

UCH-FC  
MAG-BMCN  
D452  
C.1

**SECUENCIAS PALINDRÓMICAS Y SIMETRÍA DE ESTRUCTURA  
SECUNDARIA EN ALGUNAS PROTEÍNAS DE ESTRUCTURA  
CONOCIDA**

**Tesis entregada a la Universidad de Chile en cumplimiento parcial de los requisitos  
para optar al grado de Magíster en Ciencias Biológicas con mención en Biología  
Molecular, Celular y Neurociencias  
Facultad de Ciencias**

**por**

**CLAUDIO MANUEL DÍAZ SALDAÑA**

**Julio de 2005**

**Director de tesis: DR.DAVID HOLMES**

FACULTAD DE CIENCIAS

UNIVERSIDAD DE CHILE

INFORME DE APROBACIÓN

TESIS DE MAGÍSTER

Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de

Magíster

presentada por el candidato

Claudio Manuel Díaz Saldaña

Ha sido aprobada por la Comisión de evaluación de la tesis como requisito para optar al grado de Magíster en Ciencias Biológicas con Mención en Biología Molecular, Celular y Neurociencias, en el examen de Defensa de Tesis rendido el día 06/12/ 2004

Director de tesis: Dr. David Holmes

Comisión de Evaluación de la Tesis

Dr. Octavio Monasterio

Dr. Jorge Babul

Dr. Osvaldo Álvarez

Dr. Carlos Jerez

David S. Holmes

Juan Mateo



Seals and signatures of the thesis evaluation committee members.

**Esta tesis fue realizada en el Laboratorio de Bioinformática y Microbiología Molecular, Facultad de Química y Biología, Universidad de Santiago de Chile, bajo la dirección del Dr. David Holmes y fue financiada parcialmente por el Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT, proyecto 1980665).**



**A Millaray, Thiago, Natalie, Yolanda, y Enzo.**







Tengo poca tolerancia a recibir un “NO” por respuesta. Así que una amalgama de porfía y desencanto han orquestado los acordes de mi personalidad. Desde que tengo memoria, me recuerdo observando a las personas, las arañas, las hormigas, los perros, las pulgas y algunas combinaciones de éstos peleando entre sí. Nací en Santiago el 13 de diciembre de 1969. Cursé mis estudios de enseñanza básica en el colegio E 382 de la comuna de Pudahuel, rodeado en esa época de lomas blancas e inhabitadas, y los de media en el Instituto Nacional José Miguel Carrera, sin duda un gran contrincante para un periférico como yo. En 1989 ingresé promisorio y con puntaje nacional en química, a la carrera de Química y Farmacia de la Facultad de Ciencias Químicas y Farmacéuticas de la Universidad de Chile. Lo promisorio me duró hasta que me enamoré de una mujer que, obviamente, ya no recuerdo, así que gustoso repetí un par de ramos. En esa época ya había comenzado a escribir poesía (gané un par de concursillos), a tomar una serie de electivos de naturaleza esotérica (como “Química Inorgánica de los Procesos Biológicos”) y a dejarme llevar por una tenue promiscuidad que, por supuesto, ya no practico. Entonces conocí a una brasileña encantadora y me casé con ella antes de aprenderme la acentuación de su apellido. Esta mujer implacable me ha obligado, señores, a formar una familia constituida mientras peleo con algunos molinos de viento, lo cual recomiendo como un excelente ejercicio mental para mantener la cordura. Me titulé de Química y Farmacia en 1997 tras realizar mi tesis en el área de electrofisiología con el Dr. Manuel Kukuljan en la Facultad de Medicina Norte de la misma Universidad. Luego entré a este magíster y, tras algunas interrupciones, terminé esta tesis con el Dr. David Holmes (al principio en la USACH y actualmente en la Universidad Andrés Bello). Tengo poca tolerancia a recibir un “NO” por respuesta.

Claudio Díaz Saldaña

## AGRADECIMIENTOS

Agradezco por su valioso apoyo a mi sinusoidal y, espero, ascendente trayectoria a Javier Díaz, Bernarda Llecún y familia, Claudio Cortés, Felipe Veloso, Gonzalo Riadi, Profesor Malo Huencho, Antonio Muñoz, Lea Araújo y Humberto Duarte, Dr. Octavio Monasterio, Dr. Carlos Medina, Dra. Daniela Seelenfreund, Dr. Manuel Kukuljan, Rodolfo Saldaño, Dr. Ariel Orellana, Dra. Marcela Wilkens y colaboradores, Julio Roblero, Dra. Victoria Gixé, Jorge Valdés, Dr. Omar Orellana, Dr. Hugo Cárdenas, Marisol e Inés, varios compañeros de Farmacias Ahumada.

En especial al Dr. David Holmes, cuya mentalidad fue para mí un amplio y verde paisaje en el cual me senté a pensar tranquilo mientras los pájaros cantaban. A la Dra. Eugenia Jedlicki, por comprender la mentalidad del Dr. David Holmes.

Muy especialmente a Alessandra Araújo, Yolanda Saldaña y Magaly Díaz, por compartir la trinchera conmigo.



## RESUMEN

El carácter avectorial de muchas de las interacciones estabilizadoras en una proteína sugiere que la misma secuencia podría, potencialmente, plegarse en una estructura similar, ya sea orientada del extremo amino al carboxilo o en el sentido invertido (palindrómico). Se conoce varios ejemplos de proteínas con secuencias invertidas. Sin embargo, no hay consenso de opinión respecto de la capacidad de las secuencias invertidas para plegarse en estructuras nativas. De este modo, se ha publicado resultados experimentales y bioinformáticos contradictorios.

En este trabajo, se estudió la relación entre la identidad de la secuencia aminoacídica y la simetría de estructura secundaria entre las dos mitades de unos segmentos palindrómicos y pseudo palindrómicos. Se encontró que las secuencias aminoacídicas con 40% de identidad o más entre las orientaciones inversa y directa exhibieron una mayor simetría de estructura secundaria que los segmentos control escogidos al azar.

Además, se encontró que los residuos no idénticos en las dos mitades de las secuencias pseudo palindrómicas tendían a estar enriquecidos en residuos que promueven la simetría del palíndrome. Estos se obtiene ya sea enfrentando residuos como P, L, V, F, R, Y en las mitades del palíndrome o substituyendo los residuos formadores de puentes de hidrógeno y estabilizadores de vueltas E, N, Q, S, H, D. Para explicar esto, se propone que la heterogeneidad de densidades de carga en la estructura del último grupo de residuos podría imponer condiciones de interacción

estereoespecíficas que son incapaces de mantenerse cuando la secuencia de la proteína se invierte. El hecho de que algunas sustituciones de residuos adquieren una mejor simetría de estructura secundaria que los residuos originales en la secuencia invertida, reduce la atinencia de usar los criterios de conservación clásicos para medir la relevancia biológica de las secuencias palindrómicas.

Finalmente, especulo respecto del posible significado biológico de las secuencias palindrómicas en las proteínas y como ellas pudieron haber sido seleccionadas positivamente durante la evolución.



## SUMMARY

The avectorial character of many of the stabilizing interactions in a protein suggest that the same sequence could, potentially, fold into a similar structure, either oriented from the amino to carboxyl end or in the inverted (palindromic) sense. Various examples of proteins with inverted sequences are known. However, there is no consensus of opinion regarding the ability of inverted sequences to fold into native structures. Indeed, contradictory experimental and bioinformatic results have been published.

In this work, the relationship between the identity of amino acid sequence and secondary structure symmetry between the two halves of palindromic and pseudo-palindromic segments was studied. It was found that amino acids sequences with 40% or greater identity between the sense and antisense orientations exhibited a higher symmetry of secondary structure than random chosen control segments.

In addition, it was found that non-identical residues in the two halves of pseudo-palindromic sequences tend to be enriched in residues that promote the symmetry in the palindrome. This is achieved either by opposing residues like P, L, V, F, R, Y in the halves of the palindrome or by substituting the hydrogen bonding and turn-stabilizing residues E, N, Q, S, H, D. To explain this, it is proposed that the heterogeneity of charge densities in the last group residue structure could impose stereospecific conditions of interactions that are unable to be maintained when the protein sequence is inverted. The fact that some residue substitutions achieve a better secondary structure symmetry than the original residues in the inverted

**sequence, reduces the adequacy of using the classical sequence conservation criteria to measure the biologic relevance of the palindromic sequences.**

**Finally, I speculate on the possible biological significance of palindromic sequences in proteins and how they might have been positively selected for during evolution.**

## ÍNDICE

	Agradecimientos	V
	Resumen	VI
	Summary	VIII
	Índice de tablas	XII
	Índice de figuras	XIII
1.	Introducción.	1
2.	Hipótesis.	12
3.	Objetivos.	13
4.	Materiales y métodos.	14
4.1.	Alineación de dos secuencias aminoacídicas.	14
4.2.	Elección de las secuencias proteicas a estudiar.	14
4.3.	Elección de los segmentos palindrómicos a estudiar.	15
4.4.	Corrección del porcentaje de identidad arrojado por blast 2 sequences.	15
4.5.	Asignación de estructura secundaria a las 34 secuencias seleccionadas.	17
4.6.	Preferencia de composición aminoacídica en las secuencias palindrómicas.	17
4.7.	Significación estadística de la relación entre la frecuencia de eventos de simetría estructural de los palíndromos y intervalo de porcentaje de identidad antiparalela de secuencia: controles al azar de simetría estructural.	21
4.8.	Asociación entre la presencia de residuos idénticos o distintos y la identidad de estructura secundaria observada en el intervalo 40-100% de identidad antiparalela.	22
4.9.	Relación entre las propiedades fisicoquímicas del residuo y la simetría de estructura secundaria al enfrentar un residuo idéntico o distinto al lado opuesto.	24
4.10.	Origen nucleotídico de las secuencias palindrómicas aminoacídicas.	27
5.	Resultados.	28
5.1.	Demostrar que existe una relación significativa entre el porcentaje de identidad antiparalelo de las mitades del palíndromo y su simetría de estructura secundaria, por sobre lo observado en segmentos de la misma longitud escogidos al azar.	28
5.2.	Demostrar que los eventos en que se enfrentan residuos idénticos se asocian significativamente con los eventos de simetría de estructura secundaria en los palíndromos.	32
5.3.	Demostrar que entre los residuos idénticos y los distintos existen preferencias de residuos por sobre lo esperado por azar.	36
5.3.1.	Caracterización de la población de residuos observada en las secuencias de origen con un muestreo tipo ventana móvil.	37
5.3.2.	Determinación de las preferencias de residuos que más se observa entre los residuos idénticos o distintos respecto de las poblaciones caracterizadas.	38
5.4.	Demostrar que tanto entre los residuos idénticos como en los distintos, existen propiedades fisicoquímicas que se asocian con el mantenimiento de la simetría de la estructura secundaria.	40
5.4.1.	Perfil de los residuos que al mantenerse idénticos muestran estructuras	40

	secundarias más simétricas.	45
5.4.2.	Perfil de los residuos que al enfrentarse a un residuo distinto muestran estructuras secundarias más simétricas.	47
6.	Discusión.	53
6.1.	Simetría de estructura secundaria e identidad de secuencia entre las mitades de un palíndromo.	54
6.2.	Propiedades de un residuo cualquiera para lograr simetría de estructura secundaria al ser enfrentado por residuo idéntico/distinto.	56
6.3.	Preferencias de sustitución y simetría de estructura secundaria.	63
6.4.	Interpretación del comportamiento observado: residuo-soluto (no interactivo), residuo interactivo de campo homogéneo, residuo interactivo de campo heterogéneo.	66
6.5.	Residuos simétricos, residuos no simétricos.	71
6.6.	Conservación de residuos v/s identidad antiparalela.	74
6.7.	Estructura-función.	76
6.8.	Palindromía y código genético.	79
7.	Conclusiones.	81
8.	Referencias.	83
9.	Glosario.	85
10.	Apéndice.	87



**ÍNDICE DE TABLAS**

Tabla 1.	Frecuencia de identidad de estructura secundaria (simetría) de residuos palindrómicos v/s residuos control.	29
Tabla 2.	Comparación de porcentajes de identidad de estructura secundaria (simetría) de secuencias palindrómicas v/s secuencias control.	30
Tabla 3.	Identidad de estructura secundaria para palíndromos del intervalo 40-100% de identidad antiparalela de secuencia.	34
Tabla 4.	Preferencias en la composición de las 48 secuencias palindrómicas en las 34 secuencias estudiadas.	40
Tabla 5.	Preferencias en la composición de las secuencias palindrómicas del intervalo 40-100%.	41
Tabla 6.	Preferencias en la composición de las secuencias palindrómicas del intervalo 40-100%, con substracción de las asimétricas .	42
Tabla 7.	Simetría de estructura secundaria asociada a la identidad de un par de residuos enfrentados ubicado en cualquier sitio del palíndromo.	44
Tabla 8.	Relación entre la magnitud de los índices y la simetría de estructura secundaria con la identidad de los residuos.	50
Tabla 9.	Relación entre la magnitud de los índices y la simetría de estructura secundaria con la desigualdad de los residuos.	51
Tabla 10.	Efectos de la identidad sobre la simetría estructura secundaria para residuos no interactivos o interactivos de campo homogéneo y heterogéneo.	70
Tabla 11.	Modelo de interacción de los diversos residuos dependiendo de la simetría de estructura secundaria observada ante un residuo idéntico o distinto.	73

## ÍNDICE DE FIGURAS

- Figura 1. Similitud antiparalela entre secuencias aminoacídicas.
- Figura 2. Secuencia palindrómica en la rusticianina de *Acidithiobacillus ferrooxidans*.
- Figura 3. Secuencias palindrómicas y simetría estructural.
- Figura 4. Secuencias palindrómicas y simetría estructural.
- Figura 5. Alineaciones y elección de los palíndromos a analizar. 1
- Figura 6. Identidad de estructura secundaria entre las mitades del palíndromo 22-57 de la cadena "a" de la toxina bucandina y muestreo por el método de la ventana móvil. 1
- Figura 7. Clasificaciones de estructura secundaria y definiciones de residuos idénticos o distintos antiparalelos de un palíndromo. 2
- Figura 8. Dependencia del porcentaje de identidad de estructura secundaria respecto del porcentaje de identidad antiparalela de secuencia entre las dos mitades de un palíndromo. 3
- Figura 9. Relación entre la conservación de estructura secundaria y el mantenimiento de la identidad en secuencias palindrómicas con más de 40% de identidad antiparalela de secuencia entre sus dos mitades. 3
- Figura 10. Caracterización de la distribución de frecuencias de residuos en mil muestras tomadas al azar a través de una ventana móvil. 3
- Figura 11. Propiedades significativas para el mantenimiento de la simetría de estructura secundaria dados dos residuos idénticos o distintos en posiciones equivalentes opuestas del palíndromo. 5
- Figura 12. Secuencias palindrómicas no sobrepuestas y conservación en un alineación de un conjunto de proteínas isopropil malato deshidrogenasas. 7
- Figura 13. Residuos idénticos simétricos asimétricos e intermedios y su relación con la estabilización de las proteínas de un organismo termófilo. 7

## 1. INTRODUCCIÓN

La comparación de nuevas secuencias aminoacídicas con bases de datos de secuencias cuyas estructuras han sido caracterizadas experimentalmente, alineándolas desde su extremo amino al carboxilo (de manera paralela) ha sido una herramienta para predecir sus estructuras. Este criterio desestima la probabilidad que dos secuencias aminoacídicas puedan compartir la misma estructura si presentan similitud antiparalela. Sin embargo, dado que no se ha detallado la influencia que el orden de la secuencia aminoacídica puede ejercer sobre las interacciones estabilizadoras de la estructura nativa de la proteína (Dill, 1990; Fersht, 1999) es factible preguntarse si una secuencia puede presentar la misma estructura ya sea que esté orientada desde amino a carboxilo o viceversa.

La comprobación de esto último permitiría la revisión de la trascendencia de algunos aspectos estructurales en la determinación de la conformación y estabilidad proteica. La similitud antiparalela entre secuencias aminoacídicas se abordó inicialmente por Olszewski, Kolinski y col. (1996) Se usó un programa simulador del plegamiento de la secuencia invertida (retro-proteína) correspondiente al dominio B de la proteína A estafilocócica. Se observó que durante el plegamiento simulado de la proteína A estafilocócica era capaz de adoptar la mayoría de las veces la configuración de la proteína nativa (con algunas diferencias en las orientaciones de algunos residuos

laterales) como también la configuración especular de la nativa durante la simulación del plegamiento.

Estudios bioinformáticos realizados por Preissner, Goede y col. (1997) en proteínas de estructura determinada experimentalmente muestran que dos secuencias con similitud sobre el 34% de identidad antiparalela tienden a presentar un 60 a 70% de identidad de estructura secundaria cuando están dentro de la misma proteína (palindromía). Pese a reconocer el alto valor del porcentaje de identidad de estructura secundaria, Preissner, Goede y col. (1997) estimaron que esta simetría podía ser un artefacto inducido por predominancias de estructura secundaria en las proteínas que las contenían. Además, al encontrarse los segmentos con similitud antiparalela de secuencia en proteínas distintas, los porcentajes de identidad de estructura secundaria bajaban alrededor del 40%, lo que se adjudicó al azar. Estos autores llegaron a la conclusión de que el fenómeno de identidad antiparalela de secuencia no tiene una correlación significativa con el de identidad antiparalela de estructura secundaria.

Sin embargo, Preissner, Goede y col. (1997) no consideraron que los segmentos con similitud antiparalela de secuencia en proteínas distintas mostraron porcentajes de identidad tan bajos como los observados por Kabsch y Sander (1984) al comparar paralelamente pentapéptidos idénticos pertenecientes a proteínas distintas. Esto permitiría concluir que los factores no locales que determinan el plegamiento pueden alcanzar, más o menos, el mismo balance final frente a los factores locales, independientemente del sentido de la secuencia. No se puede, sin embargo, afirmar que la alta similitud antiparalela entre dos secuencias no tenga contrapeso alguno en el grado de similitud de estructura secundaria que muestren, ante los factores no locales.



Los estudios hechos por Lacroix, Viguera y col. (1998) analizaron experimentalmente este problema. Ellos transformaron separadamente cepas de *Escherichia coli* con un DNA que codificaba respectivamente las retro-proteínas correspondientes al dominio B de la proteína A estafilocócica (60 residuos), el dominio B1 de la proteína G estreptocócica (55 residuos), y el dominio SH3 de la alfa espectrina (61 residuos). Las versiones nativas de estas proteínas presentaban distintas composiciones de estructura secundaria (eminentemente hélice, mixtas  $\alpha/\beta$ , eminentemente hebra  $\beta$ , respectivamente) Las retro-proteínas no lograron expresarse o se expresaron en baja cantidad en el caso del dominio B de la proteína A estafilocócica. Esta última retro-proteína, la estafilocócica, exhibió un bajo nivel de estructura secundaria detectable por dicroísmo circular y resonancia magnética nuclear.

La discordancia entre la simulación de plegamiento del dominio B de la proteína A estafilocócica de Olszewski, Kolinski y col. (1996) con estos resultados experimentales fue atribuida por Lacroix, Viguera y col. (1998) a la exclusión en el programa de simulación de plegamiento de las restricciones estéricas del esqueleto peptídico.

A continuación, Mittl, Deillon y col. (2000) encontraron que el polímero amino acídico invertido de la proteína activadora de la transcripción de levaduras GCN4, del tipo de cierre de leucina ( $\alpha$  hélice de 35 residuos) el cual era capaz de mantener una estructura secundaria nativa con ciertas diferencias: Mientras la proteína nativa mantenía su estructura en estado monomérico y polimérico, la retro-proteína lo hace sólo en estado polimérico; mientras la proteína nativa dimeriza, la retro-proteína del activador de transcripción tetrameriza. Por lo tanto, la retro-proteína no sólo es capaz de plegarse,

sino también de mantener su capacidad de unión de otras estructuras, tal como Phan-Chan-Du, Petit y col. (2001) observaron una mayor unión de los retro-péptidos por parte de anticuerpos, respecto de los péptidos nativos que los generaron.

Sin embargo, estos resultados nos plantean que una secuencia aminoacídica invertida, pese a poseer la misma secuencia de su contrapartida nativa y, en apariencia, todo el patrón de variaciones fisicoquímicas subyacentes, muestra ciertos impedimentos, probablemente ciertos residuos incompatibles que dificultan la adquisición del plegamiento nativo.

Park, Dietmann y col. (2000) estudiaron un conjunto de estructuras determinadas experimentalmente disponibles en los archivos del *Protein Data Bank* (PDB). Ellos encontraron que algunos segmentos de entre dos y ocho residuos de largo, pertenecientes a la misma estructura, con 100% de identidad antiparalela de secuencia, exhibieron una tasa del 70% de identidad de estructura secundaria. Este valor es muy próximo a los alcanzados actualmente por los programas de predicción de estructura secundaria y concuerda con los de las secuencias palindrómicas parcialmente idénticas de los registros de Preissner, Goede y col. (1997). Park, Dietmann y col. (2000) concluyeron que el orden invertido de la secuencia no presenta diferencias marcadas respecto de la secuencia no invertida al momento de determinar la estructura secundaria. Se pudo incluso mejorar la capacidad de predicción de estructura secundaria al complementarse con los resultados arrojados por el programa usado en dicho trabajo.

Finalmente Lorenzen, Gille y col. (2003) al comparar entre proteínas distintas, determinaron que las secuencias invertidas con menos de 50% de identidad antiparalela

logran parámetros de estructura secundaria y terciaria muy diferentes a la proteína original.

Las secuencias aminoacídicas con similitud antiparalela pueden disponerse de diversas formas (FIGURA 1) según su coexistencia o no dentro de la misma proteína y su cercanía en la secuencia. En este trabajo se estudiará el arreglo en que dos segmentos consecutivos de una proteína presentan similitud antiparalela de secuencia entre sí, constituyendo una secuencia palindrómica o palíndromo (FIGURAS 1, 2, 3 y 4) Esta decisión se tomó debido a las siguientes razones:

En primer lugar, los palíndromos proteicos son sencillos de detectar, dado que basta comparar la secuencia aminoacídica consigo misma invertida (ver métodos), fijando los parámetros de rastreo a una baja exigencia, sin necesidad de entrar en la combinatoria que implicaría comparar cada una de las secuencias escogidas con las otras. En segundo lugar, al tratarse de dos segmentos consecutivos dentro de la misma proteína, debieran tener una mayor probabilidad de estar supeditados a determinantes no locales similares. Por lo tanto, la asimetría de estructura secundaria que se observe en el palíndromo debiera dar cuenta de las diferencias locales (diferencias estructurales debido a las mismas secuencias) más que a diferencias de contexto.

