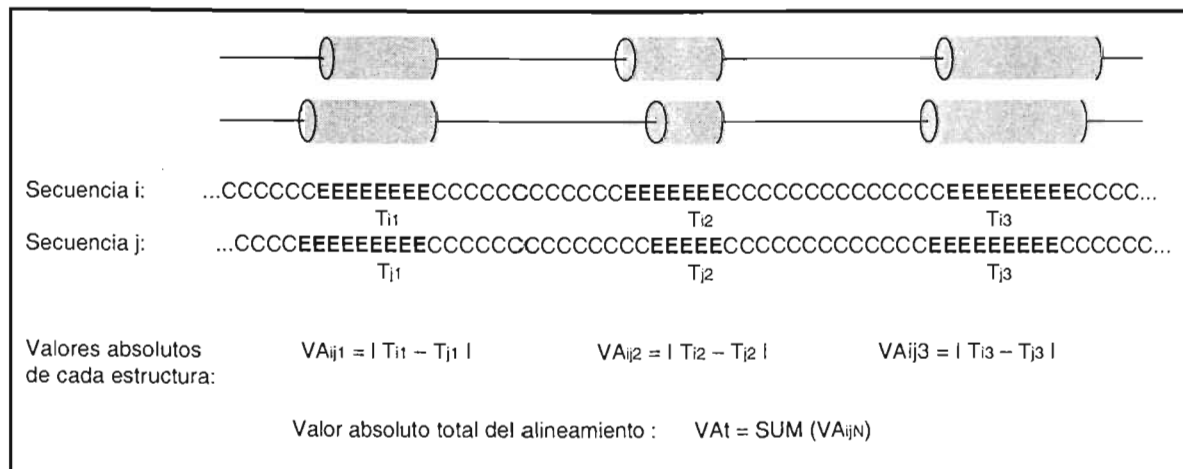


Fig. 1. Comparación de estructuras secundarias equivalentes entre proteínas homólogas. Se muestra en un alineamiento sencillo qué estructuras se comparan y cómo se obtiene el valor absoluto de cambio de tamaño de las estructuras secundarias del alineamiento. T= tamaño de la estructura secundaria, i,j,= secuencias, VA= valor absoluto del cambio de tamaño.

La relación entre el cambio de secuencia de las estructuras secundarias y el cambio de secuencia del alineamiento, se realizó de la misma forma. En este caso, para cada alineamiento sencillo se obtuvo la ID de secuencia de las hélices- α y las hojas- β , así como la ID de todo el alineamiento. El mismo análisis se realizó para las 4 clases estructurales.

Análisis de las regiones terminales de las hélices- α y las hojas- β .

El análisis de las regiones terminales se realizó con las proteínas del Dnr. Se analizaron por separado las regiones amino (CH) y carboxilo (HC) de las hélices- α y las regiones amino (CB) y carboxilo (BC) hojas- β .

a) Patrones de hidropatía.

Al igual que en trabajos anteriores ^(17, 33, 34), los patrones hidropáticos analizados consistieron en secuencias de cinco residuos polares (○) y/o no polares (●) sucesivos en las ESs. El número de patrones hidropáticos distintos que se pueden formar con secuencias de 5 residuos sucesivos es de 2^5 , es decir 32 patrones. Inicialmente se obtuvieron los patrones de hidropatía más frecuentes para las hélices- α , las hojas- β y los "coil"s. Los 32 patrones hidropáticos se clasificaron en 8 patrones típicos para cada estructura (hélices- α , hojas- β y "coil"s) y los 8 restantes se consideraron patrones atípicos, es decir, que no se hallan de forma preferente en alguna estructura en particular. Estas frecuencias fueron normalizadas de acuerdo al total de patrones hidropáticos hallados en todas las proteínas.

Normalización de los patrones hidropáticos: $\text{Log}(E_i / T_i)$

donde E = estructura secundaria (H, B o C)

T = suma de estructuras secundarias ($H_i + B_i + C_i$)

i = patrón hidropático (en total hay 32 distintos patrones)

Posteriormente se analizaron los patrones hidropáticos que presentan preferentemente las regiones terminales de las hélices- α y las hojas- β del Dnr. El análisis se hizo a lo largo de 10 a.a. en la región terminal, que corresponden a 5 a.a. de la región "coil" y 5 de la ES (Fig.2). Así, para cada región terminal se obtienen 6 patrones de hidropatía que equivalen a 6 posiciones o "ventanas", donde cada ventana corresponde a una secuencia de 5 a.a. polares y/o no polares.

Considerando todas las regiones terminales de las proteínas del Dnr, se obtuvo en cada ventana la frecuencia de los 32 patrones hidropáticos (2^5 patrones). Estos datos se normalizaron con las frecuencias generales de cada ES del Dnr. Finalmente, en cada ventana se obtuvo un valor de aparición de patrones típicos de hélice- α , hoja- β y "coil", para hacer esto, se sumaron las frecuencias normalizadas de los 8 patrones típicos de cada estructura y se obtuvo un promedio.

Se usaron dos distintos grupos de a.a. polares y no polares, ambos obtenidos de anteriores trabajos que analizaron patrones hidropáticos en proteínas. El grupo A es semejante al empleado por West y Hecht (1995) y tuvo la siguiente clasificación,

Polares (○): R, K, H, Q, E, D, N. No Polares (●): L, V, M, I, F.

La clasificación del grupo B es semejante a la usada por Mandel-Gutfreund y Gregoret (2002) y fue la siguiente,

Polares (○): S, T, R, K, H, Q, E, D, N. No Polares (●): A, G, W, P, L, V, M, I, F.

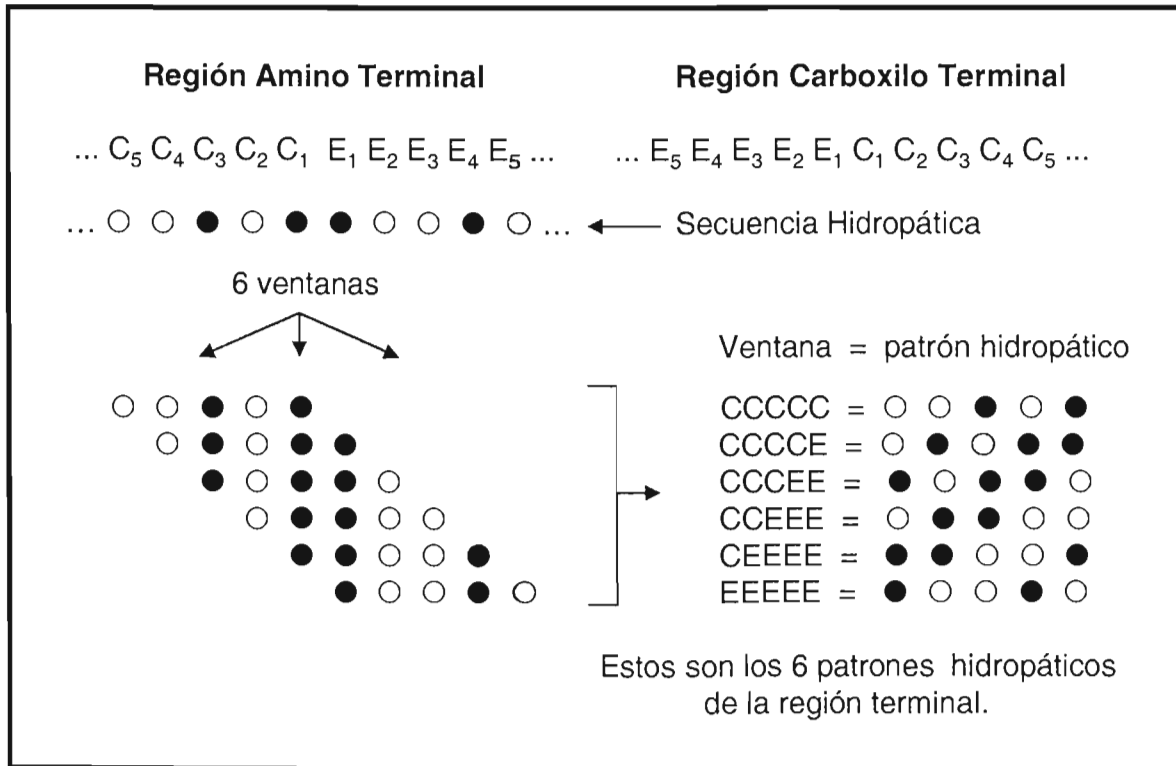


Fig.2. Obtención de los patrones de hidropatía de cada región terminal. Se obtienen un total de 6 patrones equivalentes a las 6 posibles ventanas de 5 residuos polares y/o no polares.

b) Conservación de secuencia.

Se analizó la conservación de secuencia de las regiones terminales de las proteínas del Dnr. La conservación se determinó mediante la entropía de Shannon

$$H = - \sum P_i \log_2 P_i$$

Con cada proteína del Dnr se hizo un “blast” tomando sólo aquellas proteínas con un valor de expectativa (e) de 1×10^{-20} . Las proteínas obtenidas de cada “blast” se filtraron, escogiendo sólo aquellas que tuvieran una ID >30% y se quitaron aquellas con una ID del 100% con respecto a la proteína con la que se realizó dicho “blast”. Aquellos “blast” que tuvieran menos de 20 secuencias después del filtro fueron eliminados. Con las secuencias

restantes de cada “blast”, se realizó un alineamiento múltiple con ClustalW⁽⁴¹⁾ quedando un total de 256 alineamientos múltiples. Se usó el alineamiento múltiple para obtener la frecuencia de cada a.a. (P_i) en cada posición. Con todos los valores de P_i , se obtuvo el valor de entropía de Shannon en cada posición de las regiones terminales de las ES. El valor de entropía de cada posición se normalizó de acuerdo al valor de entropía promedio de todas las posiciones del alineamiento. Cuando se encontraron menos de 20 a.a. no se obtuvo el valor de entropía. Finalmente, para analizar la conservación de secuencia se contabilizó el número de veces que cada posición tenía el valor de mayor y menor entropía en cada región terminal, para así obtener, en cada posición la diferencia entre el número de veces que tuvo la mayor y la menor entropía.

Por otro lado se usaron los datos de las frecuencias de a.a. en las 10 posiciones de las regiones terminales para determinar la polaridad general de cada posición. La frecuencia de cada a.a. en cada posición fue multiplicada por el valor de hidrofobicidad de Eisenberg de cada a.a. y fueron sumados para determinar el valor de polaridad de cada posición de las regiones terminales.

c) Patrones de “Capping”.

Los patrones de “capping” son de interés debido a que estabilizan las regiones terminales, amino y carboxilo, de las hélices- α mediante interacciones hidrofóbicas y generando puentes de hidrógeno con los grupos C=O y N-H de los a.a. en las regiones terminales. Se buscaron los patrones de capping más característicos en las hélices- α descritos anteriormente por Aurora y Rose (1998). La búsqueda se basó en la identificación de las secuencias características de cada patrón de capping, haciendo uso de los grupos polares y no polares que emplean estos autores en su trabajo. Cabe mencionar que a diferencia del trabajo de estos autores la H, K y R, siempre se consideraron dentro del grupo de a.a. polares.

Polares (○): G, S, T, R, K, H, Q, E, D, N. *Cualquier a.a. (*)*
No Polares (●): A, W, P, L, V, M, I, F. *“ β -Branched”(n): V, I, T, P.*

β -Branched : a.a. que presentan en el carbono β una ramificación más que los demás a.a.

Predicción de estructuras secundarias.

Se realizó la predicción de las ES de las proteínas del Dnr mediante el programa PsiPred⁽²⁹⁾. Al programa se le impidió usar, durante la predicción, la información proveniente de estructuras de proteínas del Protein Data Bank (PDB)⁽⁴⁴⁾. Se analizó 1) el uso de patrones hidropáticos, 2) la accesibilidad al solvente y 3) las frecuencias de a.a. con los valores de Eisenberg, en todas las posiciones de las regiones terminales de las estructuras predichas. Posteriormente se analizaron independientemente las regiones terminales predichas correcta e incorrectamente. Una predicción correcta se obtiene cuando la estructura predicha (H, B o C) en cada posición de la región terminal, es exactamente igual a la ES equivalente de la proteína de estructura conocida.

restantes de cada “blast”, se realizó un alineamiento múltiple con ClustalW ⁽⁴¹⁾ quedando un total de 256 alineamientos múltiples. Se usó el alineamiento múltiple para obtener la frecuencia de cada a.a. (P_i) en cada posición. Con todos los valores de P_i , se obtuvo el valor de entropía de Shannon en cada posición de las regiones terminales de las ES. El valor de entropía de cada posición se normalizó de acuerdo al valor de entropía promedio de todas las posiciones del alineamiento. Cuando se encontraron menos de 20 a.a. no se obtuvo el valor de entropía. Finalmente, para analizar la conservación de secuencia se contabilizó el número de veces que cada posición tenía el valor de mayor y menor entropía en cada región terminal, para así obtener, en cada posición la diferencia entre el número de veces que tuvo la mayor y la menor entropía.

