

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE QUÍMICA

**OBTENCIÓN *IN SILICO* DE UN MODELO PARA LA ESTRUCTURA
TRIDIMENSIONAL DE LA PROTEÍNA NATRXH DE *NICOTIANA ALATA***

TESIS

QUE PARA OBTENER EL TÍTULO DE

QUÍMICA FARMACÉUTICA BIÓLOGA

PRESENTA

ALEJANDRA ZAVALA CASTILLO

MÉXICO, D.F.

2009



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

JURADO ASIGNADO:

PRESIDENTE: PROFESOR: ROSARIO ADELAIDA MUÑOZ CLARES

VOCAL: PROFESOR: FELIPE CRUZ GARCÍA

SECRETARIO: PROFESOR: PERLA DEYANIRA MALDONADO JIMÉNEZ

1ER. SUPLENTE: PROFESOR: LAURA CARMONA SALAZAR

2° SUPLENTE: PROFESOR: TZVETANKA DIMITROVA DINKOVA

SITIO DONDE SE DESARROLLÓ EL TEMA:

DEPARTAMENTO DE BIOQUÍMICA, FACULTAD DE QUÍMICA, UNAM

ASESOR DEL TEMA:

DR. FELIPE CRUZ GARCÍA

SUPERVISOR TÉCNICO:

DR. ROGELIO RODRÍGUEZ SOTRES

SUSTENTANTE:

ALEJANDRA ZAVALA CASTILLO

Apoyos

Este trabajo fue apoyado por los proyectos:

CONACYT 81968

DGAPA IN205009-3

Agradecimientos

Esta tesis está dedicada a todas las personas bonitas que me han apoyado durante toda mi vida y en la realización de la tesis

En primera instancia, agradezco a mis padres, Joel y Ofelia, por haberme apoyado durante toda mi vida y por dejarme hacer lo que he querido, aunque eso ha significado regaños que al final han sido de mucha ayuda para mí.

Agradezco a mis hermanos, Ernesto y Nicté, que también me han apoyado mucho. Ernesto, ayudándome por el lado biológico del cual sigo aprendiendo y Nicté por echarme muchas porras y emocionarse por lo que hago :D

Agradezco a mis amigos, que aunque son poquititos, siempre han estado ahí conmigo, en las buenas y en las malas. Dicen, que durante la licenciatura es cuando encuentras a tus amigos, pues sí, en mi caso lo creo. Gracias a Fabian, Ezra y Nayeli.

Agradezco al amor de mi vida, John, que me ha brindado apoyo, me ha tenido mucha paciencia y siempre ha estado ahí cuando más lo he necesitado. Por eso y más te lo agradezco desde el fondo de mi corazón.

Por supuesto, por el lado académico también agradezco a muchísimas personas que me han ayudado. Primero a mi asesor el Dr. Felipe Cruz García, el cual me permitió ser parte de su grupo de trabajo y me brindó consejos, paciencia y muchas opiniones que me sirvieron para sacar adelante el proyecto.

También agradezco al Dr. León Patricio Martínez Castilla y al Dr. Rogelio Rodríguez Sotres por su ardua colaboración. Yo estoy muy consciente que sin su ayuda no hubiera podido modelar mi proteína hasta un nivel tan satisfactorio. A León le agradezco su ayuda en la parte bioinformática que aún ahora ha sido muy importante para mí. Por supuesto sus porras también han sido bien recibidas. A Rogelio toda la paciencia que ha tenido conmigo y su gran interés porque la proteína haya llegado a un modelo tan satisfactorio.

Solo me falta agradecerle a todos mis compañeros de laboratorio del 104 y de otros laboratorios, a Alejandra, Karina, Mariana, Ernesto, Rodrigo, Yuridia, Gregorio, Andrea y Claudia. Cada uno de ustedes me ha enseñado cosas muy importantes, tanto a nivel académico como personal y por ello les agradezco infinitamente.

A todos y cada uno de ustedes les doy las gracias por su ayuda y colaboración.

Índice General

ÍNDICE DE FIGURAS	iv
ÍNDICE DE TABLAS	vii
ABREVIATURAS	viii
I. INTRODUCCIÓN	1
I. 1. En busca de los genes modificadores de estilos de <i>Nicotiana alata</i>	1
I.1. 2. El cDNA-AFLP25F codifica para una tiorredoxina tipo <i>h</i>	1
I.2. Tiorredoxinas.....	3
I.2.1. Tiorredoxina reductasa.....	5
I.2.3. Tiorredoxinas vegetales.....	7
I.2.4. Tiorredoxinas tipo <i>h</i>	8
I.3. NaTrxh es una tiorredoxina tipo <i>h</i>	10
II. ANTECEDENTES	12
II.1. La secuencia de la NaTrxh contiene la información suficiente para promover su secreción.....	12
II.2. Secreción por una vía independiente al Retículo Endoplásmico/Aparato de Golgi de NaTrxh.....	14
II. 3. La región N-terminal, específicamente la región N β , contiene la información necesaria para promover la secreción de NaTrxh.....	18
II. 4. Estructura tridimensional de la proteína NaTrxh.....	18
III. HIPÓTESIS	24

IV. OBJETIVOS	25
IV.1. Objetivos generales.....	25
V. MATERIALES Y MÉTODOS	26
V.1. Programas.....	26
V.1.1. MODELLER.....	26
V.1.1.1. Comando profile.build().....	26
V.1.1.2. Comando alignment.compare_structures().....	26
V.1.1.3. Comando align2d().....	26
V.1.1.4. Comando automodel.....	28
V.1.1.5. Comando assess_dope.....	28
V.1.2. Swiss-PdbViewer.....	29
V.1.3. HyperChem.....	29
V.1.4. Visual Molecular Dynamics (VMD).....	30
VI. RESULTADOS	31
VI.1. Modelaje de la proteína NaTrxh con MODELLER.....	31
VI.1.1. Obtención de un primer modelo.....	31
VI.1.1.1. Búsqueda de las estructuras relacionadas con NaTrxh.....	32
VI.1.1.2. Selección de un molde.....	33
VI.1.1.3. Alineamiento de NaTrxh con la proteína molde (1ti3).....	37
VI.1.1.4. Construyendo los modelos de NaTrxh en base a 1ti3.....	39
VI.1.1.5. Evaluación de los modelos para la NaTrxh.....	41
VI.1.2. Presencia de un "nudo" en los modelos de MODELLER.....	44

VI.1.3. Recocimiento simulado del modelo tridimensional de la NaTrxh desanudado.....	47
VI.1.4. Simulación dinámica molecular de NaTrxh_3.....	49
VI.2. Modelaje de la secuencia de la NaTrxh con SAM-T06.....	51
VI.2. Hvtrxh2 como molde para generar un nuevo modelo con MODELLER.....	53
VII. DISCUSIÓN.....	61
VII.1. NaTrxh es una proteína importante en la Algametofítica.....	61
VII.2. Primera estructura tridimensional de la NaTrxh modelada por homología.....	62
VII.3. Otras estructuras predichas por servidores basadas en modelaje comparativo.....	66
VII.4. Un modelo estable para la NaTrxh.....	66
VIII. CONCLUSIÓN.....	69
IX. BIBLIOGRAFÍA.....	70