Hemos realizado algunos estudios preliminares de secuencias palindrómicas (no publicados) en un grupo de proteínas asociadas a cobre similares a plastocianina y rusticianina y otras secuencias aminoacídicas inspeccionadas al azar. En ellos observamos las siguientes características:

- Los palíndromos están presentes en más del 70% de las proteínas

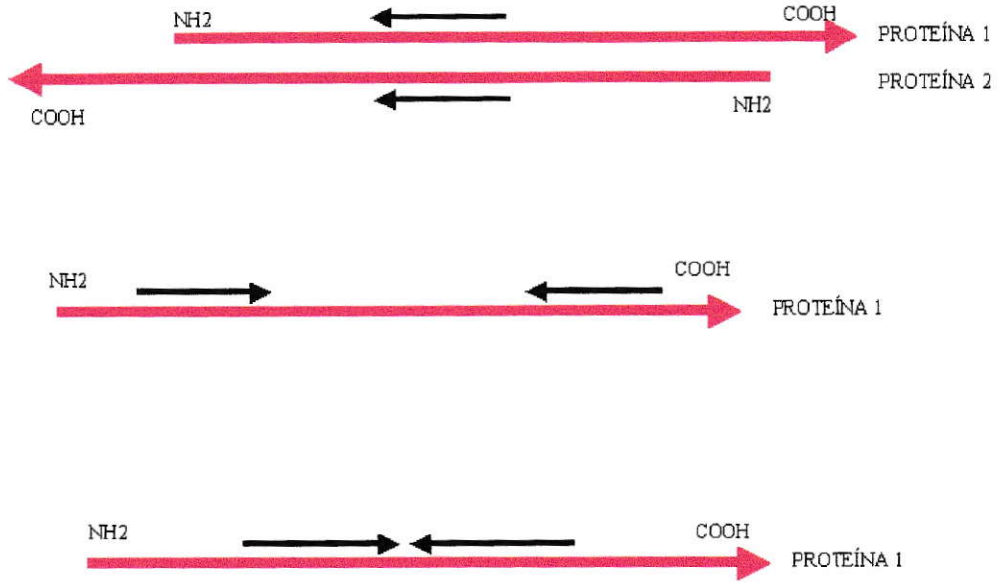


Figura 1. Similitud antiparalela entre secuencias aminoacídicas. Tres posibles ubicaciones entre secuencias aminoacídicas con similitud antiparalela (de arriba hacia abajo): En proteínas distintas (arriba) separadas dentro de la misma proteína (al medio) o consecutivas dentro de la misma proteína (abajo) En este trabajo se analiza las secuencias y estructuras de este último tipo de ejemplos. Se les define como secuencias palindrómicas.



```
>gi|2194027|pdb|1RCY| Rusticyanin (Rc) From
Acidithiobacillus Ferrooxidans
TTWKEATLPQVKAMLEKDDGKVS GDTVTYSGKT VHVVAAAVLPGFFPSPSEVHDKKNPTLEIPAGATVDV
TFINTNKGFGHSEFDITKKGPPYAVMPVIDPIVAGTGFSPVPKDGKFGYTFWHP TAGTYYYVCQIPGHA
ATGMF GKIVVK
```

```
>inv_gi.2194027.pdb.1RCY. Rusticyanin (Rc) From
Acidithiobacillus Ferrooxidans
KVVIKGFMGTAANGPIQCVYYYTGATPHWTFDTYGFKGDKVPVSPFGTGAVIPDIVPMVAYPPGKKTIDFS
HGFGKNTNIFTVDVTAGAPIELTPNKKDHVEFSPFPFGPLVAAA VVHVTKGSYTVTDGSVKGDDKELMAK
VQPLTAEKWT T
```



Blast2sequences  
Expect 1.000.000

```
Score = 28.1 bits (66), Expect = 5.5
Identities = 25/101 (24%), Positives = 41/101 (39%), Gaps = 6/101 (5%)
rusdir: 6  A T L P Q V K A M L E K D D G K V S G D T V T Y S G K T V H V V A A A V L P G F F P S P S E V H D K K N P T L E I P A G A T V D V T F I N T N K G F G H S E F D I T K K G P P Y A V M P V I D P I V A 103
A + P + H + G K + D G K + + V G P P + + H + + P P G V + + K G D + K G + H + P + A
rusinv: 103 AVIPDIVPMVAYPPGKKTIDFSHGFGKNTNIFTVDVTAGAPIELTPNKKDHVEFSPFPFGPLVAAA VVHVTKGSYTVTDGSVKGDDKELMAKVQPLTAEKWT 006
```



FIGURA 2. Secuencia palindrómica en la rusticianina de *Acidithiobacillus ferrooxidans*. Una alineación antiparalela entre la secuencia directa y la invertida de la rusticianina de *Acidithiobacillus ferrooxidans* usando la matriz *blosum62* y una expectativa de 1.000.000. En la alineación se observa algunos residuos idénticos en la posición equivalente opuesta y otros no idénticos. Se aprecia también el residuo central histidina, dada la longitud impar de la alineación (que incluye espacios más residuos).

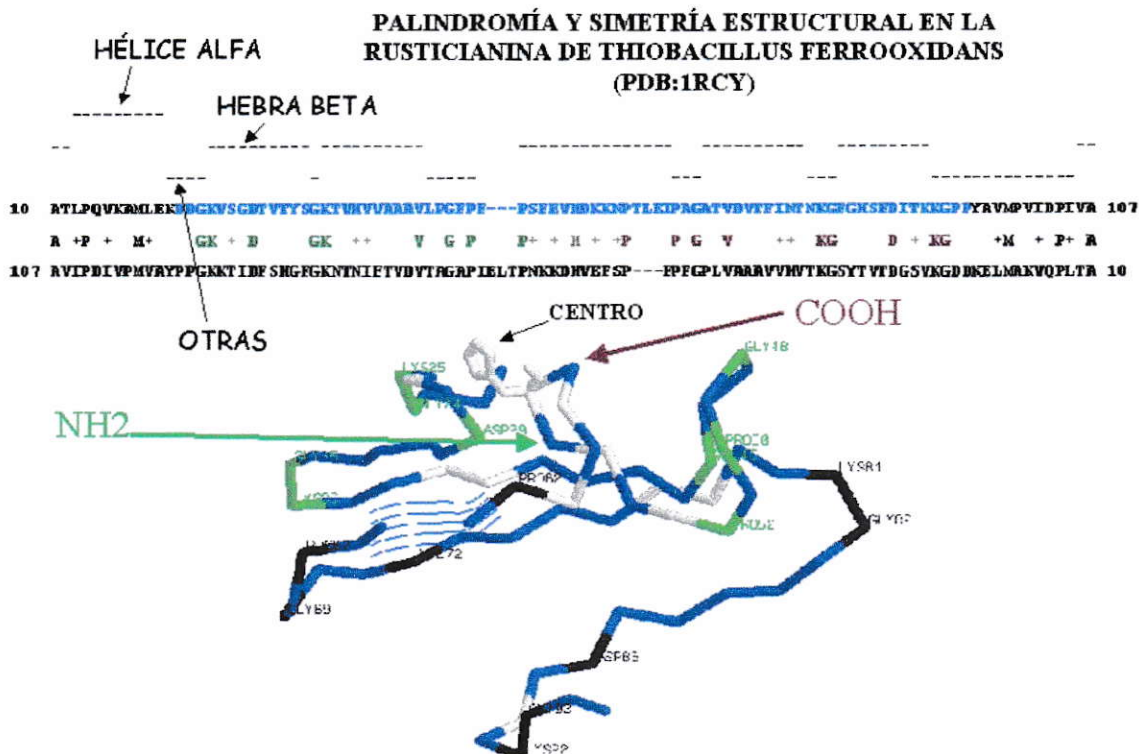


FIGURA 3. Secuencias palindrómicas y simetría estructural. Palíndromo de la rusticianina (PDB 1RCY) de *Acidithiobacillus ferrooxidans* (residuos A10-A107). La mayoría de los residuos idénticos (en pardo hacia carboxilo y en verde hacia amino respecto de la H57) poseen estructuras secundarias idénticas (57% de identidad de estructura secundaria en el segmento 22-94). Por ejemplo, los residuos G48-P50/P67-G69 y G35-K36/K81-G82 se ubican en las vueltas. Colores: azul, color por defecto del segmento palindrómico 22-94; Ciclo gris: histidina 57 (central); los otros grises, residuos conservados en la alineación, según matriz blosum62. Los residuos frente a espacios en las alineaciones se representan en hebras delgadas de color azul.

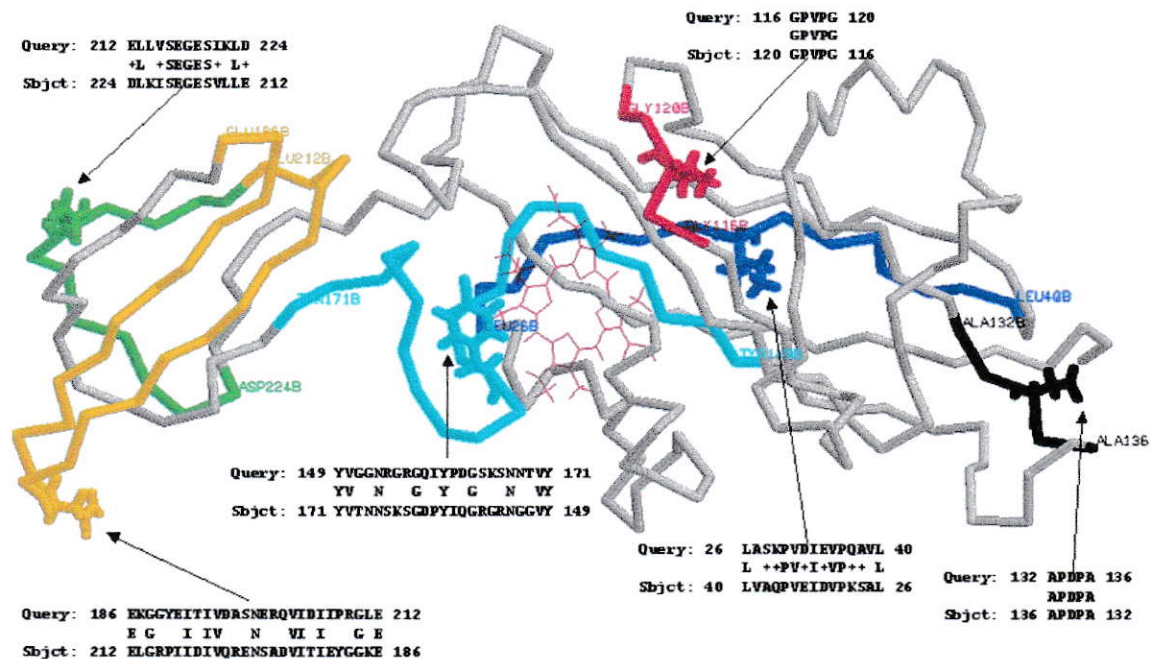


FIGURA 4. Secuencias palindrómicas y simetría estructural. Seis segmentos palindrómicos no superpuestos en el citocromo f de *Syneccocystis sp.* Se puede apreciar los segmentos palindrómicos con sus residuos centrales en la secuencia destacados con flechas y en gráfica de barras, a manera de referencia en la estructura. Estos segmentos muestran una apreciable simetría estructural.

- Poseen un amplio intervalo de extensión y porcentajes de identidad (5 a 200 residuos).
- Localización variable dentro de la secuencia.
- Es posible encontrar más de una secuencia palindrómica dentro de la proteína, como también superposiciones de ellas.
- En el caso de las proteínas asociadas a cobre de las familias de la plastocianina de *Synechococcus* sp y la rusticianina de *Acidithiobacillus ferrooxidans*, los segmentos palindrómicos no tienen una posición relativa definida en la secuencia respecto de los centros de cobre.
- Los veinte residuos aminoacídicos distintos pueden formar parte del palíndromo.
- Se ha observado que hay secuencias reordenadas al azar que son capaces de mostrar patrones palindrómicos.
- Los segmentos palindrómicos tienden a ser poco conservados.

¿Es realmente capaz de plegarse de manera nativa una retro-proteína? ¿Qué tan idénticas a la secuencia nativa deben ser las retro-proteínas para lograr plegarse de manera similar? ¿Qué características fisicoquímicas son relevantes en la disrupción/mantenimiento de la simetría estructural entre proteínas con similitud antiparalela de secuencia? El estudio de estos palíndromos no perfectos, dado que ambos extremos no presentan siempre una total identidad antiparalela en la secuencia de sus residuos, permite establecer tanto la influencia de los residuos idénticos como de los residuos distintos y las características fisicoquímicas más relacionadas con la simetría del palíndromo. El sistema palindrómico escogido no debiera diferir en cuanto a las



reglas de plegamiento de cualquier otro sistema proteico, por lo que los resultados derivados de este estudio debieran ser aplicables al campo general de plegamiento de retro-proteínas, ayudando a acercar las aparentes contradicciones de los trabajos mostrados hasta ahora.

En esta tesis se analiza la asociación entre el porcentaje de identidad y la simetría de estructura secundaria de estos segmentos en comparación con segmentos tomados al azar dentro de las mismas proteínas. Se revisa la importancia del orden de la secuencia sobre la simetría estructural. Por otro lado, se evalúa las posibles preferencias de residuos en estos segmentos y su relación con la simetría estructural. Se propone las características fisicoquímicas más determinantes de la simetría estructural. En último término se discute sobre la génesis evolutiva y posible función de este tipo de organizaciones de la estructura primaria en las proteínas.

## 2. HIPÓTESIS

Las secuencias palindrómicas dentro de las proteínas son capaces de adoptar una estructura secundaria simétrica debido a la presencia de residuos idénticos en posiciones equivalentes de sus extremos, más que a los factores vecinos que puedan inducir a mantener dicha simetría de estructura secundaria.

Dado que se sabe de ejemplos experimentales de secuencias invertidas que no han logrado plegarse de manera nativa, es probable que existan determinantes fisicoquímicos insuficientes en algunos residuos para adoptar una estructura secundaria idéntica al localizarse en la secuencia invertida de lado opuesto del palíndromo, siendo necesario su reemplazo por otro residuo para adoptar el plegamiento correcto.

### 3. OBJETIVOS

Demostrar que existe una relación significativa entre el porcentaje de identidad antiparalelo de las mitades del palíndromo y su simetría de estructura secundaria, por sobre lo observado en segmentos de la misma longitud escogidos al azar.

Demostrar que los residuos idénticos en las secuencias invertidas se asocian significativamente con los eventos de simetría de estructura secundaria en los palíndromos, por sobre los residuos distintos.

Demostrar que entre los residuos idénticos y los residuos distintos las existen preferencias de residuos por sobre lo esperado por azar.

Demostrar que tanto en los residuos idénticos y los residuos distintos, existen magnitudes de propiedades fisicoquímicas que se correlacionan con la simetría de la estructura secundaria.

## **4. MATERIALES Y MÉTODOS**

### **4.1. Alineación de dos secuencias aminoacídicas**

Cada secuencia aminoacídica fue alineada con su respectiva secuencia invertida usando el programa de Tatusova y Madden (1999) BLAST 2 SEQUENCES, ubicado en la dirección <http://www.ncbi.nlm.nih.gov/> . Los parámetros fijados fueron programa “blastp”; matriz “BLOSUM62”, expectativa de 1.000.000. El resto de los parámetros se definen por omisión. En este punto es necesario hacer notar que el uso de un determinado programa de alineación tiene como único objetivo servir de detector de segmentos aminoacídicos que tengan cierto porcentaje de identidad antiparalela sin que importe la ponderación que le otorgue la matriz de sustitución en uso. Por ello también se asignó un valor altamente permisivo al parámetro expectativa (FIGURA 2)

### **4.2. Elección de las secuencias proteicas a estudiar**

Dada la falta de recursos de programación, se restringió el análisis a un conjunto de 73 proteínas de secuencias no redundantes y funciones asociadas a la actividad de toxinas. Son de origen taxonómico heterogéneo y de estructura-función en general bien caracterizada, dado su alto interés biológico. Esto último permitiría detectar la existencia de segmentos palindrómicos coincidentes con sitios de importancia funcional en la secuencia



### **4.3. Elección de los segmentos palindrómicos a estudiar**

La alineación de la secuencia de una proteína consigo misma invertida por el método descrito en “Alineación de dos secuencias aminoacídicas” permitió obtener alineaciones de diversas longitudes y porcentajes de identidad antiparalela (FIGURAS 1, 2 y 5) como también superposiciones de ellas (no mostrado) para las 73 secuencias no redundantes estudiadas. A continuación se eliminó todas las alineaciones que tuviesen superposición entre sí (no mostrado) como también las alineaciones cuyas secuencias con similitud antiparalela no fueran palindrómicas, es decir, no estuviesen consecutivas en la misma alineación (FIGURA 1, ejemplo del medio). Terminado este proceso, quedaron 34 secuencias no redundantes que poseían 48 alineaciones palindrómicas no sobrepuestas de diversas longitudes y porcentajes de identidad asignados por el programa de alineación. Las secuencias se detallan al final de este trabajo, en el APÉNDICE.

### **4.4. Corrección del porcentaje de identidad arrojado por *blast 2 sequences***

En algunas de las 48 alineaciones palindrómicas se corrigió los porcentajes de identidad de manera de permitir que éstos expresaran exclusivamente el porcentaje de identidad entre las dos mitades con similitud antiparalela, sin considerar los gaps (segmentos donde no es posible comparar las estructuras secundarias de las secuencias) ni los residuos centrales (cuya estructura secundaria siempre será idéntica a sí misma).

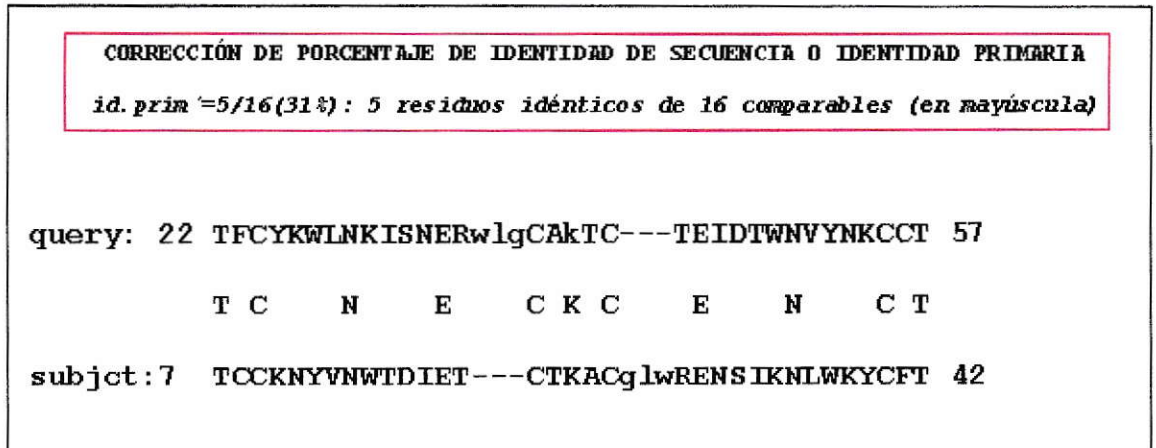


FIGURA 5. Alineaciones y elección de los palíndromos a analizar. Se muestra como ejemplo el resultado de la alineación entre la toxina bucardina y su versión invertida teórica (retro-bucandina). El procedimiento genera como soluciones varias alineaciones sobrepuestas, algunas de las cuales corresponden a palíndromos y otras no. De todas estas alineaciones, se deja exclusivamente las de segmentos palindrómicos no sobrepuestos (no mostrado) no importando que índices arroje el programa para dichas alineaciones. Finalmente, se corrige el porcentaje de identidad eliminando el residuo central (en el caso del ejemplo, la lisina central) y todos los residuos que no enfrentan otros residuos en la alineación (en minúsculas).

Ninguna de las alineaciones con más de 40% de identidad antiparalela de secuencia contenía espacios (FIGURA 6).

#### **4.5 Asignación de estructura secundaria a las 34 secuencias seleccionadas**

Dadas las coordenadas experimentales contenidas en el archivo PDB respectivo, la gráfica de las estructuras y la asignación de estructuras secundaria se realizó usando el programa RASMOL. Al igual que Park, Dietmann y col (2000), se simplificó los tipos de estructura secundaria en hélice, hebra beta, y no hélice-no hebra beta (FIGURA 6).

#### **4.6. Preferencia de composición aminoacídica en las secuencias palindrómicas**

Para expresar las preferencias de composición de residuos dentro de una secuencia, Richardson y Richardson (1988) realizaron el conteo de residuos en una posición de determinadas características. A continuación normalizaron por la frecuencia esperada correspondiente al porcentaje global de cada residuo en el total de secuencias analizadas. Esta aproximación supone que los residuos se distribuyen estocásticamente para formar las secuencias de las proteínas. Sin embargo, dado que las secuencias aminoacídicas reales han sido seleccionadas para cumplir funciones biológicas, todo el espacio teórico de configuraciones de secuencias se restringe a un espacio menor capaz de cumplir con los requerimientos estructurales y funcionales de utilidad biológica (Dill, 1990). Por lo tanto, no es suficientemente preciso suponer la total homogeneidad en la distribución de residuos para normalizar las frecuencias aminoacídicas observadas. Para hacer esta normalización, es necesario respetar los patrones fisicoquímicos en los que los

```

query: 22 TFCYKWLNKISNERwlgCAKTC---TEIDTWNVYNKCCT 57
          T C  ++N    E    C K C    E    N++  C T
subjct:7 TCCKNYVNWTDIET---CTKACGLWRENSIKNLWKYCFT 42

```

```

          T C    N    E    C K C E    N    C T
          ← IDIDDDIDDDIDGGIDCDIDIDDDIDDDIDI →
MECYRCGVSGCHLKITCSAETFCYKWLNKISNERWLGCAKTCTEIDTWNVYNKCCTTNLCNX

```

FIGURA 6. Identidad de estructura secundaria entre las mitades del palíndromo 22-57 de la cadena “a” de la toxina bucardina y muestreo por el método de la ventana móvil. En el recuadro superior se puede observar en color anaranjado los residuos en conformación de hebra beta; negro, no hebra beta ni hélice, mientras el rojo simboliza hélice (ausente en el ejemplo). 13 residuos tienen la misma estructura secundaria, de los 16 comparables que hay en cada mitad del palíndromo. En el recuadro inferior, se puede apreciar que el palíndromo (azul) constituye un patrón (en rojo) de residuos simbolizados como idénticos (I), centrales (C) y distintos (D) y de residuos frente a gaps (G, a los cuales se les omite del análisis) ubicados a ciertas distancias relativas entre sí en la secuencia proteica real. Durante el análisis de las preferencias de composición aminoacídica en los palíndromos, se respetó dichas distancias relativas usando una ventana móvil para realizar el muestreo aleatorio de residuos a lo largo de la secuencia de origen.



residuos tienden a agruparse en las secuencias. A continuación se explica cómo se aborda aquí este problema.

Un palíndromo, como se define operacionalmente en este trabajo, son dos segmentos aminoacídicos consecutivos, dentro de la secuencia de una proteína, capaces de mostrar un porcentaje de identidad de secuencia determinado, sobre 10% en este trabajo, al compararse entre sí antiparalelamente. Tanto su longitud como el porcentaje de identidad son variables (FIGURA 6).. Por lo tanto, cada uno de los 48 palíndromos representa un patrón de residuos idénticos, distintos y centrales (cuando el número de letras más espacios es impar habrá siempre un residuo central).

Para obtener una muestra aleatoria como control de cada segmento palindrómico real, se propone desplazar una ventana de la misma longitud en residuos que dicho segmento a lo largo de la secuencia de origen y, en un punto de detención aleatorio por muestra, se extrae un conjunto de residuos de la ventana viajera (FIGURA 6).

Este procedimiento conserva la posición relativa (biológica) de los residuos entre sí, por lo que la única fuente de variabilidad proviene de los patrones de composición reales de la secuencia de origen.

En nuestro caso tenemos 48 palíndromos reales de diversas longitudes distribuidos en las 34 secuencias analizadas. Por lo tanto, una muestra control de los 48 palíndromos reales consta de una ronda de 48 ventanas, cada una de longitud aminoacídica igual a sus contrapartidas reales, detenidas aleatoriamente dentro de su secuencia de origen respectiva, rescatando, sin modificar el orden de la secuencia, un conjunto de residuos clasificados arbitrariamente en las categorías central (C), distinto (D) e idéntico (I), según se ubiquen en las mismas posiciones que los residuos así

clasificados en el palíndromo real. Se elimina del conteo aquellos residuos que por posición enfrentan un espacio o *gap* (G) en el palíndromo detectado por *blast*. Este proceso se repitió 100 veces para obtener las 100 muestras control por categoría. A partir de los conteos de estas muestras se obtuvo el promedio y la desviación estándar del conteo de cada uno de los veinte residuos en la respectiva categoría.

Conocidos estos parámetros de distribución (suponiendo distribución gaussiana) es posible cuantificar cuán alejada está la composición palindrómica observada en los 48 palíndromos de la esperada por azar por la fórmula de la variable normal estándar:

$$\frac{(\text{Comp.aa.}_{\text{categ\_palíndromo}} - \text{comp.aa}_i\text{_{categ\_100 controles}})}{\text{desv.est aa}_i\text{-categ}}$$

Donde

**Comp.aa.}\_{categ\\_palíndromo}**: composición o conteo de un residuo determinado en el total de palíndromos dentro de una categoría observada.

**comp.aa}\_i\text{\_{categ\\_100 controles}}**: composición o conteo promedio de dicho residuo en las 100 muestras al azar dentro de la misma categoría asignada arbitrariamente.

**desv.est aa}\_i\text{-categ}**: desviación estándar de la composición o conteo de dicho residuo entre las 100 muestras al azar dentro de la misma categoría asignada arbitrariamente.

En el estudio de los 14 palíndromos con más de 40% de identidad antiparalela de secuencia entre sus mitades, la metodología fue la misma, salvo que se realizó 1000 muestras de 14 controles, uno para cada palíndromo, en la respectiva secuencia de origen.

Se consideró que un residuo es preferido o rechazado si el valor de la variable normal estándar es superior a 1,64 o inferior a  $-1,64$  desviaciones estándar respectivamente (5% superior e inferior de la población). Sin embargo, para efectos del estudio de la asociación entre identidad/no identidad entre residuos opuestos equivalentes y las propiedades fisicoquímicas (punto 9, métodos) se definió “preferido” a los 10 valores más positivos de preferencias, y “rechazados” al resto.