Por otro lado se usaron los datos de las frecuencias de a.a. en las 10 posiciones de las regiones terminales para determinar la polaridad general de cada posición. La frecuencia de cada a.a. en cada posición fue multiplicada por el valor de hidrofobicidad de Eisenberg de cada a.a. y fueron sumados para determinar el valor de polaridad de cada posición de las regiones terminales.

c) Patrones de “Capping”.

Los patrones de “capping” son de interés debido a que estabilizan las regiones terminales, amino y carboxilo, de las hélices- α mediante interacciones hidrofóbicas y generando puentes de hidrógeno con los grupos C=O y N-H de los a.a. en las regiones terminales. Se buscaron los patrones de capping más característicos en las hélices- α descritos anteriormente por Aurora y Rose (1998). La búsqueda se basó en la identificación de las secuencias características de cada patrón de capping, haciendo uso de los grupos polares y no polares que emplean estos autores en su trabajo. Cabe mencionar que a diferencia del trabajo de estos autores la H, K y R, siempre se consideraron dentro del grupo de a.a. polares.

Polares (○): G, S, T, R, K, H, Q, E, D, N. *Cualquier a.a.* (*)
No Polares (●): A, W, P, L, V, M, I, F. *“ β -Branched”*(n): V, I, T, P.

β -Branched : a.a. que presentan en el carbono β una ramificación más que los demás a.a.

Predicción de estructuras secundarias.

Se realizó la predicción de las ES de las proteínas del Dnr mediante el programa PsiPred ⁽²⁹⁾. Al programa se le impidió usar, durante la predicción, la información proveniente de estructuras de proteínas del Protein Data Bank (PDB) ⁽⁴⁴⁾. Se analizó 1) el uso de patrones hidropáticos, 2) la accesibilidad al solvente y 3) las frecuencias de a.a. con los valores de Eisenberg, en todas las posiciones de las regiones terminales de las estructuras predichas. Posteriormente se analizaron independientemente las regiones terminales predichas correcta e incorrectamente. Una predicción correcta se obtiene cuando la estructura predicha (H, B o C) en cada posición de la región terminal, es exactamente igual a la ES equivalente de la proteína de estructura conocida.

RESULTADOS

Comparación de estructuras secundarias.

Al comparar el cambio de tamaño entre las ESs de proteínas homólogas, se observa que al ir disminuyendo la ID de los alineamientos, las hélices- α tienen un mayor cambio de tamaño que las hojas- β (Fig.3a). La misma tendencia se observa cuando se agrupan las proteínas por clases estructurales, quedando la curva de las proteínas All Alpha por arriba de las proteínas All Beta (Fig.3b). Debido a que las proteínas Alpha/Beta y Alpha+Beta presentan ambos tipos de ESs era de esperarse que sus respectivas curvas quedaran entre las curvas de las proteínas All Alpha y las All Beta. Cabe mencionar que si se analiza el cambio de tamaño de las hélices- α y las hojas- β presentes en las proteínas Alpha/Beta y Alpha+Beta se observa el mismo resultado que en la Fig. 3a.

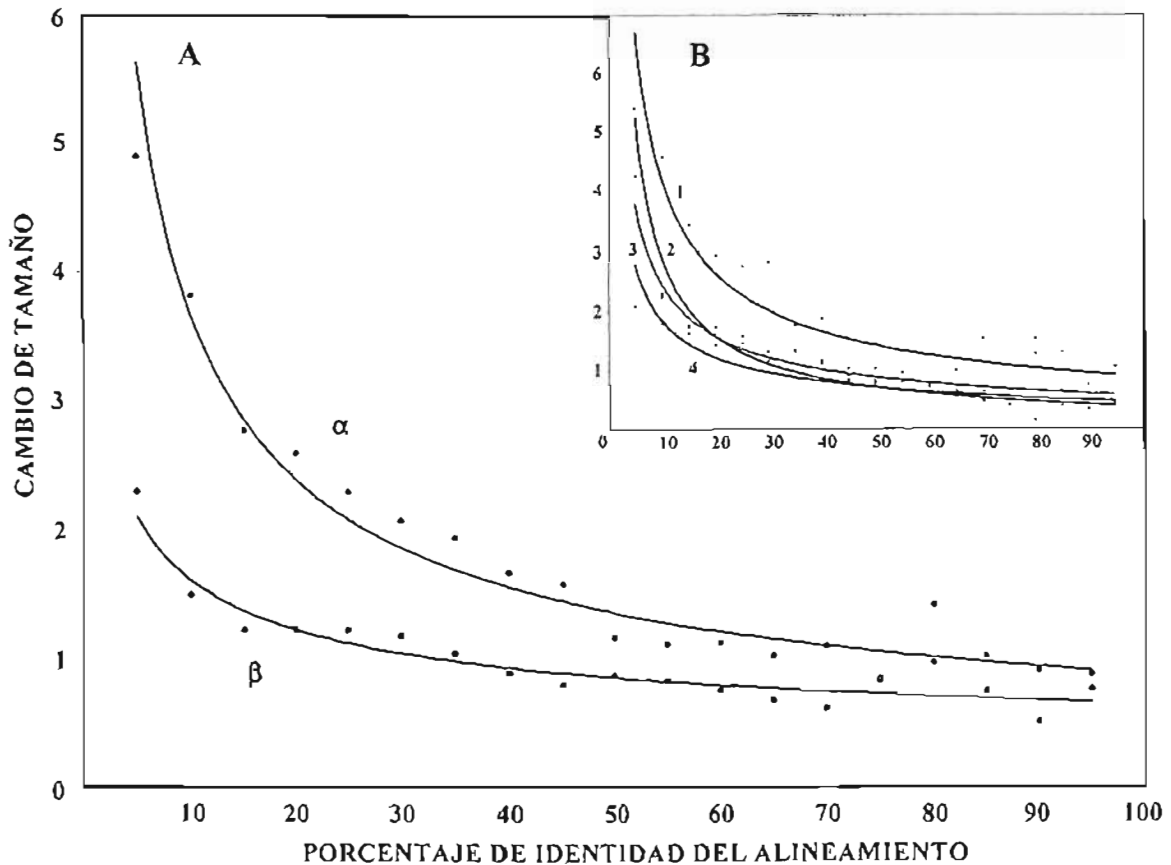


Fig. 3. Relación del cambio de tamaño de hélices- α y hojas- β con la identidad del alineamiento. A) Diferencia entre hélices- α y hojas- β de todos los alineamientos. B) Diferencia del cambio de tamaño de las estructuras secundarias presentes en las proteínas de las distintas clases estructurales: 1. All Alpha, 2. Alpha+Beta, 3. Alpha/Beta y 4. All Beta. En todos los casos se ajustó una curva de potencial. Cada punto representa el promedio en un intervalo de 5% en la identidad del alineamiento.

Las hélices- α tienden a formar estructuras más grandes (en número de a.a.) que las hojas- β , razón por la cual podrían cambiar de tamaño en mayor magnitud. Para averiguar lo anterior, se realizó el mismo análisis (cambio de tamaño vs. ID) comparando únicamente aquellas estructuras que tuvieran tamaños dentro de un intervalo específico. Se halló que no importa el intervalo de tamaño que se escoja, siempre las hélices- α cambian en mayor medida que las hojas- β (Fig.4). Lo anterior muestra que el mayor cambio de tamaño de las hélices- α no es producto de que formen estructuras con un mayor número de a.a.

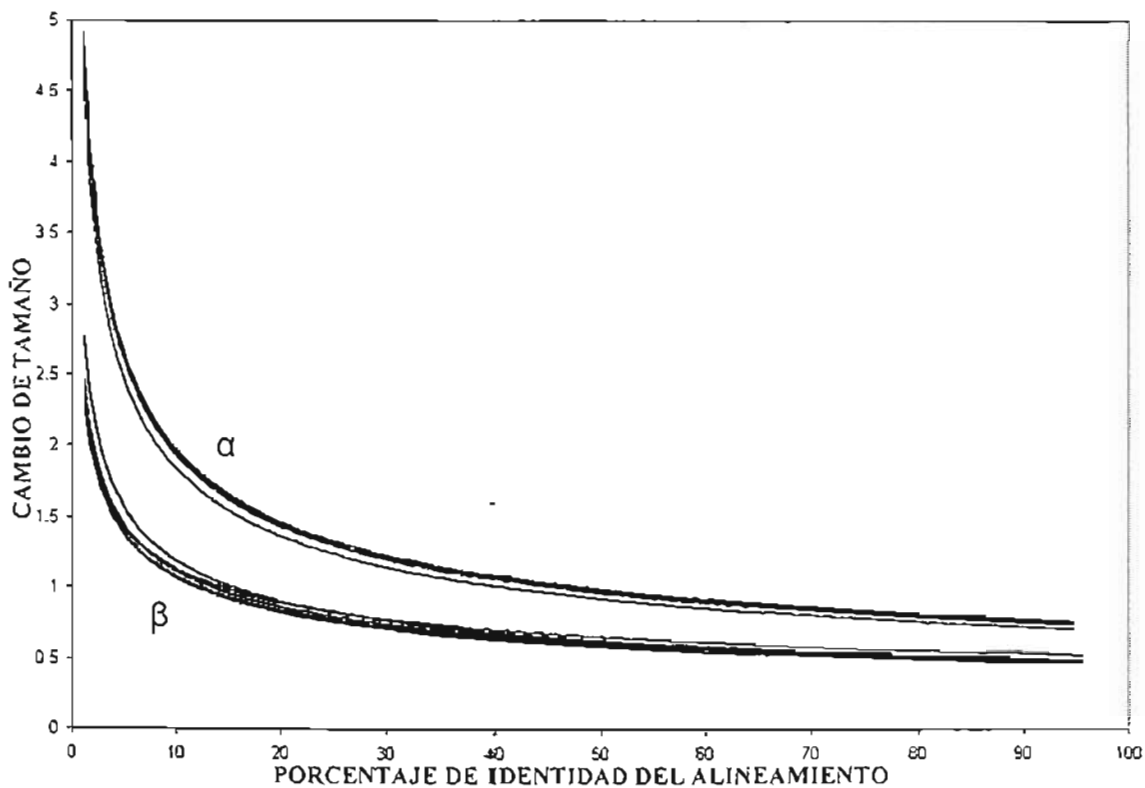


Fig. 4. Comparación del cambio de tamaño contra la identidad del alineamiento para hélices- α y hojas- β de distintos intervalos de tamaño de hasta 10 residuos: 4 a 13, 5 a 14, 6 a 15, 7 a 16, y 8 a 17. Sólo se comparaban las estructuras que estuvieran en un mismo intervalo de tamaño. Estos 5 intervalos se analizaron en ambas estructuras secundarias y se agrupan los de las hélices- α arriba y los de las hojas- β abajo. Cabe mencionar que se analizaron otros intervalos más pequeños y más grandes que 10 residuos y siempre se halló el mismo resultado. En todos los casos se ajustó una curva de potencial.

En el trabajo de Flores T, et al.(1993), quienes analizaron 90 pares de proteínas homólogas, se menciona que la conservación de secuencia de las ESs es igual entre hélices- α y hojas- β . En nuestro caso, al comparar la relación entre la ID de secuencia de las ESs y la ID de secuencia del alineamiento, se observa una mayor conservación de secuencia en las hojas- β (Fig.5). Se ejemplifica claramente que al empezar a divergir las proteínas, las hojas- β mantienen una mayor conservación de secuencia que las hélices- α .

Al ir disminuyendo la ID del alineamiento, la conservación de secuencia de las hélices- α decae más rápido que el de las hojas- β . Estos resultados muestran que el mayor cambio de tamaño de las hélices- α va a acompañando de una menor conservación de su secuencia, mientras que para las hojas- β sucede lo contrario.