Índice de Figuras

Figura 1. Secuencia del cDNA-AFLP25F.	2
Figura 2. Representación esquemática del sistema de reducción del disulfuro de una proteína blanco que es dependiente del NADPH por el sistema de Trx.	6
Figura 3. Análisis filogenético de Trx vegetales.	11
Figura 4. Colocalización de la NaTrxh y de la Sc ₁₀ -RNasa en la matriz extracelular del tejido de transmisión del estilo de <i>N. alata</i> (Juárez-Díaz <i>et al.</i> , 2006a).	12
Figura 5. La proteína de fusión NaTrxh:GFP se secreta al apoplasto de las hojas de <i>N. benthamiana</i> y de <i>A. thaliana</i> .	16
Figura 6. Mecanismos hipotéticos para la secreción independiente del RE/Golgi de la NaTrxh en las células del tejido de transmisión (TT) de estilos de <i>N. alata</i> .	17
Figura 7. Secuencia de aminoácidos de NaTrxh en formato PYR y los campos para MODELLER	33
Figura 8. Resultado obtenido después de introducir el comando build.profile() en el MODELLER.	35
Figura 9. Comparación de semejanza entre secuencias.	36
Figura 10. Alineamiento de NaTrxh con la Trx tipo <i>h</i> del <i>P. tremula</i> (PtTrxh1).	38
Figura 11. Modelo de la estructura tridimensional de 1ti3 (PtTrxh1, Trx de <i>P. tremula</i>).	39

Figura 12. Comparaciones de los modelos tridimensionales de NaTrxh.	40
Figura 13. Selección del modelo más apropiado. Gráficas del parámetro de coincidencia estructural generado por <code>assess_dope()</code> para los modelos.	42
Figura 14. Presencia de un nudo al final del carboxilo terminal en el modelo de la NaTrxh_1d.	46
Figura 15. Superposiciones de la NaTrxh_1d con otros modelos.	47
Figura 16. NaTrxh_3 luego de un recocido simulado en condiciones de vacío, con los átomos del esqueleto fijos para los aminoácidos (31-139).	48
Figura 17. Superposición de NaTrxh_3 con NaTrxh_2 donde se observa que hay un cambio en la estructura del amino terminal de la NaTrxh_3.	49
Figura 18. Modelo de NaTrxh_5 mostrando dos hebras betas en la región Na y los aminoácidos expuestos al solvente en la región Nβ.	50
Figura 19. Superposición de NaTrxh_5 con 1ti3.	51
Figura 20. Modelo de la secuencia de NaTrxh_6 propuesta por el servidor SAM-T06.	52
Figura 21. Alineamiento de NaTrxh, HvTrxh2 y PtTrxh con el programa ClustalX.	54
Figura 22. Archivo PDB de HvTrxh2 (2wit).	55
Figura 23. Modelo NaTrxh_7 obtenido con el MODELLER utilizando como molde para el núcleo a HvTrxh2 y como molde al extremo amino y carboxilo de NaTrxh_5.	56
Figura 24. Modelo NaTrxh_8 después de la minimización con AMBER	56

95 en HyperChem.

Figura 25. Calificación de NaTrxh con el programa Rosetta-HMMer. 57

Figura 26. Modelo final y estable de la NaTrxh (NaTrxh_9). 59

Figura 27. Superposición entre los modelos NaTrxh_8 y NaTrxh_9. 60

Índice de Tablas

Tabla 1. Comando del MODELLER para buscar alineamientos relacionados con NaTrxh.	27
Tabla 2. Comando del MODELLER alignment.compare_structures() .	27
Tabla 3. Comando del MODELLER align2d() .	28
Tabla 4. Comando del MODELLER automodel.	28
Tabla 5. Comando del MODELLER assess dope	29

Abreviaturas

1ti3	Código PDB de la proteína PtTrxh1
2wit	Código PDB de la proteína HvTrxh2
AC	Autocompatible
AI	Autoincompatible/Autoincompatibilidad
AFLP	Amplificación de fragmentos polimórficos en longitud
CaMV35s	Promotor constitutivo 35s del virus del mosaico del tabaco
Cβ3	Hebra beta conformada por los aminoácidos Pro-140 a Ala-143 en el modelo NaTrxh_9
FAD/FADH ₂	Flavín adenín dinucleótido oxidado / reducido
FTR	Tiorredoxina reductasa dependiente de ferredoxina
GFP	Proteína verde fluorescente
HvTrxh2	Tiorredoxina tipo <i>h</i> de <i>Hordeum Vulgare</i>
NADP ⁺ /NADPH	Nicotinamida adenín dinucleótido fosfato oxidado / reducido
Nα	Región compuesta por los aminoácidos Met-1 a la Ala-16 de la proteína NaTrxh
NaTrxh	Tiorredoxina tipo <i>h</i> de <i>Nicotiana alata</i>
NaTrxh_1	Modelo obtenido de MODELLER usando como proteína modelo a 1ti3

NaTrxh_2	Modelo sin nudo
NaTrxh_3	Modelo obtenido después de un recocido simulado en condiciones de vacío
NaTrxh_5	Modelo de la NaTrxh después de una dinámica molecular
NaTrxh_6	Modelo de la NaTrxh propuesto por el servidor SAM-TO6
NaTrxh_7	Modelo obtenido por MODELLER usando como proteína modelo a 2wit
NaTrxh_8	Modelo después de la minimización con AMBER 95 en el programa HyperChem
NaTrxh_9	Modelo final que presenta la región Nβ expuesta al solvente y además una lámina beta conformada en las regiones amino y carboxilo terminal de la NaTrxh
Nβ	Región compuesta por los aminoácidos Ala-17 a Pro-27
Nβ1	Hebra beta formada por los aminoácidos Ser-7 a Gly-13 en el modelo NaTrxh_9
Nβ2	Hebra beta formada por los aminoácidos Ala-17 a Ser-21 en el modelo NaTrxh_9
NTR	Tiorredoxina reductasa dependiente de NADPH
PDB	Banco de datos de proteínas (Protein Data Bank)
PtTrxh1	Tiorredoxina tipo <i>h</i> de <i>P. tremula</i>
RE/Golgi	Reticulo endoplásmico y aparato de Golgi
RMS	Root Mean Square

S-RNasa	Glicoproteína con actividad de ribonucleasa del <i>locus-S</i> de Solanaceae (determinante femenina del sistema AI gametofítico)
SLF	Proteína con caja F del <i>locus-S</i> de Solanaceae (determinante masculina del sistema de AI gametofítico)
Trx	Tiorredoxina
Trx-S ₂	Tiorredoxina oxidada
Trx-(SH) ₂	Tiorredoxina reducida
VMD	Programa de visualización dinámica molecular

I. Introducción

I. 1. En busca de los genes modificadores de estilos de *Nicotiana alata*

En nuestro grupo de investigación se han estado identificando genes modificadores en estilos de plantas angiospermas que aceptan o rechazan el polen SC₁₀, por medio de una técnica de cDNA-AFLP. Se utilizaron 200 combinaciones de oligonucleótidos que permitieron la comparación de 14,400 cDNAs. Dentro de los fragmentos que presentaron una mayor expresión en *Nicotiana alata* BT comparado con *Nicotiana plumbaginifolia*, se identificó al cDNA-AFLP25 (Cruz-García, sin publicar).