#### **4.7. Significación estadística de la relación entre la frecuencia de eventos de simetría estructural de los palíndromos y intervalo de porcentaje de identidad antiparalela de secuencia: Controles al azar de simetría estructural**

Se seleccionó al azar, a modo de control, un segmento de igual longitud de cada palíndromo detectado, en la misma secuencia de origen. Se eliminó de la comparación de estructura secundaria aquellos residuos que en el palíndromo original enfrentaban gaps, residuos centrales, posiciones con letras distintas al código aminoacídico de una letra, se separó estos segmentos control según los porcentajes de identidad de los palíndromos originales, se midió el porcentaje de identidad de estructura secundaria entre las mitades de los segmentos en tres intervalos de porcentajes de identidad de secuencia: [10%, 30%[, [30%, 40%[ y [40%, 100%[. Se evaluó la significación de la asociación entre la simetría estructural y el intervalo de porcentaje de identidad

antiparalela de secuencia usando la prueba del  $\chi^2$  (nivel de significación=0,05,  $\chi^2_{\text{crítico}}=3,84$ ).

#### **4.8. Asociación entre la presencia de residuos idénticos o distintos y la identidad de estructura secundaria observada en el intervalo 40-100% de identidad antiparalela**

En este punto se pregunta si, dados dos residuos enfrentados en el palíndromo, ya sea idénticos o distintos, qué relación con la simetría de estructura secundaria se observa entre ellos.

Dentro del intervalo con mayor significación de simetría de estructura secundaria, el intervalo 40-100 % de identidad antiparalela de secuencia, y dado el bajo número de datos, no era posible un análisis significativo de la simetría de la estructura secundaria para cada uno de los veinte residuos al ser enfrentado al lado opuesto del palíndromo por un residuo idéntico o distinto. Por lo tanto se fusionó los eventos individuales en las categorías idénticos y distintos, las que representan respectivamente cuando un residuo cualquiera está enfrentado a un residuo idéntico o diferente al lado opuesto del palíndromo, siempre que la comparación estructural sea pertinente (ver asignación de estructura secundaria en métodos y FIGURA 7). Para cada una de estas dos categorías de residuos idénticos/distintos, se contabiliza los eventos de simetría de la estructura secundaria (igual o distinta estructura secundaria). Esta clasificación de los datos genera una tabla de cuatro categorías: residuos idénticos con igual o distinta estructura, y residuos distintos con igual o distinta estructura secundaria. La significación de la asociación entre identidad del residuo y simetría de estructura se evaluó usando las



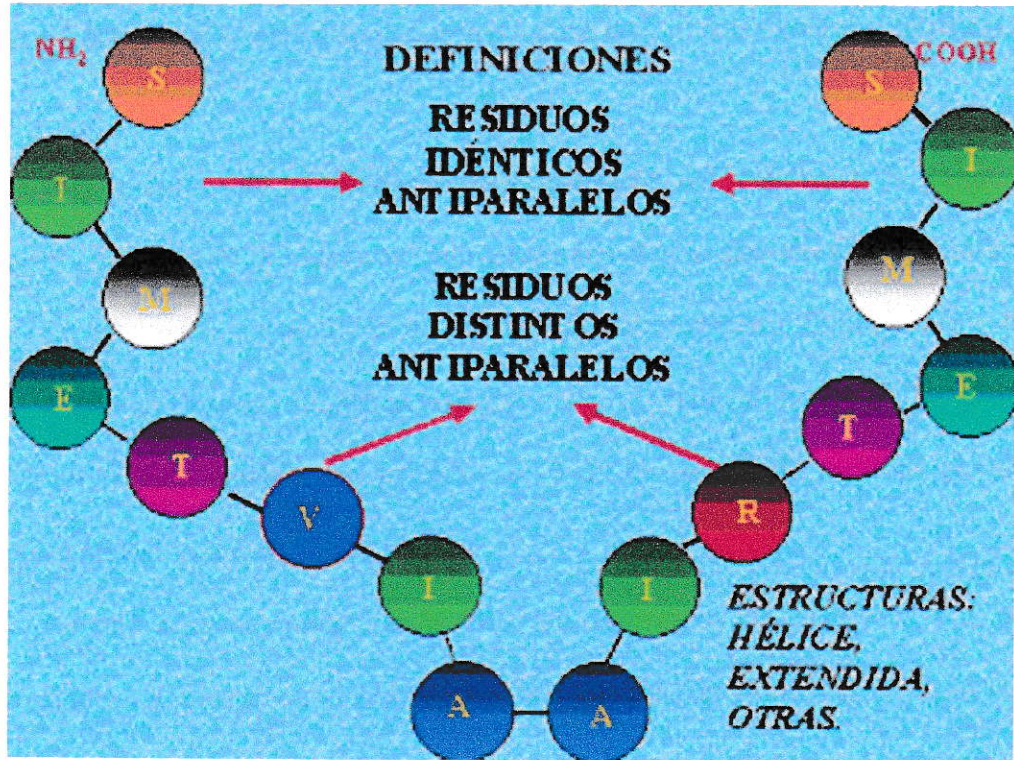


FIGURA 7. Clasificaciones de estructura secundaria y definiciones de residuos idénticos o distintos antiparalelos de un palíndromo. Se denomina residuos idénticos antiparalelos cuando el mismo residuo ocupa sitios equivalentes y opuestos dentro del palíndromo; residuos distintos antiparalelos, cuando residuos distintos ocupan lugares equivalentes y opuestos dentro del palíndromo. Se define simetría cuando dos sitios equivalentes y opuestos poseen la misma estructura secundaria.

pruebas de  $\chi^2$  y de probabilidad exacta de Fischer con la condición de que en ambas resulte por sobre el 5% requerido para considerarse que la asociación es significativa.

#### **4.9. Relación entre las propiedades fisicoquímicas del residuo y la simetría de estructura secundaria al enfrentar un residuo idéntico o distinto al lado opuesto**

En esta parte se pregunta si hay ciertas características fisicoquímicas de los residuos que se asocien significativamente con el mantenimiento de la estructura secundaria al lado opuesto del palíndromo, dado un evento de residuos idénticos o distintos antiparalelamente.

En este caso, se contabiliza la totalidad de eventos en que frente a un residuo determinado, no central en el palíndromo, se mantiene la estructura secundaria en la secuencia invertida, cualquiera sea la identidad del residuo que ocupe el sitio equivalente al lado opuesto.

Cada residuo en sí puede representarse por varios parámetros fisicoquímicos: tamaño, ramificaciones, propensión de estructura secundaria, carga, hidrofobicidad, y otras características que podrían influir en el mantenimiento de la simetría de estructura secundaria si dicho residuo es reemplazado o mantenido al lado opuesto del palíndromo.

[http://www.genome.ad.jp/dbget-bin/show\\_advanced\\_dbget?aaindex](http://www.genome.ad.jp/dbget-bin/show_advanced_dbget?aaindex), es la base de datos de índices aminoacídicos Aaindex de Shuichi, Ogata y col (1999) se encuentran tabulados aproximadamente 400 índices. De estas 400 tablas, se eliminó aquellas con datos faltantes para algún aminoácido. Se ordenó los veinte residuos de menor a mayor según sus respectivos valores para un determinado índice. A continuación se reordenó del mismo modo los datos de simetría versus mantenimiento de identidad. Los diez

residuos con valores menores de dicho índice y sus respectivos datos fusionados de simetría y mantenimiento de identidad, se definen con una magnitud baja de dicho índice o parámetro; los diez siguientes, con una magnitud alta de dicho parámetro. En esta etapa se eliminó las tablas en que el décimo residuo tuviese el mismo valor que el undécimo, para evitar la asignación arbitraria de un residuo dentro del grupo de magnitud baja o alta.

Sea un grupo de residuos con magnitud alta para un índice cualquiera. Para considerar que dicho grupo posee una mayor capacidad de mantener la simetría de estructura secundaria al encontrarse un residuo idéntico al lado opuesto del palíndromo, es necesario, como mínimo, que el número de eventos simétricos sea superior al de eventos no simétricos de este grupo. Simultáneamente, el grupo de residuos con baja magnitud para el mismo índice debiera exhibir un número de eventos de simetría inferior a los casos de no simetría. Por lo tanto, debiera cumplirse que:

$$(\text{simetría alto índice} + \text{asimetría bajo índice}) > (\text{simetría bajo índice} + \text{simetría alto índice})$$

Se definirá como +1 si el índice alto es más “simétrico” para pares de residuos idénticos o distintos, y -1 de lo contrario.

Donde:

simetría alto índice: total de eventos en que los residuos con índice alto presentan igual estructura secundaria al lado opuesto equivalente del palíndromo, ya sea que se enfrenten residuos idénticos o distintos.

asimetría bajo índice: total de eventos en que los residuos con índice bajo presentan distinta estructura.



simetría bajo índice: total de eventos en que los residuos con índice bajo presentan igual estructura.

asimetría bajo índice: total de eventos en que los residuos con índice bajo presentan distinta estructura.

La presencia de pares de residuos distintos a ambos lados del palíndromo se analiza del mismo modo.

El sentido de esta desigualdad (+1, -1) nos indica el sentido de este balance, pero no su significación estadística. Se evaluó la significación estadística de la relación entre la magnitud del índice en estudio y el mantenimiento de estructura secundaria, dado dos residuos idénticos o distintos enfrentados en el palíndromo. Se aplicó las pruebas de  $\chi^2$  y de Fischer en todos los casos, considerando significativos aquellos correspondientes a valores de p menores o iguales al 0,05 para ambas pruebas. En este punto, se consideró meritorio realizar una excepción. En los casos de los parámetros de preferencia para residuos iguales/distintos obtenidos en este mismo trabajo (ver métodos) no todos los parámetros (a saber: preferencias de residuos idénticos, preferencias de residuos intercambiados, preferencias centrales) lograron ser significativos por la prueba de Fischer y  $\chi^2$  simultáneamente. En este caso se aceptó que, por lo menos, superaran una de las dos pruebas para incluirlos en la discusión. Esta concesión se hizo en consideración de que los datos de las preferencias se obtuvieron independientemente, y que, por lo tanto, cualquiera de las pruebas estadísticas que fuera superada sería un indicador confiable de asociación significativa.

#### **4.10. Origen nucleotídico de las secuencias palindrómicas aminoacídicas**

Se obtuvo de la dirección <http://www.ncbi.nlm.nih.gov/> los segmentos nucleotídicos genómicos que especifican las mitades de los segmentos palindrómicos del intervalo 40-100% de identidad antiparalela de secuencia. Con el fin de medir si el DNA que codifica una mitad del palíndromo podría, por complementariedad de bases, codificar la otra mitad, se evaluó la hibridación entre ambos segmentos codificantes, en la dirección <http://www.bioinfo.rpi.edu/applications/mfold/old/dna/form1.cgi>. usando el programa de DNA de Zuker (2003).

## 5. RESULTADOS

### **5.1 Demostrar que existe una relación significativa entre el porcentaje de identidad antiparalelo de las mitades del palíndromo y su simetría de estructura secundaria, por sobre lo observado en segmentos de la misma longitud escogidos al azar**

Dado que Preissner, Goede y col. (1997) postularon que el alto porcentaje de identidad de estructura secundaria observado en los segmentos palindrómicos es causado por la presión del contexto estructural en que se encuentran incluidos, se comenzó por evaluar esta suposición. Si es verdadera, los palíndromos debieran mostrar porcentajes de identidad de estructura secundaria comparables a los que poseería un conjunto aleatorio de segmentos aminoacídicos extraídos de las mismas secuencias de origen.

Por lo tanto, para cada uno de los 48 segmentos palindrómicos, se eligió aleatoriamente un segmento control de la misma longitud en la misma proteína de origen y se comparó sus porcentajes de identidad de estructura secundaria por intervalos de porcentaje de identidad antiparalela del palíndromo, tal como se observa en la TABLA 1 y 2.

En la FIGURA 8 es posible apreciar que al sumar los datos de todas las secuencias pertenecientes al intervalo de porcentaje de identidad antiparalela de 40-100%, el número de veces que dos sitios equivalentes y opuestos poseen la misma estructura secundaria es significativamente mayor en los palíndromos que en las secuencias control respectivas ( $\alpha=0,05$ ,  $\chi^2$  crítico = 3,84, FIGURA 8).



TABLA 1

FRECUENCIA DE IDENTIDAD DE ESTRUCTURA SECUNDARIA (SIMETRÍA) DE RESIDUOS PALINDRÓMICOS V/S RESIDUOS CONTROL

Chi-2 observado	Rango % identidad antiparalela secuencia	Total eventos estudiados	Total eventos simetría controles	Total eventos simetría palíndromos	Significación diferencia alfa=0,05 chi2crít=3,84
0,45	[10-30[	309	135	117	NO
0,04	[30-40[	188	94	96	NO
5,38	[40-100]	64	39	51	SÍ

TABLA 2

COMPARACIÓN DE PORCENTAJES DE IDENTIDAD DE ESTRUCTURA SECUNDARIA (SIMETRÍA)  
DE SECUENCIAS PALINDRÓMICAS V/S SECUENCIAS CONTROL

<b>Rango % identidad antiparalela secuencia</b>	<b>Secuencias control más simétricas</b>	<b>Igual simetría controles y palíndromos</b>	<b>Secuencias palíndromo más simétricas</b>	<b>Total comparaciones palíndromo v/s controles</b>
[0-30[	6	3	8	17
[30-40[	7	2	8	17
[40-100]	2	5	7	14

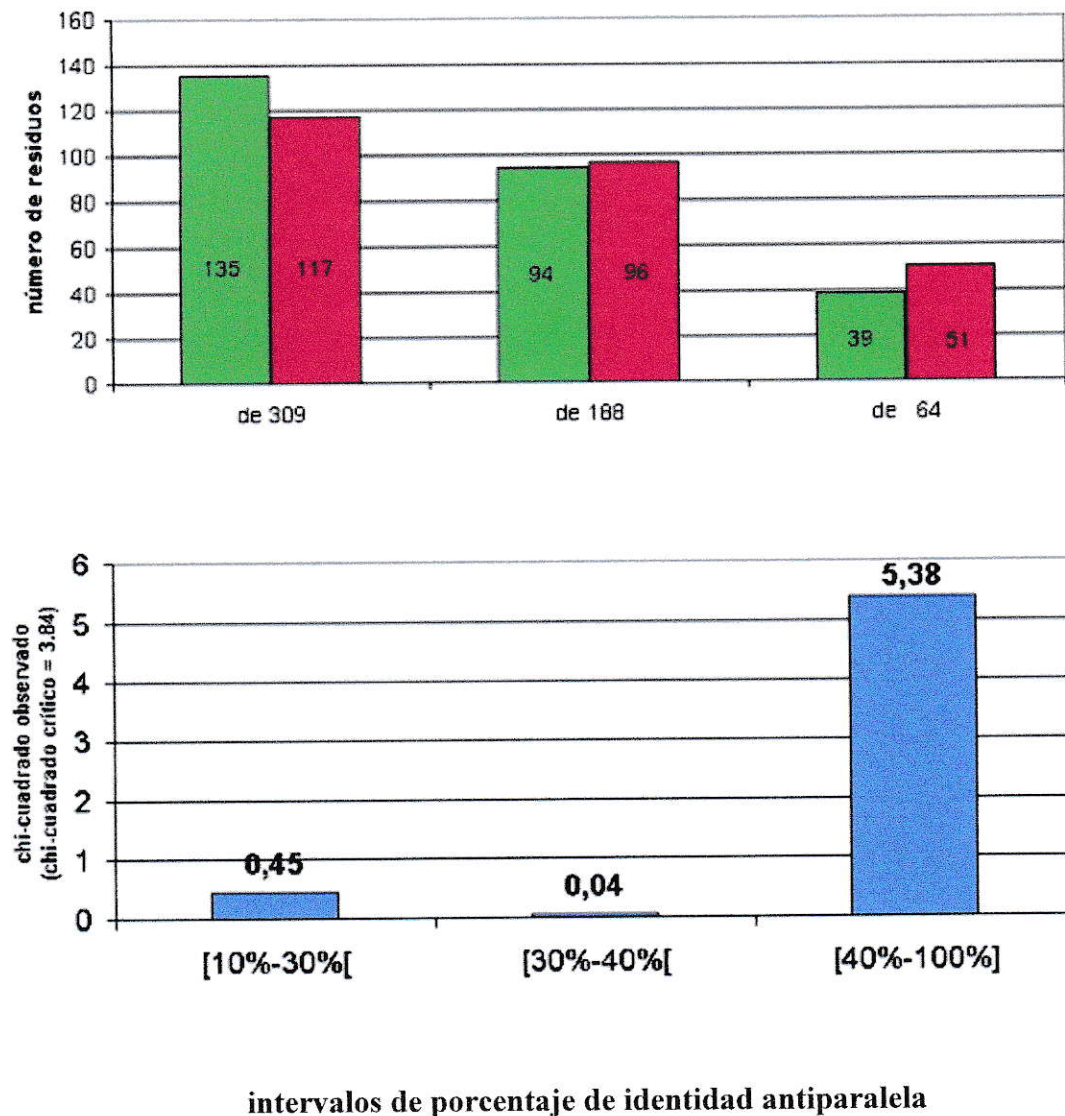


FIGURA 8. Dependencia del porcentaje de identidad de estructura secundaria respecto del porcentaje de identidad antiparalela de secuencia entre las dos mitades de un palíndromo. Gráfico superior: Se compara la estructura secundaria de residuos ubicados en posiciones exactamente opuestas dentro del palíndromo (columnas rojas) respecto del mismo parámetro observado en segmentos de la misma longitud y la misma secuencia de origen tomados al azar (columnas verdes) a distintos intervalos de porcentaje de identidad antiparalela de secuencia. Gráfico inferior: Se puede apreciar que, en el intervalo [40%-100%], la simetría de estructura secundaria es significativamente mayor en los palíndromos que en los controles según la prueba del  $\chi^2$  (nivel de significación = 0,05,  $\chi^2$  crítico = 3,84).

Esta diferencia significativa no debiera ser producto de un aporte excepcionalmente numeroso de residuos simétricos por parte de unas pocas secuencias palindrómicas ya que, según se observa en la TABLA 2, baja consistentemente la cantidad de secuencias control con simetrías superiores a la de sus respectivos palíndromos dentro del intervalo [40-100%].

Se puede concluir que la palíndromía en el intervalo 40-100% en sí es una condición que se relaciona significativamente con el aumento de la simetría de estructura secundaria de los segmentos palindrómicos por sobre lo observado en segmentos aminoacídicos de la misma longitud escogidos al azar. Esto implica que el contexto estructural que rodea un segmento palindrómico en el intervalo 40-100% no posee la suficiente capacidad de inducir el simetría de estructura secundaria que puede inducir en conjunto con la condición de palindrómico del intervalo 40-100% de dicho segmento.

## **5.2 Demostrar que los eventos en que se enfrentan residuos idénticos se asocian significativamente con los eventos de simetría de estructura secundaria en los palíndromos**

Dentro de los modelos de predicción de la estructura secundaria de las proteínas, existe una corriente iniciada por Nishikawa y Ooi (1982) que atribuye principalmente a la composición de residuos de una proteína las proporciones de estructura secundaria de dicha proteína. Esto implica que dos proteínas de la misma composición de residuos mostrarán la misma composición de estructura secundaria aunque sus secuencias no sean las mismas.

Desde este punto de vista, es factible que en las secuencias palindrómicas del intervalo 40-100%, sea la convergencia en una composición aminoacídica similar entre las mitades del palíndromo lo que induce la simetría significativa en la estructura secundaria y no el hecho de conservar un orden invertido de sus secuencias. Así, la condición de idéntico o distinto entre residuos de posición equivalente a los lados opuestos del palíndromo no debiera ser una característica que discrimine significativamente entre residuos de igual y distinta estructura secundaria. Es decir, si el orden invertido de la secuencia es irrelevante respecto de la composición para determinar la simetría de los palíndromos, los eventos de simetría estructural antiparalela debieran repartirse de manera parecida entre los eventos de residuos idénticos y los eventos de residuos distintos.

Para evaluar si hay asociación significativa entre identidad de estructura secundaria y mantenimiento de identidad, se contó los eventos de simetría de estructura secundaria en el grupo de residuos idénticos y distintos y se midió la significación estadística a través de las pruebas de  $\chi^2$  y de la probabilidad exacta de Fischer (TABLA 3).

Para la prueba del  $\chi^2$ , se fijó un nivel de significación de 0,05, lo que determinó un  $\chi^2$  crítico = 3,84 y un  $\chi^2$  observado de 6,3. Por la prueba de la probabilidad exacta de Fischer se fijó también un  $p = 0,05$ , el cual fue superado por los datos, que presentaron un  $p = 0,0219$  (FIGURA 9).

Se puede observar que los residuos idénticos presentan significativamente un mayor número de eventos de identidad de estructura secundaria al lado opuesto del palíndromo que cuando son residuos distintos (FIGURA 9).



TABLA 3

**IDENTIDAD DE ESTRUCTURA SECUNDARIA PARA PALÍNDROMOS DEL INTERVALO 40-100% DE IDENTIDAD ANTIPARALELA DE SECUENCIA**

	Simétrica		Asimétrica	
	IDÉNTICO	DISTINTO	IDÉNTICO	DISTINTO
A	3	2	-	-
C	-	2	-	-
D	3	-	2	-
E	1	2	-	-
F	-	-	-	-
G	4	-	-	-
H	1	1	-	-
I	-	1	-	3
K	3	2	-	-
L	2	1	-	1
M	1	-	-	-
N	2	1	1	-
P	6	1	-	2
Q	-	-	-	-
R	-	-	-	-
S	1	1	1	-
T	2	-	-	1
V	-	3	-	2
W	-	2	-	-
Y	2	-	-	1
<b>TOTALES</b>	<b>31</b>	<b>19</b>	<b>4</b>	<b>10</b>

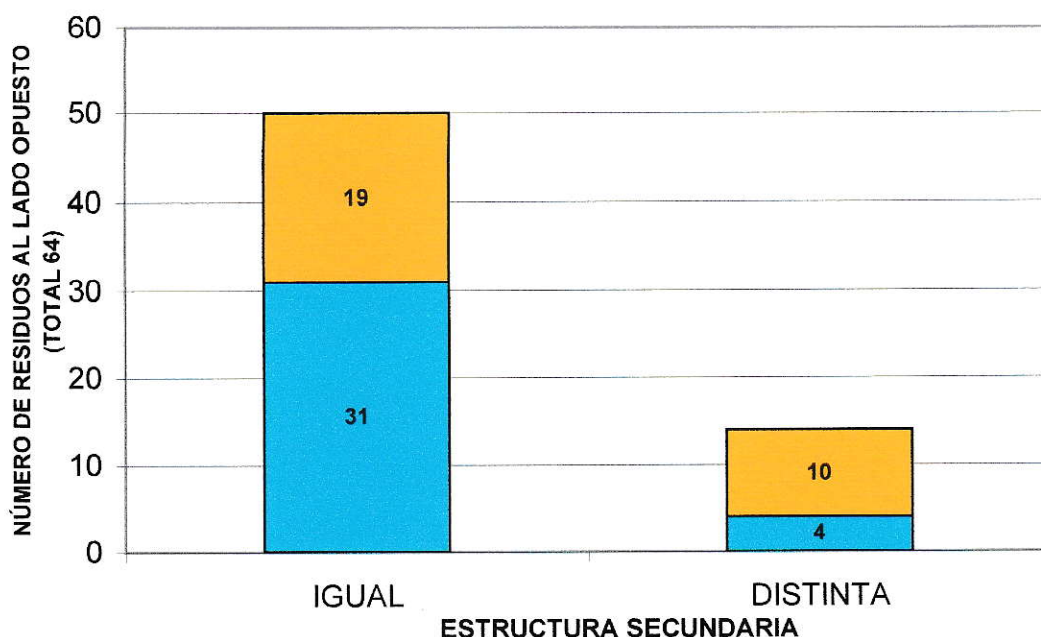


FIGURA 9. Relación entre la conservación de estructura secundaria y el mantenimiento de la identidad en secuencias palindrómicas con más de 40% de identidad antiparalela de secuencia entre sus dos mitades. Se encontró que un residuo perteneciente a una de las mitades de un palíndromo presenta significativamente una mayor simetría de estructura secundaria respecto del residuo opuesto del palíndromo cuando son idénticos (■) que cuando son distintos (■) ( $\chi^2$  crítico = 3.84 <  $\chi^2$  observado = 6,3 , nivel de significación = 0,05; por prueba de Fischer, 0,05 > p = 0,0219).