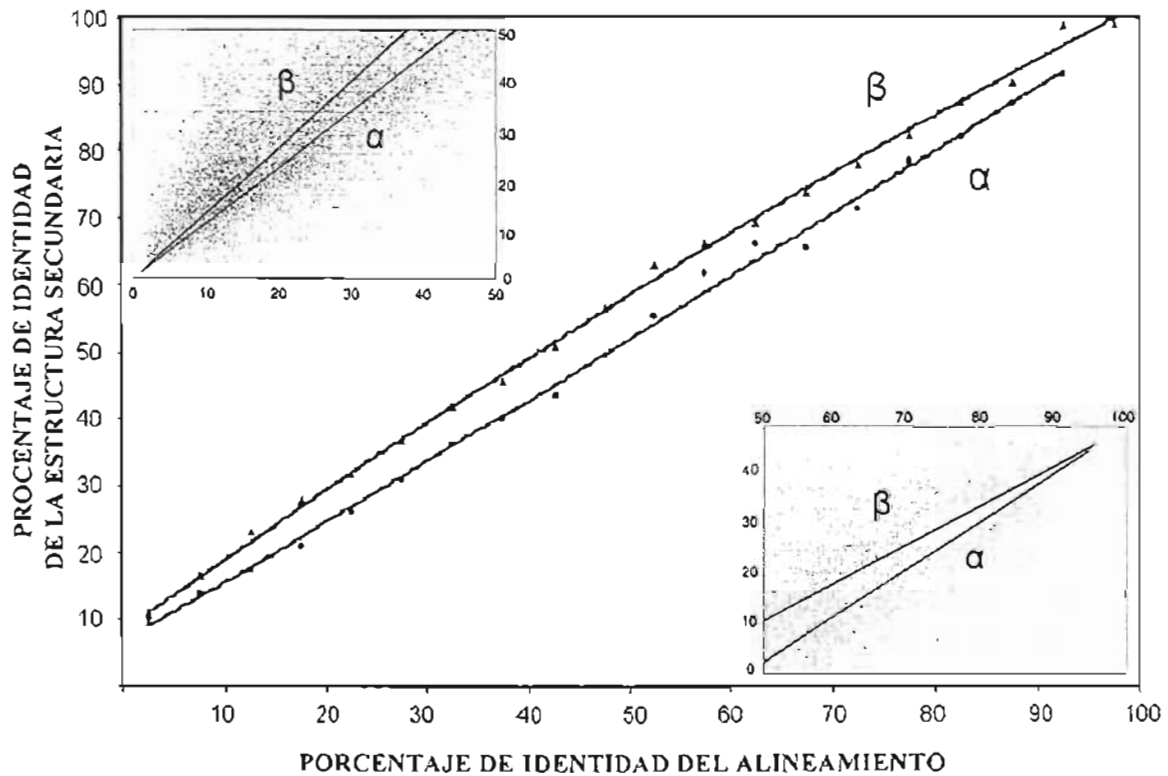


Fig. 5. Relación entre la identidad de secuencia de las hélices- α y las hojas- β contra la identidad del alineamiento. En la gráfica central se ajustaron curvas polinómicas de 2º grado y en las gráficas superior izquierda e inferior derecha se ajustaron rectas de modo que fuera notoria la diferencia entre los datos de las hélices- α y las hojas- β desde el inicio al final de los valores del porcentaje de identidad del alineamiento.

En el presente trabajo no se consideraron a las hélices 3_{10} como parte de las hélices- α . En trabajos previos se ha mencionado que las hélices 3_{10} , son estructuras de transición entre el estado “coil” y las hélices- α ⁽³²⁾. Aun considerando las hélices 3_{10} como parte de las hélices- α , no se observan cambios en los anteriores resultados (datos no mostrados). Esto se debe en gran medida a que en los alineamientos estructurales de las familias, cuando aparece una hélice 3_{10} , se encuentra conservada en la misma posición en la mayoría de las proteínas de la familia. Por otro lado, es importante resaltar que la ID de cada par de proteínas comparadas en los anteriores análisis, se obtuvo a partir del alineamiento estructural. Si se alinean las mismas secuencias con ClustalW ⁽⁴¹⁾ y se realiza el mismo análisis, el resultado final es el mismo. Por lo tanto, la diferencia entre hélices- α y hojas- β , no se debe a un efecto de los alineamientos estructurales.

Análisis de regiones terminales.

El análisis de cambio de tamaño vs. ID, se repitió considerando cada región terminal de las hélices- α y de las hojas- β por separado. Como se muestra en la Fig.6, ambas regiones terminales de las hélices- α (CH y HC) cambian en igual magnitud, y además cambian en mayor medida que las regiones terminales de las hojas- β (CB y BC). Cabe mencionar que si bien las dos regiones de las hélices- α cambian en igual magnitud, el número de veces que cada una cambia varía en la siguiente proporción, CH = 43.5 % y HC = 56.5%, en el caso de las hojas- β sus dos regiones cambian prácticamente el mismo número de veces (CB = 50.2% y BC = 49.8%). Este mismo análisis se realizó por separado para las regiones terminales con y sin accesibilidad al solvente y se obtuvo el mismo resultado. Por consiguiente, con el fin de establecer si el cambio de tamaño se debía a una menor estabilidad evolutiva producto de una menor conservación de distintas propiedades de la secuencia de a.a., se analizaron las características de las regiones terminales de ambas ESs. Estas características fueron los patrones de hidropatía, la conservación de secuencia, el uso de a.a. (solubilidad y accesibilidad al solvente) y patrones de capping. En este sentido, una menor estabilidad evolutiva significa que en la región terminal se halla una conservación de secuencia baja y aparecen patrones que no favorecen la formación de la ES. Los siguientes resultados se obtuvieron con el dataset no redundante (Dnr).

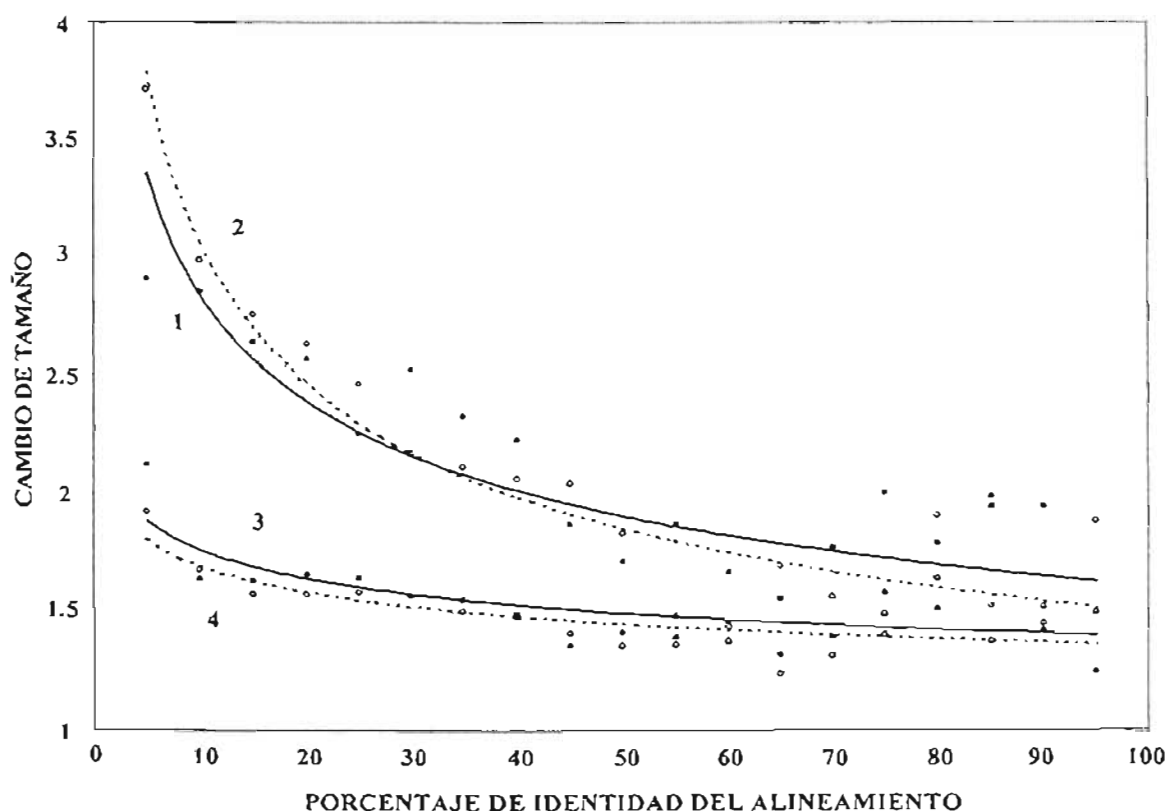


Fig. 6. Relación del cambio de tamaño de las regiones terminales de hélices- α y hojas- β contra el porcentaje de identidad del alineamiento. 1. Región CH, 2. Región HC, 3. Región CB y 4. Región BC. Cada punto representa el promedio en un intervalo de 5% en la identidad del alineamiento.

a) Patrones de hidropatía.

El primer paso consistió en obtener las frecuencias de los patrones de hidropatía de las hélices- α , las hojas- β y las regiones "coil" (ver métodos). Las frecuencias de los patrones de hidropatía (tabla 1) no difieren mucho entre el grupo A y el grupo B de a.a. polares (○) y no polares (●). En el grupo A, los patrones hidropáticos de las hélices- α son muy parecidos a los reportados ⁽¹⁷⁾. Para el caso de los patrones de las hojas- β del grupo A, los que se encontraron entre los primeros 7 lugares de preferencia son muy diferentes, en este caso los patrones hallados son por lo general más no polares en comparación con los reportados previamente.

Tabla 1. Patrones de Hidropatía característicos de cada estructura para ambos grupos (A y B) de a.a. polares (○) y no polares (●). En gris se muestran aquellos patrones diferentes entre el grupo A y B para cada estructura.

GRUPOS HIDROPÁTICOS		
	B	A
HELIX		
H1	●○○●●	●○○●●
H2	○○●●○	●●○○●
H3	○●●○○	●●○○●
H4	●●○○●	○○●●○
H5	●●○○●	○○●○○
H6	○○●○○	○○●○○
H7	○●●○●	●○●●○
H8	●○○●○	○●○○●
"COIL"		
C1	○○○○○	○○○○○
C2	●○○○○	●○○○○
C3	○○○○●	○○○○●
C4	○●○○○	●○●○○
C5	○○○●○	○●○○○
C6	○●○○○	●●○○○
C7	○○○●●	○○○●●
C8	●○○○○	○○○●○
BETA		
B1	●●●●●	●●●●○
B2	○●●●●	●●●●●
B3	●●●●○	○●●●●
B4	●○●○●	●○●○●
B5	●●●○●	●●●○●
B6	○●○○○	●●○○○
B7	○●○○●	○●○○●
B8	●○●●●	○●○○○

Para el caso del grupo B, los 7 primeros patrones hidropáticos de las hojas- β son prácticamente los mismos y los primeros 3 de las hélices- α son los mismos a los reportados por otros autores⁽³⁴⁾. Cabe mencionar que estos autores encuentran diferencias entre el uso de patrones de hojas- β paralelas y antiparalelas. Nuestros resultados se asemejan más a los patrones usados por hojas- β paralelas, las cuales se caracterizan por estar en el core de las proteínas, aun así se hallan patrones típicos de hojas antiparalelas. Debido a lo anterior, los patrones hallados en las hojas- β para ambos grupos (A y B) son marcadamente no polares.

Los resultados con ambos grupos polares (A y B) son semejantes en gran medida porque se realizó una normalización de los resultados (ver métodos). Lo anterior es importante porque entre los trabajos de Mandel-Gutfreund y Gregoret (2002) y West y Hecht (1995), los resultados difieren en gran medida por las diferencias de los grupos A y B. En nuestro caso, nos interesaba conocer aquellos patrones específicos usados por cada estructura y que al mismo tiempo no usan las demás estructuras, por esta razón se realizó la normalización, la cual por otro lado trajo como consecuencia que nuestros resultados con ambos grupos fueran similares entre sí, sin importar las diferencias de agrupación de a.a. polares y no polares. Por otro lado, la normalización es necesaria debido a la diferencia entre el número de hélices- α y hojas- β y por la diferencia entre la longitud de ambas estructuras que provoca que haya un mayor número de patrones para las hélices- α .

Posteriormente, se analizaron los patrones que aparecen más frecuentemente en las regiones terminales de ambas estructuras secundarias. La tabla 2 muestra que existe una mayor tendencia en las hojas- β a presentar patrones típicos de las hojas cerca del fin de su estructura y patrones típicos de "coil" lejos de esta zona (CB y BC es igual esta tendencia). Para las hélices- α en cambio, se hallan patrones típicos de "coil" muy cerca de las regiones terminales de las hélices- α , siendo la región CH, más propensa a este fenómeno. Lo anterior puede ser un indicio que las hojas- β tienden a ser menos propensas a cambiar su longitud, lo contrario se espera en las hélices- α , sobre todo en la región CH. Los mismos resultados se observan con ambos grupos hidropáticos (A y B), aunque por lo general se obtienen más casos para el grupo B por tener más a.a. catalogados como polares y no polares (ver métodos).

Tabla 2. Uso de patrones hidropáticos en las regiones terminales de hélices- α y hojas- β . En cada ventana (columna) se muestra el promedio de las frecuencias de los 8 patrones característicos de cada estructura (H, B, C y *, este último representa los patrones no característicos de ninguna estructura). Las casillas sombreadas son las que tuvieron el mayor valor por ventana e indican que los patrones hidropáticos de esa estructura (H, B, C o *) aparecen en esa ventana de forma preferente.

