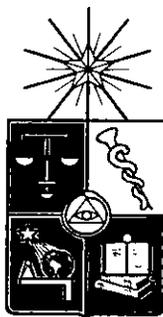


UCH-FC
MAG-B
F95Y
C.1



**“ESTUDIO DE LAS CARACTERÍSTICAS
ESTRUCTURALES Y ENERGÉTICAS DE LOS
SITIOS DE UNIÓN DE NAD Y NADP
MEDIANTE POTENCIALES ESTADÍSTICOS”**

Tesis entregada a la Universidad de Chile
en cumplimiento parcial de los requisitos
para optar al grado de
Magíster en Ciencias Biológicas

Facultad De Ciencias
Universidad de Chile

Por

Matías Sebastián Fuentealba Valenzuela

Santiago, 2015

Director de Tesis Dr. Ricardo Cabrera

**FACULTAD DE CIENCIAS
UNIVERSIDAD DE CHILE
INFORME DE APROBACIÓN
TESIS DE MAGÍSTER**

Se informa a la Escuela de Postgrado de la Facultad de Ciencias de la Universidad de Chile que la Tesis de Magíster presentada por el candidato

MATÍAS SEBASTIÁN FUENTEALBA VALENZUELA

Ha sido aprobada por la comisión de Evaluación de la tesis como requisito para optar al grado de Magíster en Ciencias Biológicas, en el examen de Defensa Privada de Tesis rendido el día 15 de Enero del 2015

Director de Tesis:
Dr. Ricardo Cabrera

Comisión de Evaluación de la Tesis

Dr. Octavio Monasterio

Dra. Victoria Guixé



Octavio Monasterio
Victoria Guixé

Dedicado a mi familia



Me llamo Matías Sebastián Fuentealba Valenzuela. Nací en Ovalle, Chile el 28 de Abril de 1986. Desde muy joven, me sentí atraído por la biología y la computación, lo que años mas tarde me impulsaría a estudiar la carrera de Ingeniería en Bioinformática en la Universidad de Talca. Durante mis estudios de pregrado, quede fascinado por el mundo de las proteínas, lo que me llevó tiempo después a llegar al Laboratorio de Bioquímica y Biología Molecular de la Universidad de Chile, donde inicie mi verdadera vocación como científico y en donde desarrollé este trabajo de tesis.

Agradecimientos

En primer lugar, me gustaría agradecer al Dr. Ricardo Cabrera por permitirme formar parte de su grupo de laboratorio, por su apoyo incondicional y por su confianza en mis ideas. Su capacidad de guiar mi trabajo, ha sido un aporte invaluable no solo en el desarrollo de esta tesis, sino también un ejemplo en mi formación como científico y persona.

A mi Madre, Janine Valenzuela y mis abuelos, Silvia Lazo y Max Valenzuela, por apoyarme en todo momento, por los valores que me han inculcado y por haberme entregado todo su cariño durante el transcurso de mi vida.

Por supuesto, agradezco a todos los miembros del Laboratorio de Bioquímica y Biología Molecular o los que alguna vez lo fueron, entre ellos a Luis Matsuda, Pablo Maturana, Rodrigo Muñoz, Erick Armingol, JC Cáceres, Eduardo Tobar, Pablo Villalobos, Ricardo Zamora, María José Abarca, Cesar Ramírez, Alejandra Herrera, Diego Quiroga, con quienes he compartido momentos de alegría y fiesta e interesantes conversaciones (Disculpas si olvide a alguien).

Agradezco también al Dr. Francisco Melo y sus alumnos Alex Slater e Ignacio Ibarra, por transmitirme sus conocimientos y contribuir intelectualmente al desarrollo este trabajo.

Finalmente, me gustaría agradecer a todos aquellos que hicieron posible este trabajo a través de su apoyo financiero. Entre ellos, quiero agradecer al proyecto 1121170 del Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) y al proyecto Anillo ACT 1107. También quiero agradecer a la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) por su financiamiento parcial de mis estudios de Magíster. Y por último, a Vicerrectoría de Asuntos Académicos de la Universidad de Chile por entregarme una beca para viajar a Brasil a realizar experimentos de cristalización y cristalografía.

ÍNDICE DE CONTENIDOS

| | |
|--|-----------|
| 1. Introducción..... | 1 |
| Dinucleótidos de nicotinamida adenina como coenzimas | 1 |
| Función celular de NAD(H) y NADP(H)..... | 4 |
| Cambios de especificidad por NAD(H) y NADP(H) | 6 |
| Base de datos de estructuras de proteínas | 8 |
| Potenciales Estadísticos | 11 |
| 1.1 Hipótesis | 13 |
| 1.2 Objetivos..... | 14 |
| 1.2.1 Objetivo general..... | 14 |
| 1.2.2 Objetivos específicos | 14 |
| 2. Materiales y Métodos | 17 |
| 2.1 Bases de datos..... | 17 |
| 2.2 Potenciales Estadísticos | 21 |
| 2.2.1 Tipos de átomos..... | 32 |
| 2.3 Caracterización de los sitios de unión de NAD(P)..... | 35 |
| 2.4 Evaluación y validación de los potenciales estadísticos..... | 37 |
| 3. Resultados..... | 39 |
| 3.1 Entrenamiento de potenciales estadísticos | 39 |
| 3.1.1 Generación de una base de datos de estructuras de complejos proteína-NAD(P) | 39 |
| 3.1.2 Aminoácidos más frecuentes en los sitios de unión de NAD(P)..... | 42 |
| 3.1.3 Potenciales estadísticos: perfiles de pseudo-energía | 47 |
| 3.2 Evaluación de los potenciales estadísticos..... | 56 |
| 3.3 Validación de los potenciales estadísticos | 60 |
| 4. Discusión..... | 76 |
| 5. Conclusiones | 86 |
| 6. Referencias..... | 88 |
| Apéndice..... | 95 |
| Anexos..... | 99 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| Figura 1. Coenzimas más utilizadas en reacciones catalíticas dentro del metabolismo de <i>Escherichia coli</i> . El número de reacciones fue calculado a partir del modelo metabólico iJO1366 de <i>Escherichia coli</i> | 2 |
| Figura 2. Estructura tridimensional de las coenzimas NAD(P). En cada coenzima se nombran los grupos químicos que componen su estructura. | 3 |
| Figura 3. Coenzimas más frecuentes dentro de las estructuras de la base de datos PDB. | 9 |
| Figura 4. Esquema de interacciones no-locales entre átomos de la coenzima y la proteína a una distancia de corte y tamaño de ventana determinados. El círculo azul representa un átomo de la coenzima de un tipo determinado y los círculos rojos representan átomos de la proteína de tipo carbono-aromático a una distancia de corte determinada. | 21 |
| Figura 5. Frecuencia relativa para obtener una cara en tres simulaciones computacionales del lanzamiento de una moneda 1.000 veces. | 23 |
| Figura 6 Aproximación de Sippl para la función $f^j(r)$. Las líneas azul y roja muestran los valores que toman en la aproximación los términos que acompañan a $f(r)$ y $g^j(r)$, respectivamente. Se señala en el gráfico mediante una línea punteada el número de observaciones a la que los términos tienen el mismo peso utilizando un $\sigma=0,02$ | 28 |
| Figura 7. Perfiles de pseudo-energía para las variable calculadas durante el desarrollo de los potenciales estadísticos. Los perfiles están calculados utilizando una distancia de corte de 7,0 Å con un tamaño de ventana de 0,2 Å. Una descripción detallada de cada gráfico puede ser encontrada en las Páginas 30 y 31. | 32 |
| Figura 8. Definición de tipos de átomos SYBYL para los 20 aminoácidos proteínogénicos. En la tabla se identifica el número que le corresponde a cada definición de tipo de átomo, el cual es asignado en las estructuras moleculares de los aminoácidos. | 34 |
| Figura 9. Definición de tipos de átomos SYBYL para los átomos de las coenzimas NAD(P). En la tabla se identifica el número que le corresponde a cada definición de tipo de átomo, el cual es asignado en las estructuras moleculares de las coenzimas. | 35 |

| | |
|--|----|
| Figura 10. Distribución del número de átomos de la coenzima en las estructuras de la base de datos PDB..... | 40 |
| Figura 11. Extensión de la coenzima (a) y resolución (b) de las estructuras en la base de datos PDB..... | 41 |
| Figura 12. Esquema de la creación de la base de datos de entrenamiento. Los criterios de filtro utilizados se indican adyacente a la flecha. | 42 |
| Figura 13. Residuos aminoacídicos más comunes en los sitios de unión de NAD(P). Cada círculo representa la cadena lateral de un aminoácido. El tamaño del círculo indica el porcentaje de estructuras de la base de datos de entrenamiento que presenta este aminoácido interactuando con cada región de la coenzima..... | 46 |
| Figura 14. Perfiles de pseudo-energía para interacciones de tipo puentes de hidrógeno. A la izquierda se muestra el perfil de pseudo-energía para los tipos de átomos indicados (parte superior-central de cada gráfico) pertenecientes a la coenzima (C) y la proteína (P). A la derecha se muestra un ejemplo de cada tipo de interacción, encontrado en la estructura cuyo código PDB se indica en la parte superior de cada imagen. | 51 |
| Figura 15. Perfiles de pseudo-energía para interacciones electrostáticas. A la izquierda se muestra el perfil de pseudo-energía para los tipos de átomos indicados (parte superior-central de cada gráfico) pertenecientes a la coenzima (C) y la proteína (P). A la derecha se muestra un ejemplo de cada tipo de interacción, encontrado en la estructura cuyo código PDB se indica en la parte superior de cada imagen. | 53 |
| Figura 16. Perfiles de pseudo-energía para la interacción catión- π entre la adenina de las coenzimas y cadenas laterales de arginina. A la izquierda se muestra el perfil de pseudo-energía para los tipos de átomos indicados (parte superior-central de cada gráfico) pertenecientes a la coenzima (C) y la proteína (P). A la derecha se muestra un ejemplo de cada tipo de interacción, encontrado en la estructura cuyo código PDB se indica en la parte superior de cada imagen. | 55 |
| Figura 17. Perfiles de energía para la interacción aromática entre el anillo de nicotinamida en las coenzimas y residuos aromáticos. A la izquierda se muestra el perfil de pseudo-energía para los tipos de átomos indicados (parte superior-central de cada gráfico) pertenecientes a la coenzima (C) y la proteína (P). A la derecha se muestra un ejemplo de cada tipo de | |

- interacción, encontrado en la estructura cuyo código PDB se indica en la parte superior de cada imagen. 56
- Figura 18. Evaluación de los potenciales estadísticos. Se muestra el coeficiente de correlación obtenido (eje y) para cada distancia de corte (eje x) y tamaño de ventana analizado (colores). En el gráfico interno se muestra el análisis de correlación para los parámetros con la mejor correlación. 59
- Figura 19. Sitio de unión de NAD en la enzima 3- α -hidroxiesteroide deshidrogenasa de *Comamonas testosteroni*, representado como cadenas laterales de los residuos y estructuras secundarias (a) o superficie, donde los colores representan la pseudo-energía de cada átomo (b). El color de la superficie oscila desde azul, para pseudo-energías negativas, a rojo, para pseudo-energías positivas, cubriendo ± 3 kcal/mol (barra de colores). En cada imagen se señalan con flechas los residuos mencionados en el análisis. 62
- Figura 20. Pseudo-energía de los sitios de unión de NAD a la enzima Acido 2,5-diceto-D-glucónico reductasa de *Corynebacterium sp.* (a), y a la mutante F22Y/K232G/ R238H/A272G (b). El color de la superficie oscila desde azul, para pseudo-energías negativas, a rojo, para pseudo-energías positivas, cubriendo ± 3 kcal/mol (barra de colores). En cada imagen se señalan con flechas los residuos mutados. 65
- Figura 21. Pseudo-energía de los sitios de unión de NAD a la enzima Gliceraldehído-3-fosfato deshidrogenasa de *Bacillus stearothermophilus* (a) y a la mutante D32G/L187A/P188S (b). El color de la superficie oscila desde azul, para pseudo-energías negativas, a rojo, para pseudo-energías positivas, cubriendo ± 3 kcal/mol (barra de colores). En cada imagen se señalan con flechas los residuos mutados. 68
- Figura 22. Pseudo-energía de los sitios de unión de NADP en la enzima Gliceraldehído-3-fosfato deshidrogenasa de *Bacillus stearothermophilus* (a) y a la mutante D32G/L187A/P188S (b) El color de la superficie oscila desde azul, para pseudo-energías negativas, a rojo, para pseudo-energías positivas, cubriendo ± 3 kcal/mol (barra de colores). En cada imagen se señalan con flechas los residuos mutados. 70
- Figura 23 Análisis de correlación para los potenciales estadísticos con mejor evaluación (puntos negros) para NAD (a) y NADP (b), incluyendo los casos de validación de los potenciales estadísticos (puntos rojos). 72
- Figura 24. Pseudo-energía de los sitios de unión de NAD (arriba) y NADP (abajo) para la mutación por Asp en la posición de la 42 a la 51 en la

estructura (a). La línea punteada en cada gráfico representa el valor de pseudo-energía de la enzima silvestre en cada complejo. En un rectángulo en rojo se destaca la posición donde el residuo de Asp produce una pseudo-energía más negativa con NAD y más positiva con NADP, y sobre cada gráfico se muestra la superficie de pseudo-energía para esta mutante. Gráfico de la dependencia de la velocidad inicial a diferentes concentraciones de NAD (puntos blancos) o NADP (puntos negros) (b). La velocidad inicial medida como $\mu\text{mol}/\text{min}$ (U) fue dividida por los miligramos de proteína utilizados en la cubeta de reacción, para normalizar aquellos casos en que se usaron cantidades diferentes de enzima. 74

Figura 25. Correlación entre la pseudo-energía del sitio de unión de NADP, en la enzima Glutación reductasa de *Escherichia coli* y mutantes, y el cambio en energía libre del estado basal del complejo calculada a partir de las constantes de Michaelis-Menten (rojo) o el cambio en la energía libre calculado a partir de las constantes de especificidad (azul). 98

ÍNDICE DE TABLAS

| | |
|--|----|
| Tabla 1. Resumen de los cambios de especificidad por NAD(P)(H). | 7 |
| Tabla 2. Identificador químico de los dinucleótidos de nicotinamida adenina y número de estructuras con estas coenzimas, en las base de datos PDB. | 18 |
| Tabla 3. Tipos de átomos de la proteína asociados a aminoácidos proteínogénicos. | 36 |
| Tabla 4. Energía libre y pseudo-energía con NAD a la enzima 3- α -hidroxiesteroide deshidrogenasa de <i>Comamonas testosteroni</i> y a mutantes de esta enzima diseñadas para cambiar su especificidad por esta coenzima. En la parte inferior de la tabla se muestra el coeficiente de correlación de Pearson (R) entre las energías libres y las pseudo-energías, magnitudes que fueron calculadas utilizando una temperatura de 298 K..... | 63 |
| Tabla 5. Energía libre y pseudo-energía de la unión de NAD a la enzima Acido 2,5-diceto-D-glucónico reductasa de <i>Corynebacterium sp.</i> y a mutantes de esta enzima diseñadas para cambiar su especificidad por esta coenzima. En la parte inferior de la tabla se muestra el coeficiente de correlación de Pearson (R) entre las energías libres y las pseudo-energías, magnitudes que fueron calculadas utilizando una temperatura de 298 K..... | 66 |
| Tabla 6. Energía libre y pseudo-energía de la unión de NAD o NADP a la enzima Gliceraldehído-3-fosfato deshidrogenasa de <i>Bacillus stearothermophilus</i> y a mutantes de esta enzima diseñadas para cambiar su especificidad por esta coenzima. En la parte inferior de la tabla se muestra el coeficiente de correlación de Pearson (R) entre las energías libres y las pseudo-energías, magnitudes que fueron calculadas utilizando una temperatura de 298 K..... | 67 |
| Tabla 7. Constantes de especificidad con NAD y NADP para la enzima glucosa-6-fosfato deshidrogenasa de <i>Leuconostoc mesenteroides</i> y la mutante A45D. | 75 |
| Tabla 8. Energía libre y pseudo-energía para la unión de NADP a la enzima Glutación reductasa de <i>Escherichia coli</i> y mutantes diseñadas para cambiar la especificidad por coenzimas en esta enzima. | 97 |

LISTA DE ABREVIATURAS

| | |
|-----------------------------|--|
| NAD | Nicotinamida adenina dinucleótido oxidado |
| NADP | Nicotinamida adenina dinucleótido fosfato oxidado |
| NAD(H) | Nicotinamida adenina dinucleótido oxidado y reducido |
| NADP(H) | Nicotinamida adenina dinucleótido fosfato oxidado y reducido |
| NAD(P)(H) | NAD, NADH, NADP y NADPH |
| NAD(P) | NAD y NADP |
| NAD(P)H | NADH y NADPH |
| NADH | Nicotinamida adenina dinucleótido reducido |
| NADPH | Nicotinamida adenina dinucleótido fosfato reducido |
| ADP | Adenosín difosfato |
| PDB | Protein Data Bank |
| PDB ID | Código identificador dentro de la base de datos PDB |
| K_D | Constante de disociación |
| ΔG^0 | Energía libre |
| K_M | Constante de Michaelis-Menten |
| k_{cat} | Constante catalítica |
| k_{cat}/K_M | Constante de especificidad |
| $\Delta\Delta G_b$ | Cambio en energía libre del estado basal |
| $\Delta\Delta G_b^\ddagger$ | Cambio en energía libre del estado de transición |

Resumen

El reconocimiento molecular de los dinucleótidos de nicotinamida adenina, NAD y NADP es una característica fundamental de un número importante de enzimas que participan en procesos del metabolismo redox y señalización intracelular. En la actualidad, la gran cantidad de estructuras cristalográficas determinadas con estas coenzimas permite la posibilidad de desarrollar potenciales estadísticos para estudiar las características estructurales y energéticas de los sitios de unión de NAD y NADP en proteínas. En este trabajo, usamos los potenciales estadísticos para transformar las frecuencias observadas de interacciones entre átomos de la coenzima y la proteína, en magnitudes relacionadas con energía, denominadas pseudo-energías. Se analizaron las interacciones más frecuentes y sus perfiles de pseudo-energía en una base de datos de 359 estructuras en el caso de NAD y 330 en el caso de NADP. Una característica distintiva en los complejos con NAD es la presencia de un carboxilato de aspartato o glutamato en la proteína que forma un puente de hidrógeno con la ribosa adyacente a la adenina. En los complejos con NADP una cadena lateral de una arginina forma al mismo tiempo una interacción electrostática con el fosfato-2' y una interacción catión- π con la adenina. Para ambas coenzimas, grupos hidroxilo en la proteína

forman puentes de hidrógeno con los grupos fosfato, mientras que el bolsillo alrededor de la nicotinamida muestra frecuentemente una cadena lateral aromática y además un puente de hidrógeno entre el nitrógeno en su grupo amida y el oxígeno en las cadenas principales de los residuos. Adicionalmente, utilizamos un grupo de enzimas con estructura y constantes de disociación conocida con el objetivo de optimizar los potenciales estadísticos para obtener una mejor correlación entre las pseudo-energías calculadas y la información experimental. Las pseudo-energías calculadas a partir de estos potenciales estadísticos optimizados se correlacionan con los cambios en la energía libre de unión en mutantes por alanina y otros residuos utilizando tanto estructuras cristalográficas como modelos *in silico*. Finalmente, utilizamos los potenciales estadísticos para proponer una mutación para cambiar la preferencia por coenzima en la enzima glucosa-6-fosfato deshidrogenasa de *Leuconostoc mesenteroides*. Resultados *in vitro*, al medir las velocidades iniciales a diferentes concentración de NAD o NADP, demuestran que la mutante Ala45Asp cambia la preferencia de esta enzima desde NADP a NAD.

Abstract

The molecular recognition of the nicotinamide adenine dinucleotides, NAD and NADP is a fundamental characteristic in a large number of enzymes involved in redox metabolism processes and intracellular signaling. Nowadays, the high amount of crystallographic structures with these coenzymes allow the possibility to develop statistical potentials to study the structural characteristics and energetics of the NAD and NADP binding sites in proteins. In this work, we use statistical potentials to transform the observed interactions frequencies between atoms in the coenzyme and the protein, in magnitudes related with energies, called pseudo-energies. We analyze the most frequent interactions and their pseudo-energies profiles in a database of 359 structures in the case of NAD and 330 in the case of NADP. One distinctive characteristic in the NAD complexes is the presence of a carboxylate from aspartate or glutamate in the protein establishing a hydrogen bond with the ribose adjacent to the adenine. In the NADP complexes an arginine side chain forms, at the same time, an electrostatic interaction with the 2'-phosphate and a cation- π with the adenine. For both coenzymes, hydroxyl groups in the protein form an hydrogen bonds with the phosphate groups, meanwhile the pocket around the nicotinamide often

shows an aromatic side chain and also a hydrogen bond between its amide nitrogen and the main chain oxygen in the residues. Additionally, we use a group of enzymes with structure and known dissociation constant with the objective to optimize the statistical potentials to obtain a better correlation between the pseudo-energies calculated and the experimental data. The pseudo-energies calculated from this optimized statistical potentials correlates with the binding free energy changes in alanine and other residues mutants using crystallographic structures and *in silico* models. Finally, we use the statistical potentials to propose a mutation to change the coenzyme preference of the enzyme glucose-6-phosphate dehydrogenase from *Leuconostoc mesenteroides*. *In vitro* results based on the quantification of the initial velocities of the enzyme at different NAD or NADP concentrations demonstrate that the mutant Ala45Asp change the preference of this enzyme from NADP to NAD.

1. Introducción

Dinucleótidos de nicotinamida adenina como coenzimas

La propiedad de las enzimas de catalizar reacciones químicas con gran eficiencia y especificidad permite el funcionamiento del metabolismo de todos los organismos vivos. En estas reacciones, el sustrato se une al sitio activo de la enzima y mediante un mecanismo de reacción, se transforma en producto. Muchas enzimas, para realizar esta función, requieren de la presencia de co-sustratos especiales denominados coenzimas, que son regenerados en una subsecuente reacción catalizada por otra enzima (Bryce 1979). Dentro de las vías metabólicas de todos los organismos, los dinucleótidos de nicotinamida adenina, NAD(H) y NADP(H), son coenzimas que participan en cientos de reacciones, como se muestra de manera particular para el metabolismo de *Escherichia coli* en la Figura 1 (Orth et al. 2011).

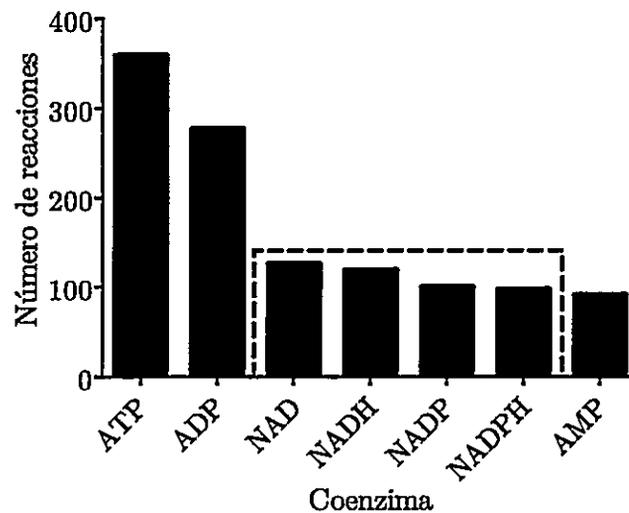


Figura 1. Coenzimas más utilizadas en reacciones catalíticas dentro del metabolismo de *Escherichia coli*. El número de reacciones fue calculado a partir del modelo metabólico iJO1366 de *Escherichia coli*.