Sobre la base de este resultado, se puede afirmar que cuando hay dos residuos idénticos en sitios equivalentes y opuestos del palíndromo, dichos residuos tienden a tener simetría de estructura secundaria.

### **5.3 Demostrar que entre los residuos idénticos y los distintos existen preferencias de residuos por sobre lo esperado por azar**

Según lo expuesto en la introducción, se ha observado que algunas retrosecuencias sintetizadas artificialmente no tienen, en la mayoría de los casos, la capacidad de mantener la misma estructura que muestra la secuencia nativa. Esto podría explicarse por la existencia de algunos residuos con incompatibilidad para mantener la misma estructura secundaria en la secuencia invertida. Por otro lado, los palíndromos con mayor porcentaje de identidad antiparalela presentan una mayor simetría de estructura secundaria. Se podría plantear una explicación intermedia a estos dos resultados aparentemente contradictorios. Es posible que no todos los residuos posean la misma capacidad de mantener una estructura secundaria igual en la secuencia invertida. Es decir, podrían existir residuos capaces de mantener la simetría y residuos incapaces de mantener la simetría de estructura secundaria en una secuencia invertida. Si esta simetría de estructura secundaria es una característica biológicamente necesaria, también se puede pensar que exista una selección biológica que tienda a privilegiar la existencia de pares de residuos idénticos o distintos que permitan esta simetría de estructura secundaria dentro de los segmentos palindrómicos.

Sin embargo, hay algunos elementos que conducen a pensar que la formación de palíndromos en las secuencias podría ser un proceso casual en el que probablemente

participe, de manera indistinta, cualquiera de los veinte residuos. Por ejemplo, se puede observar superposiciones de palíndromos en secuencias de proteínas reales, como, además, la existencia de palíndromos en secuencias construidas al azar.

Por todo lo anterior se hizo necesario evaluar en este objetivo si la composición de residuos observada en los segmentos palindrómicos es la misma que la esperable por azar dada la composición y la distribución de residuos en las proteínas que los contienen.

En este punto es preciso, según se detalla más extensamente en los métodos, definir los términos “composición” y “distribución”. En el caso de la composición, si el proceso de formación de palíndromos es al azar, es comprensible concluir que los residuos más abundantes dentro de las proteínas en estudio tenderán a estar mayoritariamente presentes en la composición de los palíndromos respectivos. Sin embargo, la distribución de residuos en las secuencias aminoacídicas no constituye una combinatoria homogénea, sino que presenta ciertos patrones que persiguen la utilidad estructural y funcional de la proteína (Dill, 1990). Así que, para evaluar la participación del azar en las secuencias palindrómicas, es necesario definir en términos operacionales qué es azar dentro de las 48 secuencias aminoacídicas reales, que no han sido construidas estrictamente al azar.

### **5.3.1 Caracterización de la población de residuos observada en las secuencias de origen con un muestreo tipo ventana móvil**

Como resultado del procedimiento realizado en los métodos, (“Preferencia de composición de residuos en las secuencias palindrómicas”), se logró la determinación de la media y desviación estándar de la distribución de frecuencias de cada uno de los



veinte residuos por categoría (centrales, distintos, idénticos, ver ejemplo en FIGURA 10). Respecto de esta población se pudo comparar la frecuencia observada de residuos por categoría dentro de los palíndromos reales.

Para determinar si las preferencias a encontrarse entre los residuos idénticos y distintos se asocian con una mayor simetría de estructura secundaria, en el objetivo siguiente, se considerará rechazados a los diez residuos con menor preferencia (C, I, G, N, Y, R, Q, W, F, S, en el caso de los residuos idénticos del intervalo 40-100%) y preferidos al resto, según se adelanta en los Métodos, y se desarrolla a continuación.

### **5.3.2 Determinación de las preferencias de residuos que más se observa entre los residuos idénticos o distintos respecto de las poblaciones caracterizadas**

En esta sección, se muestra tres tablas con los resultados. Básicamente, responden con qué preferencia (parámetro z normalizado) los residuos tienden a encontrarse entre los pares de residuos idénticos o entre los pares de residuos distintos en las secuencias palindrómicas en estudio. La TABLA 5 corresponde a los datos de preferencias de residuos según mantenimiento de identidad dentro del intervalo 40-100%, es decir el intervalo de palíndromos que mostró mayor simetría. Como control, se sustrajo a los valores de esta tabla las preferencias observadas en el total de 48 palíndromos (TABLA 4) dado que estos globalmente no presentan una simetría significativamente mayor que los segmentos control al azar. Los resultados de esta sustracción están en la TABLA 6, y reflejan, por lo tanto, si hay preferencias dentro de las categorías de residuos centrales, distintos e idénticos que podrían ser responsables del mayor grado de simetría respecto de las secuencias palindrómicas no simétricas.



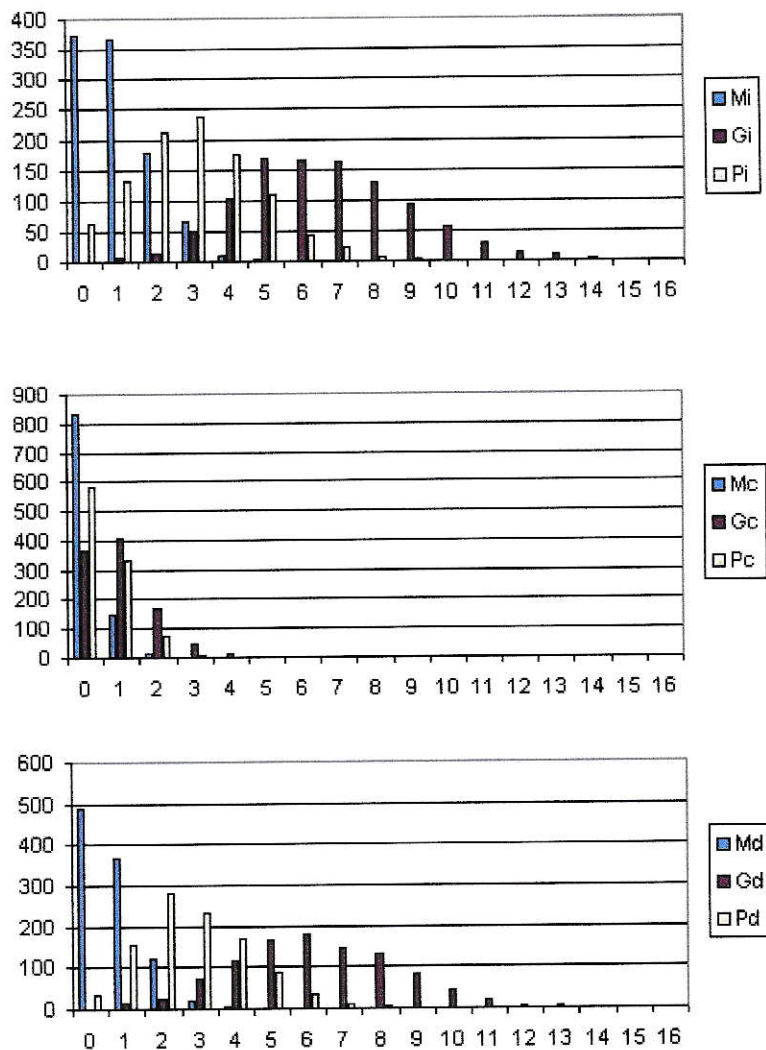


FIGURA 10. Caracterización de la distribución de frecuencias de residuos en mil muestras tomadas al azar a través de una ventana móvil. A modo de ejemplo, se puede observar la distribución de un residuo de baja frecuencia (metionina, M), uno con frecuencia intermedia (prolina, P), y alta frecuencia (glicina, G) en el conteo de residuos usando ventanas móviles moldeados con los patrones de residuos idénticos/distintos/ las ventanas móviles (preferidos), en verde, los que están a 1,64 desviaciones por debajo (rechazados) figurando el resto con calipso (indiferentes). centrales de los palíndromos del intervalo 40-100%. Esta ventana se detiene al azar en la misma secuencia que generó el palíndromo respectivo para extraer en cada una de las mil detenciones una selección arbitraria de residuos (subíndices) i: idénticos, c: centrales, d: distintos.

TABLA 4

**PREFERENCIAS EN LA COMPOSICIÓN DE LAS 48 SECUENCIAS  
PALINDRÓMICAS EN LAS 34 SECUENCIAS ESTUDIADAS**

Preferencia del residuo a clasificar como <b>IDÉNTICO</b>		Preferencia del residuo a clasificar como <b>CENTRAL</b>		Preferencia del residuo a clasificar como <b>DISTINTO</b>		
-2,74	V	-1,61	C	-2,90	C	
-2,36	E	-1,53	A	-2,11	G	
-2,01	L	-1,52	D	-1,41	Y	
-1,93	T	-1,22	V	-1,29	P	
-1,34	F	-0,94	I	-1,16	A	
-1,02	K	-0,59	M	-0,25	N	
-0,94	Q	-0,50	G	-0,19	T	
0,84	A	-0,27	P	-0,02	H	rechazado
0,70	W	0,21	L	0,07	F	indiferente
0,60	I	0,14	R	0,17	M	preferido
0,56	S	0,14	Q	0,25	Q	
0,55	R	0,03	H	0,39	D	
0,21	M	0,14	Y	0,56	K	
0,12	H	0,34	F	0,62	W	
1,56	Y	0,61	W	0,84	I	
1,83	D	0,68	N	0,88	E	
2,01	N	0,98	E	0,99	S	
2,06	G	1,76	T	1,10	L	
3,02	C	1,93	K	1,39	R	
4,41	P	1,84	S	1,42	V	

Las composiciones observadas de los residuos clasificados como residuos idénticos, distintos y centrales muestran desviaciones respecto de la composición esperada al tomar segmentos aleatorios donde se asignan arbitrariamente por posición las mismas clasificaciones. El código de colores muestra cuándo un residuo muestra un valor sobre 5% superior (preferido) o inferior (rechazado) de la población de controles observada para su respectiva categoría (1,64 desviaciones estándar, suponiendo distribución normal). Se puede observar que no todos los residuos tienen la misma probabilidad de aparecer dentro de un palíndromo. Es posible que dentro de estos residuos primen los que usualmente forman parte de repeticiones, no necesariamente relacionados con la formación de palíndromos simétricos.

TABLA 5

**PREFERENCIAS EN LA COMPOSICIÓN DE LAS SECUENCIAS  
PALINDRÓMICAS DEL INTERVALO 40-100%**

Preferencia del residuo a clasificar como		Preferencia del residuo a clasificar como		Preferencia del residuo a clasificar como	
<b>IDÉNTICO</b>		<b>CENTRAL</b>		<b>DISTINTO</b>	
-2,29	I	-0,89	L	-1,86	G
-2,03	V	-0,89	A	-1,64	N
-1,68	F	-0,87	T	-1,30	S
-1,54	Q	-0,84	D	-0,87	K
-1,51	R	-0,81	V	-0,87	M
-1,39	C	-0,81	I	-0,84	Y
-1,07	W	-0,76	Y	-0,74	F
-0,77	E	-0,62	F	-0,60	D
-0,68	S	-0,58	Q	-0,58	Q
-0,63	L	-0,49	H	-0,17	H
-0,27	T	-0,42	M	-0,08	A
0,42	A	-0,40	W	0,14	I
0,49	H	-0,39	C	0,16	E
0,52	Y	0,07	G	0,35	L
0,58	K	0,25	N	0,81	T
0,58	G	0,61	E	0,91	C
0,60	N	0,72	P	1,12	R
1,03	M	0,97	R	1,47	P
3,59	D	2,38	S	1,66	W
5,32	P	2,67	K	4,08	V

rechazado  
indiferente  
preferido

Preferencias en la composición de las 14 secuencias palindrómicas del intervalo 40-100% de identidad antiparalela en las 34 secuencias estudiadas. Igual código de colores que en la TABLA 4.



TABLA 6

**PREFERENCIAS EN LA COMPOSICIÓN DE LAS SECUENCIAS  
PALINDRÓMICAS DEL INTERVALO 40-100%, CON SUBSTRACCIÓN DE  
LAS ASIMÉTRICAS**

Preferencia del residuo a clasificar como <b>IDÉNTICO</b>		Preferencia del residuo a clasificar como <b>CENTRAL</b>		Preferencia del residuo a clasificar como <b>DISTINTO</b>	
-5,21	C	2,63	T	-2,29	S
-1,69	I	-1,01	W	-1,43	K
-1,48	G	-0,96	F	-1,39	N
-1,41	N	-0,90	Y	-1,04	M
-1,04	Y	-0,68	L	-0,99	D
-0,96	R	-0,46	H	-0,83	Q
-0,60	Q	-0,44	Q	-0,81	F
-0,37	W	-0,43	N	-0,75	L
-0,34	F	-0,29	E	-0,72	E
-0,12	S	0,13	I	-0,70	I
0,37	H	0,17	M	-0,27	R
0,71	V	0,41	V	-0,15	H
0,91	P	0,54	S	0,25	G
1,24	M	0,57	G	0,57	Y
1,26	A	0,64	A	1,00	T
1,38	L	0,68	D	1,04	W
1,59	E	0,84	K	1,08	A
1,60	K	0,99	P	2,66	V
1,66	T	1,11	R	2,76	P
2,56	D	1,22	C	3,81	C

rechazado  
indiferente  
preferido

Diferencia entre las preferencias de composición de los 14 palíndromos del intervalo 40-100% de identidad antiparalela de secuencia (simétricos estructuralmente) respecto de las preferencias observadas en los 48 palíndromos (sin simetría global significativa). Igual código de colores que las dos TABLAS anteriores. Residuos como C, P, V, A, W, T, Y, G, H, R tienden a intercambiarse entre sí con alta frecuencia. Análogamente, residuos como D, T, K, E, L, A, M, P, V, H, tienden a encontrar un idéntico al lado opuesto con alta frecuencia.

#### **5.4 Demostrar que tanto entre los residuos idénticos como en los distintos, existen propiedades fisicoquímicas que se asocian con el mantenimiento de la simetría de la estructura secundaria**

Dentro de nuestro sistema palindrómico (FIGURA 7) un residuo idéntico es un residuo que se encuentra al lado opuesto del palíndromo; uno distinto, cuando dicho residuo es reemplazado por otro distinto al lado opuesto. Es posible que algunos residuos sean incompatibles con la misma estructura secundaria en la secuencia invertida debido a ciertas características fisicoquímicas más marcadas. Suichi, Ogata y col. (1999) publicaron una base de datos con aproximadamente 400 índices fisicoquímicos para los veinte residuos. Estos índices, más los índices calculados en el objetivo anterior de Preferencias entre residuos idénticos y entre residuos distintos y el de Preferencia de Centrales en Secuencias Palindrómicas del intervalo [40-100%] (tres índices en total) fueron utilizados en el objetivo presente.

En la TABLA 7 se puede apreciar los datos de simetría (simétrica/asimétrica) para el mantenimiento de la identidad (residuos idénticos, residuos distintos) y el tipo de residuo en código de una letra. Claramente, el número de eventos recopilados en este trabajo es insuficiente para detectar de manera directa el residuo ideal para el mantenimiento de la simetría de estructura secundaria del palíndromo. De todos los residuos, la prolina parece ser uno de los residuos más adecuados para mantener la simetría de estructura secundaria con el mantenimiento del residuo al lado opuesto. Sin embargo, todavía el número de eventos es insuficiente para determinar directamente el efecto de los residuos idénticos y distintos para cada uno de los veinte residuos.



TABLA 7

**SIMETRÍA DE ESTRUCTURA SECUNDARIA ASOCIADA A LA IDENTIDAD  
DE UN PAR DE RESIDUOS ENFRENTADOS UBICADO EN CUALQUIER  
SITIO DEL PALÍNDROMO**

	ESTRUCTURA SECUNDARIA			
	SIMÉTRICA		ASIMÉTRICA	
	RESIDUO IDÉNTICO	RESIDUO DISTINTO	RESIDUO IDÉNTICO	RESIDUO DISTINTO
A	6	4	0	0
C	0	2	0	0
D	6	2	4	0
E	2	2	0	1
F	0	1	0	0
G	8	2	0	0
H	2	1	0	0
I	0	1	0	3
K	6	2	0	0
L	4	2	0	3
M	2	0	0	0
N	4	1	2	0
P	12	2	0	3
Q	0	1	0	0
R	0	4	0	0
S	2	2	2	0
T	4	1	0	4
V	0	5	0	5
W	0	3	0	0
Y	4	0	0	1

Para compensar esta carencia de datos, se procedió a fusionarlos en dos categorías, las con residuos de índice alto y bajo. Los veinte residuos son ahora dos residuos representativos para cada índice, uno con índice alto y otro con índice bajo. Los datos de simetría estructural para los residuos idénticos y distintos también están fusionados.

A continuación se determinó qué parámetros se relacionan significativamente con la simetría para un par de residuos idénticos o distintos, para lo cual se estableció un nivel de significación sobre el 5% con las pruebas de Fischer ( $p = 0,05$ ) y  $\chi^2$  (valor crítico = 3,84,  $gl = 1$ ).

#### **5.4.1 Perfil de los residuos que al mantenerse idénticos muestran estructuras secundarias más simétricas**

Se puede observar en la TABLA 8, que hay un conjunto de 14 propiedades significativas para el mantenimiento de la simetría de estructura secundaria cuando se da un evento de identidad entre residuos ubicados en posiciones equivalentes opuestas. Entre éstas, hay doce separaciones distintas (no redundantes), entre las cuales no se repite el mismo grupo de residuos formando parte del índice bajo o alto.

Cuando el número de hidratación (H114) es elevado en un residuo determinado, es significativamente más probable la asimetría de estructura secundaria (-1) con la identidad que cuando dicho índice es bajo.

Así, se puede apreciar que un residuo grande (+1/P059), de baja polaridad (-1/H400), presente comúnmente en la hebras beta (+1/A274), de alto punto isoeléctrico (+1/H401), de baja mutabilidad (-1/P065,-1/P135), de alta energía libre de transferencia

desde el vapor al agua neutra (+1/H392) tenderá a mantener más la simetría de estructura secundaria al ser idéntico. Se puede apreciar también que mientras más preferido sea este residuo entre los residuos que forman pares de residuos distintos (+1/Preferencias entre residuos distintos del intervalo 40-100% de identidad antiparalela de secuencia, objetivo anterior en este trabajo), es significativamente más probable el mantenimiento de la estructura secundaria con la identidad. No hubo una asociación importante con los otros dos índices de preferencia obtenidos en este trabajo.

Por lo tanto, la primera aproximación de estos resultados sugiere que un residuo del tipo hidrofóbico tenderá a mantener la misma estructura secundaria al ser idéntico en la secuencia invertida del lado opuesto del palíndromo. Consecuentemente, cuando dicho tipo de residuos forma pares de residuos distintos, el sistema tiende a mantener (+1/Preferencias entre Distintos) en el residuo reemplazante el alto índice de hidrofobicidad.

Por otro lado, dada la propensión de estos residuos hidrofóbicos a formar hebras beta (+1/A274), es esperable que esta clase de estructura secundaria pueda mantenerse de manera significativa en la secuencia invertida y, si uno de sus residuos forma pares de residuos distintos, el residuo distinto a localizarse al lado opuesto será del tipo hidrofóbico también (+1/Preferencias entre residuos distintos).

Para cada índice tenemos diez residuos con índice alto y diez con índice bajo. Dado que cada uno de estos ordenamientos cumple con un nivel de significación previamente elegido (en nuestro caso el 5% superior), los doce ordenamientos diferentes que hay dentro de las 14 propiedades correlacionadas son equivalentes desde el punto de vista de su asociación con la simetría de estructura secundaria de los residuos idénticos.

No hay dentro de estas 12 propiedades, en rigor, ninguna con mayor significación que otra. A modo de aproximación gruesa a la identificación de cuáles residuos presentan índices (altos o bajos) permisivos de la simetría de estructura secundaria al mantener la identidad, es posible cuantificar qué residuos aparecen más frecuentemente exhibiendo índices altos o bajos favorables a la identidad simétrica en el total de las doce separaciones distintas:

**12P > 11G > 10L > 9V, A > 8K, F > 7Y, R, C > 6W, I > 5M > 4T, E > 3Q, H > 1N > OS, D.**

El residuo prolina, por ejemplo, aparece con valores permisivos de la identidad simétrica en las doce separaciones distintas, mientras que los residuos serina y aspártico no aparecen en ninguna de ellas. Si seguimos con el esquema de separaciones de diez, los residuos P, G, L, V, A, K, F, Y, R y C debieran mostrar una alta capacidad para mantener la misma estructura secundaria al mantenerse idénticos en la secuencia invertida, mientras que los residuos restantes probablemente muestren lo contrario.

#### **5.4.2 Perfil de los residuos que al enfrentarse a un residuo distinto muestran estructuras secundarias más simétricas**

Como se muestra en la TABLA 9 se encontró 34 propiedades significativas. De ellas, 25 correspondían a separaciones distintas. Cuando un residuo presenta baja preferencia en el extremo amino de  $\alpha$ -hélices (+1/H326) bajo número de hidratación (+1/H114), de baja energía libre ( $\Delta G$ ) de transferencia a la superficie (+1/H014), baja frecuencia en las vueltas o final de hebras beta (+1/O374, +1/P282), baja presencia en hélices alfa (+1/A259), baja polaridad (+1/H400), dicho residuo distinto probablemente presentará también una menor simetría de estructura secundaria en la secuencia



palindrómica al enfrentarse a otro residuo distinto al lado opuesto del palíndromo. En la misma situación se encontraría un residuo grande (-1/P399), frecuente en estructuras extendidas o hebras beta (-1/B368, -1/B020, -1/B172, -1/B345, -1/A274, -1/B102), o de alta retención en HPLC (-1/H179, -1/H178).

Cabe destacar que los residuos con una alta Preferencia entre residuos con identidad antiparalela (índice obtenido en este trabajo) tienen correspondencia con el tipo de residuos cuyo enfrentamiento a un residuo distinto deriva en una baja probabilidad de mantener la misma estructura secundaria. Esta relación alcanza a ser significativa con la prueba de Fischer ( $p = 0,04172 < 0,05$ ) pero no con la de  $\chi^2$  ( $\chi^2$  observado = 3,4714 < 3,83). Por otro lado, los residuos con una alta Preferencia entre los Residuos Centrales corresponden al tipo de residuos cuyo enfrentamiento a un residuo distinto aumenta la probabilidad de simetría de estructura secundaria ( $p$  Fischer significativo = 0,03909 < 0,05;  $\chi^2$  observado no significativo = 3,5536 < 3,83).

Nuevamente, se puede proponer el siguiente bosquejo de cuáles residuos presentan más frecuentemente simetría al enfrentarse con residuos distintos, por el conteo de los residuos que aparecen más frecuentemente exhibiendo índices favorables (ya sea altos o bajos) en el total de las veinticinco separaciones no redundantes:

**22K, D > 21E > 19N, C, A > 18H > 16Q, G > 15S, R > 11W > 8M > 7F > 6Y > 5L > 4V, T > 3P > 1I**

En esta serie se observa que en 22 de las 25 separaciones no redundantes aparecen la lisina y el aspartato con índices permisivos para la formación de pares de residuos distintos con simetría de estructura secundaria, mientras la prolina en tres y la isoleucina en uno. Es decir, al enfrentarse a residuos distintos, para I, P, T, V, L, Y, F, M, W, R (apartando arbitrariamente S), debe ser más difícil encontrar entre los 19



residuos restantes un reemplazante que logre mantener la simetría de estructura secundaria.