Región CH

Región HC

	CCCC	CCCCH	CCCHH	CCHHH	CHHHH	HHHHH	HHHHH	HHHHC	HHHCC	HHCCC	HCCCC	CCCC
H	0.89	1.01	0.64	1.02	0.88	1.17	1.51	1.35	1.15	1.00	0.91	0.82
C	1.35	1.29	1.63	1.81	1.54	1.10	0.65	0.86	0.98	1.08	1.21	1.29
B	0.95	0.88	0.99	0.65	0.75	0.78	0.66	0.63	0.76	0.78	0.92	1.03
*	0.89	0.77	1.09	0.63	0.95	0.94	0.91	0.89	1.04	1.03	1.02	0.91

Región CB

Región BC

	CCCC	CCCCB	CCCBB	CCBBB	CBBBB	BBBBB	BBBBB	BBBBB	BBBBB	BBBBB	BBBBB	BBBBB
H	0.86	0.99	0.87	0.87	0.66	0.58	0.65	0.80	0.86	0.86	0.85	0.93
C	1.47	1.31	0.95	0.64	0.60	0.51	0.51	0.66	0.85	1.15	1.43	1.60
B	0.84	0.87	1.10	1.40	1.78	2.06	1.85	1.43	1.09	0.84	0.74	0.68
*	0.96	0.84	1.00	0.90	0.77	0.68	0.68	0.67	0.91	0.98	0.91	0.95

b) Conservación de secuencia.

Los valores de entropía para las distintas posiciones de las regiones terminales, son claramente distintos entre las hélices- α y las hojas- β (Fig. 7). En el caso de ambas regiones terminales de las hojas- β , las posiciones más conservadas son por lo común las de hojas- β y las menos conservadas son las posiciones de "coil". Para el caso de las hélices- α , en la región CH las posiciones menos conservadas son por lo común las de hélice- α y las más conservadas las de "coil". La región HC de las hélices- α muestra una gran variación de la conservación en las distintas posiciones. Lo anterior muestra, al igual que los patrones hidropáticos, que las regiones terminales de las hojas- β serían menos propensas a cambiar de tamaño, mientras que las regiones terminales de las hélices- α , en especial la región CH, cambiaría más fácilmente de tamaño.

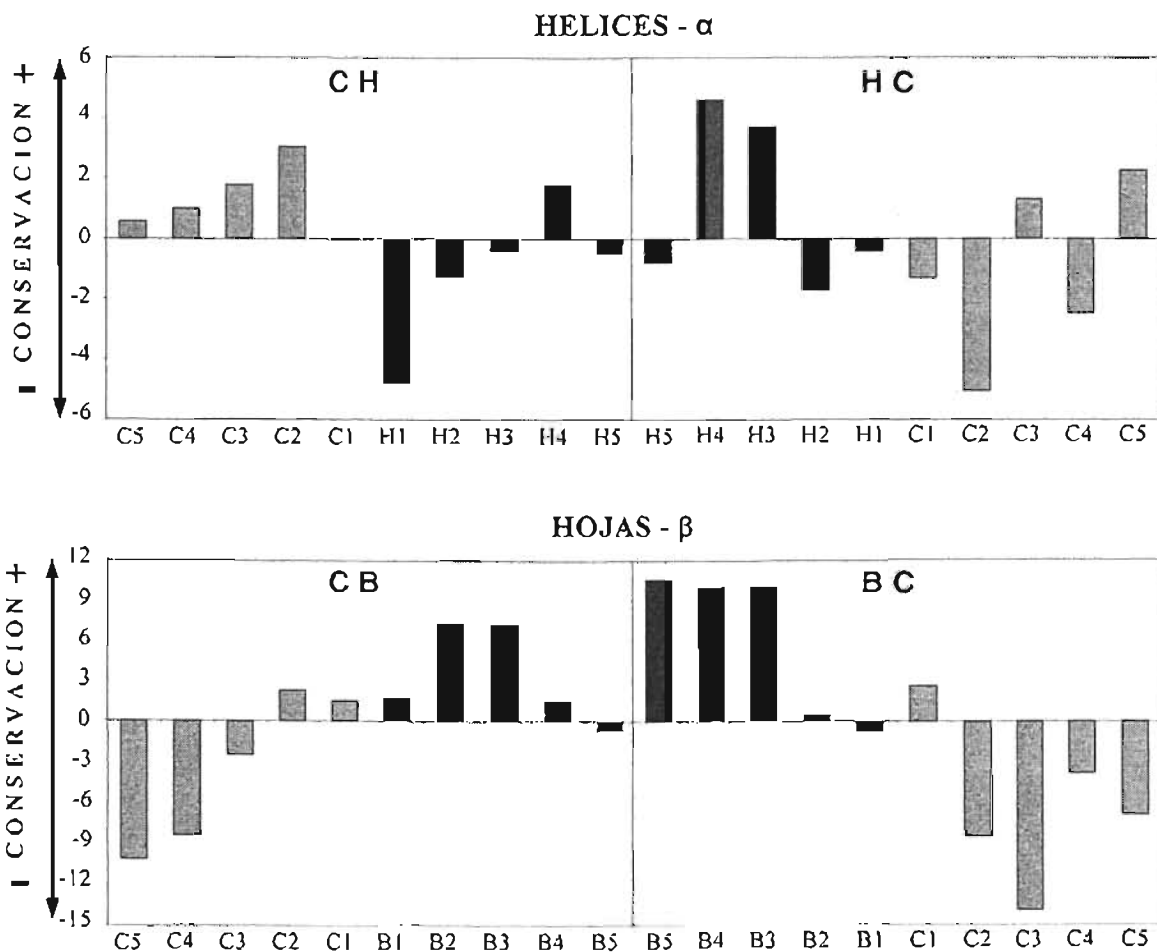


Fig.7. Conservación de secuencia en las distintas posiciones de las regiones terminales de las hélices- α y las hojas- β . Cada barra representa la diferencia entre el número de veces que en esa posición apareció el mayor y el valor menor de entropía de todas las regiones terminales.

Las frecuencias de a.a. en las fronteras (tabla 3), muestran las preferencias de determinados a.a. para hélices- α y para hojas- β . Por lo general, se observa que los a.a. característicos de hélices- α , hojas- β y "coil", aparecen preferentemente en estas regiones. Frecuencias obtenidas en otros trabajos muestran estas mismas características, lo cual es una muestra de la representatividad de Dnr. Por otro lado los valores de hidrofobicidad de Eisenberg relacionados con estas frecuencias de a.a. en cada posición (Fig.8) muestran que las regiones terminales de las hojas- β son marcadamente no polares en todas sus posiciones y polares en las posiciones de "coil". En cambio las regiones terminales de las hélices- α presentan una alternancia de posiciones polares y no polares que guarda relación con la disposición de a.a. polares y no polares de las hélices anfipáticas. Cabe mencionar que los valores de hidrofobicidad de Eisenberg asociados a las frecuencias de a.a., son un mejor indicador de la polaridad de la posición, puesto que se toman en cuenta las frecuencias de todos los a.a. de cada posición. Si se considerara únicamente la polaridad del a.a. más frecuente se obtendría una distribución de polaridades en las posiciones, que no guardaría relación con los patrones de hidropatía y de capping (ver más adelante).

TABLA 3a.

Frecuencias de aminoácidos en las regiones terminales de las hélices- α . El sombreado muestra el residuo con la frecuencia más alta.

	Región CH										Región HC									
	C ₆	C ₄	C ₃	C ₂	C ₁	H ₁	H ₂	H ₃	H ₄	H ₅	H ₅	H ₄	H ₃	H ₂	H ₁	C ₁	C ₂	C ₃	C ₄	C ₅
-	1.48	1.36	1.09	0.91	0.82	0.76	0.71	0.71	0.63	0.69	0.61	0.71	0.93	1.05	1.10	1.21	1.56	1.59	1.57	1.48
A	0.71	0.68	0.78	0.74	0.48	1.37	1.36	1.07	1.40	1.66	1.63	1.29	1.46	1.56	1.20	0.97	0.64	0.70	0.59	0.50
C	1.02	0.57	1.03	0.97	0.47	0.86	0.64	1.01	0.55	0.66	0.76	1.56	1.25	1.52	1.04	1.29	0.62	0.52	0.70	0.62
D	1.19	1.11	1.06	0.75	2.15	1.08	1.21	1.95	0.37	1.10	0.82	0.74	0.72	0.67	0.70	1.01	0.81	1.10	1.33	1.67
E	0.91	0.87	0.99	0.70	0.80	1.87	2.57	2.16	0.67	1.61	1.33	1.29	1.39	1.09	1.13	0.86	0.88	0.85	0.86	0.88
F	1.02	1.16	0.88	1.01	0.42	0.58	0.80	0.74	1.14	1.06	1.08	1.23	0.88	0.84	1.08	0.69	0.59	0.95	0.60	1.33
G	1.86	1.35	1.81	1.36	1.09	0.58	0.68	0.47	0.52	0.42	0.34	0.43	0.34	0.41	0.37	1.48	1.89	1.35	1.07	1.10
H	1.13	1.20	0.97	0.64	1.18	0.70	0.72	0.73	0.84	0.66	0.83	0.73	1.36	1.42	1.79	1.38	1.21	1.78	1.06	1.69
I	0.59	0.50	0.61	0.91	0.22	0.71	0.54	0.63	1.68	1.06	1.31	1.34	1.11	1.31	0.71	0.36	0.46	0.61	0.84	0.85
K	1.07	1.37	0.84	0.76	0.67	1.22	1.33	0.82	0.82	1.28	1.06	1.01	1.29	1.42	1.43	1.59	1.72	1.30	1.29	0.83
L	0.96	0.66	0.73	1.27	0.47	0.82	0.50	0.99	1.76	1.29	1.33	1.74	1.53	1.36	1.48	1.08	1.06	1.18	0.84	0.76
M	0.57	0.87	0.44	1.77	0.30	1.06	0.68	1.14	2.04	1.42	1.62	1.31	1.90	1.68	1.47	1.09	0.61	0.86	0.74	0.59
N	1.08	1.15	1.12	1.12	1.95	0.43	0.80	0.76	0.58	0.76	0.75	0.69	0.70	0.89	1.07	1.57	1.32	1.09	1.16	1.54
P	1.59	2.10	1.78	1.10	1.66	2.39	0.68	0.32	0.10	0.07	0.23	0.11	0.15	0.09	0.13	0.21	1.64	1.66	1.64	1.65
Q	1.04	0.86	0.98	0.85	0.71	1.27	1.68	2.11	1.21	1.20	1.21	1.18	1.08	1.14	1.07	1.20	1.07	1.13	0.68	0.76
R	0.69	1.33	0.83	0.72	0.51	0.68	1.04	0.83	1.74	1.13	1.00	1.04	1.53	1.21	1.14	1.11	1.02	0.87	1.02	1.09
S	0.80	0.94	1.08	1.09	2.55	0.75	1.10	0.62	0.52	0.88	0.81	0.71	0.63	0.77	1.11	0.98	0.89	0.90	0.93	1.10
T	1.06	0.84	0.88	1.09	2.43	0.86	1.15	1.31	0.60	0.57	0.82	0.48	0.67	0.70	0.97	0.87	0.86	0.79	1.08	0.95
V	0.53	0.59	0.74	1.08	0.34	0.93	0.53	0.81	1.49	0.82	1.07	1.03	0.72	0.75	0.68	0.41	0.51	0.57	1.10	0.71
W	0.74	0.96	0.85	1.27	0.35	1.14	1.10	0.54	0.87	1.33	0.82	1.18	1.38	1.12	0.48	0.41	0.32	0.38	0.48	0.46
Y	0.71	0.99	1.12	0.89	0.33	0.77	0.69	1.01	0.83	0.78	1.02	1.24	0.88	0.73	1.29	1.24	0.49	0.84	1.21	0.68

TABLA 3b.

Frecuencias de aminoácidos en las regiones terminales de las hojas- β . El sombreado muestra el residuo con la frecuencia más alta.