Las enzimas dependientes de NAD(P)(H) son en general altamente específicas por su coenzima. La presencia de un grupo fosfato unido al carbono-2' en la ribosa adyacente a la adenina, es la única característica estructural que diferencia a NADP de NAD (Figura 2). Por lo tanto, la capacidad de discriminar entre estas coenzimas es un impresionante ejemplo del poder de reconocimiento molecular exhibido por estas enzimas.

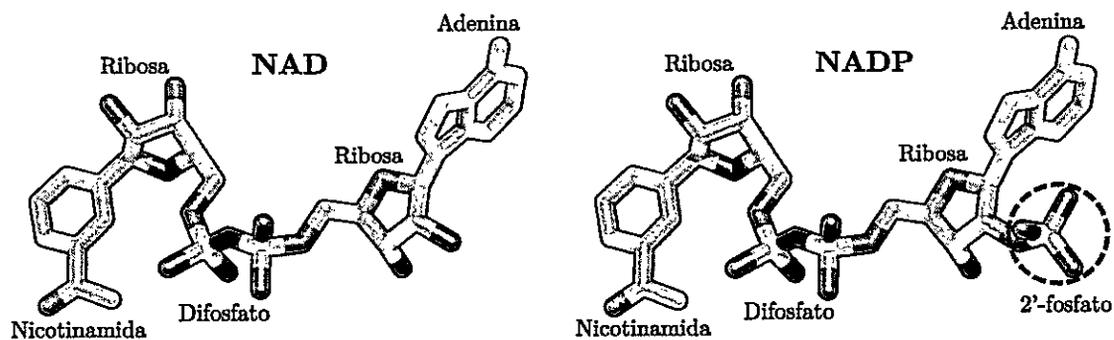


Figura 2. Estructura tridimensional de las coenzimas NAD(P). En cada coenzima se nombran los grupos químicos que componen su estructura.

El reconocimiento de los grupos químicos de la coenzima está determinado por los residuos aminoacídicos que componen el sitio activo de la enzima. Una vez que la coenzima está unida, la suma de las energías de los distintos tipos de interacciones que se forman con la enzima dan como resultado la energía unión del complejo. Las diferencias entre las energías de unión de NAD y NADP a lo largo de la coordenada de la reacción afectarán sus constantes de especificidad, y por lo tanto, la preferencia de la enzima por alguna de estas coenzimas. De esta forma y en relación con su grado especificidad por coenzimas serían posibles tres tipos de enzimas: aquellas que prefieren NADP, aquellas preferentes por NAD y las denominadas duales, que pueden usar NAD o NADP con similar eficiencia. En general las enzimas presentan preferencia por uno u otro nucleótido de nicotinamida adenina, siendo en pocos casos duales, lo que indica que existen

características estructurales claves en el sitio de unión de las enzimas que permiten discriminar por alguna de estas coenzimas.

Función celular de NAD(H) y NADP(H)

La selectividad de las enzimas que unen a estas coenzimas resulta aún más interesante cuando vemos que sus equivalentes reducidos, NAD(P)H, presentan funciones diferentes dentro del metabolismo celular, a pesar de poseer el mismo potencial de óxido-reducción. Mientras NADH transfiere sus hidruros, ya sea a la cadena transportadora de electrones en la respiración celular o hacia aceptores orgánicos terminales en la fermentación, NADPH los transfiere a diferentes aceptores de las rutas biosintéticas (síntesis de lípidos, colesterol, ácidos grasos) o aceptores involucrados en la atenuación del estrés oxidativo. Por otra parte, la selectividad por las formas oxidadas NAD(P) se observa en las deshidrogenasas de las vías centrales, aunque de manera dependiente del organismo. Por ejemplo, en *Escherichia coli*, la Isocitrato deshidrogenasa usa NADP como co-sustrato, mientras que en organismos incapaces de crecer en acetato, se observan enzimas NAD-dependientes (Miller et al. 2006).

En otras reacciones, los dinucleótidos de nicotinamida adenina no participan como coenzimas de óxido-reducción. Por ejemplo, la ADP-ribosilación (Schreiber et al. 2006) y la deacetilación (Liou et al. 2005), donde participa NAD y no NADP. Además de funciones donde ambas coenzimas participan, como ser precursor directo de moléculas involucradas en la movilización intracelular de Ca^{2+} como ADP-ribosa, ADP-ribosa cíclica y O-acetil-ADP-ribosa (en el caso de NAD) (Schuber & Lund 2004) o Acido nicotínico adenina dinucleótido fosfato y ADP-ribosa fosfato cíclica (en el caso de NADP) (Guse & Lee 2008).

Dadas las diferencias mencionadas respecto a las funciones que estas coenzimas presentan dentro del metabolismo celular, desde hace casi 20 años (Lesk 1995) se han venido realizando estudios para identificar los determinantes estructurales de la especificidad por NAD y NADP en las proteínas que los unen. Como se describe a continuación, este conocimiento es particularmente relevante en relación con el diseño racional de la especificidad por estas coenzimas.

Cambios de especificidad por NAD(H) y NADP(H)

A comienzo de la década del 90' se da a conocer el primer estudio enfocado en el cambio de especificidad por estas coenzimas, realizado en la enzima Glutación reductasa de *Escherichia coli*. En este caso se logró cambiar la especificidad de esta enzima desde NADPH a NADH mediante la introducción de siete mutaciones en el sitio de unión de la coenzima (Scrutton et al. 1990). Desde ese entonces, han habido múltiples estudios publicados sobre el rediseño de la especificidad por estas coenzimas en enzimas (Tabla 1) y algunos de ellos han logrado modificar la especificidad con gran eficiencia, como lo demuestra el trabajo realizado sobre la enzima Isocitrato deshidrogenasa de *Escherichia coli* donde se logró cambiar la especificidad de 7000-veces a favor de NADP a 200-veces a favor de NAD mediante la introducción de 6 mutaciones (Chen et al. 1995).

Una de las estrategias comúnmente utilizadas para cambiar la especificidad de estas enzimas es el análisis de las estructuras tridimensionales y/o de secuencia. Esto permite identificar posibles mutaciones que posteriormente son introducidas en la enzima mediante mutagénesis sitio-dirigida y son analizadas mediante

1. INTRODUCCIÓN

ensayos de cinética enzimática. Sin embargo, también se han utilizado otras estrategias como el reemplazo de un segmento compacto de la enzima (Yaoi et al. 1996) o el uso de evolución dirigida (Morimoto et al. 2014), particularmente el esquema semi-racional llamado saturación combinatoria del sitio activo (CASTing) (Liang et al. 2007).

Tabla 1. Resumen de los cambios de especificidad por NAD(P)(H).

| Especie | Enzima | Cambio de Especificidad | Referencia |
|------------------------------------|---|-------------------------|----------------------------|
| <i>Candida tenuis</i> | Xilosa reductasa | NADPH → NADH | (Petschacher et al. 2005) |
| <i>Corynebacterium</i> | Acido 2,5-diceto-D-glucónico | NADPH → NADH | (Banta et al. 2002) |
| <i>Escherichia coli</i> | Glutatión reductasa | NADPH → NADH | (Scrutton et al. 1990) |
| <i>Escherichia coli</i> | Ceto-ácido reductoisomerasa | NADPH → NADH | (Rane & Calvo 1997) |
| <i>Neurospora crassa</i> | Nitrato reductasa | NADPH → NADH | (Shiraishi et al. 1998) |
| <i>Pichia stipitis</i> | Xilosa reductasa | NADPH → NADH | (Liang et al. 2007) |
| <i>Pseudomonas fluorescens</i> | p-hidroxibenzoato hidroxilasa | NADPH → NADH | (Eppink et al. 1999) |
| <i>Rattus norvegicus</i> | Citocromo p450 reductasa | NADPH → NADH | (Elmore & Porter 2002) |
| <i>Saccharomyces cerevisiae</i> | 17β-hidroxiesteroide deshidrogenasa | NADPH → NADH | (Kristan et al. 2007) |
| <i>Sinorhizobium morelense</i> | 1,5-anhidro-D-fructosa reductasa | NADPH → NADH | (Dambe et al. 2006) |
| <i>Anabaena. sp.</i> | Ferredoxina:NADP ⁺ reductasa | NADP → NAD | (Medina et al. 2001) |
| <i>Escherichia coli</i> | Isocitrato deshidrogenasa | NADP → NAD | (Chen et al. 1995) |
| <i>Thermus thermophilus</i> | Isocitrato deshidrogenasa | NADP → NAD | (Yaoi et al. 1996) |
| <i>Vibrio harveyi</i> | Aldehído deshidrogenasa | NADP → NAD | (Zhang et al. 1999) |
| <i>Bacillus stearothermophilus</i> | L-lactato deshidrogenasa | NADH → NADPH | (Holmberg et al. 1999) |
| <i>Rattus norvegicus</i> | Citocromo b5 reductasa | NADH → NADPH | (Marohnic et al. 2003) |
| <i>Thermus thermophilus</i> | β-isopropilmalato deshidrogenasa | NADH → NADPH | (Miller et al. 2006) |
| <i>Bacillus stearothermophilus</i> | D-lactato deshidrogenasa | NAD → NADP | (Bernard et al. 1995) |
| <i>Bacillus stearothermophilus</i> | G3P-deshidrogenasa | NAD → NADP | (Clermont et al. 1993) |
| <i>Gluconobacter oxydans</i> | Xilitol deshidrogenasa | NAD → NADP | (Ehrensberger et al. 2006) |
| <i>Homo sapiens</i> | Enzima malica | NAD → NADP | (Hsieh et al. 2006) |
| <i>Pichia stipitis</i> | Xilitol deshidrogenasa | NAD → NADP | (Watanabe et al. 2005) |
| <i>Pseudomonas stutzeri</i> | Fosfito deshidrogenasa | NAD → NADP | (Woodyer et al. 2003) |
| <i>Saccharomyces cerevisiae</i> | Formato deshidrogenasa | NAD → NADP | (Serov et al. 2002) |
| <i>Thermus thermophilus</i> | Isopropilmalato deshidrogenasa | NAD → NADP | (Chen et al. 1996) |
| <i>Tramitichromis intermedius</i> | Leucina deshidrogenasa | NAD → NADP | (Galkin et al. 1997) |

A pesar de los casos listados en la Tabla 1, a la fecha no existe una herramienta capaz de proponer mutaciones para cambiar la especificidad por estas coenzimas en las enzimas, basándose en los rasgos generales de los bolsillos de unión de estos ligandos. En este trabajo se pretende crear una herramienta para identificar residuos y las posiciones que éstos deberían tener dentro del bolsillo de unión, para incrementar la energía de unión de una coenzima sobre otra y así poder modificar la especificidad.

Base de datos de estructuras de proteínas

Desde 1971 las estructuras cristalográficas de proteínas y ácidos nucleicos han sido depositadas en una base de datos de dominio público denominada Protein Data Bank (PDB), creada por investigadores del Laboratorio Nacional Brookhaven en Estados Unidos (Bernstein et al. 1977). Debido a los avances tecnológicos como la creación de aceleradores de partículas, equipos automatizados de cristalización y programas computacionales, esta base de datos ha logrado un crecimiento acelerado, llegando hoy en día a tener más de 100.000 estructuras depositadas. NAD(P) son unas de las coenzimas más comunes dentro de las estructuras presentes en esta base de datos (Figura 3), lo que abre la

posibilidad de utilizar esta información para encontrar los determinantes del reconocimiento y selectividad por estos dinucleótidos de nicotinamida adenina. De hecho, esta base de datos fue analizada por Carugo & Argos en 1997, quienes en dos publicaciones sucesivas entregaron luces sobre las interacciones conservadas entre proteínas y estas coenzimas, así como la variabilidad conformacional que estas últimas tendrían dentro del sitio de unión (Carugo & Argos 1997a; Carugo & Argos 1997b). Sin embargo, en ese entonces el número de estructuras cristalográficas disponible era limitado, analizándose un grupo de 21 estructuras con NAD y 11 estructuras con NADP.

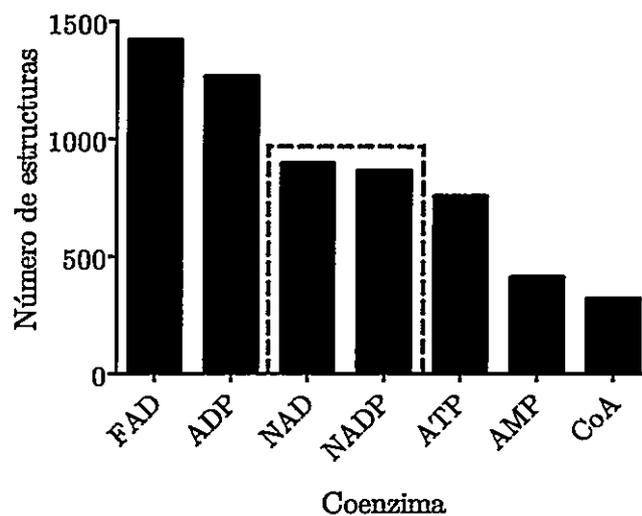


Figura 3. Coenzimas más frecuentes dentro de las estructuras de la base de datos PDB.

Posteriormente, varios investigadores se avocaron a identificar los residuos que permitían unir a estas coenzimas. Es así como se lograron identificar motivos

asociados a la unión de diferentes regiones de la coenzima como la adenina (Denessiouk et al. 2001), el difosfato (Kleiger & Eisenberg 2002) y la ribosa adyacente a la nicotinamida (Kallberg et al. 2002). A pesar de estos avances, estos trabajos estuvieron enfocados en determinar estos motivos en familias o plegamientos de proteínas específicos, sin hacer uso de toda la información estructural disponible.

Otras estrategias como la predicción de sitios de unión de estas coenzimas también han sido abordadas previamente, ya sea utilizando máquinas de vectores de soporte integrados con matrices de posición específicas (Ansari & Raghava 2010) o mediante la comparación de interacciones entre fragmentos de proteínas (Kasahara et al. 2010). Estos estudios y los algoritmos utilizados, sin embargo, son verdaderas cajas negras y no permiten la extracción de las características estructurales presentes en los sitios de unión.

Hoy en día, el número de estructuras de proteínas como complejo con estas coenzimas se ha incrementado en más de 50 veces desde el primer estudio de Carugo & Argos en 1997. Por lo tanto, actualmente es posible utilizar

aproximaciones matemáticas basadas en la estadística, para transformar frecuencias de interacciones entre los átomos de las coenzimas y las proteínas, en magnitudes relacionadas con energía, denominadas pseudo-energías. Esta aproximación se denomina potenciales estadísticos y las pseudo-energías implícitamente integran todos los componentes de la energía de interacción. Estas frecuencias y pseudo-energías, al ser calculadas en base a todos los complejos proteína-coenzima presentes en la base de datos PDB, permitirán a su vez extraer las características estructurales y energéticas que estos bolsillos de unión presentan de manera general en las proteínas.

Potenciales Estadísticos

El primer trabajo desarrollado con el propósito de calcular estadísticas sobre estructuras cristalográficas, relacionaba la frecuencia de interacción entre distintos tipos de residuos con sus energías, utilizando relaciones simples entre las pseudo-energías y las constantes de equilibrio (Tanaka & Scheraga 1976). A este estudio le sucedió el desarrollo de un potencial estadístico de contacto entre residuos usando aproximaciones denominadas cuasi-químicas (Miyazawa & Jernigan 1985). Posteriormente, en 1990 se publica una nueva forma de

potencial estadístico dependiente de la distancia entre carbonos- α , esta vez basada en la suposición de que las distribuciones de distancia de diferentes tipos de residuos son de tipo Boltzmann (Sippl 1990). Utilizando esta aproximación, en un gran número de subsecuentes trabajos, se han descrito diferentes potenciales estadísticos evaluados y validados con diversos propósitos. Por ejemplo, la predicción de errores en modelos de proteínas (Melo & Feytmans 1997) y la determinación de energías de unión para complejos proteína-proteína (Jiang et al. 2002) o proteína-ligando (Mitchell et al. 1999).

Por lo anteriormente mencionado, en este trabajo proponemos el cálculo de potenciales estadísticos entre las proteínas y los dinucleótidos de nicotinamida adenina, basados en todas las estructuras disponibles con estas coenzimas en la base de datos PDB. Esto con el objetivo de que puedan ser utilizados para describir las características estructurales y energéticas de los sitios de unión de estas coenzimas, así como una herramienta para determinar cambios en la energía libre asociados a mutaciones y de esta manera identificar mutaciones que pueden cambiar la especificidad por estas coenzimas en enzimas.

1.1 Hipótesis

Las frecuencias de interacción entre NAD(P) y proteínas pueden transformarse en pseudo-energías que se correlacionan con parámetros determinados experimentalmente para estos complejos



1.2 Objetivos

1.2.1 Objetivo general

Derivar potenciales estadísticos a partir de estructuras de proteínas como complejo con NAD(P) para calcular pseudo-energías y correlacionarlas con parámetros experimentales como energías de unión.

1.2.2 Objetivos específicos

1. Generar bases de datos de estructuras de proteínas con las coenzimas NAD o NADP para el entrenamiento, evaluación y validación de los potenciales estadísticos:
 - a. Base de datos de entrenamiento: Estructuras de proteínas a partir de las cuales se derivarán los potenciales estadísticos.
 - b. Base de datos de evaluación: Estructuras utilizadas para determinar los parámetros con que son calculados los potenciales estadísticos que entregan una mejor correlación entre datos teóricos y experimentales.
 - c. Base de datos de validación: Estructuras o modelos *in silico* de enzimas silvestres y mutantes con la especificidad por coenzimas modificada para

1. INTRODUCCIÓN

validar la capacidad de los potenciales estadísticos de representar el cambio en la energía libre en mutantes.

2. Derivar potenciales estadísticos de los sitios de unión de NAD o NADP.
 - a. Cuantificar en la base de datos de entrenamiento la frecuencia de interacciones entre tipos de átomos presentes en los residuos de las proteínas y en las coenzimas.
 - b. Transformar las frecuencias de interacciones en pseudo-energías.
3. Evaluar los potenciales estadísticos mediante su correlación con datos experimentales.
 - a. Calcular la pseudo-energía de los sitios de unión en las estructuras de la base de datos de evaluación.
 - b. Correlacionar las pseudo-energías para los sitios de unión con las energías libres determinadas experimentalmente para estos complejos.
4. Validar la capacidad de los potenciales estadísticos de representar el efecto de mutaciones sobre la preferencia por dinucleótidos de nicotinamida adenina.

1. INTRODUCCIÓN

- a. Correlacionar la pseudo-energía de los sitios de unión de las estructuras en la base de datos de validación con parámetros experimentales.
- b. Proponer una mutación que cambie la preferencia por coenzima en la enzima glucosa-6-fosfato deshidrogenasa de *Leuconostoc mesenteroides*.

2. Materiales y Métodos

2.1 Bases de datos

Durante la creación de los potenciales estadísticos es necesario realizar tres procesos: El entrenamiento, durante el cual se derivan los potenciales estadísticos. La evaluación, mediante la que se determinan los parámetros que maximizan la correlación con datos experimentales y finalmente, la validación, donde se pone a prueba la capacidad predictiva de los potenciales estadísticos. Cada uno de estos procesos requiere el uso de una base de datos independiente, lo que implica que estructuras utilizadas para el entrenamiento, no sean utilizadas posteriormente en la evaluación o validación.

Base de datos de entrenamiento: Se realizó una búsqueda en la base de datos PDB de aquellas estructuras de rayos-x co-cristalizadas con las coenzimas NAD(H) o NADP(H), mediante la realización de una búsqueda avanzada¹ utilizando como parámetro de búsqueda el identificador químico correspondiente a las coenzimas (Tabla 2). A pesar de que cada coenzima puede indicarse por

¹ <http://www.pdb.org/pdb/search/advSearch.do>

2. MATERIALES Y MÉTODOS

dos nombres, uno para su estado reducido y otro para el oxidado, en este escrito utilizaremos NAD para referirnos a NAD y NADH, y NADP para referirnos a NADP y NADPH.

Tabla 2. Identificador químico de los dinucleótidos de nicotinamida adenina y número de estructuras con estas coenzimas, en las base de datos PDB.

| Coenzima | Identificador Químico | Número de estructuras |
|----------|-----------------------|-----------------------|
| NAD | NAD | 890 |
| NADH | NAI | 132 |
| NADP | NAP | 816 |
| NADPH | NDP | 505 |

Para el caso de residuos o coenzimas con posiciones alternativas dentro de las estructuras, se eliminaron aquellas con la menor ocupancia y en caso de que tuvieran la misma ocupancia, se eliminaron aquellas con mayor factor- β . Esto se llevó a cabo utilizando el módulo DockPrep presente en el programa Chimera 1.9 (Pettersen et al. 2004). Este módulo también fue utilizado para modificar los residuos de selenio-metionina por metionina, al cambiar el átomo de selenio por un átomo de azufre y corrigiendo los largos de enlace entre el carbono- γ /azufre y carbono- ϵ /azufre a 1,81 y 1,78 Å, respectivamente.

2. MATERIALES Y MÉTODOS

Posteriormente, tanto de la base de datos de entrenamiento de NAD como de NADP, se eliminaron aquellas estructuras con menos de 3 Å de resolución y se conservaron solo aquellas estructuras que tuvieran la coenzima completa, o sea con 44 átomos en el caso de NAD y 48 átomos en el caso de NADP, y en una conformación extendida, vale decir, con una distancia entre el carbono-6 de la adenina y el carbono-2 de la nicotinamida mayor a 12 Å (Kuppuraj et al. 2011).

Para no tener estructuras redundantes, se agruparon en cada una de las bases de datos las estructuras con más de 70% de identidad de secuencia y se seleccionó como representante de cada grupo la estructura con la mejor resolución cristalográfica. Adicionalmente, se excluyeron las estructuras con enlaces covalente entre la proteína y la coenzima, junto a aquellas estructuras utilizadas para evaluar y validar posteriormente los potenciales estadísticos.

Base de datos de evaluación: Para evaluar los potenciales estadísticos se creó una base de datos de proteínas con estructura conocida y constante de disociación determinada para la coenzima presente en la estructura, para ello se buscó los casos que cumplieran con estas características dentro de la base de

datos Binding MOAD (Hu et al. 2005), cuyo objetivo es tener una colección curada manualmente de estructuras de enzimas, junto a los datos experimentales para los ligandos que une. Del total de casos encontrados, se seleccionaron aquellos con una estructura a una resolución mayor o igual a 3,0 Å que tuvieran la coenzima con todos sus átomos y en una conformación extendida. Además se excluyeron las estructuras que presentaran enlaces covalentes entre la proteína y la coenzima.

Base de datos de validación: Para validar los potenciales estadísticos se creó una base de datos de estructuras de enzimas silvestres y mutantes con estructura conocida y constantes de disociación determinadas para la coenzima en la estructura. En caso de no presentar estructura las mutaciones fueron modeladas con MODELLER 9.13 (Eswar et al. 2007) y posteriormente optimizadas mediante protocolos de minimización y simulación molecular (Feyfant et al. 2007). Además de estas estructuras, se utilizó la estructura de la glucosa-6-fosfato deshidrogenasa de *Leuconostoc mesenteroides* (Naylor et al. 2001) para validar la capacidad de los potenciales estadísticos de proponer mutaciones que invirtieran la especificidad en esta enzima desde NADP a NAD preferente.