TABLA 8

RELACIÓN ENTRE LA MAGNITUD DE LOS ÍNDICES

Y LA SIMETRÍA DE ESTRUCTURA SECUNDARIA CON LA IDENTIDAD DE LOS RESIDUOS

fischer	$\chi^2$ simet- identidad v/s ordenación	Identidad Reduce (-1) Aumenta(1) Simetria si Índice es Alto	DESCRIPCIÓN DEL INDICE
0,00065847	6,77419355	-1	H373 TANS770108 Normalized frequency of Zr (Tanaka-Scheraga, 1977)Macromolecules 10, 9-20 (1977)
0,00384651	4,78178368	-1	H114 HOPA770101 Hydration number (Hopfinger, 1971), Cited by Charton-Charton (1982) Wiley, New York (1977)
0,00384651	4,78178368	-1	H400 ZIMJ680103 Polarity (Zimmerman et al., 1968)J. Theor. Biol. 21, 170-201 (1968)
0,00384651	4,78178368	-1	P065 DAYM780201 Relative mutability (Dayhoff et al., 1978a)In "Atlas of Protein Sequence and Structure", National Biomedical Research Foundation
0,00384651	4,78178368	-1	P135 JOND920102 Relative mutability (Jones et al., 1992)CABIOS 8, 275-282 (1992)
0,00384651	4,78178368	-1	O173 MAXF760103 Normalized frequency of zR (Maxfield-Scheraga, 1976)Biochemistry 15, 5138-5153 (1976)
0,0051	4,7818	1	A274 QIAN880117 Weights for beta-sheet at the window position of -3 (Qian-Sejnowski, 1988)J. Mol. Biol. 202, 865-884 (1988)
0,00512868	4,78178368	1	H392 WOLR810101 Hydration potential (Wolfenden et al., 1981)Biochemistry 20, 849-855 (1981)
0,00512868	4,7817	1	Preferencias entre residuos distintos en secuencias palindromicas con mas de 40% de identidad antiparalela de secuencia
0,00512868	4,26523297	1	P059 COHE430101 Partial specific volume (Cohn-Edsall, 1943)"Protein, Amino Acid, and Peptides", Reinhold, New York (1943)
0,00629445	5,36290323	1	H401 ZIMJ680104 Isoelectric point (Zimmerman et al., 1968)J. Theor. Biol. 21, 170-201 (1968)
0,00641085	4,26523297	-1	H110 GRAR740101 Composition (Grantham, 1974)Science 185, 862-864 (1974)
0,00641085	4,26523297	-1	H371 TANS770106 Normalized frequency of chain reversal D (Tanaka-Scheraga, 1977)Macromolecules 10, 9-20 (1977)
0,00894916	3,97834189	-1	C301 RACS820105 Average relative fractional occurrence in E0(i) (Rackovsky-Scheraga, 1982)Macromolecules 15, 1340-1346 (1982)

TABLA 9

RELACIÓN ENTRE LA MAGNITUD DE LOS ÍNDICES

Y LA SIMETRÍA DE ESTRUCTURA SECUNDARIA CON LA DESIGUALDAD DE LOS RESIDUOS

Fischer	$\chi^2$	Pares distint.	Reduce (-1)	Aumenta(1)	Simetria	DESCRIPCIÓN DEL INDICE
P (obs+ext)	Simet-pares	distint.	Simetria	si Índice	es Alto	
Pares	v/s	Ordenación				
distint.						
5,055E-05	8,033241	1	H326	RICJ880105	Relative preference value at N2 (Richardson-Richardson, 1988)Science 240, 1648-1652 (1988)	
5,2308E-05	8,033241	-1	H179	MEEJ800102	Retention coefficient in HPLC, pH2.1 (Meek, 1980)Proc. Natl. Acad. Sci. USA 77, 1632-1636 (1980)	
5,2308E-05	8,033241	-1	P078	FAUJ880101	Graph shape index (Fauchere et al., 1988)Int. J. Peptide Protein Res. 32, 269-278 (1988)	
0,00010239	6,86842105	1	H014	BULH740101	Transfer free energy to surface (Bull-Breese, 1974)Arch. Biochem. Biophys. 161, 665-670 (1974)	
0,00010239	6,86842105	1	H388	WOEC730101	Polar requirement (Woese, 1973)Naturwiss. 60, 447-459 (1973)	
0,00010635	8,033241	-1	H178	MEEJ800101	Retention coefficient in HPLC, pH7.4 (Meek, 1980)Proc. Natl. Acad. Sci. USA 77, 1632-1636 (1980)	
0,00010635	6,86842105	-1	P399	ZIMJ680102	Bulkiness (Zimmerman et al., 1968)J. Theor. Biol. 21, 170-201 (1968)	
0,00020952	5,81453634	-1	B368	TANS770103	Normalized frequency of extended structure (Tanaka-Scheraga, 1977)Macromolecules 10, 9-20 (1977)	
0,00040086	8,033241	-1	B020	BURA740102	Normalized frequency of extended structure (Burgess et al., 1974)Isr. J. Chem. 12, 239-286 (1974)	
0,00040086	6,86842105	-1	H012	BROC820101	Retention coefficient in TFA (Browne et al., 1982)Anal. Biochem. 124, 201-208 (1982)	
0,00040086	6,86842105	-1	H402	ZIMJ680105	RF rank (Zimmerman et al., 1968)J. Theor. Biol. 21, 170-201 (1968)	
0,00225159	4,04955679	1	O370	TANS770105	Normalized frequency of chain reversal S (Tanaka-Scheraga, 1977)Macromolecules 10, 9-20 (1977)	
0,00255689	6,19579463	1	H329	RICJ880108	Relative preference value at N5 (Richardson-Richardson, 1988)Science 240, 1648-1652 (1988)	
0,00285536	6,19579463	-1	A310	RACS820114	Value of theta(i-1) (Rackovsky-Scheraga, 1982)Macromolecules 15, 1340-1346 (1982)	
0,00345368	4,04955679	1	H114	HOPA770101	Hydration number (Hopfinger, 1971), Cited by Charton-Charton (1982) Wiley, New York (1977)	
0,00345368	4,04955679	1	O051	CHOP780214	Frequency of the 3 <sup>rd</sup> residue in turn (Chou-Fasman, 1978b)Adv. Enzymol. 47, 45-148 (1978)	
0,00347186	5,0576555	-1	C201	NAKH900105	AA composition of mt-proteins from animal (Nakashima et al., 1990)PROTEINS 8, 173-178 (1990)	
0,00347186	5,0576555	-1	C208	NAKH900112	Transmembrane regions of mt-proteins (Nakashima et al., 1990)PROTEINS 8, 173-178 (1990)	
0,00347186	4,04955679	-1	H392	WOLR810101	Hydration potential (Wolfenden et al., 1981)Biochemistry 20, 849-855 (1981) (CONTINÚA EN PÁG. SIGUIENTE)	

0,0048646	6,19579463	1	A231 PALJ810109 Normalized frequency of alpha-helix in alpha/beta class (Palau et al., 1981)Int. J. Peptide Protein Res. 19, 394-401 (1981)
0,00486472	4,04955679	-1	B172 MAXF760102 Normalized frequency of extended structure (Maxfield-Scheraga, 1976)Biochemistry 15, 5138-5153 (1976)
0,00577258	5,0576555	1	O374 TANS770109 Normalized frequency of coil (Tanaka-Scheraga, 1977)Macromolecules 10, 9-20 (1977)
0,00799058	4,88711779	-1	B344 ROBB760105 Information measure for extended (Robson-Suzuki, 1976)J. Mol. Biol. 107, 327-356 (1976)
0,00799058	4,88711779	-1	B345 ROBB760106 Information measure for pleated-sheet (Robson-Suzuki, 1976)J. Mol. Biol. 107, 327-356 (1976)
0,00799058	4,04955679	-1	P096 GARJ730101 Partition coefficient (Garel et al., 1973)J. Chromatogr. 78, 381-391 (1973)
0,00799487	4,88711779	-1	H202 NAKH900106 Normalized composition from animal (Nakashima et al., 1990)PROTEINS 8, 173-178 (1990)
0,00825277	5,0576555	1	P282 QIAN880125 Weights for beta-sheet at the window position of 5 (Qian-Sejnowski, 1988)J. Mol. Biol. 202, 865-884 (1988)
0,01267865	7,48555162	-1	A274 QIAN880117 Weights for beta-sheet at the window position of -3 (Qian-Sejnowski, 1988)J. Mol. Biol. 202, 865-884 (1988)
0,01267865	4,88711779	-1	B102 GEIM800106 beta-strand indices for beta-proteins (Geisow-Roberts, 1980)Int. J. Biol. Macromol. 2, 387-389 (1980)
0,02893868	5,0576555	1	A097 GEIM800101 a-helix indices (Geisow-Roberts, 1980)Int. J. Biol. Macromol. 2, 387-389 (1980)
0,02893868	5,0576555	1	A186 NAGK730101 Normalized frequency of alpha-helix (Nagano, 1973)J. Mol. Biol. 75, 401-420 (1973)
0,02893868	5,0576555	1	A223 PALJ810101 Normalized frequency of alpha-helix from LG (Palau et al., 1981)Int. J. Peptide Protein Res. 19, 394-401 (1981)
0,02893868	5,0576555	1	A304 RACS820108 Average relative fractional occurrence in AR(i-1) (Rackovsky-Scheraga, 1982)Macromolecules 15, 1340-1346 (1982)
0,02965838	6,15177945	1	A259 QIAN880102 Weights for alpha-helix at the window position of -5 (Qian-Sejnowski, 1988)J. Mol. Biol. 202, 865-884 (1988)
0,02965838	4,88711779	1	A343 ROBB760104 Information measure for C-terminal helix (Robson-Suzuki, 1976)J. Mol. Biol. 107, 327-356 (1976)
0,02983066	5,0576555	-1	H152 LAWE840101 Transfer free energy, CHP/water (Lawson et al., 1984)J. Biol. Chem. 259, 2910-2912 (1984)
0,0417168	5,0576555	1	A160 LEVM780101 Normalized frequency of alpha-helix, with weights (Levitt, 1978)Biochemistry 17, 4277-4285 (1978)
0,0417168	5,0576555	1	A250 PRAM900102 Relative frequency in alpha-helix (Prabhakaran, 1990)Biochem. J. 269, 691-696 (1990)
0,0427045	4,04955679	1	H400 ZIMJ680103 Polarity (Zimmerman et al., 1968)J. Theor. Biol. 21, 170-201 (1968)



## 6. DISCUSIÓN

Esta parte del trabajo se inicia contrastando los resultados observados en el análisis de las secuencias palindrómicas con los ya disponibles en la literatura del mismo tópico (6.1) para continuar con un resumen del papel de la polaridad/hidrofobicidad, el volumen, la propensión por ciertas estructuras secundarias y el papel de las cargas en el mantenimiento de la simetría de estructura secundaria en la secuencia invertida. Dado que el problema de la simetría en los palíndromos se relaciona con problema general de la estabilidad de la estructura secundaria al invertir el orden de la secuencia de una proteína, aquel resumen deriva en el análisis de ciertos tópicos generales sobre la estabilidad de proteínas de la literatura (6.2). En una sección aparte se trata las preferencias de residuos acusadas en los palíndromos y cómo estas preferencias parecen perseguir la simetría de estructura secundaria (6.3). Se propone un modelo físico no descriptivo, que intenta, a través de una vía principalmente conceptual, predecir los comportamientos de mantenimiento o la ruptura de la simetría de estructura secundaria en una secuencia invertida a partir de ciertas características moleculares de un residuo aminoacídico (6.4). Una vez diseñado este modelo, se tipifica en el contexto de este modelo a los veinte residuos aminoacídicos según su comportamiento de mantenimiento/ruptura de la simetría de estructura secundaria observado en los datos (6.5). Se podrá apreciar que los diversos comportamientos de los residuos en los datos, como también las estructuras y topología de cargas de dichos residuos se adhieren

aceptablemente a los elementos estructurales y topológicos particulares que en el modelo se requieren para configurar el comportamiento específico observado.

Finalmente se extiende brevemente la discusión a ciertos puntos sobre conservación de secuencias (6.6) el mantenimiento de la función en secuencias invertidas (6.7) y los mecanismos de evolución que pudieron dar origen a las secuencias palindrómicas (6.8).

### **6.1 Simetría de estructura secundaria e identidad de secuencia entre las mitades de un palíndromo**

Se analiza en esta tesis segmentos aminoacídicos palindrómicos que presentan porcentajes variables de identidad antiparalela de secuencia. El objetivo fue evaluar la simetría de estructura secundaria entre secuencias que presentan el mínimo esperable de variación de contexto (son consecutivas) y, por otro lado, dejar abierta la posibilidad de determinar si, para una mayor simetría de estructura secundaria, será siempre necesario un residuo idéntico en la secuencia opuesta. La existencia de este tipo de residuos distintos pero más simétricos que el residuo idéntico en la secuencia invertida podría explicar la irregularidad con que las secuencias invertidas a veces se pliegan y otras, como las secuencias invertidas de Lacroix, Viguera, y col. (1998) no logran plegarse de forma nativa.

Los resultados de este trabajo indican que las secuencias palindrómicas con más de un 40% de identidad antiparalela de secuencia entre sus mitades mostraron un número significativamente mayor de eventos de identidad antiparalela de estructura secundaria (51/64) que en segmentos escogidos al azar de las mismas secuencias de origen (39/64). Expresado fraccionalmente ( $51/64=0,79$ ) este resultado es concordante

con el promedio de 74% de identidad antiparalela de estructura secundaria exhibido en los resultados de Preissner, Goede y col. (1997) provenientes de secuencias de longitud variable con “auto-similitud de secuencia invertida” (aquí llamadas palindrómicas) con más de 34% de identidad antiparalela de secuencia. Sin embargo, podría ser superior al 60-70% de identidad antiparalela de estructura secundaria de Park, Dietmann y col. (2000) observado en secuencias de longitud variable con 100% de identidad antiparalela de secuencia. Esta diferencia, de ser significativa, podría explicarse por dos razones. En primer lugar, por tratarse de secuencias con 100% de identidad antiparalela, cabe la posibilidad de la inclusión de residuos polares, como veremos más adelante, en las secuencias comparadas por Park, Dietmann y col. (2000) induzca las diferencias de identidad antiparalela de estructura secundaria. Esta explicación también podría aplicarse a secuencias independientes, como en el caso de Lacroix, Viguera y col. (1998) en que pese a la total identidad antiparalela de secuencia de péptidos expresados *in vivo*, sólo aquellos nativos adquirieron la estructura secundaria nativa (por dicroísmo circular) mientras que los reversos (sintetizados) no la adquirieron totalmente o no se expresaron. La segunda explicación es la influencia del contexto estructural que rodea a estas secuencias. Al revisar el método de Park, Dietmann y col. (2000), éste no especifica si incluye tanto segmentos consecutivos con identidad antiparalela de secuencia, como segmentos separados dentro de la secuencia. La inclusión de segmentos con identidad antiparalela de secuencia separados dentro de la estructura primaria de una proteína podría reducir ostensiblemente la probabilidad de simetría de estructura secundaria dada la diversidad de contextos en que estos segmentos podrían estar insertos. Es tan importante el factor del contexto que su influencia se hace patente en el caso de los



segmentos aminoacídicos que se ubican en proteínas distintas: pese a presentar 100% de identidad paralela (Kabsch y Sander, 1984, 16 secuencias) o sobre el 34% de identidad antiparalela de secuencia (Preissner, Goede y col., 1997, 17 secuencias) dichos segmentos aminoacídicos exhiben porcentajes de identidad de estructura secundaria que oscilan alrededor del 40%. Dado que en ambos casos se compara segmentos que forman parte de secuencias distintas, es muy probable que el contexto estructural sea responsable de anular el papel que pueda aportar la similitud de secuencia.

## **6.2 Propiedades de un residuo cualquiera para lograr simetría de estructura secundaria al ser enfrentado por residuo idéntico/distinto**

El bajo número de datos recopilado indujo a fusionar las categorías de los veinte residuos y sus respectivos veinte índices en dos grupos o categorías de alto y bajo índice. De este modo, se elude la baja capacidad de resolver directamente los comportamientos de simetría de cada residuo al ser enfrentados a residuos idénticos o distintos. Pero, simultáneamente, se otorga más consistencia a la identificación de las propiedades fisicoquímicas más asociadas con el mantenimiento de la simetría de estructura secundaria por parte de una secuencia invertida. Así, como resultado de la fusión de datos, las propiedades fisicoquímicas significativas en el mantenimiento de la simetría de estructura secundaria para los residuos que formaron pares idénticos fueron consistentes con las de los que formaron pares distintos.



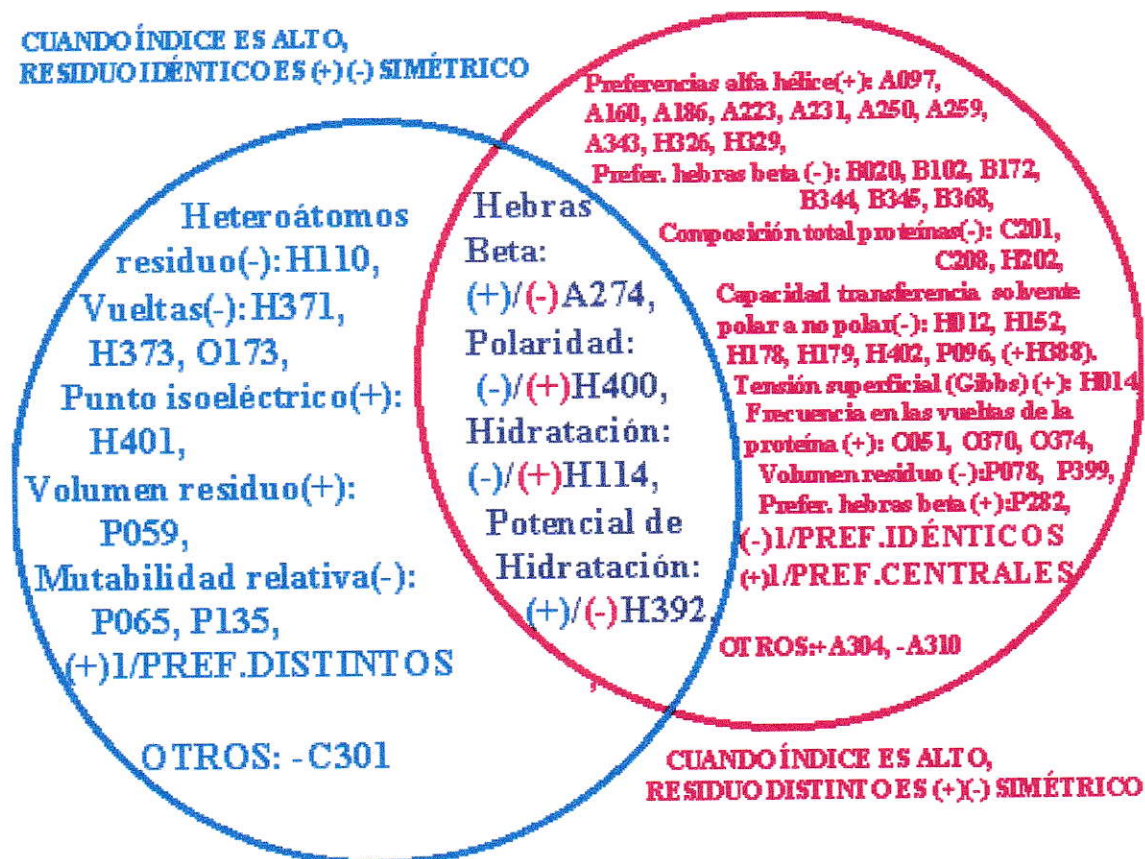


FIGURA 11. Propiedades significativas para el mantenimiento de la simetría de estructura secundaria dados dos residuos idénticos o distintos en posiciones equivalentes opuestas del palíndromo. En azul, a la izquierda, se aprecia las propiedades asociadas significativamente con las identidades que muestran simetría de estructura secundaria. A la derecha, en rojo, las propiedades asociadas con los residuos que al ser enfrentados con otro distinto mantienen la simetría de estructura secundaria. En morado, las propiedades significativas tanto para ambos casos. El signo indica la mayor (+) o menor (-) simetría de estructura secundaria cuando la propiedad está en su valor más positivo. OTROS: propiedades no identificadas o no inteligibles. Ejemplo: un alto número de heteroátomos en el residuo (H110) se asocia significativamente con una menor identidad de estructura secundaria, es decir, menos (-) simétrica cuando el residuo se mantiene idéntico. Por otro lado, una alta capacidad de hidratación del residuo (H114) se asocia simultáneamente con una menor (-) simetría al mantener la identidad y con una mayor (+) simetría al ser enfrentado a un residuo distinto.

**Polaridad-hidrofobicidad:**

Los resultados de la TABLA 8 y la FIGURA 11 apuntan a que un residuo que se enfrenta a un residuo idéntico tenderá a mantener la simetría de estructura secundaria cuando presente un bajo número de hidratación (-1/H114) o baja polaridad definida en función de la carga neta del residuo y su momento dipolar (-1/H400). Del mismo modo, una baja composición de heteroátomos dentro del residuo (-1/H110) puede reducir el número de enlaces dipolares, y con ello configurar residuos menos polares o con una mayor energía libre (desfavorable) de transferencia desde el vapor al agua (+1/H392) capaces de mantener la simetría con los residuos idénticos. Simplificando, los residuos más hidrofóbicos tienen tendencia a mantener la simetría de estructura secundaria cuando se mantienen en la secuencia invertida, mientras que los residuos hidrofílicos, al ser idénticos en la secuencia invertida, rompen la simetría con una frecuencia significativamente mayor que los hidrofóbicos.

Aunque analizada separadamente, el enfrentamiento de un residuo determinado a otro residuo distinto arrojó resultados consecuentes con los de la identidad (FIGURA 11). Para mantener la simetría de estructura secundaria, el primer residuo debe poseer un menor coeficiente de retención en una columna alifática de HPLC, es decir, un residuo no poseedor de cadenas laterales alifáticas o aromáticas, tiene una mayor probabilidad de mantener la misma estructura secundaria en la secuencia invertida si en el sitio equivalente opuesto se localiza un residuo distinto (-1/H179, -1/H178, -1/H012). Igualmente, un residuo que en una cromatografía de papel migra con las mezclas más acuosas de la fase líquida (+1/H388) o que tienen un menor  $R_f$  (-1/H402), menor energía libre de transferencia (más favorable) desde el vapor o fase orgánica al agua (-1/H392, -

1/H152) o mayor polaridad (+1/H400), corresponde a un residuo cuyo enfrentamiento a un residuo distinto se asocia con una mayor probabilidad de mantener la simetría de estructura secundaria. Simplificando, cuando un residuo es más polar, su reemplazo por un residuo distinto en la secuencia invertida tiende a mantener la misma estructura secundaria.

**Volumen:**

En el caso de la identidad, los residuos con volumen específico parcial alto son más simétricos con la identidad (+1/P059). Inversamente, los residuos que más mantienen la simetría de estructura secundaria al enfrentarse a uno distinto son aquellos con un menor abultamiento (-1/P399, -1/P078).

**Mutabilidad relativa:**

Los residuos de identidad más simétrica son aquellos cuya mutabilidad relativa respecto del residuo alanina es menor (-1/P065, -1/P135). Esta relación no es significativa en el caso de los residuos distintos.

**Preferencia en hebras beta:**

Comúnmente se describe a las hebras beta como una red altamente compleja en que un segmento constituido por residuos mantiene estabilizada su estructura secundaria gracias a la presencia de puentes de hidrógeno con sitios alejados en la secuencia. Esta característica más bien de estructura terciaria, debiera imprimir un alto grado de estereoespecificidad a las hebras beta, y, por lo tanto, una sensibilidad importante a los cambios de sentido de la secuencia aminoacídica. Sin embargo, según los resultados, los residuos idénticos tienen una mayor probabilidad de mantener la simetría de estructura



secundaria cuando tienen una mayor frecuencia en las hebras beta (+1/A274). Por su parte, un residuo que es enfrentado a otro distinto presenta mayor asimetría de estructura secundaria si es un residuo con alta preferencia en hebras beta (-1/A274, -1/B020, -1/B102, -1/B172, -1/B344, -1/B345, -1/B368). Es decir, la inversión de una secuencia aminoacídica que posea estructura de hebra beta no debiera romper dicha estructura al lado opuesto del palíndromo.

### **Preferencia en alfa hélice:**

La escala de preferencias en hélice no es un discriminador significativo entre residuos de mayor y menor simetría al ser idénticos, pero sí lo es cuando los residuos son distintos. En esta última se observa que los pares de residuos distintos tienden a ser más simétricas cuando al menos uno de ellos es un residuo de alta preferencia en alfa hélice (+1/A097, +1/A160, +1/A186, +1/A223, +1/A231, +1/A250, +1/A259, +1/A343, +1/H326, +1/H329).

### **Punto isoeléctrico y presencia en vueltas:**

Hay mayor simetría de estructura secundaria cuando la identidad involucra residuos de baja preferencia en las vueltas (-1/H371). Estos residuos generalmente se encuentran en las zonas del diagrama de Ramachandran  $-180^{\circ} \leq \varphi \leq 0^{\circ}$  y  $-9^{\circ} \leq \psi \leq 110^{\circ}$  y tienden a tener una estructura secundaria asimétrica con la identidad (-1/H373, -1/O173). En los pares de residuos distintos más simétricos, al menos uno de ellos posee una alta preferencia en las vueltas de la proteína (+1/O051, +1/O370, +1/O374). (+)1/P282 indica que a mayor frecuencia en la posición cinco de una ventana ubicada en



la sábana beta, mayor simetría de estructura secundaria al enfrentarse a un residuo distinto. La posición 5 referida se encuentra casi en el extremo de una ventana de trece residuos. Dado que la longitud de una hebra beta tiende a no superar los diez residuos de largo (Dill, 1990) es probable que la posición cinco refleje en realidad la composición de una vuelta, más que de una hebra propiamente tal.