	Región CB										Región BC									
	C ₅	C ₄	C ₃	C ₂	C ₁	B ₁	B ₂	B ₃	B ₄	B ₅	B ₅	B ₄	B ₃	B ₂	B ₁	C ₁	C ₂	C ₃	C ₄	C ₅
-	2.33	2.01	2.02	1.52	1.07	0.60	0.55	0.44	0.37	0.54	0.34	0.28	0.33	0.44	0.61	1.01	1.31	1.50	1.72	1.99
A	0.72	0.83	0.64	0.56	0.59	0.61	0.78	1.16	1.01	0.93	0.91	0.76	1.05	1.15	0.54	0.75	0.66	0.78	0.99	0.86
C	0.45	0.60	0.27	0.56	0.56	0.93	2.50	1.22	1.62	1.00	2.71	0.69	1.36	0.85	0.66	0.53	0.45	0.56	1.31	0.63
D	1.00	1.39	1.76	1.61	1.26	0.47	0.23	0.39	0.38	0.56	0.53	0.31	0.27	0.50	1.47	1.20	1.31	1.62	1.41	1.18
E	0.87	1.23	1.23	0.82	0.96	0.76	0.53	0.38	0.58	0.55	0.52	0.57	0.61	0.71	0.99	0.84	1.17	0.90	0.78	0.90
F	0.76	0.47	0.56	0.52	0.53	1.10	1.78	0.95	1.02	1.52	1.85	1.54	1.22	2.14	1.07	0.66	0.60	0.47	0.51	0.68
G	1.20	1.26	1.23	2.11	1.24	0.65	0.58	0.83	0.57	0.80	1.01	0.42	0.60	0.48	0.85	1.49	1.53	1.48	1.89	1.16
H	1.15	0.51	0.96	0.91	1.05	1.54	0.79	0.48	0.58	1.14	0.56	0.64	0.48	0.74	1.25	1.82	1.03	0.90	0.83	0.69
I	0.53	0.52	0.48	0.49	0.92	1.35	1.96	1.62	2.03	1.53	1.37	1.73	2.61	1.84	0.96	0.48	0.36	0.67	0.38	0.55
K	1.08	0.93	1.27	1.05	0.98	0.76	0.45	0.74	0.74	0.87	0.77	0.65	0.55	0.67	0.77	0.88	1.03	0.82	0.77	0.91
L	0.82	0.66	0.44	0.50	0.78	1.06	1.70	1.45	1.29	1.41	1.05	1.82	1.12	1.10	1.39	0.57	0.45	0.45	0.46	0.68
M	0.58	0.33	0.32	0.33	0.51	2.16	1.25	0.89	1.78	1.21	2.27	1.26	0.94	1.63	0.76	0.49	0.56	0.79	0.47	0.61
N	1.30	1.47	1.81	1.20	1.09	0.53	0.43	0.37	0.59	0.68	0.57	0.49	0.32	0.39	0.87	1.69	1.11	1.47	1.36	1.67
P	1.87	2.45	1.74	2.25	2.07	0.62	0.39	0.25	0.18	0.51	0.42	0.12	0.29	0.19	1.24	0.76	3.00	2.11	1.21	1.40
Q	0.89	0.73	1.33	0.90	0.99	2.10	0.56	0.76	0.72	0.95	0.66	0.67	0.61	0.51	0.72	0.65	0.69	0.78	1.05	1.18
R	0.89	0.94	0.87	0.81	1.23	0.84	0.48	1.27	0.63	0.82	0.68	0.95	0.76	1.17	0.86	0.74	0.81	0.72	0.65	1.04
S	1.29	1.35	1.32	1.06	0.99	0.94	0.54	0.86	0.63	0.89	0.77	0.65	0.74	1.33	0.67	1.44	1.51	1.47	1.18	1.43
T	0.97	1.20	0.98	0.91	1.03	1.30	0.94	1.00	1.33	0.98	1.05	1.27	1.17	0.94	0.73	1.37	1.31	1.19	1.31	1.46
V	0.66	0.46	0.50	0.80	0.91	1.56	2.42	2.11	2.47	1.56	1.47	2.36	2.61	1.76	1.43	0.73	0.33	0.53	0.73	0.59
W	1.23	0.71	0.70	0.43	1.60	0.31	1.19	0.85	0.99	0.84	0.42	1.91	0.68	0.70	2.66	3.14	1.11	1.28	0.59	0.49
Y	1.23	0.62	0.52	0.64	0.77	1.56	0.98	1.54	0.74	1.14	1.84	1.17	0.99	0.94	1.02	0.92	0.64	0.59	1.10	0.83

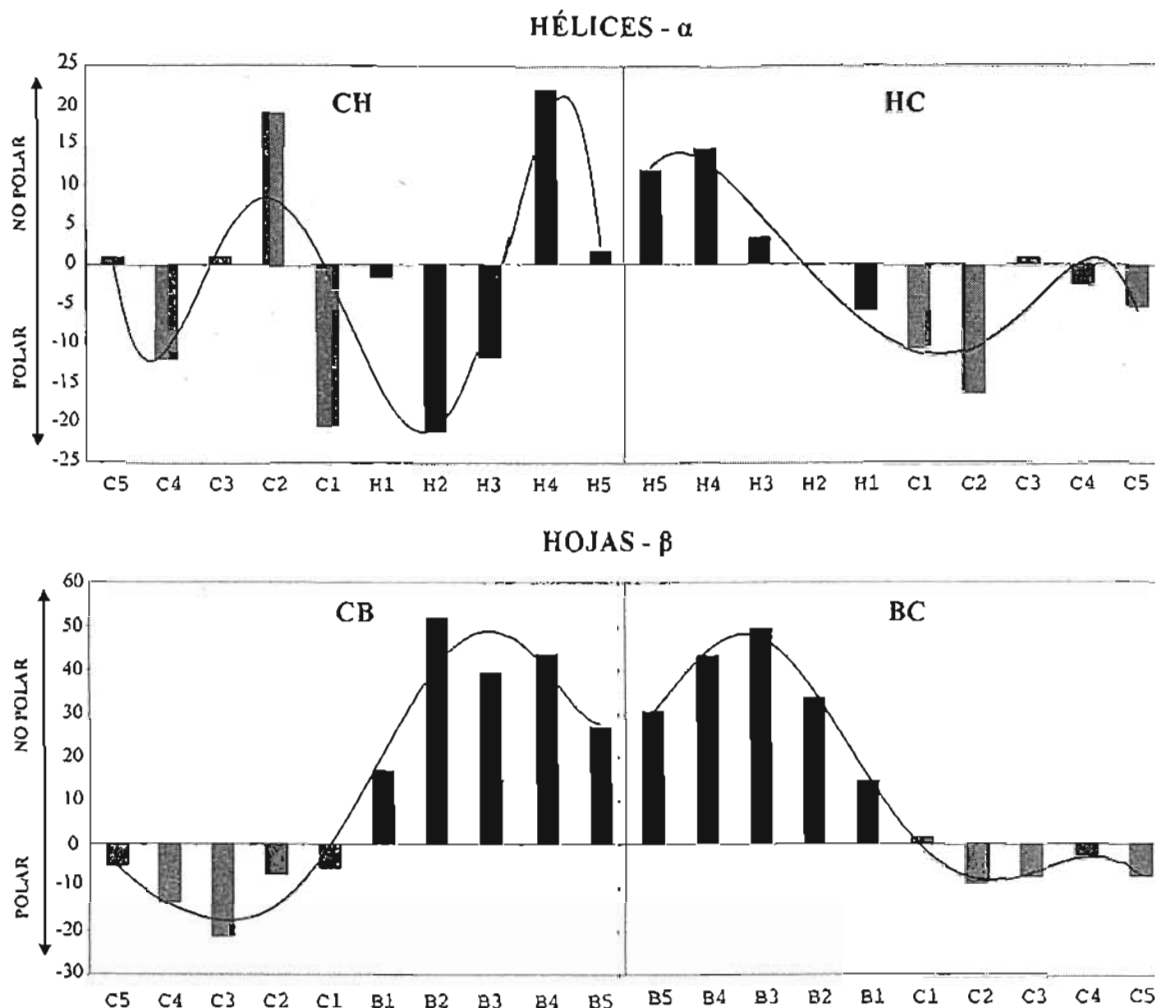


Fig. 8. Polaridad en las posiciones de las regiones terminales de hélices- α y hojas- β . Los valores se obtuvieron multiplicando las frecuencias de la tabla 3 por los valores de hidrofobicidad de Eisenberg y sumando para cada posición los valores de todos los a.a. Las curvas son las polinomiales que mejor se ajustaron a los datos.

c) Patrones de "Capping".

Los patrones de capping se caracterizan por estabilizar las regiones terminales de las hélices- α , al formar interacciones hidrofóbicas y puentes de hidrógeno con los grupos N-H y C=O libres de dichas regiones terminales. Estos patrones, descritos previamente por Aurora et al. (1998), son distintos entre las dos regiones CH y HC, debido principalmente a la orientación de las cadenas laterales en la hélice.

Tabla 4. Patrones de Capping de las dos regiones terminales de las hélices- α según Aurora, et al (1998). La zona sombreada representa los a.a. que se encuentran dentro de las hélices- α . Se muestran los resultados para los dos datasets. Total = número de regiones terminales en las cuales se podía hallar cada patrón. Casos = número de patrones hallados.

Región CH :				
Patrón de capping	Dt		Dnr	
	Total	Casos	Total	Casos
● - * <u>○ * ● *</u>	4246	392	1939	196
● - * <u>○ * ○ ●</u>	4246	366	1939	165
● ○ - * <u>○ * ● *</u>	3546	147	1596	58
● ○ - * <u>○ * ○ ●</u>	3546	116	1596	51
● ○ ○ - * <u>○ * ● *</u>	2891	41	1304	26
● ○ ○ - * <u>○ * ○ ●</u>	2891	17	1304	8
$C_4 C_3 C_2 - C_1 H_1 H_2 H_3 H_4$				
Región HC :				
Patrón de capping	Dt		Dnr	
	Total	Casos	Total	Casos
<u>● * ○ *</u> - G ●	3868	219	1784	102
<u>● * * *</u> - G ○ * ●	2330	48	1059	22
<u>● * * *</u> - G ○ * ○ ●	1783	13	805	7
<u>● * ○ *</u> - n * ●	3868	385	1408	179
<u>● * ○ *</u> - n * ○ ●	2330	165	1059	75
<u>● * * *</u> - P ○ * ●	2330	19	1059	7
<u>● * * *</u> - P ○ * ○ ●	1783	6	805	3
$H_3 H_2 H_1 C_1 - C_2 C_3 C_4 C_5 C_6$				

La frecuencia de los patrones de capping en las hélices- α , muestra que ciertos patrones tienden a aparecer más que otros (tabla 4). Para el caso de los patrones de capping característicos de la región CH, aquellos que exigen que haya un a.a. polar (H_3) antes del a.a. hidrofóbico (H_4) dentro de la hélice- α (\bullet - \ast \circ \ast \circ \bullet , \bullet \circ - \ast \circ \ast \circ \bullet y \bullet \circ \circ - \ast \circ \ast \circ \bullet), con el cual se mantiene la interacción hidrofóbica, presentan menos casos que los patrones que exigen un a.a. no polar en la posición H_3 (\bullet - \ast \circ \ast \bullet \ast , \bullet \circ - \ast \circ \ast \bullet \ast y \bullet \circ \circ - \ast \circ \ast \bullet \ast). Por lo general los patrones pequeños son los que presentan más casos. Cabe resaltar, que las frecuencias de a.a. asociadas a los valores de hidrofobicidad de Eisenberg (Fig.8) muestran que la posición H_3 y H_4 son preferentemente polar y no polar respectivamente, por lo cual se esperaría que la secuencia \ast \circ \ast \circ \bullet se viera favorecida, lo cual no sucede. Por otro lado los patrones de hidropatía en esta región (CHHHH y HHHHH) muestran así mismo una preferencia por la secuencia \ast \circ \ast \circ \bullet , en si, no existe ningún patrón hidropático típico de hélices que favorezca la secuencia \ast \circ \ast \bullet \ast . En general los patrones de hidropatía reafirman lo hallado en las frecuencias, que las posiciones C_2 y H_4 son preferentemente no polares y la posición H_3 es preferentemente polar. Por otro lado existen otras posiciones C_1 y H_2 que son claramente polares, aunque en los patrones de capping no se exige una polaridad específica. El hecho de que estas posiciones sean polares, puede deberse a que los patrones de polaridad de la estructura de una hélice- α anfipática así lo exigen.

En la región HC, se observa que los patrones que exigen la presencia de glicina o de prolina en la posición C_2 presentan pocos casos a excepción de \bullet \ast \circ \ast - G \bullet . Los patrones que exigen la presencia de 'n' en la posición C_1 , son los que aparecen mayoritariamente. Cabe mencionar que para el total de hélices- α del Dnr, los patrones de capping aparecen en el 26% de las hélices- α en la región CH y un el 22% en la región HC. Los mismos resultados se observan en Dt y Dnr.

La posición C_2 de la región HC es muy interesante. Los a.a. más frecuentes son claramente polares, los únicos no polares son G y P, lo cual cumple con el requisito de algunos patrones de capping. Para el caso de los patrones de capping que exigen P en C_2 se hallaron muy pocos casos. A este respecto, en el trabajo de Aurora R, 1998. se muestra que este tipo de capping podría requerir la presencia de hélices 3_{10} . Aun cuando se repitió el análisis tomando en cuenta las hélices 3_{10} como parte de las hélices- α , no se encontraron más patrones de capping con prolina. El caso de los patrones con 'n' se ve favorecido porque los a.a. P, I, V y T son los menos frecuentes en dicha posición. En los casos en que se exige que el a.a. sea no polar en las posiciones C_2 , C_3 y C_4 , aparecerían preferentemente G y P.

Predicción de Estructuras Secundarias.

Se obtuvieron predicciones de ES para un total de 410 proteínas del Dnr utilizando el programa PsiPred⁽²⁹⁾. Se analizaron las frecuencias de a.a. asociadas a los valores de Eisenberg, así mismo se analizó el uso de patrones hidropáticos en las regiones terminales de las ESs predichas correcta e incorrectamente.