2.2 Potenciales Estadísticos

Este proyecto propone el desarrollo de potenciales estadísticos dependientes de la distancia incorporando interacciones no-locales. Se define como ambiente no-local de un átomo de la coenzima al grupo átomos pertenecientes a la proteína dentro de una distancia Euclidiana determinada (distancia de corte), a su vez dividido en intervalos de distancia de 0,1; 0,2; 0,5; 1,0 o 2,0 Å (tamaño de ventana) (Figura 4).

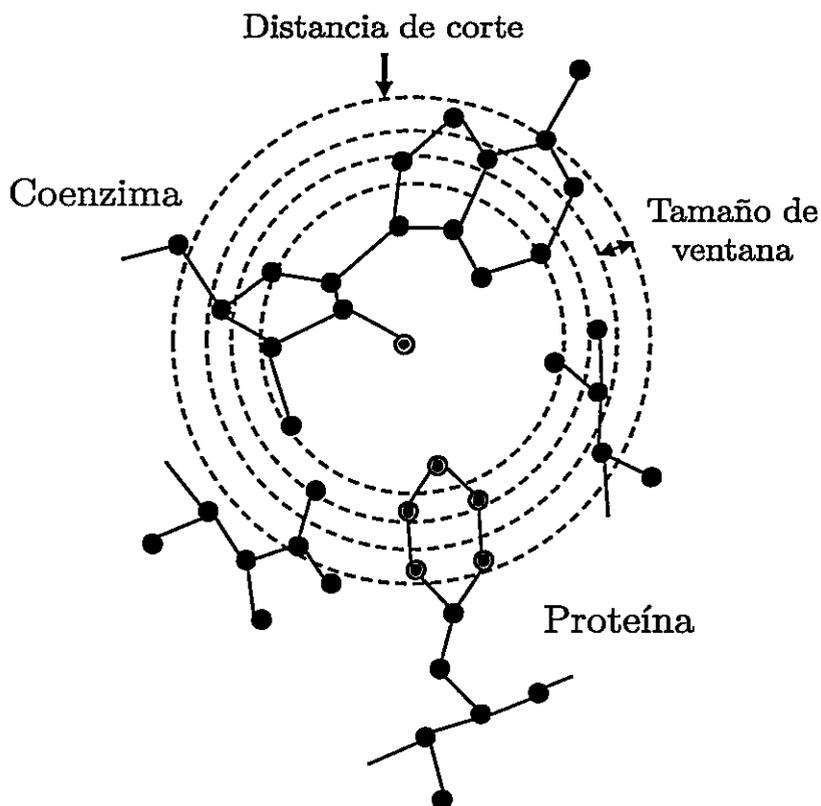


Figura 4. Esquema de interacciones no-locales entre átomos de la coenzima y la proteína a una distancia de corte y tamaño de ventana determinados. El círculo azul representa un átomo de la coenzima de un tipo determinado y los círculos rojos representan átomos de la proteína de tipo carbono-aromático a una distancia de corte determinada.

2. MATERIALES Y MÉTODOS

Para poder relacionar la frecuencia de observación de las interacciones no-locales con su energía, lo primero a considerar es que la frecuencia relativa (f) de una observación es equivalente a su probabilidad (P) cuando el número de observaciones tiende al infinito.

$$\lim_{n \rightarrow \infty} f = P \quad (1)$$

Para entender mejor esto, debemos recordar que si en un experimento se repite n veces, de las cuales se ha observado un determinado suceso A en k de estas n repeticiones, la frecuencia relativa (f) del suceso A es:

$$f = k/n \quad (2)$$

Así por ejemplo, si el suceso A fuera que al lanzar una moneda al aire esta caiga cara, la probabilidad de que esto sucediera en una moneda equilibrada sería de $\frac{1}{2}$ es decir 0,5. Por otra parte, si simulamos computacionalmente el lanzamiento de una moneda 1.000 veces podemos ver que el valor de la frecuencia relativa a un mayor número de lanzamientos tiende a 0,5 (Figura 5), lo que demuestra el postulado inicial de que a un número de observaciones grande la frecuencia relativa es equivalente a la probabilidad.

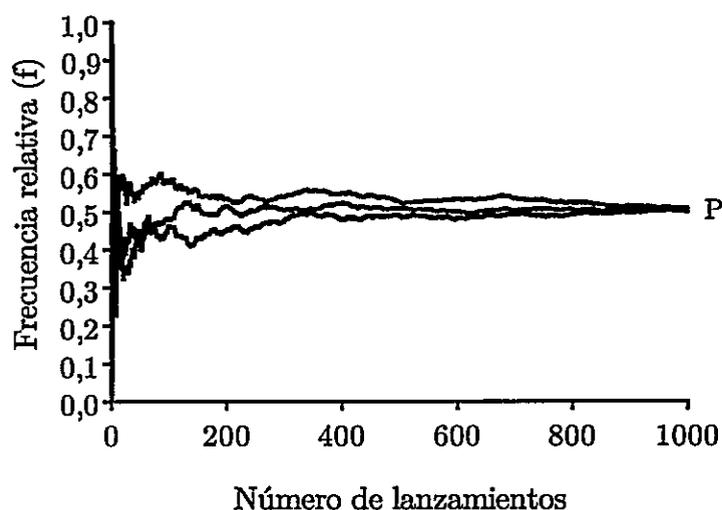


Figura 5. Frecuencia relativa para obtener una cara en tres simulaciones computacionales del lanzamiento de una moneda 1.000 veces.

Por lo tanto, la probabilidad con la que dos átomos estén interactuando a una cierta distancia r es posible calcularla a partir de la frecuencia relativa de observaciones entre estos pares de átomos a esa distancia. Esto siempre que el número de observaciones sea lo suficientemente grande, como el que podemos obtener de un grupo de cientos de estructuras presentes en la base de datos PDB.

El factor de Boltzmann derivado por primera vez por Boltzmann en 1868 (Boltzmann 1896) indica que la energía de un estado ($E(r)$), como puede ser la interacción de dos átomos a una cierta distancia, está relacionada con la

probabilidad de ese estado de estar ocupado ($P(r)$), vale decir, con la frecuencia relativa de esa interacción a esa distancia dentro de la base de datos ($f(r)$).

$$P(r) = \frac{1}{Z} \exp\left(\frac{E(r)}{kT}\right) = f(r) \quad (3)$$

Donde Z (del alemán Zustandssumme o "suma sobre lo estados") es la función de partición, que se define como la sumatoria de la energía de todos los estados posibles. k es la constante de Boltzmann (0,001987 kcal/mol/K) y T es la temperatura absoluta (K).

$$Z = \sum_{r=0}^n \exp\left(\frac{-E(r)}{kT}\right) \quad (4)$$

Integrando la ecuación de la función de partición en la ecuación que define la frecuencia relativa de una interacción, podemos obtener una forma linealizada de la energía de un estado en función de su frecuencia y la función de partición.

$$E(r) = -kT \ln f(r) - kT \ln Z \quad (5)$$

Dado que los átomos en las proteínas poseen características fisicoquímicas distintas que tienen diferentes efectos en el proceso de plegamiento, es mejor separarlos en funciones de energías independientes en vez de utilizar $f(r)$ que es un promedio. La solución a esto es usar diferentes $f(r)$ para diferentes tipos de átomos con propiedades fisicoquímicas similares (en mayor detalle en la sección 2.2.1). Así se define por ejemplo que $f^{ij}(r)$ es la probabilidad de que dos tipos de átomo i y j estén dentro de un intervalo de distancia r . Desde ahora en adelante el superíndice "ij" en los términos se refiere a un par de tipos de átomos específicos y términos sin este superíndice se referirán a un promedio.

Dada las consideraciones anteriores, la contribución de energía de un par individual $E^{ij}(r)$ al promedio de la energía $E(r)$ es llamado potencial neto y corresponde a:

$$\Delta E^{ij}(r) = E^{ij}(r) - E(r) \quad (6)$$

Si reemplazamos la función linealizada de la energía en los términos $E^{ij}(r)$ y $E(r)$ obtenemos que:

$$\Delta E^{ij}(r) = kT \ln \left(\frac{f^{ij}(r)}{f(r)} \right) - kT \ln \left(\frac{Z^{ij}(r)}{Z(r)} \right) \quad (7)$$

Dado que $Z(r)$ es en realidad un promedio de los diferentes $Z^u(r)$. Si la varianza de $Z^u(r)$ sobre $Z(r)$ no es muy grande entonces $Z^u(r)$ debería ser cercano a $Z(r)$. Es decir $Z^u(r) \approx Z(r)$ lo que implica que el cociente entre estos valores es cercano a 1, lo que hace la parte derecha de la función anterior sea cercana a 0, y por lo tanto pueda no ser considerada. Bajo esta suposición, el potencial neto de un par de tipos de átomos a una cierta distancia, puede ser calculado en función de su frecuencia relativa en la base de datos, respecto a la frecuencia relativa promedio en la base de datos entre cualquier par de tipos de átomos que se encuentren a esa distancia.

$$\Delta E^{ij} = kT \ln \left(\frac{f^{ij}(r)}{f(r)} \right) \quad (8)$$

Como se mencionó en un comienzo, las aproximaciones anteriores son satisfactorias si la base de datos de entrenamiento es suficientemente grande. Sin embargo, incluso si la base de datos de entrenamiento es grande, el número de

pares específicos de átomos puede no ser lo suficientemente grande para entregar resultados estadísticamente válidos. Para minimizar la posibilidad que esto ocurra, se debe separar la frecuencia relativa ideal de un par de tipos de átomos ($f^{ij}(r)$) de las frecuencias relativa de un par de tipos de átomos obtenida de una base de datos ($g^{ij}(r)$), considerando que en caso de que la base de datos sea suficientemente grande una buena aproximación sea que $f^{ij}(r)$ es equivalente a $g^{ij}(r)$. Una aproximación que incluye esta característica fue publicada por Sippl en 1990 y considera que $f^{ij}(r)$ en un inicio tiene un valor cercano al promedio de las frecuencias relativas ($f(r)$), y que a medida que el número de observaciones se incrementa se va dando más efecto a $g^{ij}(r)$.

$$f^{ij}(r) \approx \frac{1}{1 + M_{ij}\sigma} f(r) + \frac{M_{ij}\sigma}{1 + M_{ij}\sigma} g^{ij}(r) \quad (9)$$

$$M_{ij} = \sum_{r=0}^n f(i, j, r) \quad (10)$$

Donde M_{ij} es el número de observaciones del par de tipos de átomos i y j . Así, para valores pequeños de M_{ij} esta aproximación tiende hacia $f(r)$, mientras que si M_{ij} tiende al infinito la aproximación tiende hacia $g^{ab}(r)$ (Figura 6).



2. MATERIALES Y MÉTODOS

Por otra parte, σ es el peso estadístico que representa el número de observaciones necesario para que las variables $g^{ij}(r)$ y $f(r)$ tengan el mismo peso dentro de la función. Así por ejemplo, si utilizamos un peso estadístico de 0,02 o $1/50$ esto significa que para 50 observaciones las variables $g^{ij}(r)$ y $f(r)$ tendrán el mismo peso, es decir los términos que acompañan a $g^{ij}(r)$ y $f(r)$ se hace 0,5 (Figura 6). Sobre este número de observaciones la función tiende hacia $g^{ij}(r)$, es decir, hacia valores extraídos directamente de la base de datos, y si el número de observaciones es inferior a 50 observaciones la función tendera hacia $f(r)$, vale decir, al promedio de las frecuencias relativas para todas las interacciones.

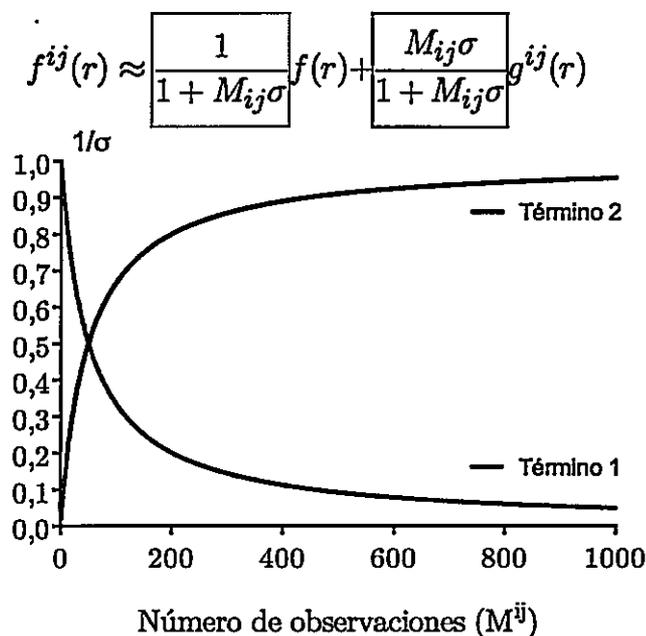


Figura 6 Aproximación de Sipl para la función $f^{ij}(r)$. Las líneas azul y roja muestran los valores que toman en la aproximación los términos que acompañan a $f(r)$ y $g^{ij}(r)$, respectivamente. Se señala en el gráfico mediante una línea punteada el numero de observaciones a la que los términos tienen el mismo peso utilizando un $\sigma=0,02$.

Finalmente, integrando la aproximación de Sippl dentro de la fórmula para el cálculo del potencial neto de un par de tipos de átomos, se obtiene una fórmula que permite relacionar las frecuencias calculadas a partir de estructuras con energías.

$$\Delta E^{ij}(r) = kT \ln (1 + M_{ij}\sigma) - kT \ln \left(1 + M_{ij}\sigma \frac{g^{ij}(r)}{f(r)} \right) \quad (11)$$

En esta fórmula, $g^{ij}(r)$ es la frecuencia relativa del par de tipos de átomos i y j a una distancia específica r . Es decir, el número de observaciones para el par de tipos de átomos i y j a una distancia específica, respecto al número de observaciones para ese mismo par de tipos de átomos i y j a cualquier distancia (M_{ij}).

$$g^{ij}(r) = \frac{f(i, j, r)}{M_{ij}} \quad (12)$$

Por otra parte, $f(r)$ es la frecuencia relativa de cualquier par de tipos de átomos a una distancia específica r , es decir, el número de observaciones para cualquier par de tipos de átomos a una distancia específica, respecto al número de observaciones para cualquier par de tipos de átomos a cualquier distancia.

$$f(r) = \frac{\sum_{i=0}^n \sum_{j=0}^n f(i, j, r)}{\sum_{i=0}^n \sum_{j=0}^n \sum_{r=0}^n f(i, j, r)} \quad (13)$$

Para entender mejor como es que se aplican estas fórmulas podemos ver los perfiles que tienen estas funciones (Figura 7). Dado que $g^{ij}(r)$ representa la frecuencia relativa de un par de tipos de átomos específicos en las estructuras de la base de datos. Esta función tiene comúnmente un perfil que va en incremento a mayor distancia, dado que a mayor distancia de un átomo mayor es la cantidad de átomos presentes, y poseerá un máximo a la distancia donde es más probable encontrar este par de átomos interactuando (Figura 7-a). Por otra parte, $f(r)$ que es la frecuencia relativa de observar cualquier par de tipo de átomos i y j , esta tiene un perfil que va en incremento lineal a medida que la distancia aumenta (Figura 7-b). Cuando hacemos la razón de los valores de $g^{ij}(r)$ y $f(r)$ obtenemos una función que representa la diferencia en número de veces entre la frecuencia relativa de observar un par de tipos de átomos i y j específicos, respecto a la frecuencia promedio de una observación. Por lo tanto, si este par de tipo de átomos es más frecuente observarlo que el promedio a una distancia determinada, tendrá un cociente superior a cero en dicha distancia

(Figura 7-c). Al integrar este cociente en la formula de Sippl da como resultado una función de energía determinada por la distancia a la que los tipos de átomos i y j se encuentran. Dicha función tendrá un valores positivos a las distancias donde no se observan interacciones, seguido de un mínimo energético a la distancia donde dichas interacciones son más frecuentes observarlas, respecto al promedio, para finalizar en una función que tiende a 0 para distancias donde el número de observaciones entre los tipos de átomos i y j es similar al promedio observado para cualquier par de tipos de átomos (Figura 7-d). Esto último debido a que si las frecuencias relativas $g^{ij}(r)$ y $f(r)$ son similares, el cociente entre estos es cercano 1 lo que integrado en la formula de Sippl da una energía cercana a 0.

A pesar de que esta aproximación es estadística en naturaleza y la coherencia con interacciones físicas reales no es aun clara (por eso es que son llamadas pseudo-energías), su utilidad y consistencia ha sido probada en múltiples publicaciones (Jones et al. 1992; Sippl 1993a; Sippl 1993b).

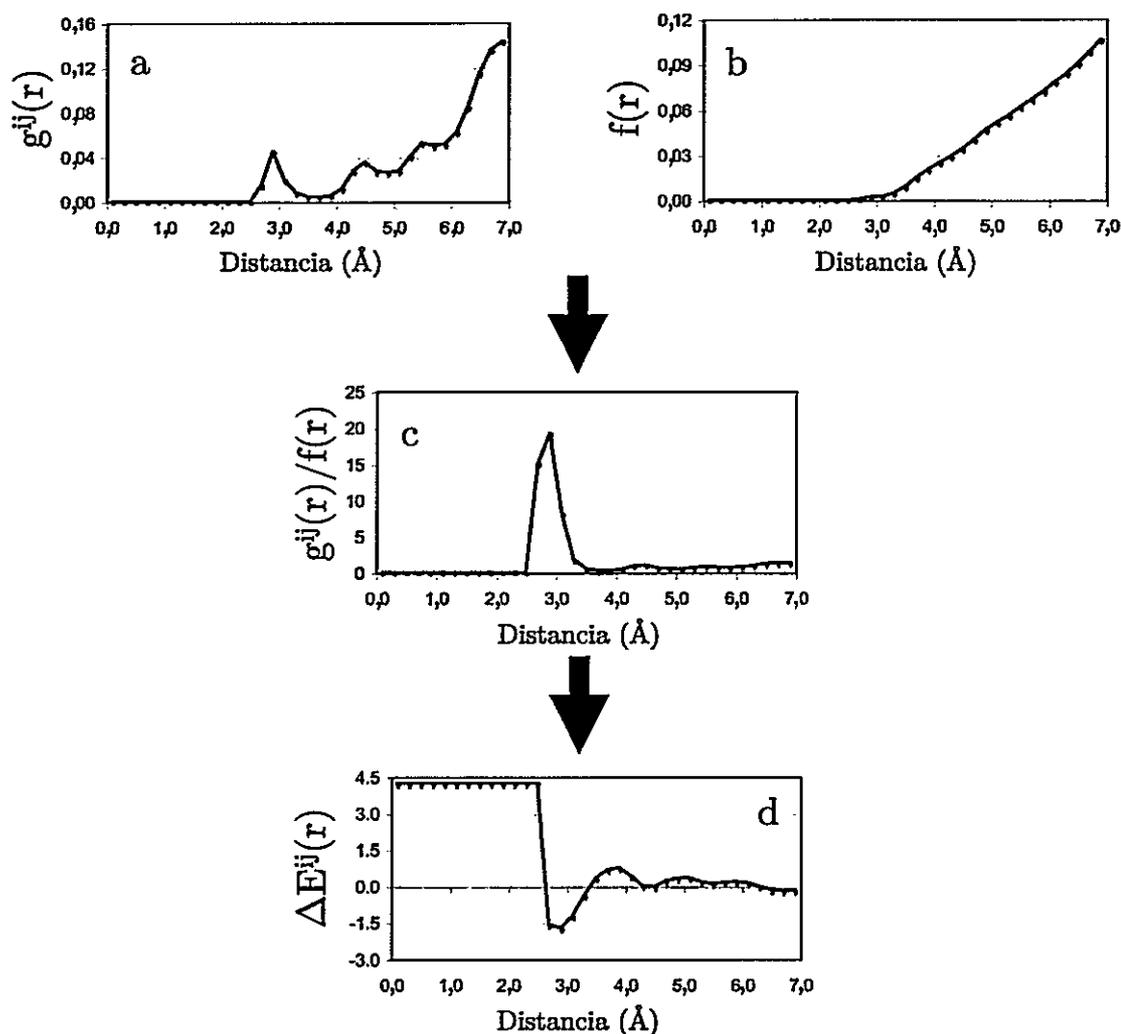


Figura 7. Perfiles de pseudo-energía para las variable calculadas durante el desarrollo de los potenciales estadísticos. Los perfiles están calculados utilizando una distancia de corte de 7,0 Å con un tamaño de ventana de 0,2 Å. Una descripción detallada de cada gráfico puede ser encontrada en las Páginas 30 y 31.

2.2.1 Tipos de átomos

Como mencionamos en la sección anterior todos los átomos tienen propiedades fisicoquímicas distintas, por lo que no pueden ser considerados como diferentes

2. MATERIALES Y MÉTODOS

durante el cálculo de los potenciales estadísticos, pero tampoco podemos considerar a cada átomo como independiente (160 tipos átomos en los 20 aminoácidos y 44-48 tipos de átomos en la coenzima) dado que estas aproximaciones matemáticas precisan de un número de frecuencias cuantificadas lo más grande posible para que el cálculo sea estadísticamente significativo. La solución a esto es que tanto los átomos en la coenzima como en la proteína con propiedades fisicoquímicas similares sean agrupados en tipos de átomos. En este trabajo, los tipos de átomos en la proteína y la coenzima fueron asignados según la definición SYBYL (Clark et al. 1989) la cual los clasifica según sus propiedades fisicoquímicas y conectividad. Se utilizaron 12 tipos de átomos para los 20 aminoácidos proteínogénicos (Figura 8) y 10 tipos de átomos para las coenzimas (Figura 9).

2. MATERIALES Y MÉTODOS

| Nº | Código | Definición |
|----|--------|-------------------------------------|
| 1 | O.2 | Oxígeno sp ² |
| 2 | O.3 | Oxígeno sp ³ |
| 3 | N.2 | Nitrógeno sp ² |
| 4 | N.am | Nitrógeno de amida |
| 5 | O.co2 | Oxígeno de carboxilo |
| 6 | N.pl3 | Nitrógeno trigonal planar |
| 7 | C.cat | Carbono unido a N+ |
| 8 | N.4 | Nitrógeno sp ³ cargado + |
| 9 | C.ar | Carbono aromático |
| 10 | C.2 | Carbono sp ² |
| 11 | C.3 | Carbono sp ³ |
| 12 | S.3 | Azufre sp ³ |

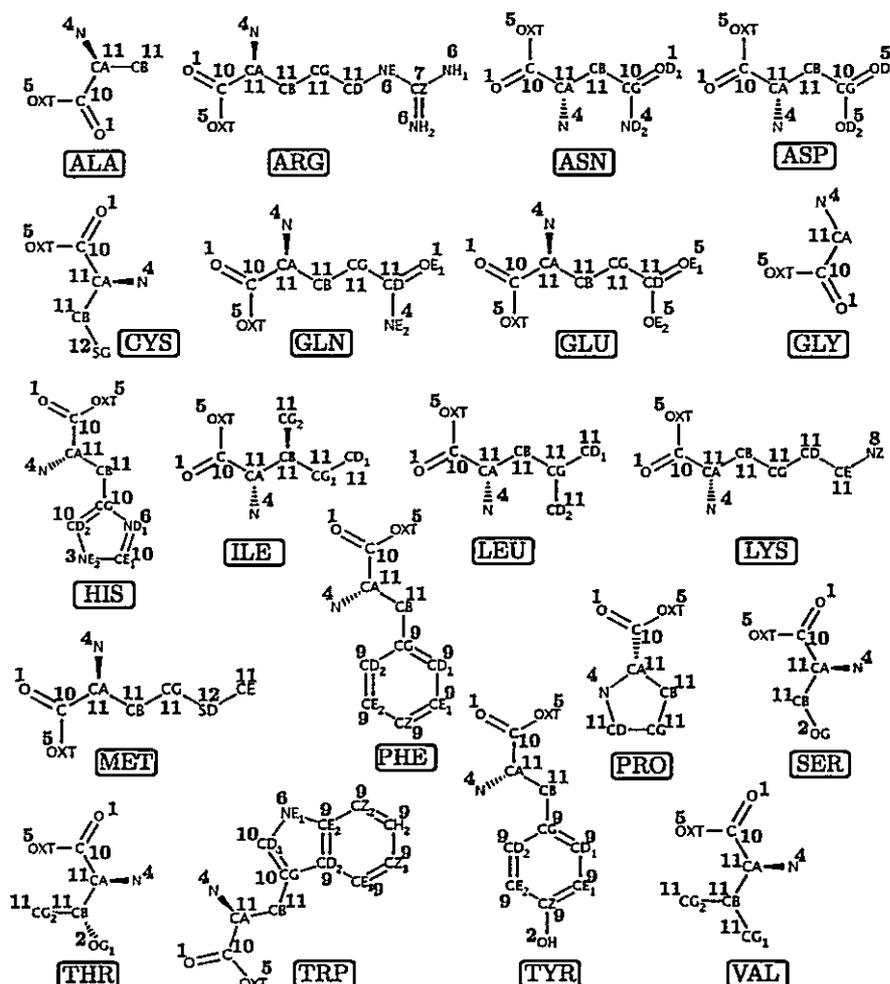


Figura 8. Definición de tipos de átomos SYBYL para los 20 aminoácidos proteínogénicos. En la tabla se identifica el número que le corresponde a cada definición de tipo de átomo, el cual es asignado en las estructuras moleculares de los aminoácidos.