Debido a que el cDNA-AFLP25 sólo era un fragmento, se recuperó su cDNA completo a partir de una biblioteca de cDNA de estilos de *N. alata* BT. Este cDNA (AFLP25-F) fue de 712 pb e incluía la secuencia del cDNA-AFLP25, por lo que se utilizó como sonda para corroborar su patrón de expresión entre *N. plumbaginifolia* y *N. alata* BT, debido a que éstos son los progenitores de la población 08681. El ensayo tipo Northern-blot y el análisis densitométrico corroboraron la expresión diferencial, es decir, los estilos maduros de *N. alata* BT expresan al menos seis veces más a *AFLP25-F* que los de *N. plumbaginifolia* (Cruz-García, sin publicar).

I.1. 2. El cDNA-AFLP25F codifica para una tiorredoxina tipo *h*

La secuencia del cDNA-AFLP25F contiene un marco de lectura abierto que codifica para una proteína de 16.8 kDa. La estructura primaria de la

proteína inferida a partir de su cDNA, contiene en la primera mitad a la secuencia típica del sitio activo de las tiorredoxinas (Trx): Trp-Cys-Gly-Pro-Cys (Figura 1).



Figura 1. Secuencia del cDNA-AFLP25F. El marco de lectura abierto denominado cDNA NaTrxh se muestra flanqueado por las flechas (desde el codón de inicio ATG hasta el de término TAA). El recuadro indica la secuencia típica de las tiorredoxinas (Trp-Cys-Gly-Pro-Cys). La numeración en la izquierda corresponde a la secuencia nucleotídica y la de la derecha a la secuencia de aminoácidos correspondiente. Tomado de Juárez-Díaz *et al.*, 2006a.

El análisis de la estructura primaria de la proteína AFLP25F dio la primera evidencia de que se trataba de una Trx citoplásmica (tipo *h*), por lo que se nombró NaTrxh (Na: *Nicotiana alata*; Trxh: tiorredoxina

tipo *h*). Al realizar un alineamiento múltiple con otras Trx, se encontró que la NaTrxh tiene una identidad del 20.1% con la de *Escherichia coli*, un máximo del 28.8% con la Trx de humano y del 55.6% con una de *Arabidopsis thaliana*, lo que demuestra que la NaTrxh está mucho más relacionada con las Trx vegetales que con la de humano y que se parece más a ésta que a la de *E. coli* (Juárez-Díaz, 2006b), tal y como ocurre con las demás Trx *h* de plantas (Schürmann y Jacquot, 2000). Además, no se encontró ningún péptido de tránsito canónico que pudiera dirigir a la proteína a algún organelo, como la mitocondria o el cloroplasto, por lo que se descartaron los tipos de Trx organelares que más adelante se describirán y se sugirió que efectivamente se trata de una Trx tipo *h* (Meyer *et al.*, 1999).

I.2. Tiorredoxinas

Las tiorredoxinas (Trx) son proteínas pequeñas en promedio de ~13 kDa, que se encuentran en todos los organismos y forman parte de la gran familia de las enzimas con actividad de oxidoreductasa tiol:sulfuro (Wollman *et al.*, 1988). Los miembros de esta familia se caracterizan por contener en su sitio activo la secuencia típica WCXXC, la cual permite la oxidación, reducción o isomerización del puente disulfuro de sus proteínas blanco (Laloi *et al.*, 2001).

El motivo WCGPC, en algunos casos WCPPC (Bréhélin *et al.*, 2000), en el sitio activo contribuye a originar el bajo potencial redox de las Trx, lo que les confiere su propiedad reductora. En su estado reducido (Trx-(SH)₂), las Trx rompen los puentes disulfuro de un gran número de proteínas blanco, participando así en la regulación enzimática, en la respuesta al estrés oxidativo, en la regulación de la transcripción y de la

traducción, entre otras. Además, las Trx participan en varias reacciones reductoras como donadoras de electrones (revisado en Balmer y Buchanan, 2002).

Las Trx descritas desde procariontes hasta eucariontes muestran del 27 al 69% de identidad a nivel de secuencia de aminoácidos con la Trx de *E. coli*. Los modelos indican que todas estas proteínas tienen estructuras tridimensionales similares a pesar de la gran variación que existe en sus secuencias aminoacídicas (Eklund *et al.*, 1991). Todas comparten un plegamiento común con una sucesión de elementos estructurales secundarios que incluyen a más del 75% de los residuos de la proteína (Holmgren, 1985): $\beta 1$, $\alpha 1$, $\beta 2$, $\alpha 2$, $\beta 3$, $\alpha 3$, $\beta 4$, $\beta 5$ y $\alpha 4$. La estructura compacta globular resultante es generada por 5 hebras beta, tres paralelas y dos antiparalelas (Holmgren, 1985), que forman una región hidrofóbica rodeada por cuatro α -hélices sobre la superficie externa (Holmgren, 1985; Powis y Montfort, 2001).

La primera cisteína del sitio activo tiene un bajo valor de pK_a y es la que realiza el ataque nucleofílico en la reducción del puente disulfuro de otras proteínas. El mecanismo involucra un intermediario mixto transitorio y un rápido intercambio tiol-disulfuro en un ambiente hidrofóbico (Figura 2). Esta reacción es reversible y la Trx puede romper o formar disulfuros dependiendo del potencial redox del sustrato. El bajo potencial redox de la Trx, por ejemplo la de *E. coli* Trx1 = -270 mV (Krause *et al.*, 1991), asegura que la Trx-(SH)₂ sea el principal reductor en el citosol (Holmgren, 1985; 1995).

Generalmente, cuando se descubre una proteína se le adjudica un nombre y una función relacionándola con algún proceso bioquímico en particular. Las funciones de las Trx en los diferentes organismos evolucionó a partir de una reacción común fundamental, a una gran

cantidad de funciones especializadas; lo cual es una prueba de su gran importancia, utilidad y especificidad. En términos muy amplios, la actividad de oxidorreductasa de las Trx participa en dos funciones bien conocidas: (1) como portadora de electrones necesarios para los ciclos catalíticos de enzimas biosintéticas y (2) como protectora de las proteínas citosólicas de la inactivación o agregación por la formación de puentes disulfuro intra o intermoleculares (Arnér y Holmgren, 2000).

I.2.1. Tiorredoxina reductasa

Para tener una idea clara de la función de las Trx se debe considerar el papel que desempeñan las tiorredoxina reductasas (NTR), las cuales son las únicas enzimas conocidas capaces de reducir el sitio activo oxidado de las Trx (Mustacich y Powis, 2000; Powis y Montfort, 2001). Por lo tanto, es posible que alteraciones en la actividad de la NTR regulen algunas de las actividades de las Trx (Mustacich y Powis, 2000).