Los resultados de los valores de Eisenberg no muestran diferencias importantes entre las regiones terminales de las estructuras predichas y las del Dnr, ni entre las estructuras predichas correcta e incorrectamente. Así mismo, el uso de patrones hidropáticos no muestra grandes diferencias entre los distintos casos considerados. Las semejanzas entre el Dnr y los resultados de las predicciones contrastan con los niveles de exactitud en la predicción para las regiones terminales. Sobresale la región CH de las hélices- α donde el nivel de predicción correcta es el mayor (60%), mientras que la región HC tiene el valor más bajo (32%). Para el caso de las regiones terminales de las hojas- β , el nivel de predicción es igual (42%). Al analizar los valores de "confianza" de la predicción que genera PsiPred se observa la misma tendencia: la región CH de las hélices- α presenta en general los valores de confianza más altos comparados con las otras regiones terminales, mientras que la región HC presenta los valores más bajos. Lo anterior se cumple tanto en las regiones terminales predichas correctamente como incorrectamente. Finalmente, en todos los casos, la última posición de la ES y la primera de "coil" de todas las regiones terminales, siempre tuvieron los valores de confianza más bajos.

DISCUSIÓN

Comparación de Estructuras Secundarias.

En el presente trabajo se muestra de una forma sencilla, que la variación de tamaño de las hélices- α es mayor que en las hojas- β entre proteínas homólogas. Lo anterior no había sido evaluado a profundidad, por no haber comparado un elevado número de proteínas homólogas de distintas familias. Así mismo, en anteriores trabajos⁽¹⁹⁾, al analizar la variación de la ID de las ESs, no se halló diferencia entre las hélices- α y las hojas- β . En nuestro caso, sí existe una clara diferencia que concuerda con los resultados del cambio de tamaño (Fig. 3), y resalta mejor el proceso en el que las hélices- α van siendo más propensas a variar que las hojas- β al ir aumentando la divergencia entre proteínas homólogas. Aun cuando este resultado puede parecer bastante intuitivo, sobretodo para investigadores relacionados con las estructuras de proteínas, estas diferencias de tamaño entre hélices- α y hojas- β pueden ser importantes para las clasificaciones estructurales⁽³⁶⁾ y para los métodos de predicción de ESs. Debido a lo anterior, se decidió analizar detalladamente las características de las regiones terminales de las hélices- α y las hojas- β y ver que relación guardan con la efectividad de la predicción de estas ES. En la tabla 5 se muestran resumidas las características analizadas de las hélices- α y las hojas- β , y de sus respectivas regiones terminales.

Los resultados de los valores de Eisenberg no muestran diferencias importantes entre las regiones terminales de las estructuras predichas y las del Dnr, ni entre las estructuras predichas correcta e incorrectamente. Así mismo, el uso de patrones hidropáticos no muestra grandes diferencias entre los distintos casos considerados. Las semejanzas entre el Dnr y los resultados de las predicciones contrastan con los niveles de exactitud en la predicción para las regiones terminales. Sobresale la región CH de las hélices- α donde el nivel de predicción correcta es el mayor (60%), mientras que la región HC tiene el valor más bajo (32%). Para el caso de las regiones terminales de las hojas- β , el nivel de predicción es igual (42%). Al analizar los valores de “confianza” de la predicción que genera PsiPred se observa la misma tendencia: la región CH de las hélices- α presenta en general los valores de confianza más altos comparados con las otras regiones terminales, mientras que la región HC presenta los valores más bajos. Lo anterior se cumple tanto en las regiones terminales predichas correctamente como incorrectamente. Finalmente, en todos los casos, la última posición de la ES y la primera de “coil” de todas las regiones terminales, siempre tuvieron los valores de confianza más bajos.

DISCUSIÓN

Comparación de Estructuras Secundarias.

En el presente trabajo se muestra de una forma sencilla, que la variación de tamaño de las hélices- α es mayor que en las hojas- β entre proteínas homólogas. Lo anterior no había sido evaluado a profundidad, por no haber comparado un elevado número de proteínas homólogas de distintas familias. Así mismo, en anteriores trabajos⁽¹⁹⁾, al analizar la variación de la ID de las ESs, no se halló diferencia entre las hélices- α y las hojas- β . En nuestro caso, sí existe una clara diferencia que concuerda con los resultados del cambio de tamaño (Fig. 3), y resalta mejor el proceso en el que las hélices- α van siendo más propensas a variar que las hojas- β al ir aumentando la divergencia entre proteínas homólogas. Aun cuando este resultado puede parecer bastante intuitivo, sobretodo para investigadores relacionados con las estructuras de proteínas, estas diferencias de tamaño entre hélices- α y hojas- β pueden ser importantes para las clasificaciones estructurales⁽³⁶⁾ y para los métodos de predicción de ESs. Debido a lo anterior, se decidió analizar detalladamente las características de las regiones terminales de las hélices- α y las hojas- β y ver que relación guardan con la efectividad de la predicción de estas ES. En la tabla 5 se muestran resumidas las características analizadas de las hélices- α y las hojas- β , y de sus respectivas regiones terminales.

Tabla 5. Características de hélices y hojas- β , y de sus regiones terminales. La columna del lado izquierdo muestra las características analizadas en el presente trabajo.

	HELICES - α		HOJAS - β	
Cambio de Tamaño	Mayor cambio		Menor cambio	
Conservación de Secuencia	Menor conservación		Mayor conservación	
	CH	HC	CB	BC
Cambio de Tamaño	Cambian en igual magnitud ambas regiones terminales. Mayor cambio que en las Hojas. Mayor número de cambios en HC.		Cambian en igual magnitud ambas regiones terminales. Menor cambio que en las Hélices.	
Presencia de Patrones Hidropáticos Típicos de cada ES	Baja en todas las posiciones excepto la más cercana a la Hélice	Alta en las 3 posiciones cercanas a la Hélice, baja en las 3 lejanas a la Hélice	Alta en las 4 posiciones cercanas a la Hoja, baja en las 2 lejanas a la Hoja	Alta en las 3 posiciones cercanas a la Hoja, baja en las 3 lejanas a la Hoja
Conservación de Secuencia	Menor en posiciones Hélice, mayor en "coil"	Varía en posiciones entre Hélice y "coil"	Mayor en posiciones Beta, menor en posiciones "coil"	Mayor en posiciones Beta, menor en posiciones "coil"
Polaridad en Posiciones de Región Terminal	Varia polaridad en posiciones entre Hélice y "coil"	Varia polaridad en Hélice, en "coil" es polar	Hojas son no polares, "coil" es polar	Hojas son no polares, "coil" es polar
Presencia de Capping	En 26% de los casos	En 22% de los casos	—	—
Predicción Correcta de Región Terminal	60%	32%	42%	42%

Características de las Regiones Terminales de Hélices- α y Hojas- β .

I. Estabilidad de regiones terminales.

Anteriormente, mediante dinámicas moleculares, se analizó la estabilidad de las hélices- α y las hojas- β . Dicha estabilidad fue medida por la tendencia de las regiones terminales a desplegarse debido a la pérdida de puentes de hidrógeno intracatenarios^(22, 23, 24). En el presente trabajo, la estabilidad se relaciona con 1) la presencia de patrones hidropáticos típicos de hélices- α y hojas- β en las regiones terminales, 2) con la conservación de secuencia en estas regiones y 3) con la presencia de patrones de capping en las hélices. Todas estas características de las regiones terminales se relacionan con el cambio de tamaño de las ESs por tener un efecto local (interacciones cercanas). Por lo anterior, una ES que presente patrones favorables y una alta conservación en sus regiones terminales, se puede considerar que presenta una mayor estabilidad evolutiva y por ende es menos propensa a cambiar de tamaño.

Cabe mencionar que la presencia de un patrón favorable para una ES no asegura su formación, ya que en muchos casos depende de interacciones con otras regiones de la estructura de la proteína que se encuentran lejanas en secuencia. En este sentido, la polaridad de los a.a. tiene un claro efecto en la interacción con otras partes de la proteína⁽³³⁾ y con el solvente. Por esta razón, los patrones hidropáticos analizados son indicativos de la interacción de las regiones terminales con otras regiones lejanas en secuencia. Finalmente, aun cuando no se consideraron todas las interacciones lejanas, los resultados muestran que las diferencias halladas entre las regiones terminales de hélices- α y hojas- β concuerdan con sus diferencias en el cambio de tamaño.

II. Diferencias entre regiones terminales.

Las frecuencias de aminoácidos en las distintas posiciones de las regiones terminales (tabla 3) no muestran grandes diferencias con las reportadas anteriormente⁽⁴²⁾. Cabe mencionar que la presencia de la prolina en la posición H₁ de la región CH de las hélices- α ya ha sido descrita⁽⁴²⁾ y se menciona que es favorecida debido a que sus ángulos phi y psi no afectan el inicio de la hélice- α . Además, al presentar el anillo pirrolidínico, evita la necesidad de satisfacer puentes de hidrógeno en esta región por lo que estabiliza la región terminal. En cierta medida se puede considerar que la presencia de la prolina en esta posición desempeña el mismo papel de los patrones de capping.

Los patrones de capping (tabla 4) que tienen más casos en ambas regiones terminales de las hélices- α se ven favorecidos por las preferencias de polaridad en cada posición (Fig. 8). Aun así existen posiciones con una marcada polaridad, lo cual resulta indiferente al proceso de capping. Esto se debe a que la estructura del capping sólo requiere de interacciones hidrofóbicas específicas entre a.a. a determinadas distancias⁽⁴²⁾. Los a.a. en otras posiciones estarían en tal caso involucrados en interacciones con el solvente u otras regiones de la proteína.

En el caso específico de los patrones de hidropatía (Tabla 2) y los valores de entropía (Fig. 7), se muestra que las hojas- β tienen regiones terminales más estables o menos propensas a modificar su tamaño, en comparación con las hélices- α . En las

hélices- α , la región CH muestra una menor estabilidad en comparación con la región HC. Esto último también se encontró con los valores de entropía en estas regiones. Lo anterior muestra que al nivel de secuencia de a.a., hay una mayor estabilidad en las hojas- β que en las hélices- α . Por otro lado, dinámicas moleculares y trabajos experimentales^(22, 23, 24, 25, 27) han hallado que la región HC de las hélices- α es más propensa a desplegarse que la región CH, y nuestros resultados muestran que la región HC cambia un mayor número de veces, lo cual está en concordancia con las dinámicas moleculares. Por otro lado, si los resultados de las dinámicas moleculares son correctos, resulta interesante encontrar que justamente la región HC que es más propensa a desplegarse, presenta una mayor conservación de secuencia y de uso de patrones hidropáticos típicos de hélices- α . Podría suponerse que a esta tendencia de desplegarse se le contraponen una mayor conservación. Finalmente, se ha mencionado que las hélices- α presentan diferentes tendencias a crecer en sus regiones terminales debido a los diferentes ambientes químicos entre CH y HC⁽²²⁾. Nuestros resultados muestran que ambas regiones terminales en las hélices- α cambian igual de tamaño (Fig. 6), de no existir algún sesgo por la forma en que se realizó el análisis (alineamientos estructurales, ver métodos), parecería que las diferencias en conservación entre CH y HC se contraponen a la tendencia a desplegarse de estas regiones, con lo cual se evitaría que alguna región terminal presente en promedio un mayor cambio de tamaño.

III. Patrones de hidropatía en ES.

Los patrones de hidropatía más frecuentes de hélices- α , hojas- β y “coil” muestran tendencias interesantes (tabla 1). En el trabajo de West y Hecht (1995), se menciona que los 7 patrones ideales para la formación de hélices- α surgen de analizar la distribución de a.a. en una hélice anfipática ideal (○●○●○●○●○●○●○●○●●). En el presente trabajo, analizamos los patrones más frecuentes de hélices- α , y se encontró que no solamente permiten formar estas hélices- α anfipáticas, sino que además permiten la formación de una hélice- α anfipática con un carácter más hidrofóbico (Fig. 9a). Si únicamente consideramos los patrones más usados por las hélices- α y al mismo tiempo no usados por otras estructuras, se pueden formar estos dos tipos de hélices anfipáticas; si se quisiera hacer una hélice- α con una superficie más hidrofílica o más hidrofóbica que aquellas que se muestran en la Fig. 9a, se requeriría usar patrones característicos de hojas- β y “coil”s. Lo anterior muestra que las hélices- α anfipáticas, dependiendo de la estructura de la proteína y su secuencia, podrían estar en más o menos contacto con el solvente y el core de la proteína.