2. MATERIALES Y MÉTODOS

| Nº | Código | Definición |
|----|--------|---------------------|
| 1 | O.2 | Oxígeno sp^2 |
| 2 | O.3 | Oxígeno sp^3 |
| 3 | N.3 | Nitrógeno sp^3 |
| 4 | N.am | Nitrógeno de amida |
| 5 | O.co2 | Oxígeno de fosfatos |
| 6 | P.3 | Fosforo sp^3 |
| 7 | N.ar | Nitrógeno aromático |
| 8 | C.ar | Carbono aromático |
| 9 | C.2 | Carbono sp^2 |
| 10 | C.3 | Carbono sp^3 |

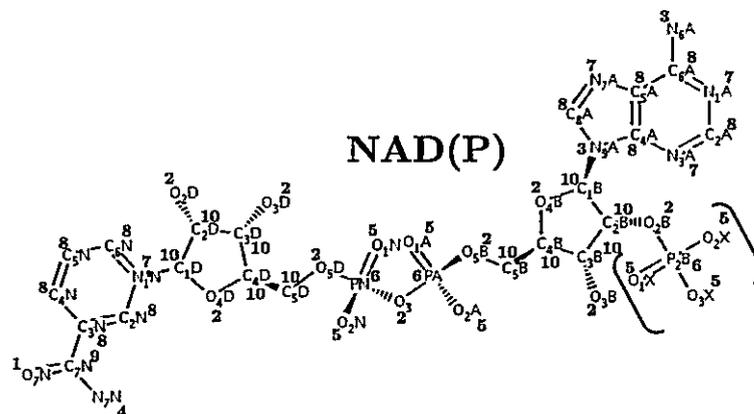


Figura 9. Definición de tipos de átomos SYBYL para los átomos de las coenzimas NAD(P). En la tabla se identifica el número que le corresponde a cada definición de tipo de átomo, el cual es asignado en las estructuras moleculares de las coenzimas.

2.3 Caracterización de los sitios de unión de NAD(P)

Con el propósito de identificar aquellos residuos aminoácidos más frecuentes y su ubicación dentro de los sitios de unión de NAD(P), cuantificamos el porcentaje de estructuras dentro de la base de datos de entrenamiento, que tuvieran tipos de átomos asociados a aminoácidos proteínogénicos interactuando con diferentes regiones de la coenzima.

Como se mencionó en el párrafo anterior los tipos de átomos de la proteína fueron utilizados como identificadores de los residuos aminoácidos presentes, así por ejemplo la presencia de un átomo de nitrógeno sp^3 cargado positivamente en la proteína indica la presencia de una lisina. Otro ejemplo es que la presencia

2. MATERIALES Y MÉTODOS

de oxígeno sp^3 indica la presencia de un residuo de treonina, serina o tirosina (Tabla 3). Para facilitar la comprensión del análisis la coenzima fue dividida en 5 regiones: adenina, ribosa adyacente a la adenina, difosfato, ribosa adyacente a la nicotinamida y nicotinamida.

Tabla 3. Tipos de átomos de la proteína asociados a aminoácidos proteínogénicos.

| Tipo de átomo | Definición | Aminoácidos |
|---------------|----------------------------|---------------|
| O.3 | Oxígeno sp^3 | Thr, Ser, Tyr |
| N.2 | Nitrógeno sp^2 | His |
| O.co2 | Oxígeno de carboxilo | Asp, Glu |
| C.cat | Carbono unido a N^+ | Arg |
| N.4 | Nitrógeno sp^3 cargado + | Lys |
| C.ar | Carbono aromático | Trp, Tyr, Phe |
| S.3 | Azufre sp^3 | Met, Cys |

Para cada uno de los aminoácidos interactuando con las diferentes regiones se asoció el porcentaje de estructuras correspondiente a la distancia máxima a la que estas interacciones puedan formarse (4 Å para puentes de hidrógeno y 6 Å para interacciones electrostáticas o aromáticas).

2.4 Evaluación y validación de los potenciales estadísticos

Utilizando los potenciales estadísticos es posible asignar a cada uno de los átomos de una proteína la pseudo-energía que aportará a la unión de la coenzima. Esta pseudo-energía dependerá del tipo de átomo al que pertenece y con qué tipos de átomos de la coenzima se encuentra a una distancia determinada. La sumatoria de la pseudo-energía de todos los átomos del sitio de unión de la coenzima a un radio determinado es lo que llamamos la pseudo-energía del sitio de unión.

$$Pseudo-energía = \sum_d^{\text{radio de corte}} \sum_i^n \sum_j^n \Delta E^{ij}(r) \quad (14)$$

El parámetro experimental con que se comparará esta pseudo-energía del sitio de unión será la energía libre calculada a partir de constante de disociación.

$$\Delta G^0 = RT \ln K_D \quad (15)$$

Se calculó la pseudo-energía de los sitios de unión de la coenzima en las proteínas de la base de datos de evaluación utilizando potenciales estadísticos obtenidos a radios de corte entre 3 y 20 Å y tamaños de ventana de 0,1, 0,2, 0,5,

2. MATERIALES Y MÉTODOS

1,0 y 2,0 Å. Estas pseudo-energías junto a las energías libres de los complejos fueron sometidas a un análisis de correlación utilizando el programa GraphPad Prism 6 (GraphPad Software, Inc), con el objetivo de determinar los parámetros del potencial estadístico (tamaño de ventana y distancia de corte) que entrega el mejor coeficiente de correlación de Pearson (R). Utilizando los potenciales estadísticos calculados con los parámetros con mejor correlación se calculó la pseudo-energía de los sitios de unión de la coenzima en las estructuras de la base de datos de validación y se correlacionó con la energía libre de estos complejos con el objetivo de validar los potenciales estadísticos.

3. Resultados

En este trabajo buscamos extraer información desde las frecuencias de interacción entre átomos de la proteína y la coenzima, observadas en bases de datos de estructuras de estos complejos para generar potenciales estadísticos. Por lo tanto, como primera etapa, se describe el proceso inicial de la generación de potenciales estadísticos que se denomina "Entrenamiento".

3.1 Entrenamiento de potenciales estadísticos

3.1.1 Generación de una base de datos de estructuras de complejos proteína-NAD(P)

Se encontraron en la base de datos PDB un total de 2429 estructuras determinadas por difracción de rayos-x (archivos en formato PDB), incluyendo estructuras redundantes, con las coenzimas NAD y NADP. De ellas, 1016 correspondieron a complejos con NAD y 1413 a complejos con NADP. Para estos conjuntos de estructuras se cuantificó el porcentaje de estructuras, en función de la cantidad de átomos en la coenzima, su extensión y la resolución cristalográfica del complejo. De esta forma, el 95% de las estructuras con NAD presentaron a la coenzima con todos sus átomos (44 átomos), mientras que el 93% de las

estructuras con NADP presentaron a la coenzima con todos sus átomos (48 átomos) (Figura 10).

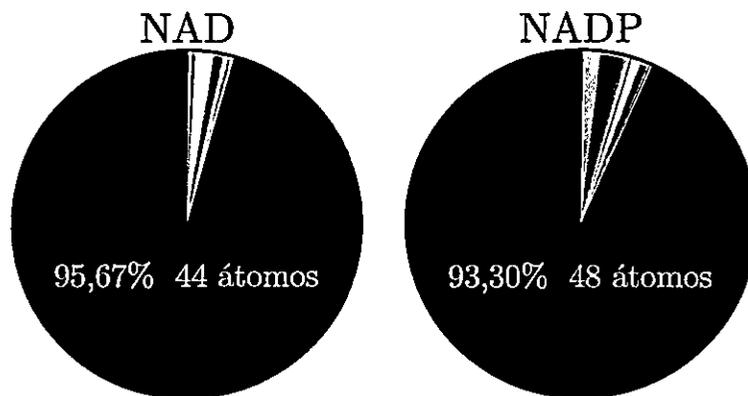


Figura 10. Distribución del número de átomos de la coenzima en las estructuras de la base de datos PDB.

Como se describió en Materiales y Métodos, la extensión de la coenzima considera la distancia entre el carbono-6 en la adenina y el carbono-2 en la nicotinamida. En los complejos con NAD las distancias más frecuentes se observaron entre 14 y 15 Å, mientras que las estructuras con NADP presentaron una mayor frecuencia a 17 Å, pero también para distancias de 14 y 15 Å (Figura 11-a). Respecto a la resolución cristalográfica, observamos que para ambas coenzimas la mayor frecuencia es 2 Å. Sin embargo, para las estructuras con NADP la distribución muestra un sesgo hacia valores menores, es decir, en general corresponden a estructuras con mejor resolución cristalográfica (Figura 11-b).

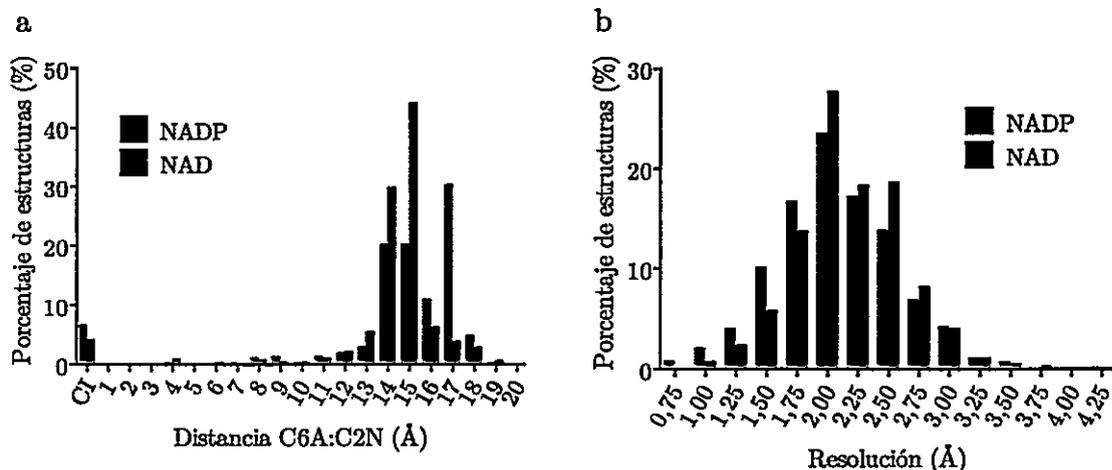


Figura 11. Extensión de la coenzima (a) y resolución (b) de las estructuras en la base de datos PDB.

Luego de filtrar por resolución, completitud y extensión el conjunto se redujo a 2065 estructuras, 920 de ellas con NAD y 1145 con NADP. La estrategia general se describe en la Figura 12. En este punto, se aplicó un filtro de redundancia para evitar tener una sobre-representación de estructuras muy similares en secuencia. Finalmente, se excluyó un conjunto de estructuras que luego fueron utilizadas para evaluar y validar los potenciales estadísticos, mientras que al mismo tiempo se descartaron las estructuras con enlaces covalentes entre la coenzima y la proteína. Así, la base de datos de entrenamiento para los potenciales estadísticos de NAD constó de 359 estructuras y la de NADP de 330 estructuras (lista disponible en Anexos). Con la generación de esta base de datos se da por completado el objetivo específico 1-a.

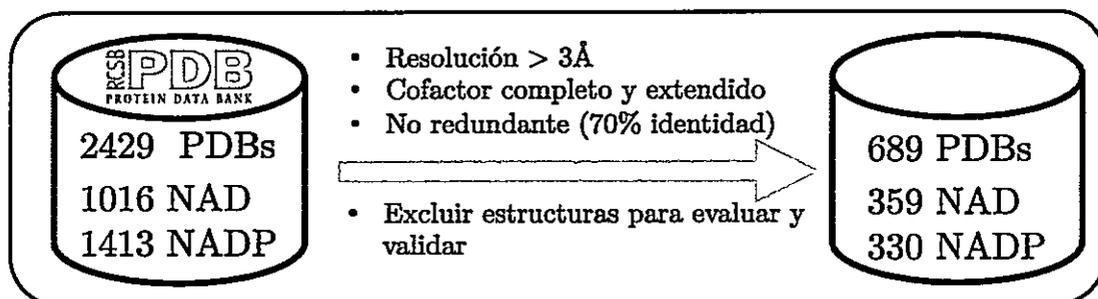


Figura 12. Esquema de la creación de la base de datos de entrenamiento. Los criterios de filtro utilizados se indican adyacente a la flecha.

3.1.2 Aminoácidos más frecuentes en los sitios de unión de NAD(P)

Las frecuencias que utilizaremos para luego calcular los potenciales estadísticos, se cuantificaron primero en términos del porcentaje de estructuras que forman una determinada interacción entre tipos de átomos en cadenas laterales de aminoácidos específicos (Tabla 3 en Materiales y Métodos) y las 5 regiones de la coenzima (adenina, ribosa adyacente a la adenina, difosfato, ribosa adyacente a la nicotinamida y nicotinamida). A continuación se describe para cada región de la coenzima, los residuos que interactúan en más de un 10% de los sitios de unión de NAD o NADP. La Figura 13 resume las observaciones para ambas coenzimas.

Adenina: En los sitios de unión de NADP, los anillos de la adenina se encuentra en un 65% de las estructuras estableciendo una interacción catión- π con argininas y en un 20% con residuos aromáticos como Trp/Tyr/Phe. Por otra parte, en los sitios de unión de NAD la presencia de un residuo de arginina que ejerce esta función es mucho menor, en comparación con los sitios de unión de NADP, ocurriendo solo en el 15% de las estructuras. En el caso de los residuos aromáticos, la frecuencia con que observamos una interacción con esta región es levemente mayor en las estructuras con NAD que con NADP.

Tanto en los sitios de unión de NAD como a NADP encontramos la presencia, en un 35-40% de los casos, de un residuo de carga negativa (Asp o Glu) que forma un puente de hidrógeno con el nitrógeno-6 de la adenina. También encontramos, en menor medida, un grupo hidroxilo de residuos de Thr/Ser/Tyr que forma un puente de hidrógeno con la adenina en ambas coenzimas. Esto es importante para el correcto posicionamiento de la coenzima en el sitio de unión.

Ribosa adyacente a la adenina: La presencia del grupo 2'-fosfato en NADP, da como resultado que generalmente encontremos una carga positiva

interactuando electrostáticamente con esta región. Siendo en su mayoría arginina (70% de las estructuras), seguido por lisina (30% de las estructuras). Cabe destacar que esta arginina es la misma que forma una interacción catión- π con el los anillos de la adenina ya descrito. Adicionalmente, se observa que el fosfato-2' forma en el 50% de las estructuras una interacción por puente de hidrógeno con residuos como Thr/Ser/Tyr. Estos residuos también forman puentes de hidrógeno con los hidroxilos 2' y 3' de NAD en un 15% de los casos.

En los sitios de unión de NAD, encontramos en el 70% de las estructuras, un residuo de carga negativa (Asp o Glu) que presenta un puente de hidrógeno bidentado con los hidroxilos de la ribosa. Solamente en el 10% de las estructuras encontramos una lisina que forma puentes de hidrógeno con estos hidroxilos.

Difosfato: En los sitios de unión de ambas coenzimas se encontró que en un 20-25% de las estructuras está presente un residuo Arg que interactúa por complementariedad electrostática, o un residuo de Thr/Ser/Tyr que forma un puente de hidrógeno con los oxígenos de esta región.

Ribosa adyacente a la nicotinamida: En los sitios de unión de ambas coenzimas encontramos con frecuencias similares residuos que forman puentes de hidrógeno con los hidroxilos 2' y 3' de esta región. Siendo más frecuente la presencia de una lisina (30-35%), luego Thr/Ser/Tyr (25-30%) y finalmente Asp/Glu (15-20%).

Nicotinamida: En ambas coenzimas, este anillo interactúa en un 35% de las estructuras con residuos aromáticos como Trp/Tyr/Phe mediante interacciones por apilamiento entre sus nubes electrónicas.

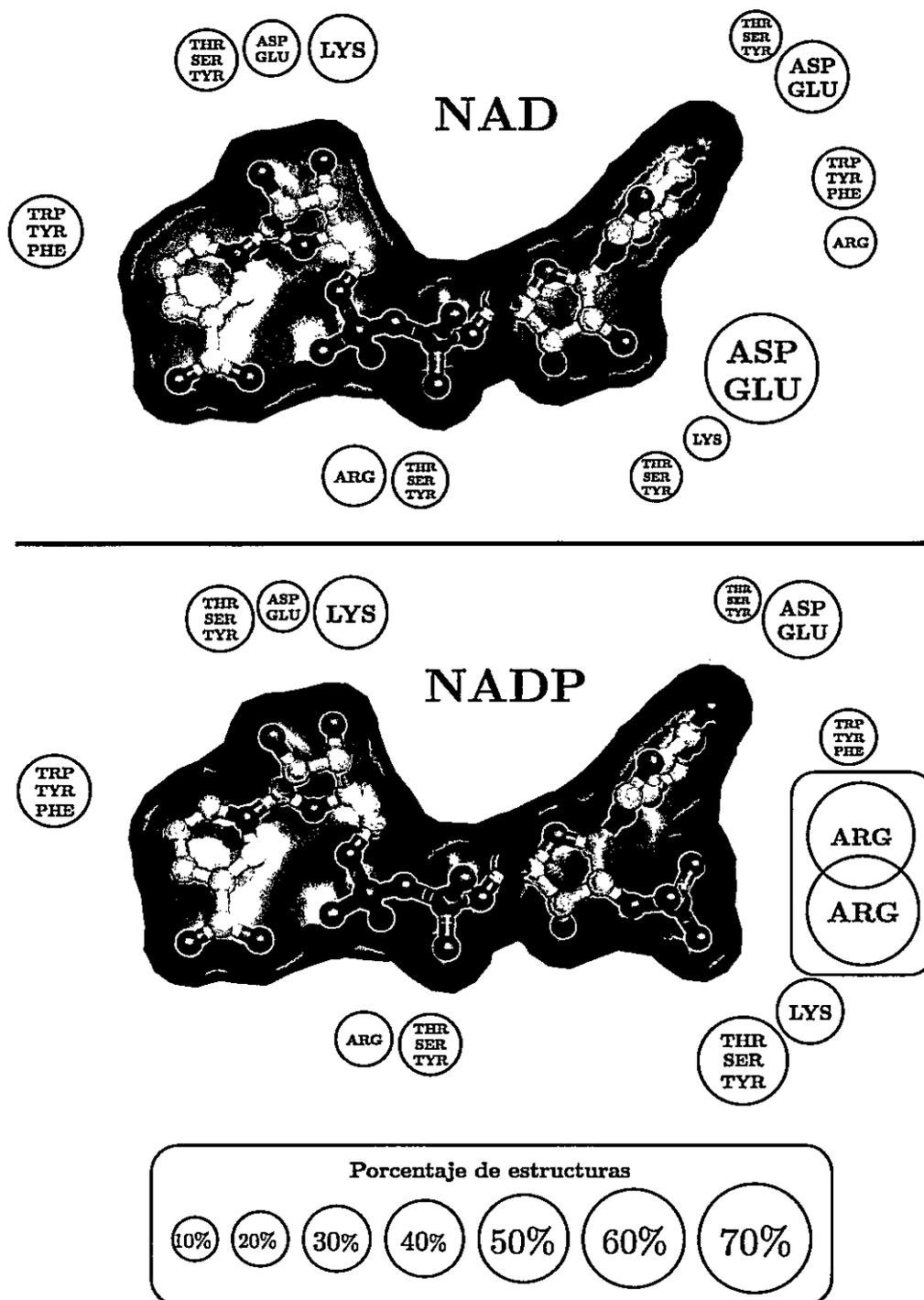


Figura 13. Residuos aminoacídicos más comunes en los sitios de unión de NAD(P). Cada círculo representa la cadena lateral de un aminoácido. El tamaño del círculo indica el porcentaje de estructuras de la base de datos de entrenamiento que presenta este aminoácido interactuando con cada región de la coenzima.

3.1.3 Potenciales estadísticos: perfiles de pseudo-energía

De acuerdo con la definición SYBYL es posible catalogar 12 tipos de átomos en los aminoácidos de las proteínas y 10 tipos de átomos en los dinucleótidos de nicotinamida adenina (Figuras 8 y 9). Por lo tanto, se calcularon perfiles de pseudo-energía para los 120 pares de interacciones posibles, para los complejos con NAD o NADP, usando la formula de Sippl (Ecuación 11) usando como parámetros de $\sigma=0,02$ y $RT=0,582$ kcal/mol (Ferrada & Melo 2007).

Para una interacción específica, los perfiles de pseudo-energía observados en las estructuras con NADP y NAD, fueron comparados con el fin de determinar los rasgos característicos de cada sitio de unión. Particularmente, analizamos interacciones de tipo puentes de hidrógeno, electrostáticas, catión- π y también interacciones entre residuos aromáticos y carbonos en grupos aromáticos en la coenzima.

Puentes de hidrógeno

Una interacción importante en la nicotinamida de las coenzimas se da entre el nitrógeno de un grupo amida (definido como N.am en la clasificación SYBYL) y

oxígenos sp^2 en la proteína (definido como 0.2 en la clasificación SYBYL) (Figura 14-a). Vemos que los perfiles de pseudo-energía son equivalentes para NAD y NADP, con un mínimo de -1,4 kcal/mol a 3 Å, lo que es clasificado como un puente de hidrógeno moderado ², seguido de un pico de energía positivo, conocido como "locking elbow", característico de puentes de hidrógeno formados por la cadena principal cuando constituyen estructuras secundarias en las proteínas (Sippl 1996), debido a que para formarse requiere de energía (para sobrepasar la barrera energética), pero una vez que está formado es más difícil de romperse.