Los residuos cuya identidad es más simétrica son aquellos con punto isoeléctrico mayor (1+/H401). Dado que el concepto de punto isoeléctrico no procede para un residuo cuyos extremos amino y carboxilo se encuentran bloqueados en una proteína, es más probable que bajo esta separación estemos realmente viendo el papel de la carga neta del residuo en el mantenimiento de la simetría de estructura secundaria. Desde este punto de vista, aquellos residuos cargados positivamente tenderían a mantener la simetría de estructura secundaria más exitosamente que los cargados negativamente. Por otro lado, esto podría deberse a que los residuos más positivos posee, simultáneamente, un mayor componente hidrofóbico que sus contrapartidas negativas, lo que les otorgaría una menor sensibilidad al cambio en la orientación de la secuencia a diferencia de los cargados negativamente (+1/H401).

En este punto es necesario especificar respecto de la función de los residuos cargados en los extremos de las hélices.

En los trabajos de Robson y Suzuki (1976) Gibrat, Garnier y col. (1987) Richardson y Richardson (1988) y Qian y Sejnowski (1988) se encontró que las alfa hélices poseen una tendencia a la existencia de residuos positivamente cargados hacia carboxilo terminal y negativamente cargados hacia amino terminal. En el caso de los residuos negativos, su papel sería el de promover el comienzo de la alfa-hélice, mientras

que los positivos actuarían como terminadores de la misma. Si fuera cierto este papel estabilizador extremo-específico de las cargas sobre las alfa-hélices, se podría esperar que, con la inversión de la secuencia el flanqueo de cargas de la hélice se invierta, desestabilizándola. La asociación de los residuos de alto punto isoeléctrico, que agrupan residuos positivos (+1/H401) con identidades simétricas podría indicarnos que un residuo positivo clásicamente terminador de hélice como histidina, lisina o arginina podría estar hacia el extremo amino de una alfa-hélice de la secuencia invertida sin desestabilizarla. A la vez, esta misma asociación indicaría que un residuo de bajo punto isoeléctrico, en cuyo extremo se ubican los residuos negativos aspartato y glutamato, al ser mantenido en la secuencia invertida (desplazándose desde el extremo amino al carboxilo de la misma alfa-hélice, por lo tanto) podría desestabilizar dicha estructura secundaria. Es decir, para la simetría de estructura secundaria entre las mitades de un palíndromo sería más importante conservar la presencia de cargas negativas hacia el extremo amino, sin tener tanta relevancia el extremo en el cual se ubican los residuos positivos. De manera concordante con estos resultados, existen antecedentes que describen un menor componente estabilizador extremo específico de los residuos cargados positivamente en las hélices que la mostrada por los residuos cargados negativamente. Fersht (1999) reseña al residuo histidina como débilmente estabilizador respecto de alanina en la posición del sombrero amino terminal (*N-cap*) pudiendo establecer puentes de hidrógeno con el tercer residuo hacia carboxilo (N+3). A la vez, el mismo autor describe a los residuos histidina, arginina y lisina como débilmente estabilizadores de la alfa-hélice, al encontrarse hacia carboxilo. En cambio, las cargas negativas provistas por aspártico y glutámico muestran un importante componente

estabilizador hacia el extremo amino, ya que, según Fersht (1999) poseen una débil estabilización hacia amino y desestabilización hacia carboxilo. Por lo tanto, según los datos de Fersht, los residuos negativos muestran un componente más claramente definido para estabilizar específicamente hacia amino, a diferencia de una especificidad menos clara por parte de los residuos positivos. Consecuentemente, si observamos la distribución de cargas que flanquean la alfa hélice en Richardson y Richardson (1988) se aprecia una alta preferencia de cargas negativas hacia amino (casi tres desviaciones estándar entre Glu y Asp) y una más leve de cargas positivas hacia carboxilo (casi dos desviaciones estándar entre Lys, His y Arg).

Por otro lado, los residuos Lys, Arg e His tienen un componente hidrofóbico mayor. Esta condición tal vez acerca su comportamiento a los residuos típicamente hidrofóbicos, lo cual justificaría su mayor capacidad de mantener la simetría de estructura secundaria al ser idénticos en la secuencia invertida. De todos modos, ya sea por la hidrofobicidad, o por la carga, o tal vez la estructura del residuo cargado, los residuos de carga negativa parecen tener una mayor tendencia a romper la simetría de estructura secundaria con la identidad, por lo que su ubicación hacia el extremo amino cobraría más importancia.

### **6.3 Preferencias de sustitución y simetría de estructura secundaria**

Dentro de los modelos predictivos de la estructura secundaria de las proteínas, existe una corriente iniciada por Nishikawa y Ooi (1982) que atribuye principalmente a las coordenadas de composición de residuos de una proteína las proporciones de estructura secundaria de dicha proteína. Esto implica que dos proteínas de la misma



composición aminoacídica mostrarán la misma composición de estructura secundaria aunque sus secuencias no sean las mismas. Desde este punto de vista, es factible que en las secuencias palindrómicas del intervalo 40-100%, sea la convergencia en una composición aminoacídica similar entre las mitades del palíndromo lo que induce la simetría significativa en la estructura secundaria y no el hecho de conservar un orden invertido de sus secuencias. Sin embargo, entre los palíndromos del intervalo 40-100% de identidad antiparalela de secuencia, aquellos residuos idénticos tienden a conservar significativamente más la estructura secundaria son más simétricos que los que son distintos, lo que indica que el orden invertido sí importa para permitir la simetría de estructura secundaria. Si no fuese así, los eventos de simetría estructural antiparalela debieran repartirse de manera indiscriminada entre los eventos de residuos idénticos y distintos.

Al revisar las preferencias de composición de los residuos dentro de los palíndromos sobre el 40% de identidad antiparalela de secuencia (significativamente más simétricos que los segmentos control) es posible observar que la composición de los palíndromos no es al azar, ya que existen preferencias composicionales alejadas por sobre o bajo el 90% central de la población de composiciones esperadas por azar (TABLA 6). Se descarta, por lo tanto, que en su formación exista una elección de residuos guiada por el azar, ya que aún considerando la composición y los diversos patrones de distribución aminoacídica existentes en las proteínas en estudio, es posible observar alejamientos significativos en la composición de los palíndromos respecto de los controles.



Los residuos que menos tienden a ser idénticos corresponden a los residuos que, al ser cambiados por otro al lado opuesto, tienen una alta probabilidad de mantener la misma estructura secundaria. En otras palabras, los residuos que al ser distintos tienen una alta probabilidad de mantener la misma estructura secundaria, son idénticos con menor frecuencia. Lo anterior pareciera indicar que lo que se intenta alcanzar es la simetría de la estructura, no importando en sí los eventos entre idénticos o distintos, sino los eventos con más simetría de estructura secundaria.

Los residuos que más forman pares distintos (+1/PREF.ENTRE DISTINTOS) entre sí son los que presentan mayor simetría con la formación de pares idénticos. Esto podría indicar que cuando un residuo con alta capacidad de mantener la simetría con la identidad enfrenta un residuo distinto en el palíndromo, se tiende a encontrar en ese sitio por lo menos otro residuo con una alta capacidad de simetría con la identidad. Este patrón podría indicarnos una cierta capacidad de intercambios simétricos sólo entre residuos de alta capacidad de identidad simétrica, los cuales, en general, corresponden a los residuos más apolares. Esta capacidad de intercambios simétricos entre residuos no polares podría explicar por qué Park, Dietmann y col. (2000) y Lorenzen, Gille y col. (2003) encontraron un muy bajo contenido, 8%, de secuencias invertidas asociadas a hebras beta (típicamente compuesta por residuos apolares): no se debería a un bajo contenido de hebras beta en los archivos del PDB (22%), ni tampoco a un mayor impacto del orden o patrón de la secuencia para la formación regular de puentes de hidrógeno entre las hebras beta, como ellos postulan; sólo que sería menos necesario en el caso de una hebra beta conservar una secuencia invertida para mantener la simetría de estructura secundaria, por la mayor capacidad de intercambios simétricos entre residuos

apolares distintos. Este resultado es consecuente con los de Gibrat, Garnier y col. (1987), los cuales muestran que los veinte residuos aminoacídicos (salvo lisina) presentan menos diferencia entre sus capacidades de afectar (positiva o negativamente) la estructura secundaria hacia amino o hacia carboxilo cuando están insertos en las conformaciones extendidas que cuando lo están en las conformaciones de hélice y vueltas. Según Fersht (1999) el núcleo hidrofóbico de la proteína es altamente tolerante a la substitución de las cadenas laterales hidrofóbicas por otras de estereoquímica distinta sin que la actividad de las proteínas se vea alterada.

Los residuos que más forman pares idénticos son los que presentan pares distintos con menor simetría de estructura secundaria (-1/PREF.ENTRE IDÉNTICOS). Adicionalmente, los residuos que usualmente se encuentran al centro de un palíndromo son los que muestran mayor simetría al estar presentes en pares distintos (+1/PREF.CENTRALES).

#### **6.4 Interpretación del comportamiento observado: residuo-soluto (no interactivo), residuo interactivo de campo homogéneo, residuo interactivo de campo heterogéneo**

Defínase una proteína como un patrón de residuos apolares principalmente embebidos y residuos polares principalmente en contacto con el medio externo, donde la polaridad es la única propiedad importante para determinar la estructura secundaria. Si el residuo reemplazante en la secuencia invertida tiene la misma polaridad que el residuo original, tanto la presencia de pares idénticos como la de pares distintos debiera asociarse a una alta probabilidad de mantener la misma estructura. En los resultados de

este trabajo se observa que la simetría de estructura secundaria se conserva con la presencia de pares idénticos entre residuos apolares y que se permite la presencia de pares distintos (+1/Preferencias entre residuos distintos, TABLA 8) siempre que la condición de baja polaridad se mantenga, de manera similar al modelo simplificado propuesto. Es decir, para los residuos apolares la inversión de la secuencia y, probablemente, la identidad de los residuos no afectan significativamente la simetría de estructura secundaria. Los residuos más polares, en cambio, no muestran con la presencia de pares idénticos la misma simetría de estructura secundaria en los resultados (FIGURA 11). ¿Por qué un residuo polar tiene menor capacidad de mantener la simetría que un residuo apolar con la formación de pares idénticos? A continuación se propone tres modalidades a través de las cuales un residuo determinado puede interactuar con los residuos cercanos. Aunque la intención inicial es explicar el comportamiento de los residuos polares en la secuencia invertida, se verá que es posible aplicar alguna de estas modalidades a todo tipo de residuos.

#### **Residuos no interactivos o residuos soluto:**

En un primer modelo, se puede suponer que los residuos polares son solutos con alta afinidad por el agua sujetos al resto de la proteína sólo a través de enlaces peptídicos e insertos en algún tipo de estructura secundaria. En este caso, la estructura secundaria estaría sustentada localmente sólo por la propiedad “polar” del residuo. La simple separación de fases entre el medio acuoso y el interior hidrofóbico bastaría para determinar la estructura proteica, y por lo tanto, sería esperable que los residuos polares de diversa topología mantengan la misma estructura secundaria en la secuencia invertida, independiente de que sea idéntico o distinto respecto al residuo polar original,



ya que su interacción con el medio carecería de restricciones estereoespecíficas importantes, de manera similar a como se comportan los residuos hidrofóbicos.

### **Residuos interactivos:**

En otro esquema, se puede enfocar a los residuos polares como parte de una trama de interacciones no covalentes adicionales a los enlaces covalentes peptídicos. De manera intrínseca a la existencia de esa trama está el concepto de la topología de los diversos campos de interacción que se verifican en la superficie del residuo. La superficie de interacción del residuo polar podría poseer,

- a) Un campo homogéneo
- b) Un campo heterogéneo

En el caso de ser un campo homogéneo cada uno de los puntos de la superficie del residuo tendría la misma capacidad de interactuar, por lo que, carecería de imponer estereo especificidad a otros residuos del medio. Así, la presencia de pares idénticos no debiera afectar mayormente la estabilización local con la que contribuye este residuo a la estructura secundaria. La presencia de pares de residuos distintos, por su parte, podría modificar a veces la estabilización de la estructura local, dadas eventuales diferencias entre las magnitudes de los campos homogéneos intercambiados (de sus densidades electrónicas).

Si, en cambio, la superficie del residuo es un campo de interacción heterogéneo, su configuración, o el tipo de estructura que estabiliza durante la interacción está condicionada a su orientación en el espacio. Esta estereoespecificidad implica que su capacidad de interacción puede verse fuertemente modificada al cambiar su orientación con la inversión de la secuencia. En otras palabras, un residuo con un campo



heterogéneo comprometido en una trama de interacciones no covalentes en las cuales se sustenta la estructura secundaria podría, al ubicarse en una secuencia invertida, perder algunas de estas interacciones por razones de estereoespecificidad, afectar la estabilidad local de la estructura secundaria y, con ello, reducir las probabilidades de mantener la simetría de estructura secundaria. La formación de pares distintos por dichos residuos mantendrá la simetría sólo si el residuo substituyente es capaz de mantener las más importantes de estas interacciones en la secuencia invertida.

En este punto es necesario destacar un aspecto fundamental para justificar que una secuencia aminoacídica compuesta por un conjunto de residuos con campo heterogéneo rompa su simetría de estructura secundaria al invertirse. La totalidad de los carbonos beta (el primer carbono del residuo no perteneciente al esqueleto peptídico) se encuentran unidos con el esqueleto peptídico a través de enlaces sigma en los veinte residuos distintos. Por otro lado, la gran mayoría de los átomos que constituyen al resto del residuo se encuentran también unidos a través de enlaces sigma. En teoría, este tipo de enlaces debiera permitir que cada segmento del residuo rote libremente. Si se configuran las rotaciones adecuadas en un residuo, y concertadamente con los residuos con los cuales interactúa, la mayoría de las interacciones que sustentan la estructura secundaria local en la secuencia original debieran poder restablecerse en la secuencia invertida. Sin embargo, cada rotación tiene, debido a problemas de impedimento estérico, un alto costo energético. Por otro lado, la necesidad de que los residuos interactivos vecinos configuren sus rotaciones de una manera concertada debiera reducir aún más la probabilidad de que los residuos con superficie de campo heterogéneo mantengan la misma estructura secundaria en la secuencia invertida. El modelo de los

TABLA 10

**EFFECTOS DE LA IDENTIDAD SOBRE LA SIMETRÍA ESTRUCTURA SECUNDARIA PARA RESIDUOS NO INTERACTIVOS O INTERACTIVOS DE CAMPO HOMOGÉNEO Y HETEROGÉNEO.**

Perfil de interacción de los residuos	pares idénticos	pares distintos
Residuo-soluto (no interactivo)	siempre simétrica	siempre simétrica
Campo homogéneo (interactivo)	siempre simétrica	a veces asimétricas
Campo heterogéneo (interactivo)	a veces asimétricas	a veces asimétricas

residuos interactivos de campo heterogéneo se ajusta más al comportamiento de varios de los residuos polares observado en los resultados de esta tesis (baja simetría con la presencia de pares idénticos, mayor simetría con pares distintos). El que la presencia de pares distintos sea simétrica para los residuos polares podría entenderse como una forma de aumentar las probabilidades de mantener la simetría de estructura secundaria al rastrear la substitución con otros residuos en la secuencia invertida. Usando este modelo, se podría inferir que los residuos polares reales debieran tener en general una participación importante en la estabilización local de la estructura secundaria, ya que, con la inversión de la secuencia, esta estabilización se vería lo suficientemente alterada como para bajar la probabilidad de simetría de estructura secundaria. Este mismo impedimento estérico en los residuos apolares, tendría, por lo tanto, una menor incidencia sobre la simetría de estructura secundaria de la secuencia invertida, dado que éstos poseerían en su superficie una mayor homogeneidad de campo, es decir, por todos los sectores de su superficie debieran ser capaces de establecer el mismo tipo de interacción con los residuos vecinos, sin requerir para esto grandes modificaciones rotacionales concertadas.

### **6.5 Residuos simétricos, residuos no simétricos**

Como una forma de bosquejar cómo se comportan cada uno de los veinte residuos con el tipo de mantenimiento de identidad, se contabilizó qué residuos poseen más frecuentemente los índices permisivos de la simetría para un tipo de mantenimiento de identidad (ver resultados):

**PARES IDÉNTICOS SIMÉTRICOS EN 12 ORDENAMIENTOS NO REDUNDANTES**

**12P > 11G > 10L > 9V, A > 8K, F > 7Y, R, C > 6W, I > 5M > 4T, E > 3Q, H > 1N > OS, D**

**PARES DISTINTOS SIMÉTRICOS EN 25 ORDENAMIENTOS NO REDUNDANTES**

**22K, D > 21E > 19N, C, A > 18H > 16Q, G > 15S, R > 11W > 8M > 7F > 6Y > 5L > 4V, T > 3P > 1I**

Se puede apreciar que los residuos P, L, V, F, Y, R (en rojo) simultáneamente se encuentran entre los diez que forman pares idénticos más simétricos y los diez que al enfrentar un residuo distinto forman pares más asimétricos, lo que los perfila como residuos interactivos de campo homogéneo.

Por otro lado, los residuos E, Q, H, N, S, D (en verde) debieran exhibir principalmente asimetría de estructura secundaria cuando se enfrentan a uno igual en la secuencia invertida. En cambio, muestran simetría si el residuo opuesto es distinto, lo que los ajusta, según el modelo, a residuos interactivos de campo heterogéneo. Es decir, probablemente estabilizan la estructura secundaria a través de configuraciones más estereoespecíficas, las cuales se debilitarían de una manera importante al formar pares idénticos en un palíndromo. Según Eswar y Ramakrishnan (1999) estos residuos más treonina estarían implicados en la estabilización no mediada por esqueleto peptídico de algunas estructuras secundarias observadas típicamente en proteínas. Tal estabilización estaría apoyada en la capacidad de formar puentes de hidrógeno por parte de las cadenas laterales de los mismos.

El resto de los residuos G, A, K, W, M, T, I, C (en anaranjado) exhibirían características de transición en el mantenimiento de la simetría de estructura secundaria con la presencia de pares de residuos idénticos o distintos. Según la TABLA 11, este



TABLA 11

MODELO DE INTERACCIÓN DE LOS DIVERSOS RESIDUOS  
DEPENDIENDO DE LA SIMETRÍA DE ESTRUCTURA SECUNDARIA  
OBSERVADA ANTE UN RESIDUO IDÉNTICO O DISTINTO

RESIDUO	IDÉNTICOS		DISTINTOS		MODELO DE INTERACCIÓN
	SIMÉTRICA	ASIMÉTRICA	SIMÉTRICA	ASIMÉTRICA	
A	X		X		SOLUTO
C	X		X		SOLUTO
D		X	X		HETEROGÉNEO
E		X	X		HETEROGÉNEO
F	X			X	HOMOGÉNEO
G	X		X		SOLUTO
H		X	X		HETEROGÉNEO
I		X		X	HETEROGÉNEO
K	X		X		SOLUTO
L	X			X	HOMOGÉNEO
M		X		X	HETEROGÉNEO
N		X	X		HETEROGÉNEO
P	X			X	HOMOGÉNEO
Q	X		X		SOLUTO
R	X			X	HOMOGÉNEO
S		X	X		HETEROGÉNEO
T		X		X	HETEROGÉNEO
V	X			X	HOMOGÉNEO
W		X		X	HETEROGÉNEO
Y	X			X	HOMOGÉNEO

grupo incluiría una combinación de residuos de comportamiento residuo-soluto y residuos con campo heterogéneo de interacción. Estos residuos presentarían una mayor capacidad de formar tanto pares de residuos idénticos como distintos manteniendo la simetría. Dentro de este grupo se comportan como residuos tipo soluto A, C, G y K, lo cual debiera indicarnos las interacciones no covalentes que estos residuos debieran aportar en la estabilización local de la estructura secundaria, debieran ser suficientemente débiles o inespecíficas para que su mantenimiento en la secuencia invertida no afecte la simetría de estructura secundaria.

#### **6.6. Conservación de residuos v/s identidad antiparalela**

Esta clasificación plantea un distinto valor de cada uno de los 20 residuos en el mantenimiento de la estructura secundaria secuencia invertida. Existiría, por lo tanto, un conjunto limitado de residuos, principalmente hidrofóbicos, aparentemente con un campo de interacción homogéneo, capaces de mantener la misma estructura secundaria. Esta simetría de estructura se vería modificada al formar pares de residuos distintos con otro residuo cualquiera, en general. Teóricamente, estos podrían ser residuos con una mayor tendencia a ser conservados para alcanzar la simetría de estructura secundaria. Sin embargo, la mayoría de los residuos no tiene este perfil de mantenimiento de identidad y simetría de estructura secundaria y, por lo tanto, es esperable un bajo grado de conservación de los arreglos palindrómicos de las secuencias, tal como se observó entre los distintos miembros de algunas familias revisadas en resultados preliminares (un ejemplo preliminar a este trabajo en FIGURA 12). Si a esto sumamos un cierto grado de intercambiabilidad entre los residuos hidrofóbicos distintos, es entendible, por lo tanto,

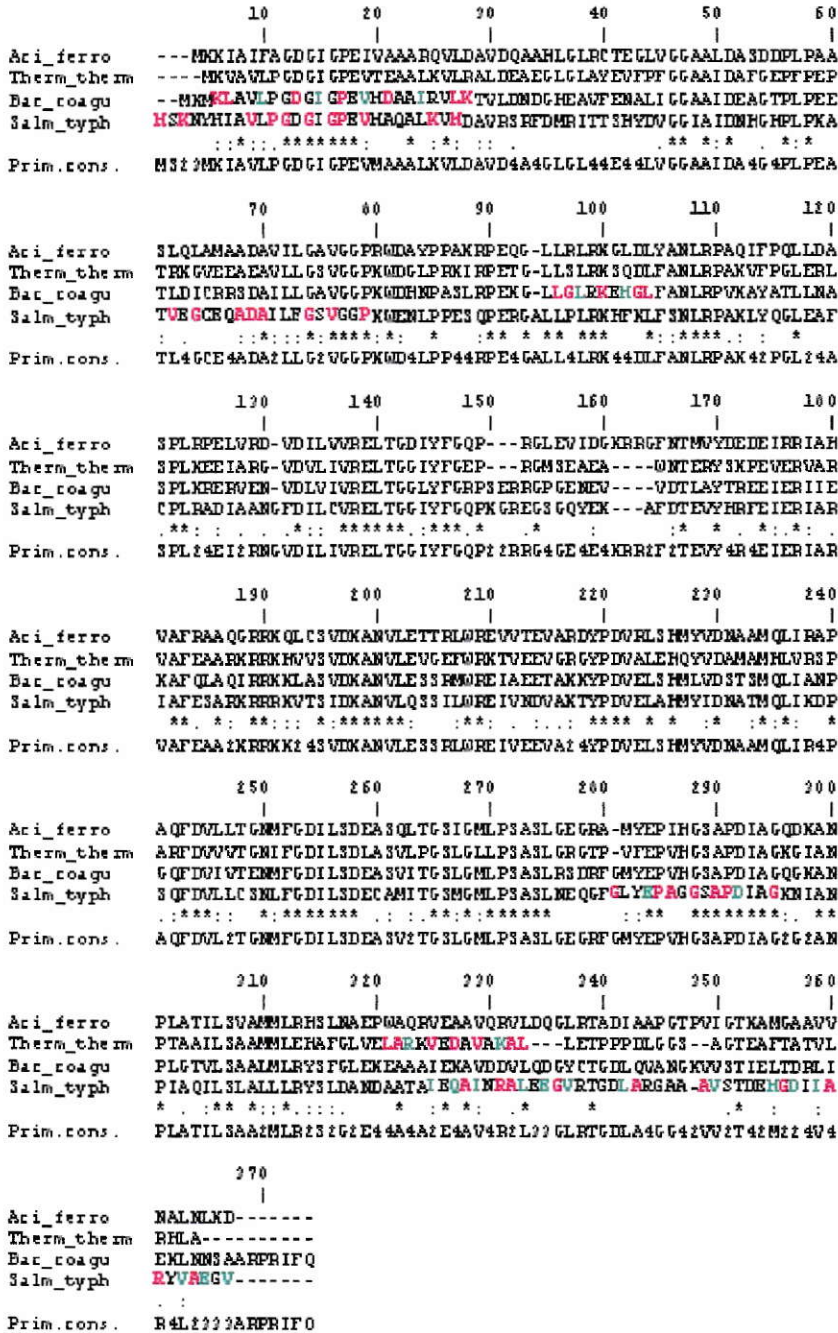


FIGURA 12. Secuencias palindrómicas no superpuestas y conservación en un alineación de un conjunto de proteínas isopropil malato deshidrogenasas. Se muestra cuatro miembros de la familia de las Isopropilmalato deshidrogenasas. En cada uno de ellos se destaca los palíndromos no superpuestos encontrados. Al analizar transversalmente la presencia de estos palíndromos, no se aprecia conservación de los residuos palindrómicos ni tampoco colocalización de los segmentos con secuencia palindrómica provenientes de los distintos miembros de estos miembros de esta familia. Es usual encontrar este comportamiento en otras alineaciones.



que el patrón palindrómico no sería en sí una organización necesaria de conservar en la secuencia para mantener la simetría de estructura secundaria. Entonces, la medición de la conservación de los residuos no sería una herramienta válida para medir la importancia biológica del fenómeno de la palindromía (FIGURA 12).