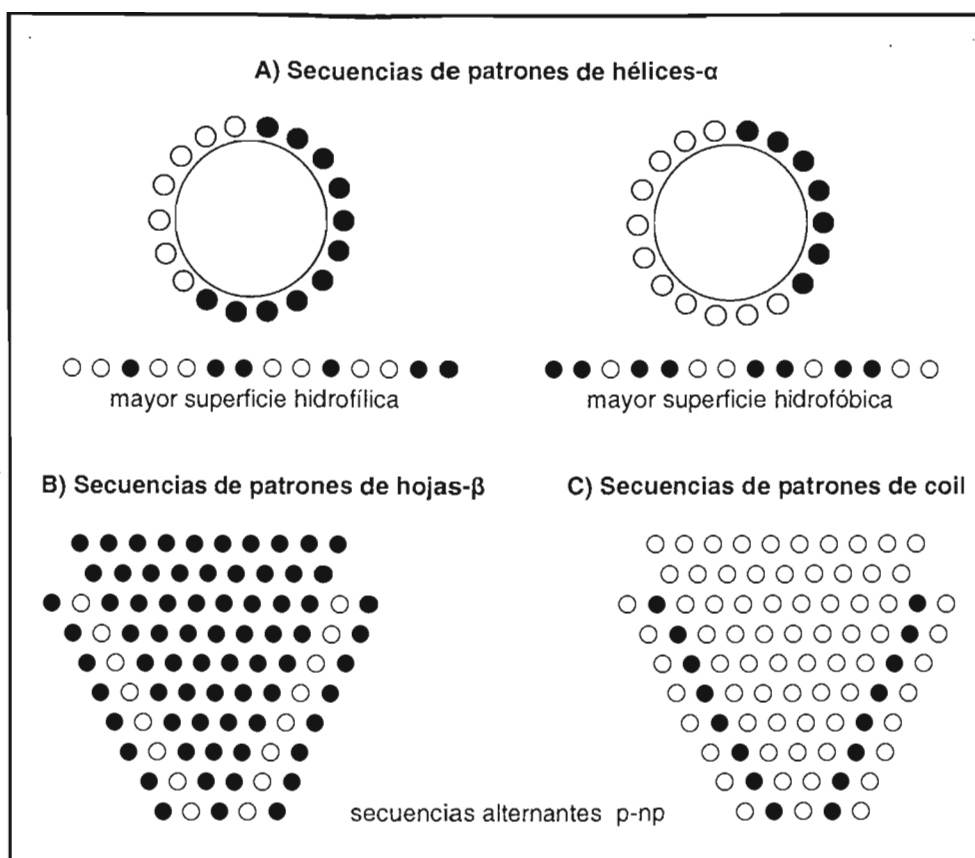


Fig. 9. Secuencias generables con los patrones hidropáticos más frecuentes de cada estructura. En las hélices- α se muestra en una representación circular la posición de los residuos polares y no polares.

En el caso de las hojas- β , los patrones más frecuentes permiten formar secuencias que van desde hojas completamente hidrofóbicas, típicas de hojas- β paralelas⁽³⁴⁾, hojas- β hidrofóbicas con a.a. polares en los extremos y secuencias alternantes de a.a. polares y no polares, típicas de hojas- β antiparalelas y extremos de hojas paralelas (Fig. 9). Exactamente lo contrario sucede para las regiones “coil”, donde se pueden formar secuencias completamente polares, con a.a. no polares en los extremos, hasta secuencias binarias alternantes (Fig. 9). Lo anterior muestra que los patrones hallados como más frecuentes, efectivamente responden a las características propias de las estructuras secundarias⁽¹⁷⁾. Por lo tanto las diferencias halladas en el uso de estos patrones en las fronteras de hélices- α y hojas- β , pueden ser de importancia para el diseño y predicción de estas regiones.

En trabajos previos se ha mencionado que el diseño de proteínas no sólo requiere que la secuencia favorezca la formación de la estructura secundaria deseada (diseño positivo), además debe desfavorecer estructuras alternativas (diseño negativo)⁽¹⁷⁾. A este respecto cabe mencionar que anteriormente no se habían analizado las preferencias en el uso de patrones hidropáticos para las regiones “coil”. En el presente trabajo se muestra que hay patrones que aparecen preferentemente en el estado “coil” y no en las hélices- α y

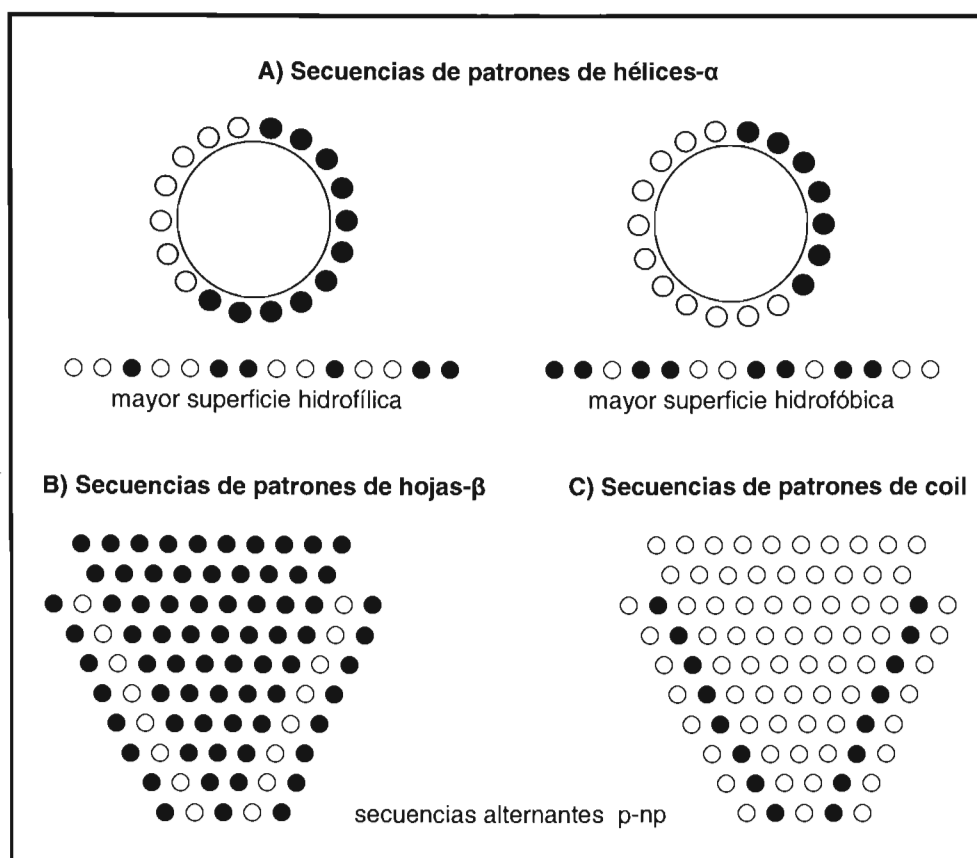


Fig. 9. Secuencias generables con los patrones hidropáticos más frecuentes de cada estructura. En las hélices- α se muestra en una representación circular la posición de los residuos polares y no polares.

En el caso de las hojas- β , los patrones más frecuentes permiten formar secuencias que van desde hojas completamente hidrofóbicas, típicas de hojas- β paralelas⁽³⁴⁾, hojas- β hidrofóbicas con a.a. polares en los extremos y secuencias alternantes de a.a. polares y no polares, típicas de hojas- β antiparalelas y extremos de hojas paralelas (Fig. 9). Exactamente lo contrario sucede para las regiones "coil", donde se pueden formar secuencias completamente polares, con a.a. no polares en los extremos, hasta secuencias binarias alternantes (Fig. 9). Lo anterior muestra que los patrones hallados como más frecuentes, efectivamente responden a las características propias de las estructuras secundarias⁽¹⁷⁾. Por lo tanto las diferencias halladas en el uso de estos patrones en las fronteras de hélices- α y hojas- β , pueden ser de importancia para el diseño y predicción de estas regiones.

En trabajos previos se ha mencionado que el diseño de proteínas no sólo requiere que la secuencia favorezca la formación de la estructura secundaria deseada (diseño positivo), además debe desfavorecer estructuras alternativas (diseño negativo)⁽¹⁷⁾. A este respecto cabe mencionar que anteriormente no se habían analizado las preferencias en el uso de patrones hidropáticos para las regiones "coil". En el presente trabajo se muestra que hay patrones que aparecen preferentemente en el estado "coil" y no en las hélices- α y

las hojas- β . Así mismo, los resultados sobre las regiones terminales pueden ser de utilidad para estos fines.

Finalmente, cabe mencionar que los patrones hidropáticos tienden a modificarse de forma distinta. Los patrones más característicos de las hojas- β y “coil”, al cambiar un a.a de una polaridad a otra en cualquiera de sus 5 posiciones, dan origen a un patrón hidropático que favorece la misma estructura. Las hélices- α por el contrario dan origen principalmente a patrones que no favorecen ninguna estructura en particular. Lo anterior también es indicio de una mayor susceptibilidad de la hélices- α a cambiar de tamaño, aunque hay que repetirlo, en este trabajo no se están tomando en cuenta las interacciones lejanas que pueden estabilizar una región de la cadena polipeptídica favoreciendo la presencia de una estructura como hélice- α aun cuando la secuencia de aminoácidos no favorezca esta estructura.

IV. Predicción y estabilidad de regiones terminales.

Los anteriores resultados de las regiones terminales de ambas ESs, mostraron claras diferencias entre estas y en el caso de las hélices- α diferencias entre sus regiones CH y HC. Debido a lo anterior se esperaba así mismo que hubiera una diferencia en los niveles de predicción en estas regiones. Se sabe que los niveles de predicción, con distintos métodos, son mejores para hélices- α que para hojas- β y regiones “coil”⁽⁴⁵⁾, pero no se tenía tan claro como variaban las predicciones en las distintas regiones terminales. Para sorpresa nuestra, la región CH, que mostró ser poco estable por su baja conservación y uso de patrones hidropáticos, resultó tener los mejores niveles de predicción, aun por arriba de las hojas- β .

Al principio del presente trabajo se encontró que ambas regiones terminales de las hélices- α cambian de tamaño en la misma magnitud, aunque el número de casos es mayor para la región HC. La predicción de ESs muestra que esta región tiene los valores de predicción más bajos, así mismo como se mencionó anteriormente es la región donde se ha hallado, mediante simulaciones, que se despliega más fácilmente la estructura de hélices. Por otro lado la región HC de las hélices- α es así mismo la región con más bajos niveles de confianza en la predicción. Dado que el método empleado en el presente trabajo forma parte de los métodos que dan mejores predicciones, este resultado nos parece muy interesante.

La mayoría de estos métodos emplea información evolutiva⁽²⁸⁾ (al igual que nuestro análisis de las diferentes usos de patrones y conservación) y algunos, como el caso de PSIPRED, se basan en métodos de redes neurales⁽²⁹⁾. A este respecto cabe mencionar que se ha observado que esta metodología puede provocar bajos niveles de predicción en las hojas- β debido al intervalo de a.a. tomados en cuenta en cada paso de la predicción. Un intervalo de entre 15-25 residuos podrá contener dos hebras de una hoja sólo si estas son adyacentes y tienen tamaños entre 5 y 6 residuos. Es por esta razón que hebras más grandes tienen bajos niveles de predicción y por ende sus regiones terminales pueden no estar bien predichas. Aun así, no parece haber alguna razón para la baja predicción en la región HC de hélices- α y la alta predicción en la CH a menos que se deba a señal evolutiva, la cual no identificamos claramente con la metodología empleada

en este trabajo. Una posibilidad se halla en las posiciones polares y no polares (Fig. 8), ya que la región CH muestra posiciones de muy alta y muy baja polaridad en comparación con la región HC.

Modificación de Tamaño de Hélices- α y Hojas- β .

Una posible razón por la cual las hélices- α cambian más de tamaño es que aquellas hélices- α exteriores a la estructura 3D (que son la mayoría) presentan la tendencia de tener un número entero de vueltas ⁽⁴³⁾. Cuando la secuencia de a.a. no favorece la formación de hélices- α en la región terminal, el cambio de tamaño es de por lo menos 3 residuos (una vuelta de la hélice), ya que cuando cambia la secuencia y no existen las condiciones adecuadas para que permanezca la conformación de hélice- α en la región terminal, es preferible sacrificar otros aminoácidos más de la estructura. Lo anterior se debe a que es preferible tener un mejor contacto, por parte de la hélice- α , con el core de la proteína ⁽⁴³⁾. De esta forma se logra que la nueva región terminal interactúe mejor con el solvente. Al mismo tiempo, es una posible razón del porque ambas regiones terminales cambian de tamaño en la misma magnitud.

Empleando la misma idea de interacciones con el solvente y otras regiones de la proteína, es posible explicar porque aparecen patrones hidropáticos tipo “coil” en las regiones terminales de las hélices- α . Aun cuando el patrón sea típico de “coil”, se podría formar una hélice- α , siempre y cuando pueda tener contactos adecuados con el core y con el solvente, es decir que existan interacciones hidrofóbicas adecuadas y que permanezca una cara polar para interactuar con el solvente. Este razonamiento implica que la estabilidad (producto de la conservación de secuencia y uso de patrones) depende en parte del ambiente fisicoquímico en el cual se encuentra la región terminal. La variación de la conservación y de patrones hallados en las hélices- α en comparación con las hojas- β es muestra de que sus regiones terminales se enfrentan a una mayor diversidad de ambientes fisicoquímicos e interacciones con la proteína y el solvente.