Otro perfil de pseudo-energía que resultó similar entre las coenzimas es el de la interacción entre oxígenos de grupos fosfato (O.co2) y el nitrógeno del grupo amidas (N.am) en la proteína (Figura 14-b). Este tipo de puente de hidrógeno se da generalmente entre el grupo difosfato de la coenzima y la cadena principal de glicinas (Kleiger & Eisenberg 2002). Los perfiles de energía muestran un mínimo

² Clasificación de los tipos de puentes de hidrógeno según su distancia (Jeffrey 1997)

- Fuertes entre 2,2-2,5 Å
- Moderados entre 2,5-3,2 Å
- Débiles entre 3,2-4,0 Å

de energía de -1,4 kcal/mol a los 3 Å, manteniendo una energía favorable hasta los 4 Å.

Por lo contrario, una de las diferencias mayores entre NAD y NADP, se da para el perfil de pseudo-energía del puente de hidrógeno entre oxígenos sp^3 (O.3) en la coenzima y oxígenos de carboxilatos (O.co2) en la proteína (Figura 14-c). Esta interacción es frecuente entre los hidroxilos-2' y 3' de las ribosas con carboxilatos de Asp y Glu. Los perfiles muestran un mínimo de pseudo-energía a una distancia de 2,5 Å, lo que es clasificado como un puente de hidrógeno fuerte ², el cual es más negativo para el potencial estadístico de NAD que de NADP (-2,5 kcal/mol y -1,6 kcal/mol, respectivamente). El valor menos negativo en la función que describe esta interacción con NADP es explicada por un valor menor en el factor M_{ij} (que simplemente consiste en el total de observaciones para un par específico de átomos - Ecuación 10). Efectivamente, en el caso de los sitios para NADP se registra un número menor de observaciones debido a la repulsión entre el fosfato-2' y las cargas negativas. Si observamos en detalle los perfiles, vemos que el potencial estadístico de NADP tiene un "locking elbow", lo que

indica que para que esta interacción sea formada se requiere de mayor energía que en los sitios de NAD.

Finalmente, el perfil de energía para el puente de hidrógeno de energía más negativa para la interacción entre estas coenzimas y las proteínas corresponde a aquel formado entre oxígenos de los grupos fosfato (O_{CO2}) en la coenzima y los oxígenos sp^3 (O_3) en la proteína (Figura 14-d). Estos últimos corresponden a hidroxilos en las cadenas laterales de Thr/Ser/Tyr. Observamos para esta interacción perfiles de energía para NAD y NADP con un mínimo de energía a una distancia de 2,5 Å de -2,5 kcal/mol y -2,8 kcal/mol, respectivamente.

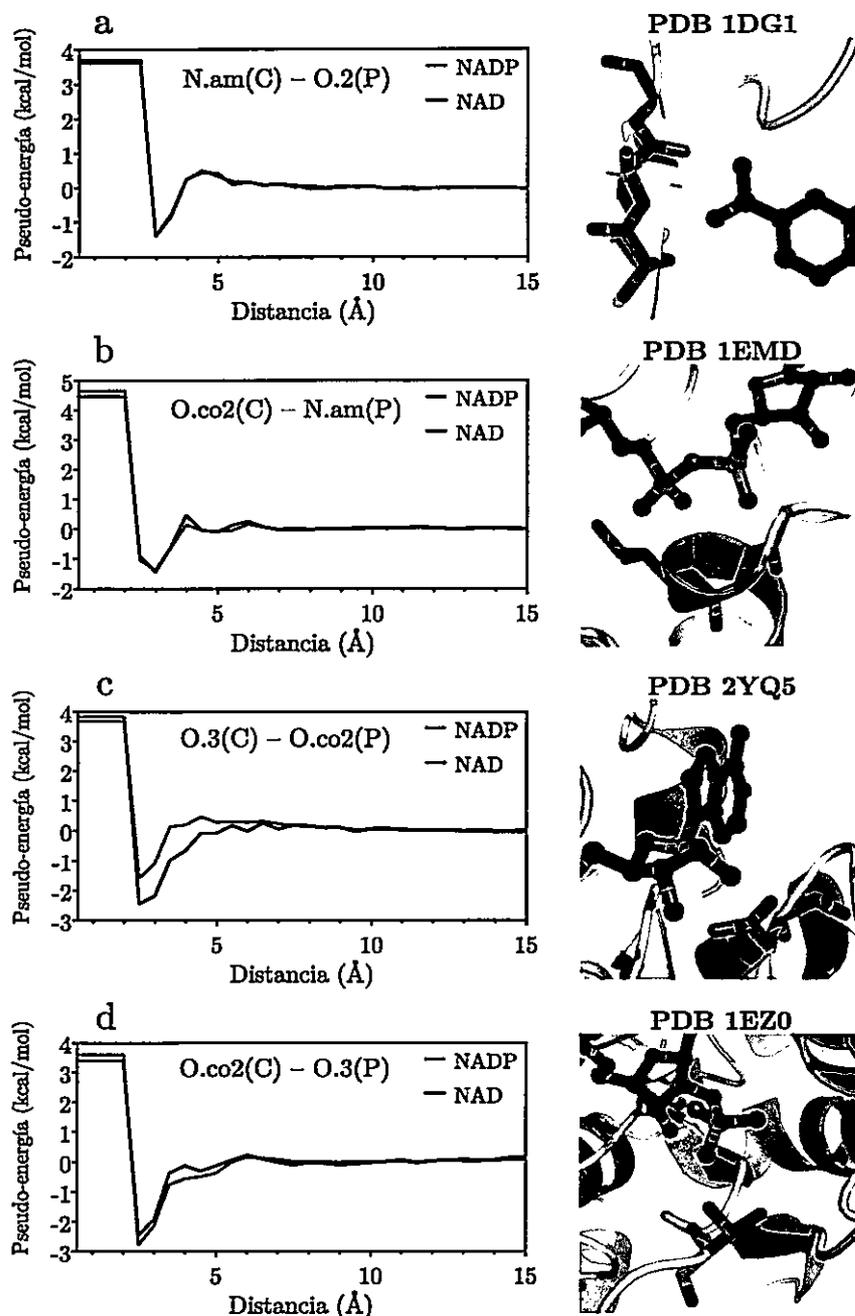


Figura 14. Perfiles de pseudo-energía para interacciones de tipo puentes de hidrógeno. A la izquierda se muestra el perfil de pseudo-energía para los tipos de átomos indicados (parte superior-central de cada gráfico) pertenecientes a la coenzima (C) y la proteína (P). A la derecha se muestra un ejemplo de cada tipo de interacción, encontrado en la estructura cuyo código PDB se indica en la parte superior de cada imagen.

Interacciones electrostáticas

Para estudiar las interacciones electrostáticas, observamos los perfiles entre tipos de átomos asociados a cargas negativas en las coenzimas y cargas positivas en las proteínas. Específicamente, observamos la interacción entre oxígenos de los grupos fosfato (O.co2) en la coenzima y, tanto el nitrógeno trigonal planar (N.pl3) como el nitrógeno sp3 (N.4), cargados positivamente en la proteína, según se muestra en la Figura 15-a y -b, respectivamente.

Para el primer caso, una interacción frecuente ocurre entre los fosfatos de la coenzima y cadenas laterales de argininas. Los perfiles de energía para esta interacción tienen un mínimo de pseudo-energía entre -2,0 y -2,2 kcal/mol, a una distancia de 2,5-3,0 Å. Adicionalmente, vemos que este tipo de interacción es de mayor alcance que los puentes de hidrógeno, ya que mantiene pseudo-energías favorables y atractivas hasta los 6 Å.

La segunda interacción ocurre entre los fosfatos de la coenzima y el nitrógeno de la cadena lateral de lisinas. Para esta interacción se observa un perfil de pseudo-energía similar a la interacción electrostática con arginina, con un mínimo de

energía entre 2,5-3,0 Å, con un valor de -2,3 kcal/mol para NAD y -2,2 kcal/mol para NADP. Los perfiles además muestran tener mayor alcance en los sitios de NADP que en los de NAD.

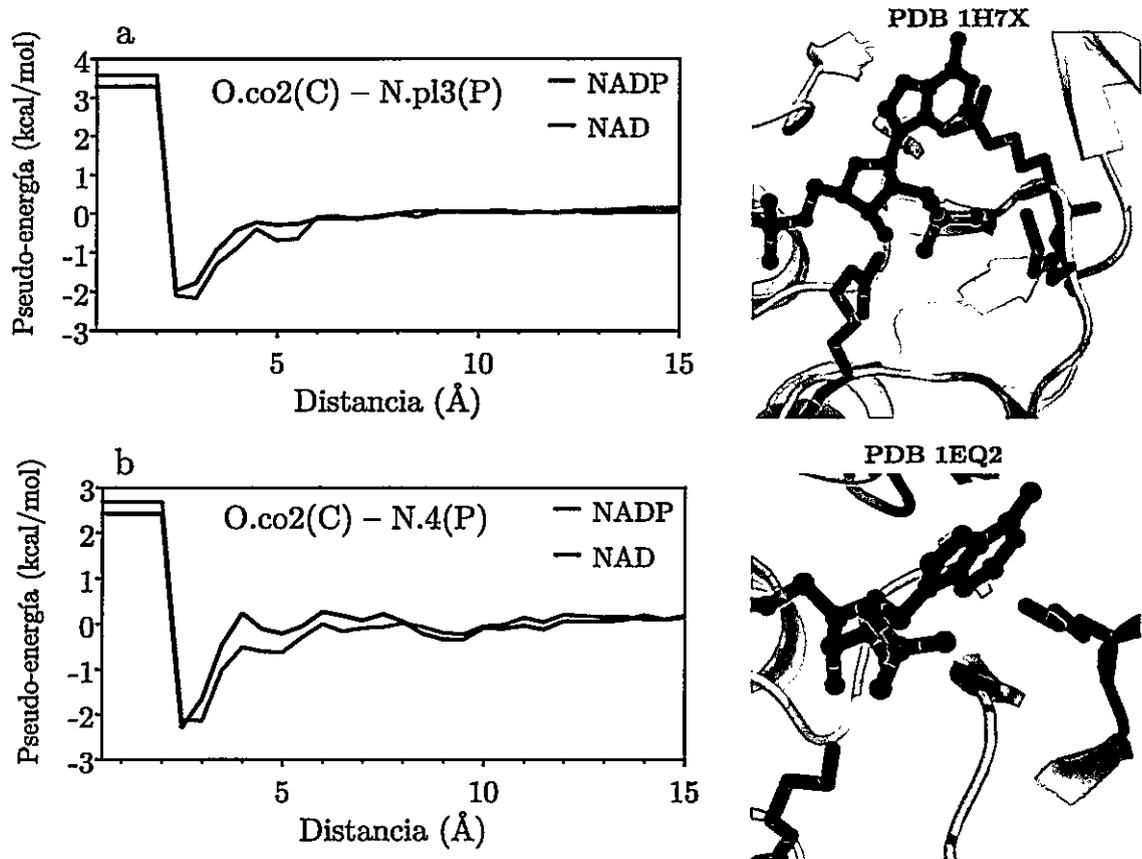


Figura 15. Perfiles de pseudo-energía para interacciones electrostáticas. A la izquierda se muestra el perfil de pseudo-energía para los tipos de átomos indicados (parte superior-central de cada gráfico) pertenecientes a la coenzima (C) y la proteína (P). A la derecha se muestra un ejemplo de cada tipo de interacción, encontrado en la estructura cuyo código PDB se indica en la parte superior de cada imagen.



Interacciones Cati3n- π

La interacci3n cati3n- π es una interacci3n molecular no covalente entre la cara de un sistema π , rico en electrones, con un cati3n. Una interacci3n com3n de este tipo se da entre la adenina de las coenzimas y cadenas laterales de argininas. Para esta observaci3n, nos focalizamos en el perfil de la interacci3n entre nitr3genos sp^3 (N.3) en las coenzimas y nitr3genos trigonal planar (N.pl3) en la prote3na. En la Figura 16 se observa un m3nimo energ3tico a 4,5 Å con valores de pseudo-energ3a m3s favorables para NADP que para NAD. Mientras el perfil para NAD muestra un m3nimo de -0,5 kcal/mol, para NADP tiene un m3nimo de -1,2 kcal/mol. Esta diferencia se debe a la menor frecuencia de esta interacci3n en los sitios de NAD, ya que en los sitios de NADP la presencia del grupo fosfato-2' conduce a un incremento de las interacciones observadas con Arg (Figura 13). Si observamos detalladamente el perfil para NADP en la Figura 16, vemos que no tiende a cero, sino que toma valores positivos despu3s de los 10 Å. Esto es com3n en las interacciones electrost3ticas en prote3nas, como por ejemplo entre el grupo guanidino de Arg y el carboxilato de Asp (Sippl 1996). Sin embargo, a3n no ha sido descrito el fen3meno molecular que explica esta caracter3stica. Por otra parte, esta funci3n toma valores negativos para ambas

coenzimas bajo los 6,5 Å, es decir, tienen un alcance mayor a una interacción de puente de hidrógeno.

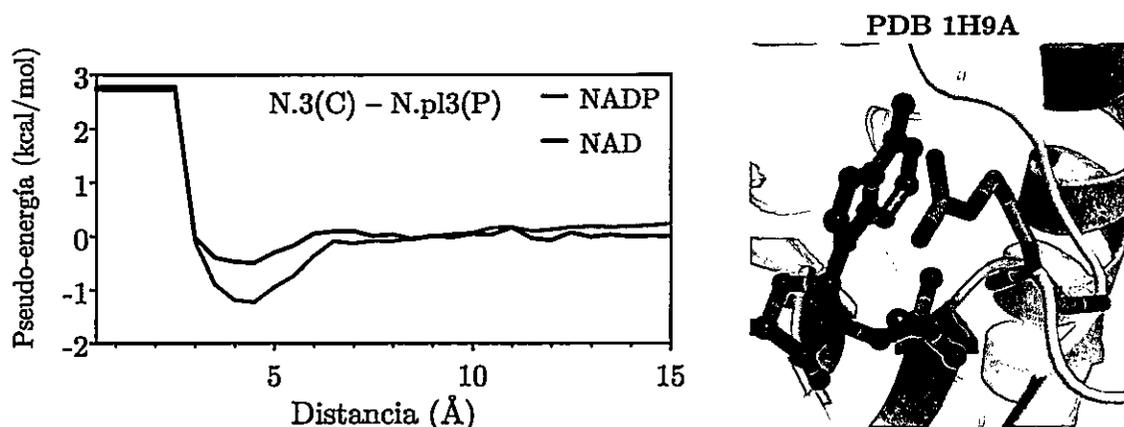


Figura 16. Perfiles de pseudo-energía para la interacción catión- π entre la adenina de las coenzimas y cadenas laterales de arginina. A la izquierda se muestra el perfil de pseudo-energía para los tipos de átomos indicados (parte superior-central de cada gráfico) pertenecientes a la coenzima (C) y la proteína (P). A la derecha se muestra un ejemplo de cada tipo de interacción, encontrado en la estructura cuyo código PDB se indica en la parte superior de cada imagen.

Interacciones aromáticas

La Figura 17 muestra los perfiles de pseudo-energía para las interacciones entre carbonos aromáticos (C.ar) en las coenzimas y carbonos aromáticos (C.ar) en las cadenas laterales de Phe, Trp y Tyr. Se observan mínimos energéticos a una distancia similar a las interacciones catión- π , pero con un valor de -0,3 kcal/mol para ambas coenzimas, es decir menor a los tipos de interacción ya descritos. Sin embargo, toma valores cercanos a cero recién alrededor de los 9 Å, por lo que

puede ser considerada como una función de mayor alcance que las demás interacciones.

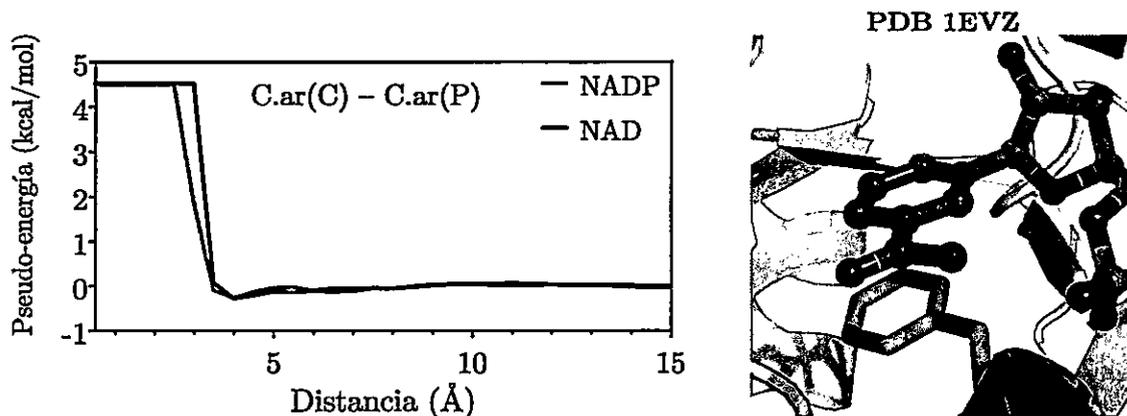


Figura 17. Perfiles de energía para la interacción aromática entre el anillo de nicotinamida en las coenzimas y residuos aromáticos. A la izquierda se muestra el perfil de pseudo-energía para los tipos de átomos indicados (parte superior-central de cada gráfico) pertenecientes a la coenzima (C) y la proteína (P). A la derecha se muestra un ejemplo de cada tipo de interacción, encontrado en la estructura cuyo código PDB se indica en la parte superior de cada imagen.

Con la derivación de estos potenciales estadísticos para NAD y NADP que incluyen perfiles de pseudo-energía para distintos tipos de interacción se da por completado el objetivo específico 2.

3.2 Evaluación de los potenciales estadísticos

En esta etapa, usando los potenciales estadísticos obtenidos para los 120 tipos de interacción posibles con cada coenzima, se calculó la pseudo-energía total de los sitios de unión de los complejos proteína-coenzima, en una base de datos

independiente a la de entrenamiento. El objetivo fue determinar los valores de parámetros de los potenciales estadísticos que entreguen la mejor correlación entre las pseudo-energías calculadas y las energías libres de unión, determinadas experimentalmente para estos complejos. De esta manera, se espera que la asignación de un determinado valor de pseudo-energía a la estructura de un sitio con su ligando unido, refleje la afinidad de dicho sitio.

Se encontraron en la base de datos Binding MOAD (Hu et al. 2005), 77 proteínas con estructura conocida y constantes de disociación determinadas para la coenzima presente en la estructura. Se filtró por resolución, completitud, extensión y exclusión de las estructuras con enlaces covalentes entre la coenzima y la proteína, del mismo modo en que se procedió con la base de datos de entrenamiento. Se obtuvo un total de 51 estructuras, 27 como complejo con NAD y 24 como complejo con NADP. Para cada estructura se calculó la pseudo-energía total, usando diferentes distancias de corte y tamaño de ventana. Se graficaron las pseudo-energías obtenidas para cada complejo, versus los correspondientes valores de cambio de energía libre de unión (ΔG^0), y se obtuvo el coeficiente de correlación de Pearson, R. En la Figura 18 se observa la

dependencia de R con la distancia de corte, para diferentes tamaños de ventana. Para los sitios de NAD se observa un coeficiente de correlación máximo de 0,76 a una distancia de corte de 15 Å, usando un tamaño de ventana de 0,5 Å. Por otra parte, los sitios de NADP muestran un coeficiente de correlación máximo de 0,74, a una distancia de corte de 4 Å con un tamaño de ventana de 0,5 Å.

Los potenciales estadísticos para NAD presentan un rápido incremento de la correlación entre 3 y 5 Å de distancia de corte. De ahí en adelante, aumenta de manera más discreta hasta llegar a un máximo en 15 Å, para disminuir paulatinamente hasta los 20 Å. Al observar el efecto de los diferentes tamaños de ventana, podemos observar que en general se obtiene una mayor correlación a un mayor tamaño de ventana. Sólo el tamaño de ventana de 2 Å posee una correlación peor que el tamaño de ventana de 0,1 Å.

A diferencia de NAD, los potenciales estadísticos para NADP, muestran que el coeficiente de correlación se incrementa rápidamente con la distancia de corte entre 3 y 4 Å, para posteriormente disminuir gradualmente, llegando a valores cercanos a cero alrededor de 12 Å (Figura 18). Luego no se observan cambios

hasta los 20 Å. Por otra parte, observamos que a diferencia de los potenciales estadísticos para NAD, en los potenciales estadísticos para NADP a menor tamaño de ventana se obtiene mayor valor de R, con excepción del tamaño de ventana de 2 Å, a distancias de corte superiores a 12 Å.

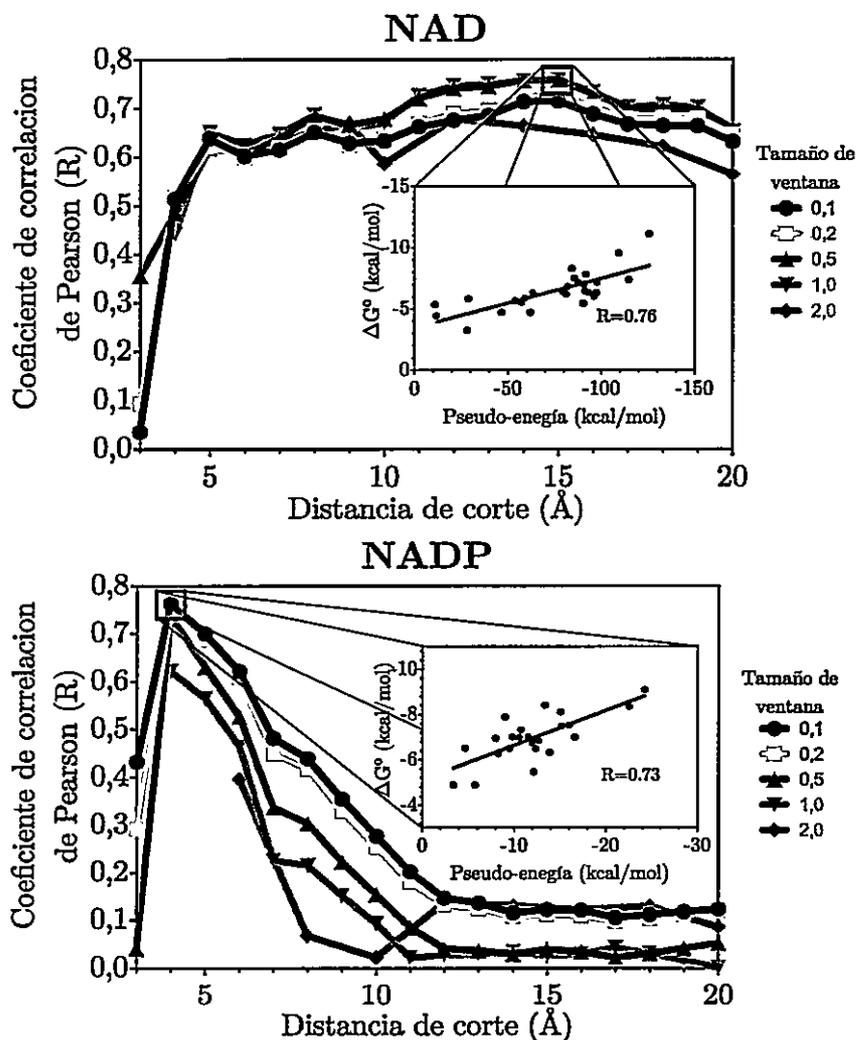


Figura 18. Evaluación de los potenciales estadísticos. Se muestra el coeficiente de correlación obtenido (eje y) para cada distancia de corte (eje x) y tamaño de ventana analizado (colores). En el gráfico interno se muestra el análisis de correlación para los parámetros con la mejor correlación.