La NTR forma parte de la familia de las disulfuro reductasas, la cual incluye flavoenzimas homodiméricas que catalizan la transferencia de electrones entre nucleótidos de piridina y compuestos disulfuro/ditio (William *et al.*, 2000). En cada monómero de las enzimas de esta familia se encuentra un grupo prostético FAD, un sitio de unión NADPH y un sitio activo que contiene un disulfuro con actividad óxido-reductora (Mustacich y Powis, 2000). De esta manera, la catálisis de la enzima se lleva a cabo vía FAD y un disulfuro activo (Williams *et al.*, 2000).

La reacción que efectúa la familia de las disulfuro reductasas involucra la transferencia de equivalentes reductores del sustrato nucleótido de piridina al sustrato disulfuro. Un hidruro pasa del NADPH a

la flavina y entonces el equivalente reductor es transferido de la flavina reducida al disulfuro del sitio activo (Williams Jr. *et al.*, 2000). Finalmente, la NTR reducida ($\text{NTR}-(\text{SH})_2$) en su sitio activo puede reducir a su sustrato (Figura 2).

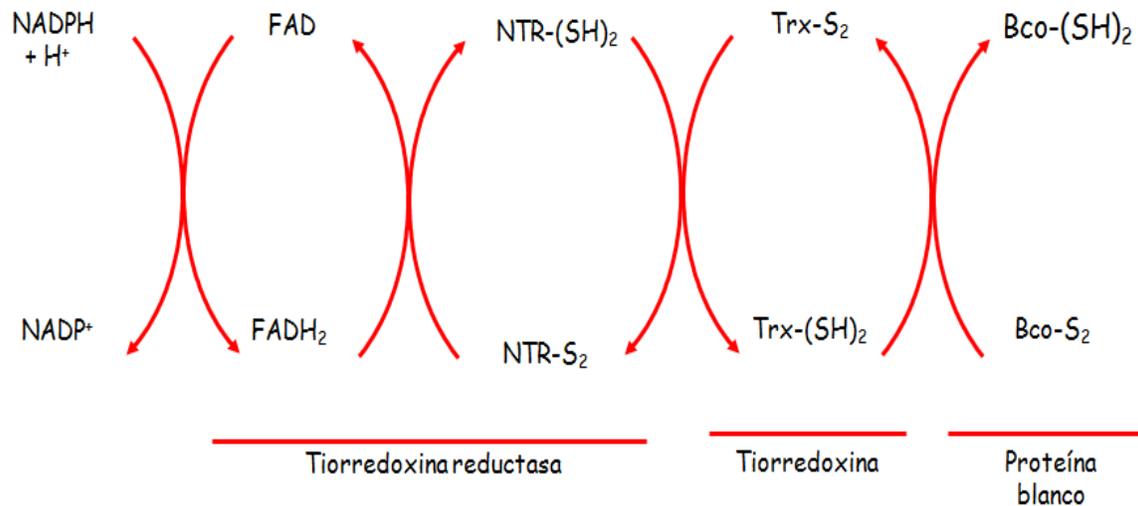


Figura 2. Representación esquemática del sistema de reducción del disulfuro de una proteína blanco que es dependiente del NADPH por el sistema de Trx. El sitio activo de la Trx reductasa oxidada (NTR-S_2) se reduce ($\text{NTR}-(\text{SH})_2$) vía FADH. El FAD es reducido (FADH_2) a partir de $\text{NADPH} + \text{H}^+$. La Trx oxidada (Trx-S_2) es entonces reducida por la $\text{NTR}-(\text{SH})_2$ y la Trx reducida ($\text{Trx}-(\text{SH})_2$) rompe el puente disulfuro de su proteína blanco, generando en ésta dos grupos tiol ($\text{Bco-S}_2 \rightarrow \text{Bco}-(\text{SH})_2$)

Una vez reducida la NTR, ésta reduce a la Trx, formando así el sistema NADPH/NTR/Trx (Figura 2), el cual es uno de los mecanismos de reducción de disulfuros más ampliamente distribuidos y presente en todos los tipos celulares. Sin embargo, en cloroplastos y en procariontes fotosintéticos oxigénicos, la reducción de la Trx no está relacionada con el NADPH, sino con la ferredoxina por la enzima heterodimérica de fierro-azufre ferredoxina-tiorredoxina reductasa (FTR). En ambos casos, la $\text{Trx}-(\text{SH})_2$ reduce puentes disulfuro específicos de sus proteína blanco,

lo que altera su actividad, o sirve como sustrato reductor en alguna reacción enzimática (Balmer y Buchanan, 2002).

I.2.3. Tiorredoxinas vegetales

En levaduras y células animales, el sistema de las Trx es más compleja que en bacterias, ya que cuentan con un segundo sistema dependiente de NADPH en mitocondria. El patrón en plantas es aún más complejo. En mamíferos sólo se han reportado dos genes que codifican para Trx, tres en levadura y dos en *E. coli*, mientras que en plantas se han detectado al menos 20 genes (Meyer *et al.*, 2002; Marchand *et al.*, 2004; Gelhaye *et al.*, 2004a).

En las plantas existen al menos 5 grupos filogenéticos principales de Trx: tres de cloroplasto, uno de mitocondria y uno citoplásmico (Figura 3). El sistema en cloroplastos contiene tres tipos de Trx s. Dos grupos principales, el tipo *m* (Trx *m*) y el tipo *f* (Trx *f*). Existe también la Trx tipo *x* (Trx *x*), que está más relacionada a las secuencias procariontes y está relativamente distante de la Trx *m* que es la considerada como bacteriana (Balmer y Buchanan, 2002). Estas Trx son reducidas por la enzima heterodimérica FTR por medio de la cadena de transporte de electrones fotosintética y se encuentran representadas por dos genes para *Trx f*, al menos cuatro para *Trx m* y uno para *Trx x* en el genoma de *A. thaliana* (Mestres-Ortega y Meyer, 1999; Laloi *et al.*, 2001).

En la mitocondria sólo se tenía evidencia de actividad tanto de Trx como de NTR, pero Laloi *et al.* (2001) identificaron un nuevo tipo de Trx, denominado tipo *o* (Trx *o*) y una NTR mitocondrial similar a la citoplásmica. Tanto el gen de la Trx *o* como el de la NTR mitocondrial

codifican para mRNAs cuyos productos contienen un péptido de tránsito hacia la mitocondria (Laloi *et al.*, 2001).

El sistema citoplásmico de las plantas consiste de las Trx tipo *h* (Trx h) y una NTR similar a la de levadura y a la de procariontes (Balmer y Buchanan, 2002). En el genoma de *A. thaliana* se encuentran 8 genes que codifican para Trx h (Rivera-Madrid *et al.*, 1995; Mestres-Ortega y Meyer, 1999; Laloi *et al.*, 2001).

A partir de análisis filogenéticos se sugiere que la Trx m tiene un origen procarionte, mientras que las Trx *f* y *h* son de tipo eucarionte (Balmer y Buchanan, 2002). Al incluir en el análisis a la Trx *o*, se encontró que las Trx se agrupan en cuatro grupos filogenéticos diferentes de acuerdo a su localización celular, donde la Trx h está relacionada filogenéticamente con la Trx *f* (Laloi *et al.*, 2001). Por otro lado, la Trx *x* forma un quinto grupo formado sólo por un miembro (Mestres-Ortega y Meyer, 1999).