Para el caso de las hojas- β es razonable pensar que las hebras externas son más susceptibles a cambiar de longitud sin alterar a las otras hebras internas de la hoja- β . Por otro lado las hebras internas tienen más restricciones estructurales para cambiar puesto que están estabilizadas por puentes de hidrógeno en ambos lados. Debido a lo anterior, el cambio de secuencia en una hebra podría no tener efecto alguno en la estructura de la hoja- β . Además, los resultados del presente trabajo muestran que las hojas- β presentan fronteras más estables.

Las cadenas laterales tienen un efecto diferente dependiendo si se encuentran en hojas- β o en hélices- α ⁽¹³⁾. Esto se debe a que la conformación de sus rótameros, impone una baja restricción estérica en las hojas- β , debido a que la conformación es, al igual que el estado “coil”, una conformación extendida y flexible ⁽⁴⁶⁾. En cambio las hélices- α si se ven más afectadas por las conformaciones de las cadenas laterales de algunos residuos, especialmente por los que presentan ramificaciones en el C ^{β} de la cadena lateral (V, I, T) ⁽⁴²⁾. Por otro lado, se han realizado trabajos sobre la formación de hojas- β , en los cuales

se muestra que una vez que la hoja- β ha alcanzado una barrera energética (no dependiente de la longitud), originada por interacciones que la estabilizan (puentes de hidrógeno y/o cargas parciales), difícilmente decrecerá del tamaño que ha alcanzado ⁽²⁴⁾. Así mismo, la mayor flexibilidad de las hojas- β , les permite soportar cambios de secuencia sin alterar su estructura, lo que no sucede en el caso de las hélices- α .

CONCLUSIÓN

En el presente trabajo se analizaron las diferencias entre las dos principales ESs de las proteínas. La idea principal era considerar las relaciones evolutivas entre familias de proteínas para llevar a cabo dicho trabajo. El problema de la diferencia de tamaño entre hélices y hojas resulta de importancia para entender la forma en que estas estructuras varían y para los métodos de predicción de estructura secundaria. Debido a que la variación estructural entre proteínas homólogas es un proceso que depende tanto de factores locales o cercanos en la secuencia como de factores lejanos, el realizar un análisis del cambio en las estructuras secundarias se antoja complicado. Debido a lo anterior, se decidió tomar una aproximación sencilla pero que resultara lo suficientemente informativa, como para distinguir que factores tienen una mayor influencia en la modificación de las ES. Los resultados muestran una mayor estabilidad evolutiva de las hojas- β sobre las hélices- α , así como claras diferencias entre sus regiones terminales, las cuales no se ven del todo reflejadas en las predicciones de ES. Lo anterior seguramente se debe a la metodología que sigue Psipred ⁽²⁹⁾ para realizar la predicción en la cual es capaz de obtener información de carácter evolutiva para determinar correctamente cada región terminal.

Finalmente, no hay que perder de vista dos características de cada estructura secundaria que participan en su modificación. Por un lado las hélices- α representan una estructura rígida ⁽⁴⁶⁾ (un cilindro de poca flexibilidad) pero con una mayor capacidad de modificación en sus extremos, debido a su baja estabilidad en las regiones terminales. Por otro lado las hojas- β presentan una estructura muy flexible ⁽⁴⁶⁾, pero con una capacidad de modificación de sus extremos muy baja. Al ser estas dos las estructuras secundarias más abundantes en las proteínas y presentar características distintas en términos de formación, crecimiento y estabilización, otorgan una mayor posibilidad de variables estructurales y formas en que la estructura 3D de la proteína se puede estabilizar.

se muestra que una vez que la hoja- β ha alcanzado una barrera energética (no dependiente de la longitud), originada por interacciones que la estabilizan (puentes de hidrógeno y/o cargas parciales), difícilmente decrecerá del tamaño que ha alcanzado ⁽²⁴⁾. Así mismo, la mayor flexibilidad de las hojas- β , les permite soportar cambios de secuencia sin alterar su estructura, lo que no sucede en el caso de las hélices- α .

CONCLUSIÓN

En el presente trabajo se analizaron las diferencias entre las dos principales ESs de las proteínas. La idea principal era considerar las relaciones evolutivas entre familias de proteínas para llevar a cabo dicho trabajo. El problema de la diferencia de tamaño entre hélices y hojas resulta de importancia para entender la forma en que estas estructuras varían y para los métodos de predicción de estructura secundaria. Debido a que la variación estructural entre proteínas homólogas es un proceso que depende tanto de factores locales o cercanos en la secuencia como de factores lejanos, el realizar un análisis del cambio en las estructuras secundarias se antoja complicado. Debido a lo anterior, se decidió tomar una aproximación sencilla pero que resultara lo suficientemente informativa, como para distinguir que factores tienen una mayor influencia en la modificación de las ES. Los resultados muestran una mayor estabilidad evolutiva de las hojas- β sobre las hélices- α , así como claras diferencias entre sus regiones terminales, las cuales no se ven del todo reflejadas en las predicciones de ES. Lo anterior seguramente se debe a la metodología que sigue Psipred ⁽²⁹⁾ para realizar la predicción en la cual es capaz de obtener información de carácter evolutiva para determinar correctamente cada región terminal.

Finalmente, no hay que perder de vista dos características de cada estructura secundaria que participan en su modificación. Por un lado las hélices- α representan una estructura rígida ⁽⁴⁶⁾ (un cilindro de poca flexibilidad) pero con una mayor capacidad de modificación en sus extremos, debido a su baja estabilidad en las regiones terminales. Por otro lado las hojas- β presentan una estructura muy flexible ⁽⁴⁶⁾, pero con una capacidad de modificación de sus extremos muy baja. Al ser estas dos las estructuras secundarias más abundantes en las proteínas y presentar características distintas en términos de formación, crecimiento y estabilización, otorgan una mayor posibilidad de variables estructurales y formas en que la estructura 3D de la proteína se puede estabilizar.

REFERENCIAS

1. Russell R, Saqi M, Sayle R, Bates P, Sternberg M. 1997. Recognition of analogous and homologous protein folds: Analysis of sequence and structure conservation. *J. Mol. Biol.* 269:423-439.
2. Wood T, Pearson W. 1999. Evolution of protein sequences and structures. *J. Mol. Biol.* 291:977-995.
3. Gan H, Perlow R, Roy S, Ko J, Wu M, Huang J, Yan S, Nicoletta A, Vafai J, Sun D, Wang L, Noah J, Pasquali S, Schlick T. 2002. Analysis of protein sequence/structure similarity relationships. *Biophys. J.* 83:2781-2791.
4. Koehl P, Levitt M. 2002. Sequence variations within protein families are linearly related to structural variations. *J. Mol. Biol.* 323:551-562.
5. Chothia C, Lesk A. 1986. The relation between the divergence of sequence and structure in proteins. *EMBO J.* 5:823-826.
6. Murzin A, Brenner S, Hubbard T, Chothia C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536-540.
7. Orengo C, Michie A, Jones S, Jones D, Swindells M, Thornton M. 1997. CATH- A Hierarchic Classification of Protein Domain Structures. *Structure.* 5:1093-1108.
8. Thorne J, Goldman N, Jones D. 1996. Combining protein evolution and secondary structure. *Mol. Biol. Evol.* 13:666-673.
9. Tourasse N, Li W. 2000. Selective constraints, amino acid composition, and the rate of protein evolution. *Mol. Biol. Evol.* 17:656-664.
10. Liu X, Zhang L, Guan S, Zheng W. 2003. Distances and classification of amino acids for different protein secondary structures. *Phys. Rev.* 67:051927 1-7.
11. Koshi J, Goldstein R. 1995. Context-dependent optimal substitution matrices. *Protein Engng.* 8:641-645.
12. Xiong H, Buckwalter B, Shieh H, Hecht M. 1995. Periodicity of polar and nonpolar amino acids is the mayor determinant of secondary structure in self-assembling oligomeric peptides. *Proc. Natl. Acad. Sci.* 92:6349-6353.
13. Chakrabarti P, Pal, D. 2001. The interrelationships of side-chain and main-chain conformations in proteins. *Prog. Biophys. Mol. Biol.* 76:1-102.
14. Roder H. 2004. Stepwise helix formation and chain compaction during protein folding. *Proc. Natl. Acad. Sci.* 101:1793-1794.
15. Zhu Z, Blundell T. 1996. The use of amino acid patterns of classified helices and strands in secondary structure prediction. *J. Mol. Biol.* 260:261-276.
16. Khoehl P, Levitt M. 1999. Structure-based conformational preferences of amino acids. *Proc. Natl. Acad. Sci.* 96:12524-12529.
17. West M, Hecht M. 1995. Binary patterning of polar and nonpolar amino acids in the sequences and structures of native proteins. *Protein Sci.* 4:2032-2039.
18. Hubbard T, Blundell T. 1987. Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modelling. *Protein Eng.* 1:159-171.
19. Flores T, Orengo C, Moss D, Thornton J. 1993. Comparison of conformational characteristics in structurally similar protein pairs. *Protein Sci.* 2:1811-1826.
20. Srinivasan R, Rose G. 1999. A physical basis for protein secondary structure. *Proc. Natl. Acad. Sci.* 96:14259-14263.
21. Mizuguchi K, Blundell, T. 2000. Analysis of conservation and substitutions of secondary structure elements within protein superfamilies. *Bioinformatics.* 16:1111-1119.
22. Young S, Brooks C. 1996. A microscopic view of helix propagation: N and C-terminal helix growth in alanine helices. *J. Mol. Bio.* 259:560-572.
23. Muñoz V, Thompson P, Hofrichter J, Eaton W. 1997. Folding dynamics and mechanism of β -hairpin formation. *Nature.* 390:196-199.

24. Cieplak M, Xuan T, Robbins M. 2002. Thermal folding and mechanical unfolding pathways of protein secondary structures. *Proteins*. 49:104-113.
25. Takano M, Yamato T, Higo J, Suyama A, Nagayama K. 1999. Molecular dynamics of a 15-residue poly(L-alanine) in water: Helix formation and energetics. *J. Am. Chem. Soc.* 121:605-612.
26. Bolin K, Millhauser G. 1999. α and 3_{10} : The split personality of polypeptide helices. *Acc. Chem. Res.* 32:1027-1033.
27. Goch G, Maciejczyk M, Oleszczuk M, Stachowiak D, Malicka J, Bierzynski A. 2003. Experimental investigation of initial steps of helix propagation in model peptides. *Biochemistry*. 42:6840-6847.
28. Rost B. 2001. Review: Protein secondary structure prediction continues to rise. *J. Struct. Biol.* 134:204-218.
29. Jones D. 1999. Protein secondary structure based on position-specific scoring matrices. *J. Mol. Biol.* 292:195-202.
30. Chandonia J, Karplus M. 1999. New methods for accurate prediction of protein secondary structure. *Proteins*. 35:293-306.
31. Deleage G, Blanchet C, Geourjon C. 1997. Protein structure prediction. Implications for the biologist. *Biochimie*. 79:681-686.
32. Aurora R, Creamer T, Srinivasan R, Rose G. 1997. Local interactions in protein foldings: Lessons from the α -Helix. *J. Biol. Chem.* 272:1413-1416.
33. Broome B, Hecht M. 2000. Nature disfavors sequences of alternating polar and non-polar amino acids: Implications for amyloidogenesis. *J. Mol. Biol.* 296:961-968.
34. Mandel-Gutfreund Y, Gregoret L. 2002. On the significance of alternating patterns of polar and non-polar residues in XXX-strands. *J. Mol. Biol.* 323:453-461.
35. Russell R, Barton G. 1994. Structural features can be unconserved in proteins with similar folds. *J. Mol. Biol.* 244:332-350.
36. Przytycka T, Aurora R, Rose G. 1999. A protein taxonomy based on secondary structure. *Nature*. 6:672-682.
37. Mizuguchi K, Deane C, Blundell T, Overington J. 1998. HOMSTRAD: A database of protein structure alignments for homologous families. *Protein Sci.* 7:2469-2471.
38. Mallika V, Bhaduri A, Sowdhamini R. 2002. PASS2: a semi-automated database of protein alignments organized as structural superfamilies. *Nucleic Acids Res.* 30:284-288.
39. Weizhong Li. CD-HIT: Sequence clustering software. <http://bioinformatics.org/cd-hit/>.
40. Henikoff S, Henikoff J. 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.* 89:10915-10919.
41. Thompson J, Higgins D, Gibson T. 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.* 22:4673-4680.
42. Aurora R, Rose G. 1998. Helix capping. *Protein Sci.* 7:21-38.
43. Penel S, Morrison R, Mortishire-Smith R, Doig A. 1999. Periodicity in α -Helix lengths and C-Capping preferences. *J. Mol. Biol.* 293:1211-1219.
44. Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I, Bourne P. 2000. The protein data bank. *Nucleic Acids Res.* 28:235-242.
45. Westhead D, Thornton J. 1998. Protein structure prediction. *Curr. Opin. Biotech.* 9:383-389.
46. Chothia C, Hubbard T, Brenner S, Barns H, Murzin A. 1997. Protein folds in the All- β and All- α classes. *Annu. Rev. Biophys. Biomol. Struct.* 26:597-627.