Luego de haber evaluado los potenciales estadísticos para NAD y NADP, y determinar los parámetros que permiten una mejor correlación entre las pseudo-energías de los sitios de unión con los datos experimentales, damos por completado el objetivo específico 3.

3.3 Validación de los potenciales estadísticos

La energía de unión de una proteína por su ligando está determinada por las interacciones establecidas con los residuos que componen el sitio de unión. Experimentalmente, el aporte de la cadena lateral de un residuo a la energía de unión de NAD(P) puede ser estimado mediante la comparación con la energía libre de unión de la mutante por Ala de dicho residuo, respecto del valor en la enzima silvestre. Por otra parte, el reemplazo de múltiples residuos en un sitio de unión de NAD(P), con el objetivo de cambiar la especificidad, puede incrementar o disminuir la energía de unión de la coenzima. Entonces, usando los parámetros de distancia de corte y tamaño de ventana obtenidos en la etapa de evaluación, se espera que los potenciales estadísticos sean capaces de representar el efecto de mutaciones en los sitios de NAD(P), sobre la energía de unión del complejo.

En los casos que se describen a continuación (base de datos de validación) analizamos el efecto de mutaciones en la pseudo-energía del sitio unión, calculada a partir de estructuras cristalográficas o modelos *in silico*. Estos valores de pseudo-energía fueron comparados con las energías libres para estos complejos, con el propósito de determinar si un incremento en la energía libre se relaciona con un incremento en la pseudo-energía del sitio de unión o viceversa.

3- α -hidroxiesteroide deshidrogenasa

Se utilizó para este análisis la enzima 3- α -hidroxiesteroide deshidrogenasa de *Comamonas testosteroni*, cuya estructura cristalográfica fue determinada (Grimm et al. 2000) y sus constantes de disociación establecidas para la enzima silvestre y las mutantes por alanina de residuos claves en la unión de NAD (Chang et al. 2010).

Para cada átomo del bolsillo de unión de NAD (Figura 19-a) se calculó la pseudo-energía y la Figura 19-b permite visualizarla como lo que llamamos superficie de pseudo-energía.

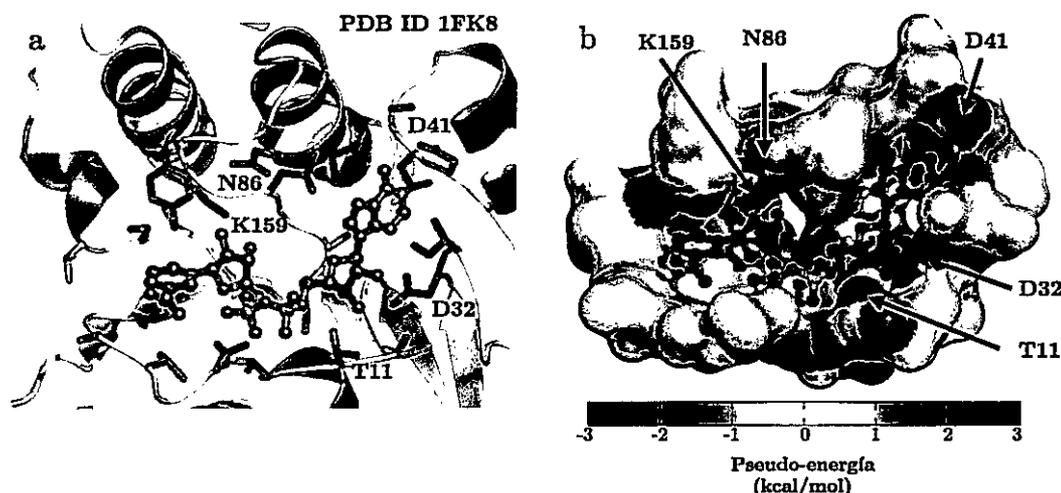


Figura 19. Sitio de unión de NAD en la enzima 3- α -hidroxiesteroide deshidrogenasa de *Comamonas testosteroni*, representado como cadenas laterales de los residuos y estructuras secundarias (a) o superficie, donde los colores representan la pseudo-energía de cada átomo (b). El color de la superficie oscila desde azul, para pseudo-energías negativas, a rojo, para pseudo-energías positivas, cubriendo ± 3 kcal/mol (barra de colores). En cada imagen se señalan con flechas los residuos mencionados en el análisis.

La Tabla 4, indica que el valor de pseudo-energía calculado para las diferentes mutantes sigue la misma tendencia de cambio que los valores experimentales de ΔG^0 . Es decir, el potencial estadístico es capaz de predecir que la pérdida de afinidad por NAD es mayor para la mutación de Lys159 que para Asn86.

Tabla 4. Energía libre y pseudo-energía con NAD a la enzima 3- α -hidroxiesteroide deshidrogenasa de *Comamonas testosteroni* y a mutantes de esta enzima diseñadas para cambiar su especificidad por esta coenzima. En la parte inferior de la tabla se muestra el coeficiente de correlación de Pearson (R) entre las energías libres y las pseudo-energías, magnitudes que fueron calculadas utilizando una temperatura de 298 K.

| | NAD | | |
|------------------|---------------------|----------------------------|------------------------------|
| | K_D (μ M) | ΔG^0 (kcal/mol) | Pseudo-energía (kcal/mol) |
| Enzima silvestre | 2,8 | -7,44 | -98,3 |
| N86A | 22 | -6,24 | -93,9 |
| K159A | 210 | -4,93 | -85,0 |

R=0,99

Por otra parte, la superficie de pseudo-energía revela la existencia de otros residuos claves en la unión del dinucleótido como Asp32, que forma un puente de hidrógeno bidentado con la ribosa adyacente a la adenina, Asp41 que forma un puente de hidrógeno con el nitrógeno-6 de la adenina y Thr11 que forma un puente de hidrógeno con el hidroxilo-3' de la ribosa adyacente a la adenina, entre otros.

Acido 2,5-diceto-D-glucónico reductasa

Hoy en día, determinar la estructura cristalográfica de una proteína puede ser una tarea a menudo difícil y que en general toma bastante tiempo. A pesar de esto, algunos estudios enfocados a cambiar la especificidad por estas coenzimas

en proteínas han utilizado estructuras cristalográficas como fundamento, tanto para proponer mutaciones, como para explicar los efectos de las mutaciones en las interacciones con las coenzimas. En la mayoría de los casos no es posible contar con una estructura cristalográfica de las proteínas silvestre o mutantes estudiadas. Esto, en principio nos impediría aplicar los potenciales estadísticos en estas enzimas con el propósito de estudiar la contribución de los residuos del sitio de unión a la energía de unión de la coenzima. Por ello, quisimos estudiar modelos *in silico* en la predicción del efecto de mutaciones sobre la energía de unión de la coenzima. Para esto analizamos la estructura de la enzima Acido 2,5-diceto-D-glucónico reductasa de *Corynebacterium sp.*, de la que sólo se dispone la estructura cristalográfica de una mutante cuádruple con la especificidad modificada a favor de la unión de NAD (Sanli et al. 2004). A partir de esta estructura se modelaron las estructuras de la enzima silvestre y las mutantes F22Y/A272G y K232G/R238H como complejo con NAD. Se determinó para la estructura cristalográfica y cada modelo la pseudo-energía del sitio de unión de NAD (Figura 20).

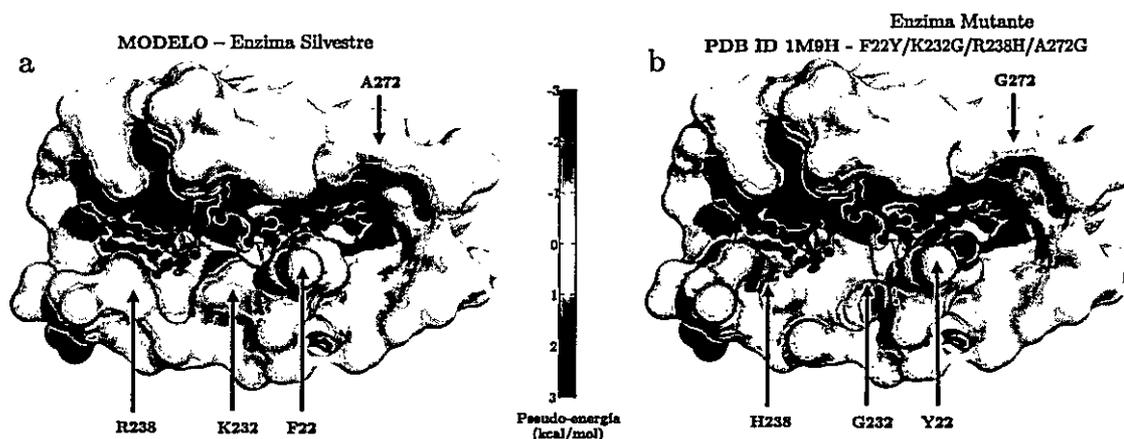


Figura 20. Pseudo-energía de los sitios de unión de NAD a la enzima Acido 2,5-diceto-D-glucónico reductasa de *Corynebacterium sp.* (a), y a la mutante F22Y/K232G/R238H/A272G (b). El color de la superficie oscila desde azul, para pseudo-energías negativas, a rojo, para pseudo-energías positivas, cubriendo ± 3 kcal/mol (barra de colores). En cada imagen se señalan con flechas los residuos mutados.

La Tabla 5 muestra que las pseudo-energías se incrementan producto de las mutaciones en la misma medida que el incremento en las energías libres calculadas a partir de las constantes de disociación. Dado que en este caso de estudio fueron utilizados 3 modelos *in silico* y una estructura cristalografía, esto sugiere que los potenciales estadísticos también pueden ser aplicados sobre modelos *in silico* para estudiar los determinantes de la unión de estas coenzimas en enzimas.

Tabla 5. Energía libre y pseudo-energía de la unión de NAD a la enzima Acido 2,5-diceto-D-glucónico reductasa de *Corynebacterium sp.* y a mutantes de esta enzima diseñadas para cambiar su especificidad por esta coenzima. En la parte inferior de la tabla se muestra el coeficiente de correlación de Pearson (R) entre las energías libres y las pseudo-energías, magnitudes que fueron calculadas utilizando una temperatura de 298 K.

| | NAD | | |
|------------------------|------------------------|----------------------------|------------------------------|
| | K _D (uM) | ΔG^0 (kcal/mol) | Pseudo-energía (kcal/mol) |
| Enzima silvestre | 2400 | -3,51 | -28,51 |
| F22Y/A272G | 1800 | -3,68 | -30,41 |
| K232G/R238H | 810 | -4,14 | -37,25 |
| F22Y/K232G/R238H/A272G | 660 | -4,26 | -39,15 |

R=0,99

Gliceraldehído-3-fosfato deshidrogenasa

En este caso observamos si los potenciales estadísticos eran sensibles a cambios en la energía de unión del dinucleótido, en mutantes diseñadas con el propósito de cambiar la preferencia por la coenzima. Para ello utilizamos la enzima Gliceraldehído-3-fosfato deshidrogenasa de *Bacillus stearothermophilus*, de la cual sus constantes de disociación para NAD y NADP han sido determinadas tanto en la enzima silvestre como en sus mutantes. Más aún las estructuras de las mutantes con NAD y NADP han sido determinadas experimentalmente (Didierjean et al. 1997).

3. RESULTADOS

Los resultados muestran que los potenciales estadísticos representan correctamente tanto el aumento como el decremento en la afinidad de NAD y de NADP, respectivamente (Tabla 6). En otras palabras, los cambios en pseudo-energía de unión detectados para las mutaciones, ocurren en magnitudes proporcionales al cambio de energía libre experimental para estos complejos.

Tabla 6. Energía libre y pseudo-energía de la unión de NAD o NADP a la enzima Gliceraldehído-3-fosfato deshidrogenasa de *Bacillus stearothermophilus* y a mutantes de esta enzima diseñadas para cambiar su especificidad por esta coenzima. En la parte inferior de la tabla se muestra el coeficiente de correlación de Pearson (R) entre las energías libres y las pseudo-energías, magnitudes que fueron calculadas utilizando una temperatura de 298 K.

| | NAD | | | NADP | | |
|----------------------------|------------------------|----------------------------|------------------------------|------------------------|----------------------------|------------------------------|
| | K _D (uM) | ΔG^0 (kcal/mol) | Pseudo-energía (kcal/mol) | K _D (uM) | ΔG^0 (kcal/mol) | Pseudo-energía (kcal/mol) |
| Enzima silvestre | 0,9 | -8,10 | -102,19 | 35 | -5,97 | -7,26 |
| L33T/T34G/D36G/L187A/P188S | 4,4 | -7,18 | -97,55 | 20,8 | -6,27 | -8,25 |
| D32G/L187A/P188S | 17 | -6,39 | -79,36 | 1,3 | -7,89 | -8,99 |
| | | R=0,93 | | | R=0,90 | |

Al observar la superficie de pseudo-energía de la enzima silvestre podemos ver que uno de los residuos con mayor contribución a la unión de NAD corresponde al residuo Asp32. Por el contrario, Leu187 y Pro188 estarían aportando pseudo-energías desfavorables a la unión de la coenzima NAD (Figura 21-a).

3. RESULTADOS

Al comparar la superficie de pseudo-energía de la enzima silvestre y la mutante que presenta un mayor efecto sobre la energía libre para NAD (D32G/L187A/P188S), vemos que al mutar el residuo Asp32 por Gly provoca una pérdida completa de la pseudo-energía que este residuo aporta (de azul a blanco, Figura 21). De manera similar, el reemplazo de Leu187 por Ala aumenta el carácter desfavorable de la interacción con NAD (aumento de intensidad de rojo, Figura 21). Por el contrario, el reemplazo de Pro188 por Ser resulta en una pseudo-energía más negativa del sitio de unión con NAD, posiblemente debido a un puente de hidrógeno adicional que se puede formar entre el oxígeno- γ de Ser con los hidroxilos de la ribosa adyacente a la adenina.

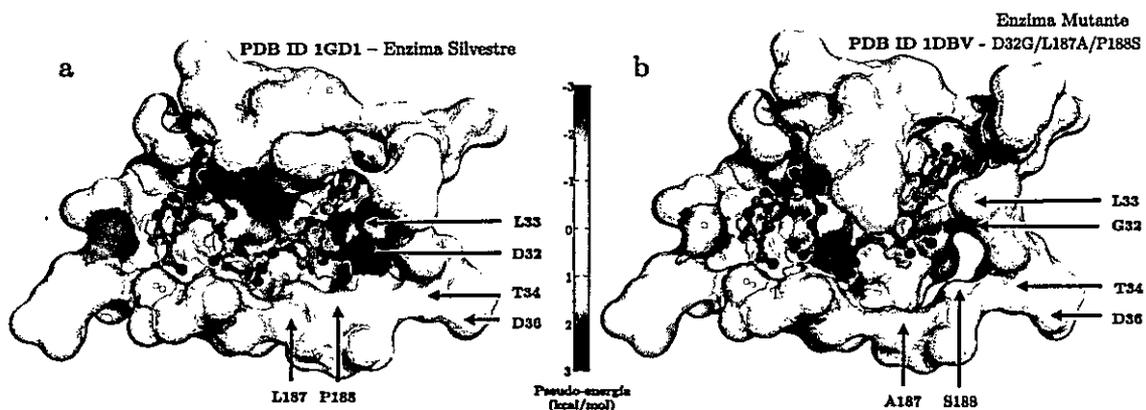


Figura 21. Pseudo-energía de los sitios de unión de NAD a la enzima Gliceraldehído-3-fosfato deshidrogenasa de *Bacillus stearothermophilus* (a) y a la mutante D32G/L187A/P188S (b). El color de la superficie oscila desde azul, para pseudo-energías negativas, a rojo, para pseudo-energías positivas, cubriendo ± 3 kcal/mol (barra de colores). En cada imagen se señalan con flechas los residuos mutados.

La cuantificación del aporte por residuo individual, indica que la magnitud del aumento en la pseudo-energía del sitio de unión ocasionada por la mutación Leu187Ala es similar a la disminución ocasionada por la mutación Pro188Ser. De esto podemos deducir que el aumento en la energía de unión con la coenzima NAD en la mutante D32G/L187A/P188S se debe fundamentalmente a la mutación Asp32Gly.

En la enzima silvestre unida a NADP, la presencia de la carga negativa de Asp32 provoca una repulsión electrostática con el 2'-fosfato que origina una energía de unión desfavorable con esta coenzima. Esto es representado en la superficie de pseudo-energía con un valor positivo en dicha posición (Figura 22-a). La mutación de este residuo por Gly conlleva a la obtención de valores negativos de pseudo-energía.

En la posición 188, la presencia de Pro y Ser son favorables para la unión con NADP. Sin embargo, la presencia de Ser da como resultado un valor de pseudo-energía más negativo, probablemente debido a que este residuo puede establecer un puente de hidrógeno entre su cadena lateral con los oxígenos del grupo

fosfato-2'. Por el contrario, la presencia de Leu o Ala en la posición 187, no determina un mayor cambio su pseudo-energía de interacción con NADP.

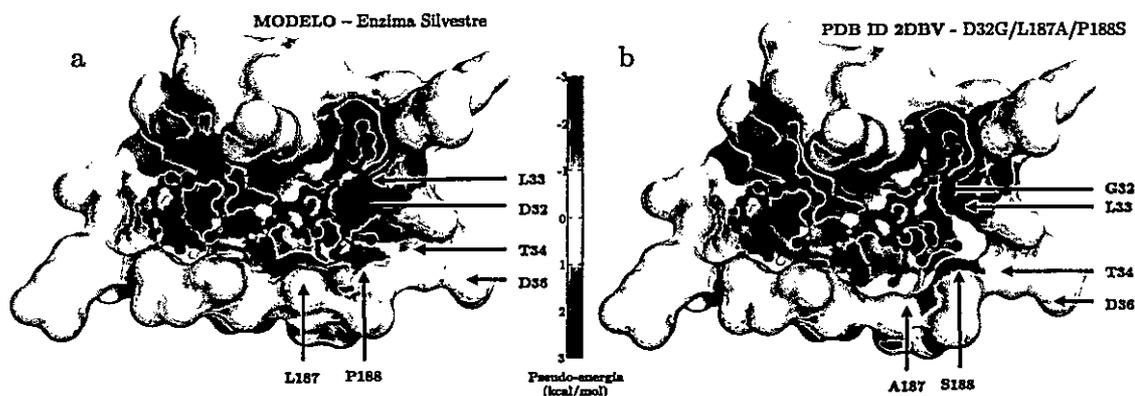


Figura 22. Pseudo-energía de los sitios de unión de NADP en la enzima Gliceraldehído-3-fosfato deshidrogenasa de *Bacillus stearothermophilus* (a) y a la mutante D32G/L187A/P188S (b) El color de la superficie oscila desde azul, para pseudo-energías negativas, a rojo, para pseudo-energías positivas, cubriendo ± 3 kcal/mol (barra de colores). En cada imagen se señalan con flechas los residuos mutados.

Los resultados anteriormente descritos indican que el residuo Asp32 es el principal determinante de la discriminación para las coenzimas, al estar involucrado en la unión de la coenzima NAD y la repulsión del 2'-fosfato de NADP. Adicionalmente, podemos teorizar que la mutación Leu187Ala provoca una disminución en la energía de unión de NAD sin un mayor cambio en la energía de unión de NADP, en tanto que la mutación Pro188Ser promueve la formación de un puente de hidrógeno adicional tanto con el hidroxilo-2' de NAD como con los oxígenos del fosfato-2' de NADP. En ausencia de datos

experimentales para las mutantes individuales en estas posiciones, nuestro análisis permite predecir su rol en la unión de las coenzimas.

Los valores de pseudo-energía y energías libres calculados en para los casos recién presentados, pueden ser incorporados sobre los casos analizados en la base de datos de evaluación y observar su efecto sobre la correlación global con los datos experimentales. Como se espera, la Figura 23-a muestra que los casos estudiados en la validación de los potenciales estadísticos de NAD, siguen la misma tendencia de los datos de la base de evaluación, logrando incrementar la correlación de 0,76 a 0,8, con un cambio en la pendiente de $3,55 \times 10^{-3}$. Por su parte, los casos evaluados para NADP, disminuyen la correlación de 0,73 a 0,71, pero se mantienen dentro de la misma tendencia (Figura 23-b).

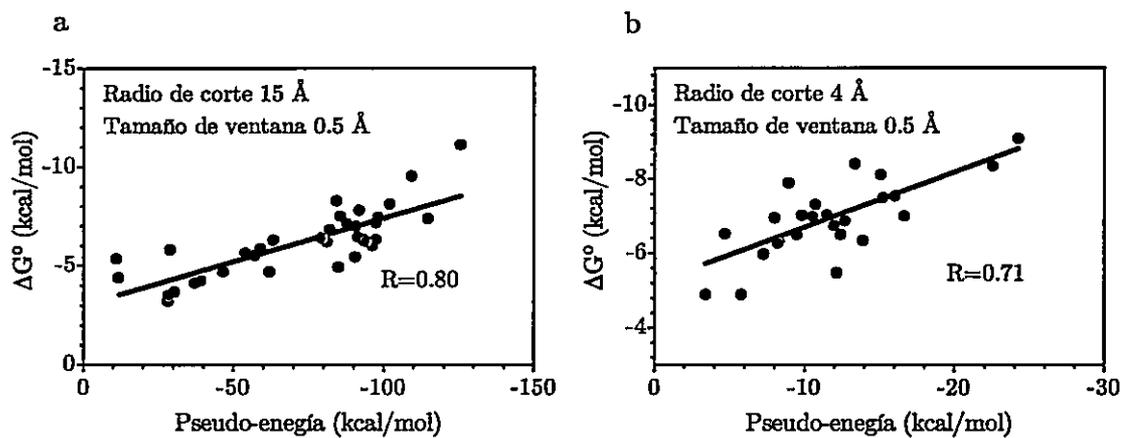


Figura 23 Análisis de correlación para los potenciales estadísticos con mejor evaluación (puntos negros) para NAD (a) y NADP (b), incluyendo los casos de validación de los potenciales estadísticos (puntos rojos).

Diseño racional mediante el uso de potenciales estadísticos

Finalmente, los potenciales estadísticos fueron utilizados para identificar mutaciones que puedan cambiar la preferencia por la coenzima en la enzima glucosa-6-fosfato deshidrogenasa de *Leuconostoc mesenteroides*. Esta es considerada como enzima dual, pero el cociente de sus constantes de especificidad indica que prefiere NADP 9,4 veces por sobre NAD (Vought et al. 2000). En la descripción de los potenciales estadísticos vimos que los residuos que desfavorecían en mayor medida la unión de NADP, favoreciendo a la vez la unión de NAD, eran Asp o Glu interactuando con los oxígenos 2' y 3' de la ribosa adyacente a la adenina (ver sección 3.1.2).

A partir de este resultado nos preguntamos entonces ¿en qué posición dentro del plegamiento debería estar posicionado un residuo de Asp para obtener la pseudo-energía más favorable para la unión de NAD y más desfavorable para la unión de NADP?. Para contestar esto generamos modelos *in silico* de la mutación de Asp en las posiciones 42 a la 51 en presencia de NAD o NADP, y calculamos la pseudo-energía de cada complejo modelado. La Figura 24-a muestra que la presencia de Asp en la posición 45 favorece en mayor medida la unión con NAD

y desfavorece la unión con NADP. La enzima fue caracterizada mediante cinética enzimática en nuestro laboratorio. Para la mutante Ala45Asp de esta enzima, la velocidad inicial depende hiperbólicamente de la concentración de coenzima, tanto con NAD como con NADP, de manera que los datos de fueron ajustados a la ecuación de Michaelis-Menten (Figura 24-b).