I.2.4. Tiorredoxinas tipo *h*

Las Trx tipo *h* tienen en común con los otros grupos de Trx la presencia del sitio catalítico conservado WC(G/P)PC. El sitio catalítico WCGPC es común en la mayoría de las Trx *h*. Aparentemente, las Trx *h* hasta ahora caracterizadas están lejos de tener algo en común; no obstante, Gelhaye *et al.*, (2004a) las clasificó en tres subgrupos diferentes: I, II y III.

Entre las Trx *h*, los miembros del primer subgrupo son los más estudiados. La reducción de las isoformas de Trx *h* de éste subgrupo por NADPH es mediado por la NTR (Dai *et al.*, 1996; Gelhaye *et al.*, 2004).

Los miembros del subgrupo I son presumiblemente citosólicos (Santandrea *et al.*, 2002).

La principal característica del subgrupo II de las Trx *h* es la presencia de una extensión hacia el amino terminal y la conservación del sitio activo WCGPC. Hasta ahora, en este subgrupo se han identificado tres Trx en *A. thaliana* (AtTRX*h2*, *h7* y *h8*) y dos en *Populus tremula* (PtTrx*h2* y *h5*) (Gelhaye *et al.*, 2002; Reichheld *et al.*, 2002; Gelhaye *et al.*, 2004a). Varios miembros de éste subgrupo se predice que tienen localización subcelular; mientras que el subgrupo III es también citosólico. Sin embargo, la localización subcelular de éstas tiorredoxinas con la extensión hacia el N-terminal aún debe ser aclarado (Gelhaye *et al.*, 2004a).

Finalmente el subgrupo III exhibe una extensión hacia el N-terminal, pero comparado con el subgrupo I, ésta contiene una cisteína conservada en la cuarta posición (Gelhaye *et al.*, 2004b), lo que en el subgrupo II no se cumple tampoco. El tamaño de estas proteínas es alrededor de 10-15 kDa con un sitio activo de la forma CxxC requerida para sus propiedades reductoras (Fomenko y Gladyshev, 2002). Grx se mantiene reducida con la ayuda del NADPH, la glutatión reductasa y el tripéptido glutatión (GSH). Varios miembros del subgrupo III, tienen un sitio activo WCXXS, como lo fue encontrado en *A. thaliana* y en *P. tremula* aunque la serina ya le impide reducir a su sitio específico (Holmgren *et al.*, 1995; Gelhaye *et al.*, 2003; 2004a, 2004b).

Las Trx *h* están involucradas en múltiples procesos como la germinación de semillas de cereales (Wong *et al.*, 2002); la protección celular contra el estrés oxidativo, en particular durante la desecación y la germinación de las semillas (Serrato *et al.*, 2003); en el metabolismo del carbono y del nitrógeno (Baumann y Juttner, 2002; Maeda *et al.*,

2003a y b); y en la autoincompatibilidad (Bower *et al.*, 1996; Cabrillac *et al.*, 2001)

Las Trx *h* carecen de un péptido señal, por lo que se asumía que eran proteínas netamente citosolicas (Meyer *et al.*, 1999). Sin embargo, en arroz (Ishiwatari *et al.*, 1995) y en muchas otras plantas las Trx *h* también están presentes en el floema. Al no haber una señal de secreción, se cree que las Trx son transportadas a través de los plasmodesmos debido a su tamaño (Ishiwatari *et al.*, 1995; Meyer *et al.*, 1999). De hecho, se ha demostrado por microinyección de una Trx *h* recombinante que la proteína misma tiene la capacidad de mediar su propio transporte de célula a célula a través de los plasmodesmos (Ishiwatari *et al.*, 1998). Al sumar esta evidencia con la que se tiene de la Trx humana (Trx-1) que es secretada masivamente a la sangre por linfocitos, hepatocitos, fibroplastos y una variedad de células cancerígenas (Powis y Montfort, 2001); se sugiere que las Trx tienen la capacidad de actuar como mensajeros en la transducción de señales (Meyer *et al.*, 1999; Powis y Montfort, 2001, Schürman y Jacquot, 2000).

I.3. NaTrxh es una tiorredoxina tipo *h*

Debido a que las Trx forman un grupo muy diverso en las plantas, fue necesario determinar para qué tipo de Trx codificaba *NaTrxh*, por lo que se hizo un análisis filogenético utilizando el método de *neighbor-joining* (Kumar *et al.*, 2001) con diferentes secuencias de aminoácidos de Trx vegetales (Figura 3). El resultado de este análisis mostró a 5 tipos específicos de Trx (*f*, *m*, *x*, *o* y *h*), de los cuales 4 corresponden a las Trx organelares, donde 3 son de cloroplastos (*m*, *f* y *x*) y una

mitocondrial (*o*), y las tipo *h* divididas en tres subgrupos (Juárez-Díaz, 2006b).