### **6.7 Estructura-función**

Hasta aquí se ha descrito los índices característicos de estos residuos y se ha discutido las razones físicas que determinarían la mayor simetría para un tipo de mantenimiento de identidad. Si a través de la conservación no es posible medir la importancia biológica de la palindromía, entonces ¿qué significado estructural tiene este tipo de residuos para la proteína? ¿qué significado funcional tienen los arreglos palindrómicos?

Estructuralmente hablando, los residuos de mayor capacidad de simetría con la la formación de pares idénticos son descritos por Qian y Sejnowski (1988), como potentes iniciadores, terminadores y quebradores de estructura secundaria. Además, corresponden según Haney, Badger y col. (1999) a los residuos más recurridos en la termoestabilización de proteínas de ciertos organismos termófilos. En este último caso, la mayor presencia de residuos grandes e hidrofóbicos reduciría la accesibilidad por parte del solvente al interior de la proteína, además de aumentar el núcleo hidrofóbico (FIGURA 13).

Funcionalmente hablando, no existe hasta la fecha un estudio sistemático que caracterice los arreglos palindrómicos en las proteínas, pero sí algunos ejemplos.



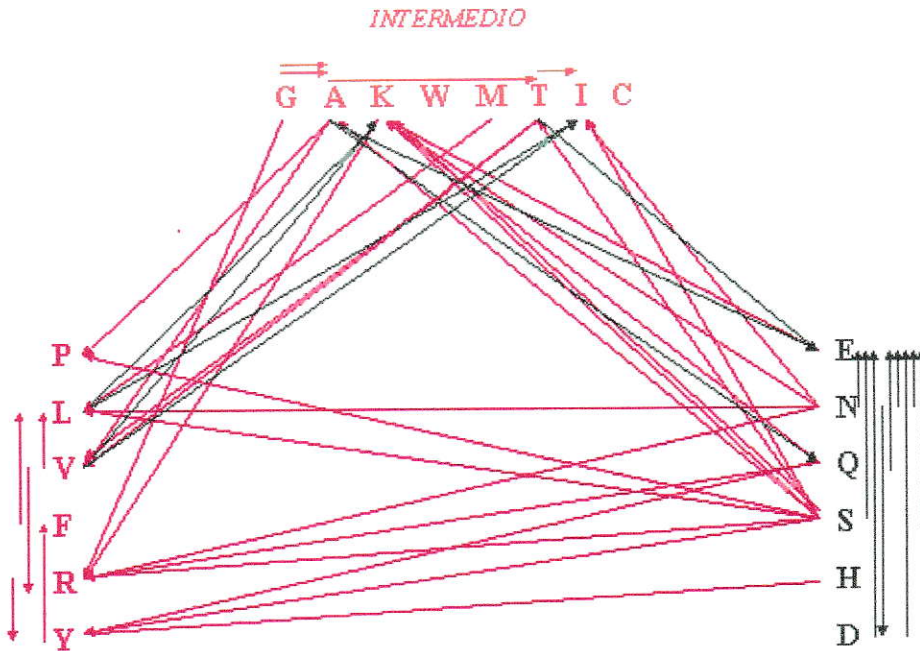


FIGURA 13. RESIDUOS IDÉNTICOS SIMÉTRICOS ASIMÉTRICOS E INTERMEDIOS Y SU RELACIÓN CON LA ESTABILIZACIÓN DE LAS PROTEÍNAS DE UN ORGANISMO TERMÓFILO. Presencia de pares de residuos idénticos simétricos (grupo de residuos en rojo) asimétricos (en verde) e intermedias (anaranjado) y su relación con su uso en la estabilización de las proteínas de un organismo termófilo. En el diagrama, las transiciones estabilizadoras analizadas por Haney, Badger y col. (1999) a través de tres métodos distintos se representa con flechas rojas largas cuando el sentido de la transición va desde PARES DE RESIDUOS IDÉNTICOS ASIMÉTRICOS → INTERMEDIOS → SIMÉTRICOS y flechas verdes largas si va en sentido contrario. Las transiciones dentro de cada grupo se representan por flechas cortas del mismo color del grupo. Se puede apreciar que la mayoría de las transiciones van desde los residuos menos simétricos a los más simétricos para los pares idénticos. Pese a que hay algunas transiciones entre el grupo intermedio y los extremos, no hay ninguna desde el grupo de residuos formadores de pares idénticos simétricos al de residuos formadores de pares distintos asimétricos. Esta relación sugiere que la mayor capacidad por parte del residuo para estabilizar la proteína a altas temperaturas se acompaña de una mayor capacidad de mantener la simetría de estructura secundaria aunque la secuencia esté invertida. En un estudio de predicción de estructura secundaria, Qian y Sejnowski (1988) determinaron que el residuo P es un fuerte rompedor de hélice alfa, mientras A, L y M son fuertes formadores de hélice, especialmente cuando están en su extremo C-terminal. Dos residuos básicos, K y R, son formadores de hélice cuando están en el extremo C-terminal, mientras que E, un residuo ácido que se propone como asimétrico en nuestros datos, apoya la estructura tipo hélice cuando está en su extremo N-terminal. Los residuos I, Y, V, y W, son fuertes formadores de hebras beta y no muestran preferencias hacia los extremos C-terminal y N-terminal.

Se ha encontrado segmentos palindrómicos de diversa longitud y simetría en las secuencias de algunas proteínas asociadas a dominios de interacción entre proteínas (Díaz, Jedlicki y col., 2000); la plastocianina de espinaca contiene un palíndromo de 67 residuos (P16 a F82) que incluye dos zonas aniónicas equidistantes del aspártico central del palíndromo (D42 a E45 y E59 a D61). Esta región está involucrada en el acoplamiento al palíndromo K16 a K23 de otro componente transportador de electrones, la subunidad psaF del fotosistema I de *Chlamydomonas reinhardtii*, por intermedio de sus respectivos palíndromos. Un palíndromo adicional entre L5 y L12 le permite a la plastocianina una mayor interacción con el citocromo F.

Tiempo antes, Suzuki (1992) también encontró que la histona H1 de esperma de orca posee un alfa-hélice palindrómico perfecto (total identidad) de 37 residuos, que tendría la función de unir el DNA nucleosomal. Esta secuencia actualmente no se encuentra en la base de datos.

Dados estos antecedentes, se está efectuando una inspección bibliográfica para determinar si los segmentos palindrómicos del intervalo 40-100% de identidad antiparalela de secuencia se encuentran asociados a sitios de unión de ligandos.

Preliminarmente, se puede decir que hay al menos dos segmentos palindrómicos del intervalo 40-100% siendo o formando parte de sitios de unión de la toxina a un ligando.

Pero no sólo se cuentan los palíndromos entre los ejemplos funcionales. Se encontró que las secuencias invertidas de lisozima humana y de citocromo c de corazón de caballo presentan secuencias similares a las secuencias directas de la betaglucanasa y la reductuctasa de tiorredoxina, respectivamente (Díaz, Díaz y col., 1998). En este

trabajo, el segmento invertido asociado al grupo heme (centro redox) del citromo c de corazón de caballo presentó un 46% de identidad con el sitio activo redox de la reductasa de tiorredoxina de *Bacillus subtilis* (contra 25% de identidad al compararse normalmente).

La secuencia invertida del segmento correspondiente al sitio catalítico de la lisozima humana es un 43,3% idéntica al sitio catalítico de la enzima betaglucanasa de *Bacillus subtilis* (contra 9,4% de identidad al compararse normalmente). La lisozima y la betaglucanasa comparten el mecanismo de catálisis, substratos e inhibidores. Lo anterior sugeriría un aspecto nuevo dentro del concepto de convergencia evolutiva.

### **6.8 Palindromía y código genético**

Una vez demostrados los rasgos estructurales, cabe preguntarse cómo surgen este tipo de secuencias en la evolución (ver métodos). Los resultados de este trabajo plantean dos explicaciones:

- a) Una vía de mutaciones independientes: por mutaciones independientes se genera una organización palindrómica a nivel de la secuencia aminoacídica.
- b) Una vía autogénica: la secuencia genómica es capaz de generar por complementariedad de bases un símil invertido de la misma secuencia aminoacídica (un mecanismo que podría ser más rápido evolutivamente hablando).

En el caso de la vía autogénica es posible que el genoma que codifica para una mitad del palíndromo participe, por complementariedad de bases, en la codificación de



la otra mitad (mutación cruzada), ya que al hibridar el segmento genómico codificador del palíndromo consigo mismo, los tripletes que determinan los residuos idénticos a los lados opuestos del palíndromo complementan sus bases, aunque no necesariamente en marcos totalmente superpuestos, como la cadena a de la toxina de la difteria.

```

  A   M   Y   E   Y   M   A
gcg atg tat gag tat atg gct

```

Structure 1

Folding bases 1 to 21 of gi|1633500|pdb|1tox|a

dG = -0.4 dH = -38.7 dS = -123.5 T<sub>m</sub> = 40.2

```

  A   M   Y
                10
-| G       T
  GC ATGTA G      E
  CG TATAT A
T^ G       G
20
  A   M   Y

```

Siendo este caso excepcional entre los datos analizados, puede suponerse que la gran mayoría de arreglos palindrómicos de los residuos son obtenidos por la vía de las mutaciones independientes durante la evolución de las proteínas.

Cualquiera de los dos mecanismos, sin embargo, abre un campo importante en el área de la evolución molecular.



## 7. CONCLUSIONES

- 1) Sobre el 40% de identidad antiparalela de secuencia existe una relación significativa entre el porcentaje de identidad antiparalelo de las mitades del palíndromo y su simetría de estructura secundaria, es significativamente superior a lo observado en segmentos control.
- 2) Este porcentaje es similar al observado en secuencias comparadas paralelamente, lo que puede sugerir balances de fuerzas de plegamiento y umbrales de nucleación similares, independientemente del sentido de la secuencia.
- 3) Los pares de residuos idénticos se asocian significativamente con los eventos de simetría de estructura secundaria en los palíndromos, por sobre los pares de residuos distintos. Esto implica que el orden invertido de la secuencia, más allá de la composición similar, sería la causa de la simetría de estructura secundaria.
- 4) Tanto entre los pares de residuos idénticos como distintos existen preferencias que escapan significativamente de lo esperado por azar, lo que sugiere que el proceso no es estocástico y sí guiado biológicamente.
- 5) Tanto entre los pares de residuos idénticos como distintos, existen magnitudes de propiedades fisicoquímicas que se asocian con la simetría de estructura secundaria. Los residuos más hidrofóbicos poseen una alta e intermedia capacidad para mantener la simetría de estructura secundaria en la secuencia invertida. La mayoría de estos residuos registran en la literatura un papel

estabilizador de proteínas en organismos termófilos, como también de fuertes ruptores, e iniciadores de estructura secundaria. Los residuos más polares, por otro lado, presentarían en general una baja capacidad para mantener la simetría en la secuencia invertida. A su vez, las preferencias por formar pares de residuos idénticos como distintos se asocian con el mantenimiento de las propiedades fisicoquímicas importantes para la simetría.

- 6) La tendencia a formar pares de residuos distintos de los residuos polares y el apreciable número de intercambios entre residuos hidrofóbicos indicarían que la conservación aminoacídica no es un imperativo de este tipo de secuencias en familias de proteínas.
- 7) La obtención evolutiva de ambas mitades de los palíndromos se realizaría a veces de manera independiente y otras a través de complementariedad entre los respectivos genomas codificantes.
- 8) La comprensión cabal de este tipo de arreglos aminoacídicos puede mejorar la capacidad de predicción de los actuales programas de modelación estructural, como también abrir un campo interesante en el área de la evolución molecular.

## 8. REFERENCIAS

Dill, K. 1990. Dominant forces in protein folding. *Biochemistry* 29: 7133-7154.

Díaz, C., Díaz, J., Cortés, C. 1988. Similitud primaria y funciones relacionadas de proteínas comparadas consecuencias invertidas de la lisozima humana y citocromo c de corazón de caballo. *Noticiero de Biología de la Sociedad de Biología de Chile* 6: R-169.

Díaz, C., Jedlicki, E., Holmes, D. 2000. Secuencias palindrómicas en las proteínas. *Biological Research* 33, R-34.

Eswar, Narayanan, Ramakrishnan, C. 1999. Secondary structures without backbone: an analysis of backbone mimicry by polar side chains in protein structures. *Protein Engineering* 12: 447-455.

Fersht, A. 1999. *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding*. W.H. Freeman and Company, Sec. Ed. 8-22.

Gibrat, J., Garnier, J., Robson, B. 1987. Further developments of protein secondary structure prediction using information theory: New parameters and consideration of residue pairs. *J.Mol.Biol.* 198: 425-443.

Haney, P., Badger, J., Buldak, G., Reich, C., Woese, C., Olsen, G. 1999. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *PNAS USA* 96: 3578-3583.

Kabsch y Sander. 1984. On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations. *PNAS USA* 81: 1075.

Lacroix, E., Viguera, A., Serrano, L. 1998. Reading protein sequences backwards. *Folding and Design* 3: 79-85.

Lorenzen, S., Gille, C., Preissner, R., Frömmel, C. 2003. Inverse sequence similarity of proteins does not imply structural similarity. *F.E.B.S. Lett.* 545: 105-109.

- Mittl, P., Deillon, C., Sargent, D., Liu, N., Klauser, S., Thomas, R., Gutte, B., Grütter, M. 2000. The retro-GNC4 leucine zipper sequence forms a stable three-dimensional structure. *PNAS USA* 97: 2562-2566.
- Nishikawa, K., Ooi, T. 1982. Correlation of the amino acid composition of a protein to its structural and biological characters. *J. Biochem (Tokyo)* 91:1821-1824.
- Olszewski, K., Kolinski, A., Skolnick, J. 1996. Does a backwardly read protein sequence have a unique native state?. *Protein Engineering* 9: 5-14.
- Park, J., Dietmann, S., Heger, A., Holm, L. 2000. Estimating the significance of sequence order in protein secondary structure and prediction. *Bioinformatics* 16: 978-987.
- Phan-Chan-Du, A., Petit, M., Guichard, G., Briand, J., Muller, S., Thong Cung, M. 2001. Structure of antibody-bound peptides and retro-inverso analogues. A Transferred nuclear overhauser effect spectroscopy and molecular dynamics approach. *Biochemistry* 40: 5720-5727.
- Preissner, R., Goede, A., Michalski, E., Frömmel, C. 1997. Inverse sequence similarity in proteins and its relation to the three-dimensional fold. *FEBS* 414: 425-429.
- Qian, N., Sejnowski, T. 1988. Predicting the secondary structure of globular proteins using neural network models. *J.Mol.Biol.* 202: 865-884.
- Richardson, J., Richardson, D. 1988. Amino acid preferences for specific locations at the ends of alpha helices. *Science* 240: 1648-1652.
- Robson, B., Suzuki, E. 1976. Conformational properties of amino acid residues in globular proteins. *J.Mol.Biol.* 107: 327-356.
- Shuichi, K., Ogata, H. Kanehisa, M. 1999. AAindex: amino acid index database. *Nucleic Acids Res.* 27: 368-369.
- Suzuki, M. 1992. DNA-bridging by a palindromic  $\alpha$ -helix. *PNAS USA* 89: 8726-8730.
- Tanaka, S., Scheraga, H. 1977. Statistical mechanical treatment of protein conformation. 5. Amultistate model for specific sequence copolymers of amino acids. *Macromolecules* 10: 9-20.
- Tatusova, T., Madden, T. 1999. Blast 2 sequences - a new tool for comparing protein and nucleotide sequences, *FEMS Microbiol Lett.* 174:247-250.
- Zucker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31: 3406-3415.



## 9. GLOSARIO

Alineación antiparalela:	Alineación de dos secuencias aminoacídicas con sentidos amino carboxílicos opuestos.
Asimetría de estructura secundaria:	Desigualdad de estructura secundaria entre sitios equivalentes de secuencias alineadas antiparalelamente o paralelamente según sea el caso.
Campo heterogéneo:	Superficie de un residuo aminoacídico que muestra topología e intensidades de campos de fuerza variables en su superficie, los cuales le confieren estereoespecificidad en su capacidad de interactuar con residuos vecinos. Típicamente asimétrico en la secuencia invertida.
Campo homogéneo:	Superficie de un residuo aminoacídico que muestra topología e intensidades de campos de fuerza homogéneos en su superficie, los cuales le restan estereoespecificidad en su capacidad de interactuar con residuos vecinos. Típicamente simétrico en la secuencia invertida.
Estocásticamente:	Aleatoriamente, a azar.
Forma nativa:	Estructura en todos sus niveles idéntica a la estructura de la proteína nativa. Especialmente estructura secundaria en este trabajo.
Heteroátomos:	Átomos distintos al carbono e hidrógeno (S, N, O). Insertos en el residuo, su electronegatividad y presencia de pares de electrones no enlazados les permite establecer puentes de hidrógeno en residuos típicamente de campo heterogéneo.
Identidad antiparalela (de secuencia):	Residuos idénticos o porcentaje de residuos idénticos en una Alineación antiparalela.
Matriz blosum 62:	Matriz de sustitución obtenida a través de la cuantificación de intercambios funcionales de residuos entre miembros de diversas familias de proteínas.
Palíndromo:	Segmento de la secuencia de una proteína en el cual una mitad de la secuencia presenta identidad antiparalela con la otra mitad de la misma.

Parámetro z normalizado:	Variable normal estándar, obtenida después de restarle a una variable el valor de la media poblacional de las variables y dividir este resultado por la desviación estándar.
Pares distintos:	Residuos que se enfrentan a residuos distintos en el sitio equivalente opuesto del palíndromo, definido según la alineación antiparalela.
Pares idénticos:	Residuos que se enfrentan a residuos idénticos en el sitio equivalente opuesto del palíndromo, definido según la alineación antiparalela.
Restricciones estéricas:	Impedimentos que la topología del residuo le impone a la capacidad de interactuar del mismo.
Retro-péptidos o retro-proteína:	Péptido o proteína a la cual se le ha invertido el orden amino-carboxílico que muestran sus residuos.
Segmentos palindrómicos:	Secuencia palindrómica que exhiben algunos segmentos de una proteína.
Simetría de estructura secundaria:	Igualdad de estructura secundaria entre sitios equivalentes de secuencias alineadas antiparalelamente o paralelamente según sea el caso.
A, C, D, E, F, G, H, I, K, L, M, N, P, Q R, S, T, V, W, Y:	Códigos de una letra para representar los residuos alanina, cisteína, aspartato, glutamato, fenilalanina, glicina, histidina, isoleucina, lisina, leucina, metionina, asparragina, prolina, glutamina, arginina, serina, treonina, valina, triptófano y tirosina, respectivamente.

## 10. APÉNDICE

A continuación se adjunta las treinta y cuatro secuencias analizadas en este trabajo. Se destaca los cuarenta y ocho palíndromos contenidos en ellas, todos en la notación que se explica más adelante.

Notación del encabezado:

>gi|10835642|pdb|1dlh|a chain a, solution structure of hanatoxin 1 (*Eukaryota, Grammostola spatulata* )

P(9-15)id.sec.=0/3(0%) id.prim'=1/3(33%)

Palíndromo encontrado entre los residuos 9 y 15 (incluidos) de la cadena “a” de la hanatoxina 1, cuyo archivo de coordenadas estructurales se identifica con el código PDB 1dlh. Sus mitades, sin incluir el centro, mostraron entre sí cero residuos, de tres comparables entre las mitades, con identidad antiparalela de estructura secundaria (id.sec. = 0%). A nivel de sus secuencias, mostraron un residuo idéntico de los tres (id.prim' = 33%) (ejemplo en FIGURA 5).

Notación de la secuencia:

En cada una de las secuencias se destaca en letra de tamaño mayor, el o los segmentos palindrómicos encontrados en ella, salvo el residuo central del palíndromo que está en tamaño menor. Dentro de cada palíndromo se subraya los residuos idénticos. En color rojo se indica cuando un residuo del palíndromo forma parte de la estructura secundaria hélice, en anaranjado los que forman parte de hebras beta, en negro no-

hélice/no-hebra beta, y en azul cuando la estructura secundaria no esta determinada para esos sitios de la secuencia. En todas las secuencias se omiten los espacios que inserta el programa de alineación usado para detectar los palíndromos.