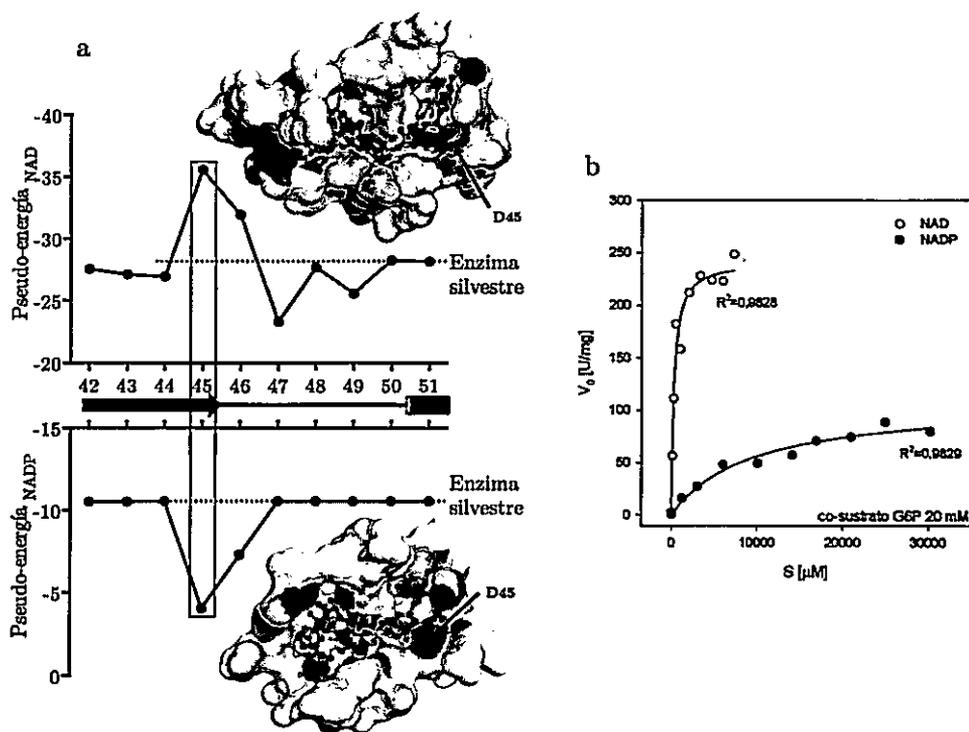


Figura 24. Pseudo-energía de los sitios de unión de NAD (arriba) y NADP (abajo) para la mutación por Asp en la posición de la 42 a la 51 en la estructura (a). La línea punteada en cada gráfico representa el valor de pseudo-energía de la enzima silvestre en cada complejo. En un rectángulo en rojo se destaca la posición donde el residuo de Asp produce una pseudo-energía más negativa con NAD y más positiva con NADP, y sobre cada gráfico se muestra la superficie de pseudo-energía para esta mutante. Gráfico de la dependencia de la velocidad inicial a diferentes concentraciones de NAD (puntos blancos) o NADP (puntos negros) (b). La velocidad inicial medida como $\mu\text{mol}/\text{min}$ (U) fue dividida por los miligramos de proteína utilizados en la cubeta de reacción, para normalizar aquellos casos en que se usaron cantidades diferentes de enzima.

3. RESULTADOS

Las constantes de especificidad de esta mutante muestran un cambio en la preferencia de 100 veces, favoreciendo a NAD sobre NADP (Tabla 7).

Tabla 7. Constantes de especificidad con NAD y NADP para la enzima glucosa-6-fosfato deshidrogenasa de *Leuconostoc mesenteroides* y la mutante A45D.

| | (k_{cat}/K_M) NADP | (k_{cat}/K_M) NAD | $\frac{(k_{cat}/K_M)_{NADP}}{(k_{cat}/K_M)_{NAD}}$ |
|------------------|-------------------------|------------------------|--|
| Enzima silvestre | 2530 | 269 | 9,4 |
| A45D | 0,63 | 50 | 0,01 |

Al validar la capacidad de los potenciales estadísticos para detectar cambios en la energías de unión de las mutantes y proponer mutaciones que cambien la preferencia de la enzima damos por completado el objetivo específico 4.

4. Discusión

En este trabajo describimos un nuevo método basado en potenciales estadísticos para caracterizar estructural y energéticamente las interacciones de NAD y NADP a sus sitios de unión en enzimas, y así representar las energías de unión de estas coenzimas.

Nuestro análisis de las estructuras en la base de datos PDB con NAD(P) determinó que la mayoría de los complejos muestran a la coenzima con todos sus átomos y más del 95% en una conformación extendida. Esto es equivalente a lo observado de Kuppuraj y colaboradores (Kuppuraj et al. 2011). Sin embargo, hasta ahora se desconocía si la distribución de frecuencias era equivalente para ambas coenzimas, como efectivamente lo muestra la Figura 11-a

Para derivar potenciales estadísticos es necesario utilizar una base de datos de entrenamiento que sea representativa y con la suficiente cantidad de estructuras que garantice una frecuencia de observaciones estadísticamente significativas, para poder transformarlas en pseudo-energías. De hecho, aunque los filtros

utilizados redujeron en un 28% la cantidad de estructuras, aún se mantuvo un número suficiente de observaciones para obtener perfiles de pseudo-energía similares a una función de energía clásica para la interacción entre pares de átomos. Es decir, con una repulsión a distancias pequeñas, seguidas de un mínimo de pseudo-energía negativa a una distancia determinada y una tendencia a cero a distancias mayores. Un potencial estadístico para interacciones no locales, es decir, entre átomos de la proteína y la coenzima no unidos covalentemente, contiene una colección completa de las posibles combinaciones para interacciones entre átomos de la proteína y la coenzima (Sippl 1996). Los perfiles de pseudo-energía obtenidos, muestran diferentes formas y mínimos energéticos dependiendo del tipo de interacción: puente de hidrógeno, electrostáticas, catión- π o aromáticas (Figuras 14-17).

En el caso de los puentes de hidrógeno, la comparación de los perfiles de pseudo-energía para NAD y NADP evidenció el fenómeno de "locking elbow" para la interacción entre el nitrógeno-7 de la nicotinamida y los oxígenos de la cadena principal. Por lo tanto, este puente de hidrógeno necesita energía tanto para formarse como romperse y junto al empacamiento de los residuos aromáticos

(Figura 14-a) permite estabilizar esta región para realizar la transferencia del hidruro. Notablemente, la ribosa adyacente a la nicotinamida presenta un “locking elbow” para el puente de hidrógeno con los oxígenos del carboxilato en el caso de los sitios de NADP (Figura 14-c). Es posible relacionar este resultado con la observación de Kuppuraj en el 2011 de que los ángulos diedros de la región ribosa-nicotinamida aparentan estar más restringidos en una base datos de complejos de 809 proteínas (Kuppuraj et al. 2011). Ya que ninguna otra parte de la coenzima presenta estos rasgos, parece ser que los sitios de unión de NAD(P), de manera general imponen una configuración especial a la región catalítica.

Kleiger & Eisenberg en el 2002 encontraron un motivo rico en glicinas que interactúa por puente de hidrógeno con el difosfato de NAD(P) (Kleiger & Eisenberg 2002), nuestros resultados indican que esta interacción posee su mínimo de energía a una distancia de 3 Å y es energéticamente similar para ambas coenzimas.

Al igual como fue descrito por Carugo y Argos en 1997 en los sitios de unión de NAD existe un puente de hidrógeno frecuente entre carboxilatos de Asp/Glu y los hidroxilos-2' y 3' de la ribosa adyacente a la adenina (Carugo & Argos 1997a). En presencia del fosfato-2' esta interacción es casi inexistente debido a la repulsión electrostática. Sin embargo, se observan frecuencias similares de interacción de Asp/Glu con la ribosa adyacente a la nicotinamida de ambas coenzimas. Otros residuos frecuentemente observados alrededor de la ribosa adyacente a la nicotinamida son lisinas y puentes de hidrógeno con Thr/Ser/Tyr. Estos resultados son consistentes con los de Kallberg en el 2002 quien hizo notar el involucramiento de residuos de Ser y el motivo YxxxK en la interacción con esta región (Kallberg et al. 2002).

En los sitios de unión de NADP, es frecuente encontrar la presencia de cargas positivas que interactúan electrostáticamente con el 2'-fosfato a una distancia máxima de 6 Å, con una energía cercana a las 2 kcal/mol. Cuando se trata de arginina, este residuo además forma una interacción catión- π con la adenina a un máximo de distancia de 6,5 Å, con una pseudo-energía similar a la de un puente de hidrógeno fuerte. El aporte de una mayor pseudo-energía a la unión de la



coenzima NADP sobre NAD, dado por la interacción catión- π con la adenina, se debe a la alta frecuencia de argininas en esta región. La presencia de este residuo de arginina con interacción dual ya había sido identificado por Carugo & Argos (1997), pero no había sido determinada su magnitud energética respecto a la unión de las coenzimas.

Nuestro análisis de la base de datos representativa para NADP indican que en el 85% de los sitios existe la presencia de una carga positiva, ya sea lisina o arginina cercana al fosfato-2', mientras que en el 20% de las estructuras ambos residuos cargados positivamente están interactuando con este grupo. Sin embargo, aún se desconoce si la selectividad por NADP está determinada por la cantidad de cargas positivas que interactúan con el fosfato-2'.

Nuestros resultados muestran que las interacciones aromáticas presentan un aporte a la energía de unión de la coenzima menor de lo que implica un puente de hidrógeno o una interacción electrostática. Esta característica en este tipo de interacciones ya ha sido reportada en el desarrollo de otros potenciales estadísticos entre residuos en las proteínas (Sippl 1996; Ferrada & Melo 2009).

Además de utilizar los potenciales estadísticos para describir los perfiles de pseudo-energía para las interacciones entre la coenzima y la proteína, optimizamos los parámetros, por los que estos fueron derivados para que permitieran obtener una mejor correlación entre las pseudo-energías y la afinidad experimental por la coenzima. La cantidad de estructuras en la base de datos de evaluación fue un 7% en comparación a la cantidad estructuras presentes en la base de datos de entrenamiento.

Para el potencial estadístico de NAD se obtuvieron correlaciones mejores usando potenciales estadísticos generados con distancias de corte más lejanas para la interacción entre pares de átomos (Figura 18). Esto es similar a lo obtenido por Li y colaboradores en el 2010, quienes estudiaron como variaba el coeficiente de correlación de Pearson entre los datos teóricos y experimentales para potenciales estadísticos dependientes de la distancia a nivel atómico para complejos proteína-proteína (Li et al. 2010). Por el contrario, los potenciales estadísticos para NADP presentan mejores coeficiente de correlación de Pearson a distancias de corte menores, disminuyendo drásticamente a distancias mayores (Figura 18).

Una posible explicación a esta diferencia en la evaluación se basa en la variabilidad conformacional que estas coenzimas presentan en los sitios de unión. Carugo & Argos (1997) indicaron que la conformación de NADP es más flexible que la de NAD, y que, a pesar que las interacciones proteína-coenzima son en gran parte conservadas en los complejos con NAD, son bastante variables con NADP. Kuppuraj posteriormente en el 2011 cuantificó 11 ángulos diedros en las coenzimas para identificar diferencias en estos ángulos, pero no destacaron la diferencia en las desviaciones estándar que estos ángulos poseen, las cuales indicaban que 9 de estos 11 ángulos diedros presentaban mayor desviación estándar con NADP que con NAD (Kuppuraj et al. 2011). Como vieron Carugo & Argos (1997), la mayor variabilidad conformacional de NADP se debe a que los residuos en el bolsillo son mas variables. Teniendo en cuenta esto, es esperable que en los sitios de unión de NAD sea más frecuente encontrar residuos similares entre las estructuras que en los bolsillos de NADP. Por lo tanto, al calcular las pseudo-energías de los sitios de unión de NAD en las estructuras de la base de datos de evaluación, será común encontrar residuos frecuentes en la base de datos de entrenamiento. Esto permite que estas frecuencias contribuyan con energías favorables tanto a cortas como largas

distancias, incrementando la correlación entre los datos teóricos y experimentales. Por el contrario, los sitios de unión de NADP al ser más variables, incrementan la posibilidad de encontrar residuos en las estructuras de la base de datos de evaluación que son poco comunes en la base de datos con la cual fueron derivados los potenciales estadísticos. Así estos residuos contribuirán con energías desfavorables, que serán menores a cortas distancias ya que interactúan con una menor cantidad de átomos de la coenzima, y que serán mayores al incluir interacciones con más átomos de la coenzima. Por esta razón observamos que a mayores distancias disminuye la correlación entre los datos teóricos y experimentales

Con anterioridad se han utilizado casos particulares de enzimas para validar la capacidad de los potenciales estadísticos (Zheng et al. 2007). En nuestro trabajo se analizaron tres casos para determinar la capacidad de los potenciales estadísticos para predecir el aporte a la energía de unión de la coenzima de: cadenas laterales individuales mutadas por alanina, múltiples mutaciones orientadas a incrementar la afinidad por NAD y múltiples mutaciones orientadas a cambiar la especificidad.

Empleando la estructura de la enzima 3- α -hidroxiesteroide deshidrogenasa de *Comamonas testosteroni* junto a las mutantes N96A y K159A en presencia de NAD determinamos que los potenciales estadísticos son competentes en detectar correctamente el aporte de cadenas laterales a la afinidad por la coenzima. Esto ya que la sumatoria de la pseudo-energía de los átomos de la cadena lateral de estos residuos (excluyendo el carbono- β) es proporcional al aumento de la energía libre de la coenzima. Nuestros resultados adicionalmente proponen a los residuos T11, D32, D41 como relevantes para la unión de la coenzima. Modelos de la enzima Acido 2,5-diceto-D-glucónico reductasa de *Corynebacterium* junto a la estructura de una mutante con la especificidad modificada de esta enzima fueron analizadas utilizando los potenciales estadísticos. Los resultados indican que los potenciales estadísticos asignan valores de pseudo-energía a partir de modelos proporcionales a los cambios energéticos experimentales en mutantes. Utilizando tanto modelos y estructuras de la enzima Gliceraldehído-3-fosfato deshidrogenasa de *Bacillus stearothermophilus* y mutantes de esta enzima diseñadas para cambiar la especificidad de esta enzima desde NAD a NADP,

determinamos que los potenciales estadísticos son apropiados para representar correctamente incrementos y disminuciones en la afinidad por las coenzimas.

Finalmente, se crearon mutantes *in silico* de la enzima glucosa-6-fosfato deshidrogenasa de *Leuconostoc mesenteroides* posicionando un aspartato en diferentes posiciones de la estructura cercana a la ribosa adyacente a la adenina con el objetivo de identificar que mutante permitía el cambio de especificidad de esta enzima desde NADP a NAD. Los resultados indicaron a la mutante A45D como aquella con la menor pseudo-energía con NAD y mayor con NADP. Los resultados experimentales indican un cambio en la especificidad en un factor 100 veces a favor NAD.

5. Conclusiones

- Para los complejos NAD(P)-proteína analizados, los perfiles de pseudo-energía de puentes de hidrógeno, interacciones electrostáticas y aromáticas representan correctamente las características esperadas para la energía de estas interacciones.
- De acuerdo con el análisis de las frecuencias de interacción por residuo, la característica mas general en los sitios de unión de NAD, es la presencia de una carga negativa que interactúa con la ribosa adyacente a la adenina de NAD, mientras que en los sitios de NADP, se encuentra una carga positiva que interactúa con el fosfato-2'.
- La mejor correlación entre los datos experimentales y pseudo-energías, se obtiene teniendo en cuenta las interacciones lejanas en el caso de NAD, y las interacciones cercanas con NADP.
- Los potenciales estadísticos optimizados (mejor correlación con datos experimentales) permiten calcular pseudo-energías que se correlacionan con las energías libres de unión de complejos de NAD y NADP.

5. CONCLUSIONES

- Los potenciales estadísticos pueden ser utilizados para proponer mutaciones para cambiar la especificidad por las coenzimas NAD(P) en glucosa-6-fosfato deshidrogenasa de *Leuconostoc mesenteroides*.

6. Referencias

- Ansari HR, Raghava GPS. 2010. Identification of NAD interacting residues in proteins. *BMC Bioinformatics* 11:1-8.
- Banta S, Swanson B a, Wu S, Jarnagin A, Anderson S. 2002. Optimizing an artificial metabolic pathway: engineering the cofactor specificity of *Corynebacterium* 2,5-diketo-D-gluconic acid reductase for use in vitamin C biosynthesis. *Biochemistry* 41:6226-6236.
- Bernard N, Johnsen K, Holbrook JJ, Delcour J. 1995. D175 discriminates between NADH and NADPH in the coenzyme binding site of *Lactobacillus delbrueckii* subsp. *bulgaricus* D-lactate dehydrogenase. *Biochem. Biophys. Res. Commun.* 208:895-900.
- Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535-542.
- Boltzmann L. 1896. Theorie des Gase mit einatomigen Molekülen, deren dimensionen gegen die Mittlere weglänge verschwinden, p. 1-230. En Barth, JA (ed.), Vorlesungen über Gastheorie. Leipzig, J. A. Barth, New York Public Library.
- Bryce C. 1979. SAM - semantics and misunderstandings. *Trends Biochem. Sci.* 4:62-63.
- Carugo O, Argos P. 1997a. NADP-dependent enzymes. I: Conserved stereochemistry of cofactor binding. *Proteins* 28:10-28.
- Carugo O, Argos P. 1997b. NADP-dependent enzymes. II: Evolution of the mono- and dinucleotide binding domains. *Proteins* 28:29-40.
- Chang Y-H, Wang C-Z, Chiu C-C, Chuang L-Y, Hwang C-C. 2010. Contributions of active site residues to cofactor binding and catalysis of 3 α -hydroxysteroid dehydrogenase/carbonyl reductase. *Biochim. Biophys. Acta* 1804:235-241.

- Chen R, Greer A, Dean A. 1995. A highly active decarboxylating dehydrogenase with rationally inverted coenzyme specificity. *Proc. Natl. Acad. Sci. U. S. A.* 92:11666-11670.
- Chen R, Greer ANN, Dean AM. 1996. Redesigning secondary structure to invert coenzyme specificity in isopropylmalate dehydrogenase. *Proc. Natl. Acad. Sci. U. S. A.* 93:12171-12176.
- Clark M, Cramer RD, Van Opdenbosch N. 1989. Validation of the general purpose tripos 5.2 force field. *J. Comput. Chem.* 10:982-1012.
- Clermont S, Corbier C, Mely Y, Gerard D, Wonacott A, Branlant G. 1993. Determinants of coenzyme specificity in glyceraldehyde-3-phosphate dehydrogenase: role of the acidic residue in the fingerprint region of the nucleotide binding fold. *Biochemistry* 32:10178-10184.
- Dambe TR, Kühn AM, Brossette T, Giffhorn F, Scheidig AJ. 2006. Crystal structure of NADP(H)-dependent 1,5-anhydro-D-fructose reductase from *Sinorhizobium morelense* at 2.2 Å resolution: construction of a NADH-accepting mutant and its application in rare sugar synthesis. *Biochemistry* 45:10030-10042.
- Denessiouk KA, Rantanen V V, Johnson MS. 2001. Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins. *Proteins* 44:282-291.
- Didierjean C, Rahuel-Clermont S, Vitoux B, Dideberg O, Branlant G, Aubry A. 1997. A crystallographic comparison between mutated glyceraldehyde-3-phosphate dehydrogenases from *Bacillus stearothermophilus* complexed with either NAD⁺ or NADP⁺. *J. Mol. Biol.* 268:739-759.
- Ehrensberger AH, Elling R a, Wilson DK. 2006. Structure-guided engineering of xylitol dehydrogenase cosubstrate specificity. *Structure* 14:567-575.
- Elmore CL, Porter TD. 2002. Modification of the nucleotide cofactor-binding site of cytochrome P-450 reductase to enhance turnover with NADH *in vivo*. *J. Biol. Chem.* 277:48960-48964.
- Eppink MH, Overkamp KM, Schreuder HA, Van Berkel WJ. 1999. Switch of coenzyme specificity of p-hydroxybenzoate hydroxylase. *J. Mol. Biol.* 292:87-96.

- Eswar N, Webb B, Marti-Renom M a, Madhusudhan MS, Eramian D, Shen M-Y, Pieper U, Sali A. 2007. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* 2:1-31.
- Ferrada E, Melo F. 2007. Nonbonded terms extrapolated from nonlocal knowledge-based energy functions improve error detection in near-native protein structure models. *Protein Sci.* 16:1410-1421.
- Ferrada E, Melo F. 2009. Effective knowledge-based potentials. *Protein Sci.* 18:1469-1485.
- Fersht A. 1999. Enzyme-substrate complementarity and the use of binding energy in catalysis, p. 349-376. En Russel, M (ed.), *Structure and Mechanism in Protein Science*. W. H. Freeman and Company, New York.
- Feyfant E, Sali A, Fiser A. 2007. Modeling mutations in protein structures. *Protein Sci.* 16:2030-41.
- Galkin A, Kulakova L, Ohshima T, Esaki N, Soda K. 1997. Construction of a new leucine dehydrogenase with preferred specificity for NADP+ by site-directed mutagenesis of the strictly NAD+-specific enzyme. *Protein Eng.* 10:687-690.
- Grimm C, Maser E, Möbus E, Klebe G, Reuter K, Ficner R. 2000. The crystal structure of 3alpha -hydroxysteroid dehydrogenase/carbonyl reductase from *Comamonas testosteroni* shows a novel oligomerization pattern within the short chain dehydrogenase/reductase family. *J. Biol. Chem.* 275:41333-41339.
- Guse AH, Lee HC. 2008. NAADP: a universal Ca²⁺ trigger. *Sci. Signal.* 1:re10.
- Holmberg N, Ryde U, Bülow L. 1999. Redesign of the coenzyme specificity in L-lactate dehydrogenase from *Bacillus stearothermophilus* using site-directed mutagenesis and media engineering. *Protein Eng.* 12:851-856.
- Hsieh J-Y, Liu G-Y, Chang G-G, Hung H-C. 2006. Determinants of the dual cofactor specificity and substrate cooperativity of the human mitochondrial NAD(P)+-dependent malic enzyme: functional roles of glutamine 362. *J. Biol. Chem.* 281:23237-23245.
- Hu L, Benson ML, Smith RD, Lerner MG, Carlson H a. 2005. Binding MOAD (Mother Of All Databases). *Proteins* 60:333-340.