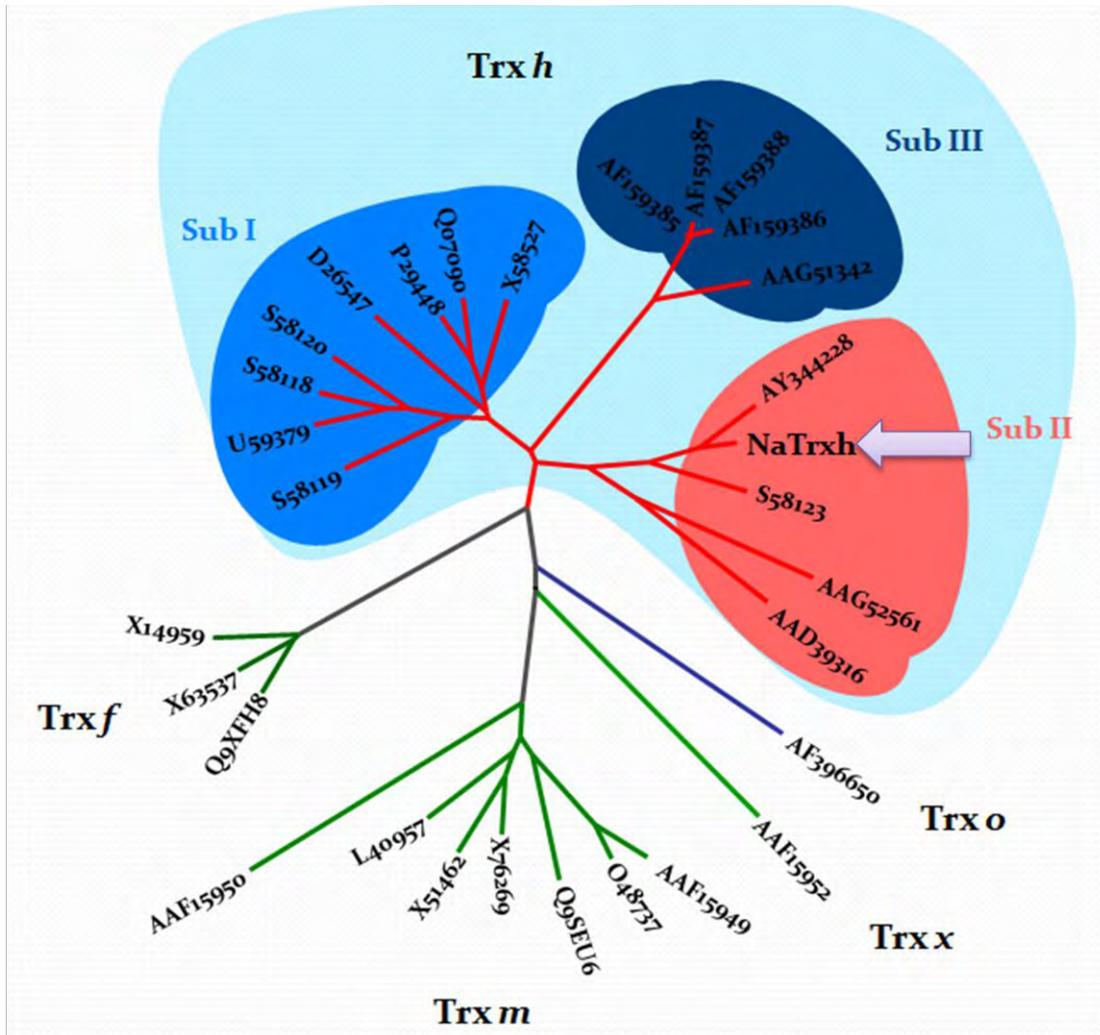


Figura 3. Análisis filogenético de Trx vegetales. Cuatro grupos (Trx *f*, Trx *m*, Trx *x* y Trx *o*) que corresponden a Trx de órganos. El quinto grupo es el tipo *h* que a su vez se encuentra dividido en tres subgrupos (I, II y III). El análisis arroja que la NaTrxh (flecha morada) corresponde al tipo *h* subgrupo II. Tomado de Juárez-Díaz *et al.*, 2006a.

II. Antecedentes

II.1. La secuencia de la NaTrxh contiene la información suficiente para promover su secreción

Experimentos de solubilidad y fraccionamiento subcelular indicaron que la NaTrxh está presente en la matriz extracelular del tejido de transmisión de los estilos de *N. alata* con las mismas características de solubilidad que las proteínas NaTTS y S-RNasa, dos proteínas extracelulares presentes en la matriz extracelular del tejido de transmisión del estilo (Anderson *et al.*, 1989; Wu *et al.*, 2000). Así mismo, ensayos de inmunocolocalización mostraron a la S_{c10}-RNasa y de la NaTrxh en la matriz extracelular del tejido de transmisión en los estilos de *N. alata* (Figura 4).

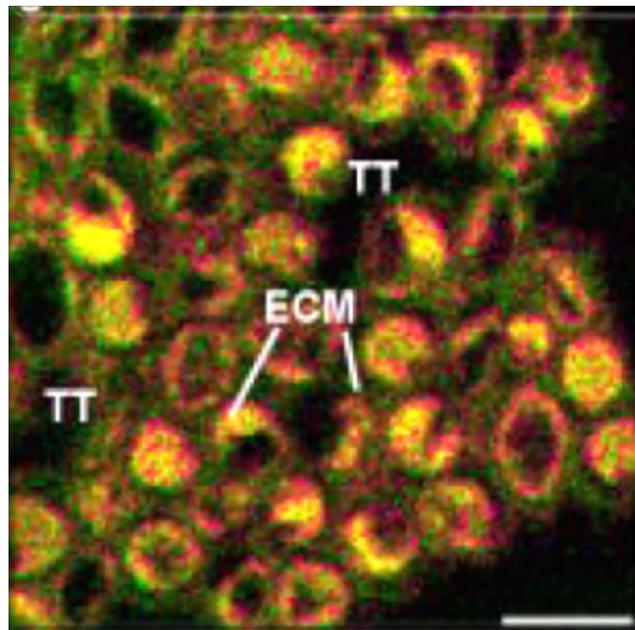


Figura 4. Colocalización de la NaTrxh y de la S_{c10}-RNasa en la matriz extracelular del tejido de transmisión del estilo de *N. alata* (Juárez-Díaz *et al.*, 2006a). TT es el tejido de transmisión y ECM es la matriz extracelular.

La NaTrxh es la primera Trx de plantas con la que se demuestra que hay Trx extracelulares. El mecanismo por el cual la región que tiene en su extremo amino dirige la secreción de la NaTrxh no está claro. Es evidente que esta secuencia actúa como una señal de secreción. No obstante, la señal es diferente a los péptidos señal descritos a la fecha (Juárez-Díaz, 2006b).

El hecho de que la NaTrxh sea secretada por las células del tejido de transmisión podría deberse al carácter secretor de este tejido. Sin embargo, cuando se expresa de manera transitoria la proteína de fusión NaTrxh:GFP en las hojas de *N. benthamiana* y de *A. thaliana* bajo el control del promotor CaMV35S (Sheen *et al.*, 1995), la secreción de la NaTrxh también ocurre en hojas, sugiriendo que la NaTrxh contiene su propia señal para dirigir su secreción, señal que es reconocida por otras plantas (Figura 5) (Juárez-Díaz *et al.*, 2006a y b). Cuando la proteína de fusión NaTrxh:GFP es expresada, la mayor parte de la expresión de GFP se acumula en el apoplasto tanto de *N. benthamiana* (Figura 5B y D) como de *A. thaliana* (Figura 5J y L). Como se esperaba, la proteína GFP sin la NaTrxh, utilizada como control, principalmente se localiza en el citoplasma (Figura 5G y H) (Juárez-Díaz *et al.*, 2006a).

Con estos resultados se demostrará claramente que la secuencia de la NaTrxh contiene la información necesaria para su secreción. Además, el hecho de que este comportamiento se observe en hojas y en dos sistemas heterólogos a *N. alata*, significa que la señal contenida en la NaTrxh es general y reconocida a pesar de ser órganos y especies diferentes (Juárez-Díaz *et al.*, 2006a y b).

II.2. Secreción por una vía independiente al Retículo Endoplásmico/Aparato de Golgi de NaTrxh

Los experimentos demuestran que la NaTrxh se secreta en la matriz extracelular (Figura 4), incluso en sistemas heterólogos a *N. alata* (Figura 5). Sin embargo, se debe considerar que la NaTrxh no posee un péptido señal de secreción canónico (Juárez-Díaz *et al.*, 2006a). Al analizar la secuencia con diferentes algoritmos, éstos arrojan resultados conflictivos (Juárez-Díaz *et al.*, 2006a y b). Mientras que el algoritmo de *redes neuronales* (Bendtsen *et al.*, 2004a) predice que la NaTrxh no es una proteína de secreción; el programa de predicción *Hidden Markov* (Bendtsen *et al.*, 2004b) indica que la NaTrxh podría ser una proteína de secreción, ya que predijo un péptido señal con una probabilidad del 95.3%. Además, predijo un sitio de corte que se encuentra entre los residuos Ala-16 y Ala-17 con una probabilidad de 59.3%. Por otro lado, el programa *Secretome 1.0* (Bendtsen *et al.*, 2004b) predice también, con una probabilidad del 87.4% que la NaTrxh es una proteína que se secreta por una vía no clásica; es decir es independiente del retículo endoplásmico y del aparato de Golgi (RE/Golgi). Esta predicción se logra porque el programa compara a la NaTrxh con la Trx-1 de humano, la cual es secretada al torrente sanguíneo. Esta Trx carece de péptido señal y se sabe que puede secretarse por una vía independiente RE/Golgi (Tanudji *et al.*, 2003).