>gi|10835642|pdb|1dlh|a chain a, solution structure of hanatoxin 1 (Eukaryota, *Grammostola spatulata* )

P(9-15)id.sec.=0/3(0%) id.prim´=1/3(33%)

XCRYLFGGCKTTSDCCKHLGCKFRDKYCAWDFTF

>gi|10120908|pdb|1f82|a chain a, botulinum neurotoxin type b catalytic domain (Bacteria, *Clostridium botulinum* )

P(218-240)id.sec.=6/11(54%) id.prim´=4/11(36%)

P(291-306)id.sec.=2/8(25%) id.prim´=2/8(25%)

PVTINNENYNDPIDNNNIIMMEPPFARGTGRIYKAFKITDRIWIIPERYTFGYKPEDFNKSSGIFNRDVCEYYDPDYLNNDKKNIFLQ  
 TMIKLFNRIKSKPLGKLEMIINGIPYLGDRRVPLEEFNTNIASVTVNKLI SNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQN  
 HFASREGFGGIMQMKFCPEYVSVFNQENKASIFNRRGYFSDPALIIMHELIHVLHGLYGIKVDDLPIVNEKFFMQSTD  
 AIQAEELYTFGGQDPSIITPSTDKSIYDKVLQNFRGIVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEKYSIDVESFDK  
 LYKSLMFGFTEETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTIIEGFNISDKDMEKEYRGONKAINKQAYEEIS

>gi|10120890|pdb|1crz|a chain a, crystal structure of the e. coli tolB protein (Bacteria, *Escherichia coli* )

P(115-137) id.sec.=3/11(27%) id.prim´=3/11(27%)

DSGVDSGRPIGVVPPQWAGPGAAPEDIGGIVAADLRNSGKFNPLDRARLPQQPGSAQEVQPAAWSALGIDAVVVGQVTPNPDGSYNVAY  
 QLVDTGAGPGTVLAQNSYKVNKQWLRYAGHTASDEVEKLTGIKGAFRTRIAVYVQTNNGGQFPYELRVSDYDGYNQFVVRHSP  
 QPLXSPAWSPDGSKLAYVTFESGRSALVIQTLANGAVRQVASFPRHNGAPAFSPDGSKLAFALSKTGSNLNLYVXDLASGQIRQVTDGRS  
 NNTEPTWFPDSQNLAFSTDAQGRPQVYKVNINGGAPQRITWEGSQNDADVSDGKFXVXVSSNGGQQHIAKQDLATGGVQVLSSTFLD  
 ETPSLAPNGTXVIYSSSQGXGSVLNLVSTDGRFKARLPATDQGVKFPAWSPYL

>gi|9955127|pdb|1f94|a chain a, the 0.97 resolution structure of buccandin, a novel toxin isolated from the malayan krait (Eukaryota, *Bungarus candidus*)

P(22-57)id.sec.=13/16(81%) id.prim´=5/16(31%)



MECYRCGVSGCHLKITCSAEETFCYKWLNKISNERWLGCAKTCTEIDTWNVYNKCCTTNLCNX

>gi|7546294|pdb|1chz|a chain a, a new neurotoxin from *buthus martensii kirsch* (Eukaryota, *Buthus martensii*)

P(49-59)id.sec.=3/5(60%) id.prim'=1/5(20%)

VRDAYIAKPHNCVYECARNEYC>NNLCTKNGAKSGYCQWSGKYGNGCWCIBLPDnVPIRVPgKCH

>gi|6980632|pdb|1cz6|a chain a, solution structure of androctonin (Eukaryota, *Androctonus australis*)

P(1-23) id.sec.=7/11(63%) id.prim'=3/11(27%)

RSVCRQIKICRRGGCYKCTNRpy

>gi|6980449|pdb|1qoy|a chain a, e.coli hemolysin e (hlye, clyA, shea) (Bacteria, *Escherichia coli*)

P(53-71) id.sec.=1/9(11%) id.prim'=1/9(11%)

P(121-129)id.sec.=4/4(100%) id.prim'=2/4(50%)

P(141-189)id.sec.=20/24(83%) id.prim'=6/24(25%)

P(287-298)id.sec.=6/6(100%) id.prim'=2/6(33%)

GSPGISGGGGIILDSMAEIVADKTVEVVKNAIETADGALDLYNKYLDQVIPWQTFDETIKELSRFKQEYSQAASVLVGDIKTLL  
MDSQDKYFEATQTVYEWCGVATQLLAAYILLFDEYNEKKAsAQKDILIKVLDDGITKLNEAQKSLLVSSQSQFNNASGKLLA  
LDSQLTNDFSEKSSYFQSQVDKIRKEAYAGAAAGVVVGPFLIISYIAAGVVEGKLIPELKNKLKSVQNFFTLSNTVKQAN  
KDIDAAKLKLTTEIAAIGEIKTETETTRFYVDYDDLMLSLLKEAAKKMINTCNEYQKRHGKKTLEFVPEV

>gi|6729953|pdb|2b3c|a chain a, solution structure of a beta-neurotoxin from the new world scorpion *centruroides sculpturatus ewing* (Eukaryota, *Centruroides exilicauda*)

P(6-33)id.sec.=6/14(42%) id.prim'=3/14(21%)

KDGYLVEKTGCKKTCYKLGENDFCNRECKWKHIIGGSYGYCYFGCYCEGLPDSTQTWPLPNKTC

>gi|6730534|pdb|1qs1|b chain b, crystal structure of vegetative insecticidal protein2 (vip2) (Bacteria, *Bacillus cereus*)

P(206-220)id.sec.=2/7(28%) id.prim'=3/7(42%)

P(423-437)id.sec.=3/7(42%) id.prim'=2/7(28%)

MKRMEGKLFMVSKKLQVVTKTIVLLSTVFSISLLNNEVIKAEQLNINSQSKYTNLQNLKITDKVEDFKEDKEKAKEWGKEKEKEWKLTAI  
 EKGKMNFLDNKNDIKTNYKEITFSMAGSFEDEIKDLKEIDRMFDRKTNLSNSIITYKNVEPTTIGFNKSLTEGNTINSDAMAQFKEQFL  
 DRDIKFDSYLDTHLTAQQVSSKERVILKVTVPSGkGSTTPTKAGVILNNSEYKMLIDNGYMHVHDVKVSKVKKGVECLQIEGTLK  
 KSLDFKNDINAEAHSWGKMKNYEAWKDLTDSQREALDGYARQDYKEINNYLRNQGGSGNEKLDQAQIKNISDALGKKPIPENITVYRWCG  
 MPEFGYQISDPLPSLKDFEEQFLNTIKEDKGYMSTSLSSERLAAFGSRKIILRLQVPKGGSTGAYLSAIGGFASEKEILLDKSKY  
 HIDKVTEVVIKGVKRYVVDATLLTN

>gi|6573609|pdb|1c2u|a chain a, solution structure of [abu3,35]shk12-28,17-32 (*Eukaryota*,  
*Stichodactyla helianthus*)

P(11-29)id.sec.=9/9(100%) id.prim'=3/9(33%)

RSXIDTIPKSRCTAFQCKHsAKYRLSFCRKTGTX

>gi|6573377|pdb|1b1z|a chain a, streptococcal pyrogenic exotoxin a1 (*Bacteria*,  
*Streptococcus pyogenes*)

P(86-98)id.sec.=0/6(0%) id.prim'=2/6(33%)

DPDPSQLHRSSLVKNLQNIYFLYEGDPVTHENVKSVDQLLSHDLIYNVSGPNYDKLKTTELKNQEMATLFDKKNVDIYGVVEYHLCYLC  
 ENAERSACIYGGVTNHEGNHLEIPKKIVVKVSIDGIQSLSFDIETNKKMVTAQELDYKVRKYLTDNKQLYTNGPSKYETGYIKFIP  
 KNKESFWDFDFPEPEFTQSKYLMYKDNETLDSNTSQIEVYLTK

>gi|6137706|pdb|1qtf|a chain a, crystal structure of exfoliative toxin b (*Bacteria*,  
*Staphylococcus aureus*)

P(237-245)id.sec.=3/4(75%) id.prim'=3/4(75%)

KEYSAEEIRKLKQKFEVPPPTDKELYTHITDNARSPYNSVGTVVFVKGSTLATGVLIGKNTIVTNYHVAREAAKNPSNIIFTPAQNRDAEK  
 NEFPTPYGKFEAEIKESPYGQGLDLAIKLPNEKGESAGDLIQPANIPDHIDIAKGDKYSLLGYPYNYSAISLYQSQIEMFNDSQYF  
 GYTEVGNSSGSGIFNLKGEIIGHSGKGGQHNLPIGVFFNRKISSLYSDNTFGDTLGNDLKKAKLDK

>gi|6137610|pdb|1bud|a chain a, acutolysin a from snake venom of agkistrodon acutus at ph  
 5.0 (*Eukaryota*, *Deinagkistrodon acutus*)

P(56-78)id.sec.=0/11(0%) id.prim'=3/11(27%)

FQRYMEIVIVVDHSMVKRYNGSDSIKAWVYEMINTITESYSYLKIDISLSGLEIWSGKDLIDVEAsAGNTLKSFGEWRAKDL  
 IHRISHDNAQLLTATDFDGATIGLAYVASMCPKRSVGVIQDHSSVNLVAITLAHEMAHNLGVSHDEGSCSCGGKSCIMSPSISDETI  
 KYFSDCSYIQCRDYISKENPPCILN

>gi|6137387|pdb|3bta|a chain a, crystal structure of botulinum neurotoxin serotype a  
 (Bacteria, Clostridium botulinum )

P(7-50)id.sec.=13/21(62%) id.prim'=5/21(23%)

P(119-194)id.sec.=3/33(9%) id.prim'=11/33(33%)

P(756-991)id.sec.=15/110(14%) id.prim'=23/110(21%)

PFVVKQFNYKDPVNGVDIAYIKIPNVGQMQPVKAFKIHNKIWVIPERDTFTNPEEGDLNPPEAKQVPVSYDSTYLS  
 TDNEKDNYLKGVTKLFERIYSTDLGRMLLTSIVRGIPFWGGSTIDTELKVIDTNCINVIQPDGSYRSEELNLVIIGPSA  
 DIIQFECKSFGHEVLNLTRNGYGSTQYIRFSPDFTFGFEESLEVDTNPLLGAGKFATDPAVTLAHELIHAGHRLYGIAIN  
 PNRVFKVNTNAYYEMSGLEVSFEELRTFGGHDAKFIDSLQENEFRLYYNKFKDIASTLNKAKSIVGTASLQYMKNVFKEKYLLEDST  
 SGKFSVDKLFKFDKLYKMLTEIYTEDNFVKFFKVLNRKTYLNFDAKAVFKINIVPKVNYTIYDGFNLRNTNLAANFNGQNTTEINNMNFTKL  
 KNFTGLFEFYKLLCVRGIITSKTKSLDKGYNKALNDLCIKVNNWDLFFSPSEDNFTNDLNKGEETSDTNIEAAEENISLDLIQQYYLT  
 FNFNEPENISIENLSSDIIGQLELMPNIEFNGKKYELDKYTMFHYLRAQEFHGHKSRIALTNVNEALLNPSRVYTFSSDYVVKV  
 NKATEAAMFLGWVEQLVYDFDTSETSEVSTTDKIADITIIIPYIGPALNIGNMLYKDDFVGALIFSGAVILLEFIPEIAIPVLGTFALVS  
 YIANKVLTVQTIDNALSKRNEKWDEVYKYIVTNWLAKVNTQIDLIRKMKKEALENQAATKAIINYQYNOYTEEEKNNINFNID  
LSSKLNESINKAMINKFLNQCSVSYLMNSMIPYGVKRLEDFASLKDALLKYIDNRGTLIGVDRLKD  
 KVNNTLSTDIPFQLSKYVDNQRLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGSKVNFDPID  
 KNQIQLFNLESSKIEVILKNAIVNSMYENFSTSFWIRIPKYFNSISLNNEYTIINCMENNSGWKVSLNYG  
EIIWTLQDTQEIKQRVVFKYSQMINISDYINRWIFVTITNRLNSKIYINGRLIDQKPISNLGNIHASNNIMFKDGCRDTHRYI  
 WIKYFNLFDKELNEKEIKDLYDNQSNNGILKDFWGDYLYQDKPYMLNLYDPNKYVDVNVNGIRGYMYLKGPRGSVMTTNIYLNSSLYR  
 GTKFIIKKYASGNKDNIVRNDRVYINVVVKNKEYRLATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNKCKMNLQDNNGN  
 DIGFIGFHQFNNAKLVASNWYNRQIERSSRTLGCSEWEIFPVDGGERPL

>gi|5822082|pdb|1kp6|a chain a, ustilago maydis killer toxin kp6 alpha-subunit  
 (Eukaryota, Ustilago maydis )

P(22-36)id.sec.=2/7(28%) id.prim'=2/7(28%)

NNAFCAGFGLSCKWEWCWTAHGTGNELRYATAAGCGDHLSKSYDARAGHCLFSDDLRNQFYSHCSSLNNNMSCRSLSK

>gi|5821818|pdb|1bi0| structure of apo- and holo-diphtheria toxin repressor (Bacteria, *Corynebacterium diphtheriae*)

P(120-138)id.sec.=6/9(66%) id.prim'=3/9(33%)

P(196-211)id.sec.=8/8(100%) id.prim'=3/8(37%)

MKDLVDTEMYLRTIYELEEEGVTPLRARIAERLEQSGPTVSQTVARMERDGLVVVASDRSLQMTPTGRTLATAVMRKHRLAERLLTDI  
 IGLDINKVHDEAXRWEHVMSDEVERRLVKVLKDVSRSRPFFNPIPGLDELGVGNSDAAAPGTRVIDAATSMPRKVRIVQINEIFQ  
 VETDQFTQLLDADIRVGSEVEIVDRDGHITLSHNGKDVELLDDLAHTTIRIEEL

>gi|5821789|pdb|1b45| alpha-conia conotoxin from conus consors, nmr, 43 structures (Eukaryota, *Conus consors*)

P(8-14)id.sec.=2/3(66%) id.prim'=1/3(33%)

GRCCHPACGKyYSCX

>limt\_mol:protein length:80 mamba intestinal toxin 1 nmr 39 structures (Eukaryota, *Dendroaspis polylepis polylepis*)

P(43-60)id.sec.=9/9(100%) id.prim'=2/9(22%)

P(61-67)id.sec.=0/3(0%) id.prim'=2/3(33%)

AVITGACERDLQCGKGTCCAVSLWIKSVRVCTFPVGTSGEDCHPASHKIPFSGQRMHHTCPCAPNLACVQTSPPKFKCLSK

>gi|4699530|pdb|1a87| colicin n (Bacteria, *Escherichia coli*)

P(33-107)id.sec.=14/35(40%) id.prim'=9/35(25%)

HGDNNSKPKPGGNSGNRGNNGDGASAKVGEITITPDNSKPGRYISSNPEYSLLAKLIDAESIKGTEVYTFHTRKGQY  
VKVTVPDSNIDKMRVDYVNWKGPKYNNKLVKRFVSQFLLEKKEKEKEALLKASELVSGMGDKLGEYLGVKYKNVAKE  
 VANDIKNFHGRNIRSNEAMASLNKVLANPKMKVNSDKDAIVNAWKQVNAKDMANKIGNLGKAFKVADLAIKVEKIREKSIEGYNTGN  
 WGPLLLEVESWIIIGGVVAGVAISLFGAVLSFLPISGLAVTALGVIGIMTISYLSFFIDANRVSNINNISSVIR

>gi|3891964|pdb|1av3| potassium channel blocker kappa conotoxin pviia from *C. purpurascens*, nmr, 20 structures (Eukaryota, *Conus purpurascens*)

P(20-26)id.sec.=3/3(100%) id.prim'=1/3(33%)

CRIXNQKCFQHLLDCCSRKCNRFNKCV

>gi|3212819|pdb|3seb| staphylococcal enterotoxin b (Bacteria, *Staphylococcus aureus*)

P(5-9)id.sec.=2/2(100%) id.prim'=2/2(100%)



ESQPDPKPELHKSSKFTGLMENMKVLYDDNHVSAINVKSIDQFLYFDLIYSIKDTKLGNYDNVRVEFKNKDLADKYKDKYVDVFGAN  
 YYYQCYFSKKTNDINSHQTDKRKTCMYGGVTEHNGNQLDKYRSITVRFVFDGKNLLSFDVQTNKKKVTAQELDYLTRHYLVKNKKLYEF  
 NNSPYETGYIKFIENENSFWYDMMPAPGDKFDQSKYLLMMYNDNMVDSKDVKIEVYLTTKK

>gi|2982092|pdb|lexf|a chain a, exfoliative toxin a (*Bacteria, Staphylococcus aureus*)  
 P(209-217)id.sec.=4/4(100%) id.prim´=2/4(50%)

EVSAAEIKKHEEKWNKYGVNAFNLPKELFSKVEKDRQKYPYNTIGNVVFVKGQTSATGVLIGKNTVLTNRHIAKFANGDPSKVSFRPS  
 INTDDNGNTETPYGEYEVKEILQEPFGAGVDLALIRLQPDQNGVSLGDKISPARIKIGTSNDLKDGDKLELIGYPFDHKVNQMRSEIELT  
 TLRGLRYYGFTVPGNSGSGIFNSNGELVGIHSSkVSHLDREHQINYGVGIGNYVKRIINEKNE

>gi|2981940|pdb|2nbt|a chain a, neuronal bungarotoxin, nmr, 10 structures (*Eukaryota, Bungarus multicinctus*)

P(3-58)id.sec.=12/26(46%) id.prim´=9/26(34%)

RTCLISPSSTPQTCPNGQD ICFLKAQC DrFCSIRG PVIEQGC VATCPQFRSNY RSLLCCTTDNCNH

>gi|2914135|pdb|lacc| anthrax protective antigen (*Bacteria, Bacillus anthracis*)

P(91-151)id.sec.=16/22(72%) id.prim´=7/22(31%)

P(371-381)id.sec.\*=3/5(60%) id.prim´=2/5(40%)

\*kasn no tiene estructura secundaria asignable.

EVKQENRLLNESESSSQGLLGYFSDLNFOAPMVVTSSTTGDLSSIPSELENIPSENQYFQSAIWSGFIVKKSDEYTFATSADNHVTM  
 WVDQEVIN kasn SNKIRLEKGRLYQIKIQYQRENPTKGLDFKLYWTDSONKKEVISSDNLQLPELKQKSSNS  
 RKKRSTSAGPTVPDRDNDGIPDSLEVEGYTVDVKNKRFLSPWISNIHEKKGLTKYKSSPEKWSTASDPYSDFEKVTGRIDKNVSPEAR  
 HPLVAAYPIVHVDMENIILSKNEDQSTQNTDSETRTISKNTSTSRTHTSEVHGNAEVHASFFDIGGSVSAGFSNSNSSTVAIDHSLSLA  
 GERTWAETMGLNTADTARLNANIRYVNTGTAPIYNVLPTTSLVLGKNQTLATIKAKENQLSQILAPNNYPSKNLAPIALNAQDDF  
 SSTPITMNYNQFLELEKTKQLRLD TDQVYGNIA TYNFENGVRVDTGSNWSEVLPQIQETTARIIFNGKDLNLVERRIA AVNPSDPLET  
 TKPDMTLKEALKIAFGFNEPNGNLQYQKDI TEFDNFQDQTSQNIKNQLAELNATNIYTVLDKIKLNKMNILIRDKRFHYDRNNAV  
 GADESVVKEAHREVINSSTEGLLLNIDKDIRKILSGYIVEIEDTEGLKEVINDRYDMLNISSLRQDGKTFIDFKKYNDKPLIYISNPY  
 KVNVAVTKENTIINPSENGDTSTNGIKKILIFSCKGYEIG

>gi|2780936|pdb|5tss|a chain a, toxic shock syndrome toxin-1: orthorhombic p222(1)  
 crystal form (*Bacteria, Staphylococcus aureus*)

P(144-164)id.sec.=4/10(40%) id.prim´=3/10(30%)

STNDNIKDLLDWYSSGSDTFNSEVLDNSLGSMRIKNTDGSISLIIIFPSPIYSPAFTKGEKVDLNTKRTKKSQHTSEGTIHFQISGVT  
 NTEKLPPIELPLKVKVHGKDSPLKYWPKFDKKQLAISTLDFEIRHQLTQIHGLYRS\_SDKTGGYwKITMNDGSTY\_QSDLSKKFE  
 YNTEKPPINIDEIKTIEAEIN

>gi|2781193|pdb|lbeq| structure of fungal elicitor, nmr, 18 structures (Eukaryota,  
*Phytophthora cryptogea*)

P(67-70)id.sec.=2/2(100%) id.prim'=2/2(100%)

P(83-89)id.sec.=2/3(66%) id.prim'=2/3(66%)

TACTATQQTAAKYKTLVSIKSDASFNQCSTDSGYSMLTAKALPTTAQYKLMCASTACNTMIKKIVTLNPPNCDLTVPTSGLVLVNV  
YsYANGFSNKCSL

>gi|2098492|pdb|lkve|a chain a, killer toxin from halotolerant yeast (Eukaryota, *Pichia*  
*farinosa*)

P(50-61)id.sec.=6/6(100%) id.prim'=3/6(50%)

WSLRWRMQKSTTIAAIAGCSGAATFGGLAGGIVGCIAGILAILQGFEVNWHNGGGDRSNPV

>gi|1827608|pdb|lpre|a chain a, proaerolysin (Bacteria, *Aeromonas hydrophila*)

P(91-101) id.sec.=1/5(20%) id.prim'=2/5(40%)

P(129-157)id.sec.=7/14(50%) id.prim'=5/14(35%)

AEPVYPDQLRFLFSLGQVCGDKYRPVNRBEAQSVKSNIVGMMGQWQISGLANGWVIMGPGYNGEIKPGTASNTWCYPTNPVTGEIPTLS  
ALDIPDgDEEVDVQWRLVHDSANFIKPTSYLAHYLGAWVGGNHSQYVGEDMDVTRDGDGWVIRGNNDGgCDGYRCGDKT  
 AIKVSNFAYNLDPDSFKHGDVTQSDRQLVKTTVGVAVNDSPTPQSGYDVTLRYDTATNWSKTNTYGLSEKVTTKNKFKWPLVGETELSI  
 EIAANQSWASQNGGSTTTLSQSVRPTVPARSKIPVKIELYKADISYPYEFKADVSYDLTSLGFLRWGGNAWYTHPDNRPNNHFTFVIG  
 PYKDKASSIRYQWDKRYIPGEVKWWDWNWTIQONGLSTMQNNLARVLRPVRAGITGDFSAESQFAGNIEIGAPVPLAADSKVRRARSVD  
 GAGQGLRLEIPLDAQELSGLGFNNVSLSVTPAANQ

>gi|1827727|pdb|ltii|d chain d, escherichia coli heat labile enterotoxin type iib  
 (Bacteria, *Escherichia coli*)

P(76-96)id.sec.=10/10(100%) id.prim'=1/10(10%)

GASQFFKDNCRNTASLVEGVELTKYISDINNNTDGMVYVVSSTGGVWRISRKADYPDNVMTAEMRKIAMAAVLSGMRVNMCASEAS  
SPNVIWAIELEAE

>gi|1827732|pdb|ltii|a chain a, escherichia coli heat labile enterotoxin type iib

P(170-182)id.sec.=6/6(100%) id.prim'=3/6(50%) (Bacteria, *Escherichia coli*)

NDYFRADSRTPDEVRRSGGLIPRGQDEAYERGTPININLYDHARGTATGNTRYNDGYVSTTTTLRQAHLLGQNLGGYNEYIYVVA  
 PNLFDVNGVLGRYSPIYSENEYAALGGIPLSQIIGWYRVSFGAIEGGMHRNRDYYRDLFRGLSAAPNEDGYRIAGFPDGFPAWE<sup>U</sup>EV<sup>P</sup>  
 WRE<sup>F</sup>AP<sup>N</sup>SCLPNNK

>gi|1827850|pdb|1xtc|d chain d, cholera toxin (Bacteria, *Vibrio cholerae*)

P(4-21)id.sec.=1/9(11%) id.prim'=2/9(22%),5/9(55%)

P(31-47)id.sec.=8/8(100%) id.prim'=2/8(25%)

TPQ<sup>N</sup>IT<sup>D</sup>LCAEYHNTQ<sup>I</sup>Y<sup>T</sup>LN<sup>D</sup>KIFSYTESLAGKREMA<sup>I</sup>ITFKNGA<sup>I</sup>FOVEVPSSQHIDSQKKAIERMKD<sup>T</sup>LRIAYL<sup>T</sup>EA  
 KVEKLC<sup>T</sup>WNNKTPHAI<sup>A</sup>AAISMAN

>gi|1633286|pdb|1apf| anthopleurin-b, nmr, 20 structures (Eukaryota, *Anthopleura xanthogrammica*)

P(1-15)id.sec.=7/7(100%) id.prim'=3/7(42%)

GVPCLCDs<sup>D</sup>GPR<sup>P</sup>RG<sup>N</sup>TL<sup>S</sup>GILWFYPSGCPGWHNCKAHGPNIGW<sup>C</sup>CKK

>gi|1633500|pdb|1tox|a chain a, diphtheria toxin dimer complexed with nad (Bacteria, *Corynebacterium diphtheriae*)

P(177-183)id.sec.=3/3(100%) id.prim'=3/3(100%)

P(425-436)id.sec.=6/6(100%) id.prim'=3/6(50%)

GADDVVDSSKSFVMENFSSYHGTKPGYVDSIQKGIQKPKSGTQGNYYDDWKGIFYSTDNKYDAAGYSVDNENPLSGKAGGVVKVYTPGLT  
 KVLALKVDNAETIKKELGLSLTEPLMEQVGT<sup>E</sup>E<sup>F</sup>IKRFGDGASRVVLSLPPAEGSSSVEYINNWEQAKALSVELEINFETR<sup>G</sup>KRGQD<sup>A</sup>  
 MYEYMA<sup>Q</sup>ACAGNRVRRSVGSSLS<sup>C</sup>INLDWDVIRDKTKTKIESLKEHGPIKNMSESPNKT<sup>V</sup>SEEKAKQYLEEFHQ<sup>T</sup>ALEHP<sup>E</sup>SELK  
 TVTGTNPVFAGANYAAWAVNVAQVIDSETADNLEKTTAALSILPGIGSVMGIADGAVHHNTEETVAQ<sup>S</sup>IALS<sup>S</sup>LMVAQA<sup>I</sup>PLV<sup>G</sup>ELVDI  
 GFAAYNFVESIINLFQVVHNSYNRPAYSPGHKTQ<sup>F</sup>FLHDGYAVSWNTVEDSII<sup>R</sup>TGFQGESGHDIKITAEN<sup>T</sup>PL<sup>P</sup>IA<sup>G</sup>VLL<sup>L</sup>PT<sup>I</sup>PG  
 KLDV<sup>N</sup>KSKTHISVNGRKIRMCRAIDGDVTF<sup>C</sup>RPKSPVYVNGVHANLHVA<sup>F</sup>HRS<sup>S</sup>SEKIHSNEISSDSIGVLGYQKTVDHTK<sup>V</sup>NSKLS  
 LFFEIKS

>gi|1431755|pdb|1bun|a chain a, hydrolase, presynaptic neurotoxin mol\_id: 1; molecule:  
 beta2-bungarotoxin; chain: a, b; synonym: phosphatide acyl-hydrolase, phosphatidylcholine  
 2-acylhydrolase; ec: 3.1.1.4; other\_details: numerous natural isoforms, this is the beta2  
 isoform (Eukaryota, *Bungarus multicinctus*)

P(50-62)id.sec.=1/6(16%) id.prim'=2/6(32%)

NLINFMEMIRYTIPCEKTWGEYADYGCGAGGSGRPIDALDRCCYVHDNCYGDAEKKKHCNPKTQSYSYKLTkRTIICYGAAGTC

ARIVCDCRТАALCFGNSEYIEGHKNIDTARFCQ