- Jeffrey GA. 1997. Weak Hydrogen Bonds, p. 1-303. En Walsh, EJ (ed.), An Introduction to Hydrogen Bonding. Oxford University Press, New York.
- Jiang L, Gao Y, Mao F, Liu Z, Lai L. 2002. Potential of mean force for protein-protein interaction studies. *Proteins* 46:190-196.
- Jones DT, Taylor WR, Thornton JM. 1992. A new approach to protein fold recognition. *Nature* 358:86-89.
- Kallberg Y, Oppermann U, Jörnvall H, Persson B. 2002. Short-chain dehydrogenases/reductases (SDRs). *Eur. J. Biochem.* 269:4409-4417.
- Kasahara K, Kinoshita K, Takagi T. 2010. Ligand-binding site prediction of proteins based on known fragment-fragment interactions. *Bioinformatics* 26:1493-1499.
- Kleiger G, Eisenberg D. 2002. GXXXG and GXXXA Motifs Stabilize FAD and NAD(P)-binding Rossmann Folds Through C α -H...O Hydrogen Bonds and van der Waals Interactions. *J. Mol. Biol.* 323:69-76.
- Kristan K, Stojan J, Adamski J, Lanisnik Rizner T. 2007. Rational design of novel mutants of fungal 17 β -hydroxysteroid dehydrogenase. *J. Biotechnol.* 129:123-130.
- Kuppuraj G, Sargsyan K, Hua Y-H, Merrill a R, Lim C. 2011. Linking distinct conformations of nicotinamide adenine dinucleotide with protein fold/function. *J. Phys. Chem. B* 115:7932-7939.
- Lesk AM. 1995. NAD-binding domains of dehydrogenases. *Curr. Opin. Struct. Biol.* 5:775-783.
- Li X-L, Hou M-L, Wang S-L. 2010. A residual level potential of mean force based approach to predict protein-protein interaction affinity. *Adv. Intell. Comput. Theor. Appl.* 6215:680-686.
- Liang L, Zhang J, Lin Z. 2007. Altering coenzyme specificity of *Pichia stipitis* xylose reductase by the semi-rational approach CASTing. *Microb. Cell Fact.* 6:1-11.
- Liou G-G, Tanny JC, Kruger RG, Walz T, Moazed D. 2005. Assembly of the SIR complex and its regulation by O-acetyl-ADP-ribose, a product of NAD-dependent histone deacetylation. *Cell* 121:515-527.

- Marohnic CC, Bewley MC, Barber MJ. 2003. Engineering and characterization of a NADPH-utilizing cytochrome b5 reductase. *Biochemistry* 42:11170-11182.
- Medina M, Luquita A, Tejero J, Hermoso J, Mayoral T, Sanz-Aparicio J, Grever K, Gomez-Moreno C. 2001. Probing the determinants of coenzyme specificity in ferredoxin-NADP⁺ reductase by site-directed mutagenesis. *J. Biol. Chem.* 276:11902-11912.
- Melo F, Feytmans E. 1997. Novel knowledge-based mean force potential at atomic level. *J. Mol. Biol.* 267:207-222.
- Miller S, Lunzer M, Dean A. 2006. Direct demonstration of an adaptive constraint. *Science* 314:458-461.
- Mitchell JBO, Laskowski RA, Alex A, Thornton JM. 1999. BLEEP-potential of mean force describing protein-ligand interactions: I. Generating potential. *J. Comput. Chem.* 20:1165-1176.
- Miyazawa S, Jernigan RL. 1985. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* 18:534-552.
- Morimoto Y, Honda K, Ye X. 2014. Directed evolution of thermotolerant malic enzyme for improved malate production. *J. Biosci. Bioeng.* 117:147-152.
- Naylor CE, Gover S, Basak AK, Cosgrove MS, Levy HR, Adams MJ. 2001. NADP⁺ and NAD⁺ binding to the dual coenzyme specific enzyme *Leuconostoc mesenteroides* glucose 6-phosphate dehydrogenase: different interdomain hinge angles are seen in different binary and ternary complexes. *Acta Crystallogr. D. Biol. Crystallogr.* 57:635-648.
- Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BØ. 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. *Mol. Syst. Biol.* 7:1-9.
- Petschacher B, Leitgeb S, Kavanagh KL, Wilson DK, Nidetzky B. 2005. The coenzyme specificity of *Candida tenuis* xylose reductase (AKR2B5) explored by site-directed mutagenesis and X-ray crystallography. *Biochem. J.* 385:75-83.

Petterson EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605-1612.

Rane MJ, Calvo KC. 1997. Reversal of the Nucleotide Specificity of Ketol Acid Reductoisomerase by Site-Directed Mutagenesis Identifies the NADPH Binding Site. *Arch. Biochem. Biophys.* 338:83-89.

Sanli G, Banta S, Anderson S, Blaber M. 2004. Structural alteration of cofactor specificity in *Corynebacterium* 2,5-diketo-D-gluconic acid reductase. *Protein Sci.* 13:504-512.

Schreiber V, Dantzer F, Ame J-C, de Murcia G. 2006. Poly(ADP-ribose): novel functions for an old molecule. *Nat. Rev. Mol. Cell Biol.* 7:517-528.

Schuber F, Lund FE. 2004. Structure and enzymology of ADP-ribosyl cyclases: conserved enzymes that produce multiple calcium mobilizing metabolites. *Curr. Mol. Med.* 4:249-261.

Scrutton NS, Berry A, Perham RN. 1990. Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature* 343:38-43.

Serov AE, Popova AS, Fedorchuk V V, Tishkov VI. 2002. Engineering of coenzyme specificity of formate dehydrogenase from *Saccharomyces cerevisiae*. *Biochemistry* 367:841-847.

Shiraishi N, Croy C, Kaur J, Campbell WH. 1998. Engineering of pyridine nucleotide specificity of nitrate reductase: mutagenesis of recombinant cytochrome b reductase fragment of *Neurospora crassa* NADPH:Nitrate reductase. *Arch. Biochem. Biophys.* 358:104-115.

Sippl M. 1990. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* 213:859-883.

Sippl M. 1993a. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comput. Aided. Mol. Des.* 7:473-501.

- Sippl M. 1993b. Recognition of errors in three-dimensional structures of proteins. *Proteins Struct. Funct. Bioinforma.* 17:355-62.
- Sippl M. 1996. Helmholtz free energy of peptide hydrogen bonds in proteins. *J. Mol. Biol.* 260:644-648.
- Tanaka S, Scheraga HA. 1976. Medium- and Long-Range Interaction Parameters between Amino Acids for Predicting Three-Dimensional Structures of Proteins. *Macromolecules* 9:945-950.
- Vought V, Ciccone T, Davino M, Fairbairn L. 2000. Delineation of the roles of amino acids involved in the catalytic functions of *Leuconostoc mesenteroides* glucose 6-phosphate dehydrogenase. *Biochemistry* 39:15012-15021.
- Watanabe S, Kodaki T, Makino K. 2005. Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc. *J. Biol. Chem.* 280:10340-10349.
- Woodyer R, van der Donk WA, Zhao H. 2003. Relaxing the nicotinamide cofactor specificity of phosphite dehydrogenase by rational design. *Biochemistry* 42:11604-11614.
- Yaoi T, Miyazaki K, Oshima T, Komukai Y, Go M. 1996. Conversion of the coenzyme specificity of isocitrate dehydrogenase by module replacement. *J. Biochem.* 119:1014-1018.
- Zhang L, Ahvazi B, Szittner R, Vrielink A, Meighen E. 1999. Change of nucleotide specificity and enhancement of catalytic efficiency in single point mutants of *Vibrio harveyi* aldehyde dehydrogenase. *Biochemistry* 38:11440-11447.
- Zheng S, Robertson TA, Varani G. 2007. A knowledge-based potential function predicts the specificity and relative binding energy of RNA-binding proteins. *FEBS J.* 274:6378-6391.

Apéndice

Correlación con parámetros cinéticos

Muchos de los estudios orientados a cambiar la preferencia de sitios de unión por las coenzimas NAD y NADP no determinan constantes de disociación sino que calculan constantes de especificidad, es decir el cociente entre la constante catalítica (k_{cat}) y la constante de Michaelis-Menten (K_M), para observar el efecto de mutaciones. A su vez, de acuerdo con Fersht (Fersht 1999), la k_{cat}/K_M está relacionada con la energía del estado de transición a la enzima, mientras que la K_M está relacionada con la energía libre del estado basal. Por lo tanto, quisimos determinar si se correlacionaba la pseudo-energía con el cambio en la energía libre del estado de transición calculado a partir del cociente entre las constantes de especificidad entre una enzima silvestre y una mutante (Fersht 1999; Banta et al. 2002).

$$\Delta\Delta G_b^\ddagger = RT \ln \frac{k_{cat}/K_M(\text{enzima silvestre})}{k_{cat}/K_M(\text{enzima mutante})} \quad (16)$$

En este caso, utilizamos como modelo de estudio la enzima Glutación reductasa de *Escherichia coli* preferente por NADP, la cual cuenta con su estructura determinada (Scrutton et al. 1990). Dado que a la fecha no existía estructura para las mutantes construidas en dicho estudio, y que en un caso analizado con anterioridad (Acido 2,5-diceto-D-glucónico reductasa de *Corynebacterium sp.*) se demostró que los modelos *in silico* son susceptibles de ser analizados por potenciales estadísticos, se decidió modelar las mutantes mediante mutagénesis *in silico*.

En la Tabla 8 se muestran los valores de K_M y k_{cat}/K_M para NADP como co-sustrato, de la enzima silvestre y 10 mutantes, junto a las estimaciones de la pseudo-energía de los sitios de unión en los respectivos modelos. La Figura 25 muestra que las energías del estado de transición calculadas a partir de constantes de especificidad y las pseudo-energías se correlacionan con un coeficiente de Pearson de 0,83. Sin embargo, existe una menor correlación (0,37) al tomar como parámetro la constante de Michaelis-Menten para calcular la energía del estado basal.

Tabla 8. Energía libre y pseudo-energía para la unión de NADP a la enzima Glutación reductasa de *Escherichia coli* y mutantes diseñadas para cambiar la especificidad por coenzimas en esta enzima.

| | NADP | | | | |
|---|-------------|--------------------|--------------------------|-----------------------------|----------------|
| | K_M | $\Delta\Delta G_b$ | k_{cat}/K_M | $\Delta\Delta G_b^\ddagger$ | Pseudo-energía |
| | (μM) | (kcal/mol) | ($\mu M^{-1}min^{-1}$) | (kcal/mol) | (kcal/mol) |
| Enzima silvestre | 21,7 | 0,00 | 704 | 0,00 | -7,45 |
| R204L | 450 | -1,76 | 12,8 | 2,33 | -6,37 |
| R198M | 570 | -1,90 | 8,5 | 2,57 | -6,24 |
| R204L/R198M | 480 | -1,80 | 1,6 | 3,54 | -5,17 |
| A179G | 12 | 0,34 | 650 | 0,05 | -7,45 |
| R204L/A179G | 723 | -0,70 | 172 | 0,82 | -6,37 |
| R198M/A179G | 72,6 | -2,04 | 21,6 | 2,03 | -6,24 |
| R204L/A179G/R198M | 27 | -0,13 | 18 | 2,13 | -5,17 |
| R204L/R198M/A179G/A183G | 140 | -1,09 | 12,9 | 2,33 | -5,17 |
| R198M/A179G/A183G/R204P | 80,3 | -0,76 | 15,6 | 2,22 | -5,02 |
| R198M/A179G/A183G/R204P/ A197G/K199F/H200D | 220 | -1,35 | 3 | 3,18 | -5,02 |

A pesar de la correlación observada con las energías derivadas de las constantes de especificidad, podemos ver que para los casos donde el cambio en la energía libre del estado de transición es cercana a 2 kcal/mol los potenciales asignan valores de pseudo-energía entre -6,37 y -5,02, lo cual implica que no discriminan correctamente las diferencias asociadas a estas mutaciones. A pesar de esto los potenciales estadísticos permiten asignar las pseudo-energías más desfavorables a

las 2 mutaciones (R204L/R198M y R198M/A179G/A183G/R204P/A197G/
K199F/H200D) con una constante de especificidad más baja.

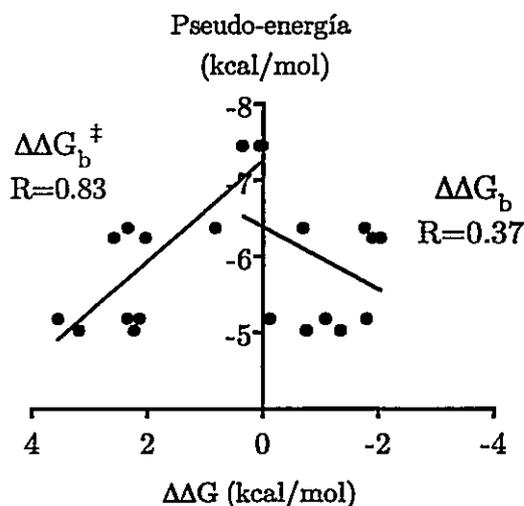


Figura 25. Correlación entre la pseudo-energía del sitio de unión de NADP, en la enzima Glutación reductasa de *Escherichia coli* y mutantes, y el cambio en energía libre del estado basal del complejo calculada a partir de las constantes de Michaelis-Menten (rojo) o el cambio en la energía libre calculado a partir de las constantes de especificidad (azul).

Anexos

Lista de PDB ID de la base de datos de entrenamiento:

NAD(H)

1A5Z, 1ARZ, 1B8V, 1BDB, 1BMD, 1BPW, 1BXS, 1C1D, 1CDO, 1D7O, 1DHR, 1DLJ, 1DXY, 1E3I, 1EBF, 1EE9, 1EJ2, 1EK6, 1EMD, 1EVJ, 1EVZ, 1EZ4, 1F0Y, 1F8F, 1FK8, 1FMC, 1GEE, 1GEG, 1GEU, 1GR0, 1GUZ, 1GY8, 1HDG, 1HEX, 1HYH, 1I24, 1ICI, 1IY8, 1J49, 1J5P, 1JQ5, 1K4M, 1K6X, 1KAE, 1KEW, 1KOL, 1KQN, 1L7E, 1LC3, 1LDN, 1LJ8, 1LLD, 1LSS, 1LVL, 1LW7, 1M9H, 1MEW, 1MI3, 1MX3, 1N2S, 1NFB, 1NFF, 1NHW, 1NUU, 1NVM, 1O04, 1O0S, 1O6Z, 1OBB, 1OC2, 1OG3, 1OJZ, 1OMO, 1ORR, 1P1J, 1P9L, 1PJ3, 1PJC, 1PJS, 1PL8, 1PZH, 1QS2, 1QSG, 1R6D, 1RKX, 1ROZ, 1S20, 1S7G, 1SB8, 1SBY, 1SC6, 1SG6, 1T2D, 1T90, 1TAE, 1U8X, 1UDB, 1UP7, 1UR5, 1UWK, 1UXT, 1V9L, 1VBI, 1VKO, 1VL0, 1VM6, 1WDK, 1WNB, 1WPQ, 1WWK, 1WXH, 1X14, 1X7D, 1X87, 1XAH, 1Z2I, 1Z45, 1ZBQ, 1ZEM, 1ZJY, 1ZMD, 2A5F, 2A9K, 2AG5, 2B4R, 2B69, 2BI4, 2C5A, 2CFC, 2CZC, 2D1Y, 2D4E, 2D8A, 2DC1, 2DFD, 2DPH, 2DT5, 2DTE, 2DVM, 2EER, 2EJV, 2EKL, 2EKP, 2EP7, 2EQ7, 2EWM, 2FKN, 2FM3, 2FOI, 2G5C, 2G76, 2G82, 2G8Y, 2GAG, 2GDZ, 2GRU, 2GSD, 2GWL, 2H4H, 2H7I, 2HMT, 2HSD, 2HUN, 2I65, 2I9P, 2IXA, 2IZZ, 2J5N, 2J6L, 2JHF, 2NPX, 2O23, 2O4C, 2P5Y, 2PD4, 2PD6, 2PH5, 2PV7, 2PZM, 2Q1S, 2Q1W, 2Q2V, 2Q3E, 2QJO, 2RAB, 2RC3, 2VHX, 2W2L, 2WN7, 2WSB, 2WYV, 2X06, 2X0I, 2XXJ, 2YQ5, 2YVF, 2YY7, 3A28, 3ABI, 3ADP, 3AJR, 3AK4, 3AMZ, 3B1F, 3B4W, 3B6J, 3B82, 3BTS, 3C7A, 3CE6, 3CEA, 3CGD, 3CIN, 3CLH, 3CMC, 3CPS, 3D4P, 3D64, 3DOC, 3E18, 3EC7, 3EFV, 3EHE, 3ENK, 3FHM, 3GD4, 3GEG, 3GGO, 3GNQ, 3GVH, 3GVP, 3H9E, 3H9U, 3HAZ, 3HJA, 3HL0, 3I0P, 3IDS, 3IJR, 3JSA, 3JU8, 3JV7, 3JYO, 3JZD, 3KB6, 3KEO, 3KFT, 3KO8, 3LN3, 3LT0, 3LVF, 3M2T, 3M6I, 3MW9, 3N58, 3N7U, 3NRC, 3NT2, 3O9Z, 3OA2, 3OET, 3OIG, 3OJO, 3OKF, 3OND, 3ORF, 3OX4, 3P2O, 3PGX, 3PLR, 3PQD, 3PVZ, 3PXX, 3PYM, 3Q2I, 3Q3C, 3QBD, 3QJ5, 3QV1, 3QVS, 3R3S, 3RF7, 3RFV, 3RIY, 3RUF, 3RWB, 3S2E, 3S55, 3SX2, 3SXP, 3SYT, 3T4E, 3T7C, 3TNL, 3TSC, 3U31, 3UH1, 3UKO, 3UVE, 3UXY, 3V1U, 3V9L, 3VPS, 3W2G, 3W8E, 3WMX, 3ZCX, 3ZOK, 3ZU3, 3ZWC, 4B79, 4BBJ, 4BKU, 4C3S, 4DYD, 4E5N, 4EGB, 4EUE, 4F3X, 4FN4, 4FRY, 4G6H, 4GH5, 4GKV, 4GLO, 4H03, 4H8A, 4HAN, 4I1I, 4I3X, 4I4B, 4I5I, 4I8P, 4ID9, 4IHI, 4IIN, 4ILK, 4IM7, 4IS3, 4IYM, 4J1T, 4J4H, 4JNK, 4K28, 4KQX, 4L2O, 4LIS, 4LRS, 4LSM, 4LTN, 4M89, 4MP8, 4MPY, 4NBT, 4NBU, 4NBW, 4NK5, 4O1M, 5MDH

NADP(H)

1A27, 1A4I, 1AE1, 1AOE, 1C3V, 1CF2, 1CIV, 1CYD, 1D1G, 1D4O, 1DIH, 1DJL, 1DQA, 1E1M, 1E5Q, 1E7S, 1E7W, 1EDO, 1EQ2, 1EZ0, 1F06, 1G0O, 1GET, 1GIR, 1GQ2, 1GUF, 1GVE, 1H5Q, 1H6D, 1H7X, 1HDO, 1HYE, 1I36, 1IYZ, 1J3K, 1JA1, 1JA9, 1JAY, 1JUV, 1K0J, 1K8C, 1KC1, 1KEV, 1KMV, 1LQA, 1LUA, 1MXH, 1N7H, 1NVT, 1NYT, 1O2D, 1OAA, 1OG6, 1P0F, 1PGO, 1PNO, 1PQU, 1PS9, 1PZ0, 1PZ1, 1Q0Q, 1QFZ, 1QOR, 1QP8, 1QYV, 1RM4, 1RPN, 1RYD, 1S1P, 1SNY, 1T2A, 1TCS, 1TLL, 1U2G, 1US0, 1UXN, 1V3V, 1V9N, 1VBJ, 1VL8, 1VLJ, 1VP5, 1W6U, 1WMA, 1XG5, 1XHL, 1XKQ, 1XQ6, 1XU9, 1Y7T, 1YB5, 1YNQ, 1YQD, 1YS4, 1YVE, 1Z82, 1ZDL, 1ZGD, 1ZH8, 1ZK4, 2A87, 2AE2, 2AG8, 2AHR, 2AZN, 2B4Q, 2B5W, 2BD0, 2BGS, 2BH9, 2BKA, 2BL9, 2BSA, 2C0C, 2C29, 2CDC, 2CF6, 2CVZ, 2CWH, 2CY0, 2D2I, 2D5N, 2DBQ, 2EHQ, 2F1K, 2FR1, 2GCG, 2GGS, 2GN4, 2GQ2, 2GV8, 2H63, 2HK9, 2HXV, 2I3G, 2I76, 2IZ0, 2J8Z, 2JAH, 2JL1, 2O7P, 2O7S, 2Q0L, 2QE0, 2QHX, 2R6J, 2RAF, 2RCY, 2RHC, 2RIR, 2V6G, 2VN8, 2VQ3, 2VUU, 2W3W, 2WM3, 2WN6, 2WOW, 2WZW, 2X6T, 2X99, 2X9G, 2XVH, 2XW7, 2Y5D, 2YDX, 2YJZ, 2YLY, 2YUT, 2YW9, 2YYY, 2Z1M, 2ZAT, 2ZB4, 2ZCV, 2ZZC, 3A06, 3AFN, 3AI3, 3ASV, 3AU8, 3AXK, 3BAZ, 3BUV, 3C1O, 3CGE, 3CSE, 3D3F, 3D3W, 3DFR, 3DJJ, 3DTT, 3E8X, 3EAU, 3ETD, 3F8R, 3GQV, 3GWF, 3GY0, 3H2S, 3H4G, 3HJ3, 3HSH, 3HWR, 3I6I, 3IA4, 3ICC, 3IJP, 3ING, 3INV, 3IUP, 3IX9, 3JYN, 3JZ4, 3KBO, 3KGY, 3KRB, 3KVO, 3LNS, 3MJE, 3NGL, 3NRR, 3NX4, 3NZB, 3O03, 3O26, 3OID, 3OJF, 3OP4, 3P19, 3PDU, 3PEF, 3PHI, 3PYX, 3PZR, 3Q6J, 3QG2, 3QWB, 3QWF, 3RC1, 3RG9, 3RHL, 3RKU, 3S9U, 3SC6, 3SEF, 3SJ7, 3SJU, 3SLK, 3SQY, 3TOX, 3TQ8, 3TQH, 3TRI, 3TWO, 3U4C, 3UCE, 3UF0, 3UOY, 3W6Z, 3WBB, 3WCZ, 3WG6, 3ZBR, 3ZHB, 3ZQA, 4A03, 4A0S, 4AG3, 4ALK, 4AOS, 4AUB, 4B4O, 4B63, 4B7X, 4BMV, 4BVA, 4DMM, 4DPL, 4E5M, 4E5Y, 4EIL, 4EJO, 4EJM, 4ESO, 4FC7, 4FDA, 4G5D, 4G5H, 4GCM, 4GI2, 4GIE, 4GMG, 4GVX, 4H4U, 4H8N, 4HA9, 4HFM, 4HSH, 4HP8, 4HXY, 4I5E, 4IDC, 4IUU, 4IJR, 4IMP, 4IMR, 4IQG, 4ITB, 4J1Q, 4J1T, 4J2O, 4JB1, 4JBI, 4JRO, 4K6F, 4K7Z, 4KJJ, 4KL9, 4KOA, 4KP7, 4KQW, 4KWH, 4L04, 4L4X, 4L8V, 4M7U, 4N5N, 4NHE, 4OL9

Lista de PDB ID de la base de datos de evaluación:

NAD(H)

1DBV, 1GZF, 1H9A, 1HZZ, 1NM5, 1R37, 1TOX, 2AQ8, 2AQH, 2AQI, 2AQK, 2C8H, 2DKN, 2FZW, 2JHG, 2OOR, 2VHW, 3DBV, 3HAD, 3OEW, 3OEY, 3OF2, 3Q9O, 3RJ5, 3RJ9, 3UQ8, 4EMI

NADP(H)

1H9A, 1Q7B, 1TI7, 1XJB, 1YJQ, 2DBV, 2EXX, 2GZ1, 2OFF, 2VNH, 2VNJ, 2W9G, 2W9S, 2XNJ, 3CAQ, 3GR6, 3H7R, 3H7U, 3LZW, 3LZX, 3RHQ, 3RHR, 3ULK, 4DBV