Juárez-Díaz (2006b) propone diversos mecanismos hipotéticos para la secreción independiente del RE/Golgi de la NaTrxh (Figura 6) basados en diferentes proteínas que carecen de un péptido señal canónico, pero que son secretadas al ser capaces de cruzar la membrana plasmática (Zhu *et al.*, 2001), proponiendo que esos mecanismos pudieran llevarse a cabo también en plantas.

El tejido de transmisión estilar tiene un carácter secretor, debe tener una actividad muy alta en cuanto a producción y transporte de vesículas de secreción del aparato de Golgi hacia la membrana plasmática. Una posibilidad es que la NaTrxh tenga cierta afinidad a las vesículas de secreción y se adhiera a ellas para alcanzar el espacio extracelular junto con las otras proteínas de secreción que participan en las interacciones polen-pistilo (Figura 6, situación A) (Juárez-Díaz, 2006b). También es posible que en el reciclamiento de endosomas en las células del tejido de transmisión pueda haber integración de la NaTrxh a uno de estos compartimientos para salir al apoplasto (Figura 6, situación B). Sin embargo, aunque se requiere mucho más evidencia genética y molecular, tampoco se excluye la posibilidad de que existan factores que medien la secreción de la NaTrxh, ya sea en forma específica, interaccionando **por ejemplo con algún "transportador" específico y exclusivo** (Figura 6, situación C) o de manera inespecífica, es decir, interaccionando con ortólogos hipotéticos de *NCE101* y *NCE102* que son genes involucrados en la exportación no clásica de proteínas como la galectina de mamíferos (Cleves *et al.*, 1996; Nombela *et al.*, 2006) o con proteínas que realicen funciones similares (Figura 6, situación D). Sin embargo, en esta última propuesta existe el problema de que no se han reportado genes homólogos a los *NCE* de levadura en plantas (Juárez-Díaz, 2006b).

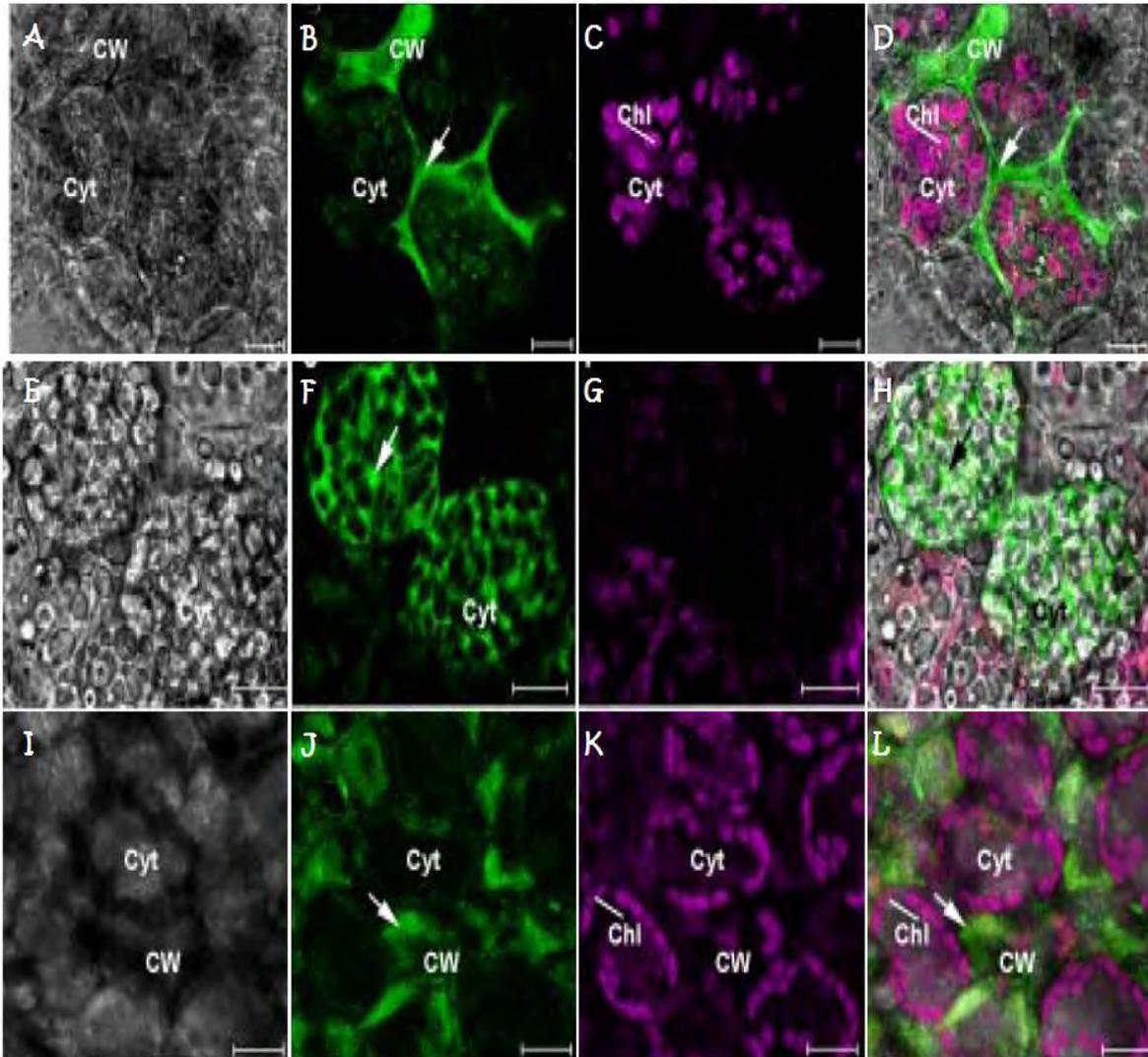


Figura 5. La proteína de fusión NaTrxh:GFP se secreta al apoplasto de las hojas de *N. benthamiana* y de *A. thaliana*. **A-D**. Expresión de GFP en hojas de *N. benthamiana* como control de expresión y localización. **E-H**. Expresión de la proteína de fusión NaTrxh:GFP en hojas de *N. benthamiana*. **I-L**. Expresión de la proteína de fusión NaTrxh:GFP en hojas de *A. thaliana*. Después de la transfección, las hojas se analizaron bajo microscopía de contraste de fases (**A, E, I**) y bajo microscopio de fluorescencia para GFP (**B, F, J**) y para la clorofila (**C, G, K**). Las imágenes **D, H** y **L** corresponden a la superposición de las tres imágenes de cada experimento (Juárez-Díaz *et al.*, 2006a).

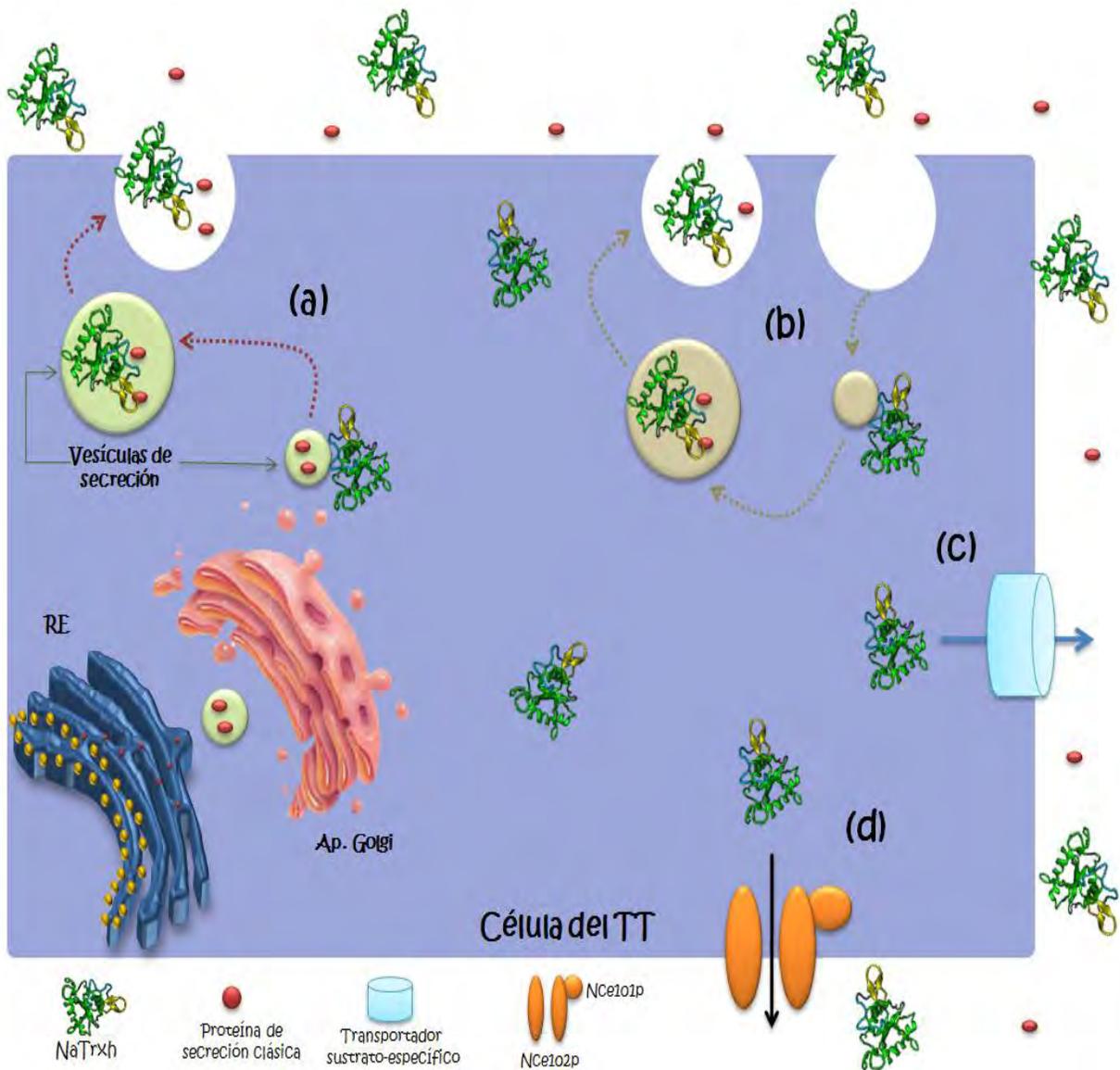


Figura 6. Mecanismos hipotéticos para la secreción independiente del RE/Golgi de la NaTrxh en las células del tejido de transmisión (TT) de estilos de *N. alata*. Se propone que la NaTrxh (verde) puede tener cierta afinidad hacia las vesículas de secreción producidas por el sistema RE/Golgi (a) o a los subcompartimientos endosomales en su proceso de reciclamiento (b) que le permite adherirse y/o incorporarse a ellos para ser secretada. En el primer caso (a), la NaTrxh sería secretada junto con otras proteínas de secreción clásica (rojo). Por otro lado, pueden existir transportadores específicos (azul claro) que permitan la secreción de la NaTrxh (c) o proteínas homólogas de Nce101p y Nce102p de levaduras en plantas (naranja) que sean las responsables de su exportación (d). RE: retículo endoplásmico; Ap. Golgi: aparato de Golgi. Tomado de Juárez-Díaz, 2006b.

II. 3. La región N-terminal, específicamente la región N β , contiene la información necesaria para promover la secreción de NaTrxh

Un dato interesante encontrado buscando una señal de secreción en el extremo amino de la NaTrxh, fue que la secuencia de aminoácidos responsable de la salida de NaTrxh de la célula estaba en la región N β . Lo primero que se hizo fue eliminar los primeros 16 residuos del N-terminal (NaTrxh Δ Na o región Na), que es la región que se predice como péptido señal por el *Hidden Markov* (Bendtsen *et al.*, 2004b). También se generó otra mutante, la NaTrxh Δ N β (región N β) con 27 residuos faltantes en su amino terminal. Los resultados mostraron que la eliminación de los primeros 16 residuos no modifica la secreción de la proteína de fusión NaTrxh Δ Na:GFP. No obstante, cuando se transformó con la NaTrxh Δ N β fusionada a la GFP, la capacidad de secreción de la fusión se perdió. Esto significa que la señal que dirige la secreción de la NaTrxh en hojas de *A. thaliana* no se encuentra entre los primeros 16 residuos, sino en un dominio interno que aparentemente estaría en la secuencia contenida entre los residuos Ala-17 y Pro-27. Por lo que se concluyó que la región N β , tiene la suficiente información para promover la secreción de la NaTrxh (Juárez-Díaz y Cruz-García, no publicado).

II. 4. Estructura tridimensional de la proteína NaTrxh

Como se ha visto, Juárez-Díaz (2006b) propuso diferentes mecanismos (Figura 6) que implican la unión de la NaTrxh con alguna proteína, sabiendo que ésta tiene una extensión hacia el N-terminal y habiéndose constatado que dicha secuencia tiene la información suficiente para su secreción. La estructura tridimensional podría mostrar

si efectivamente ésta región puede ser un dominio expuesto en la superficie de la molécula, que pudiera funcionar en interacciones proteína-proteína involucradas en la secreción de la NaTrxh.

Desde la publicación de la primera estructura tridimensional en 1998 (Zarembinski *et al.*, 1998), en donde la estructura resuelta experimentalmente por medio de cristalografía de rayos X de una proteína hipotética reveló su función molecular, muchos estudios similares han aparecido (Zhang y Kim, 2003) hasta el día de hoy. Los resultados de estos estudios piloto son convincentes (Campbell, 2002). En el presente hay millones de secuencias proteicas incluyendo la de la NaTrxh (Petsko, 2006), pero sólo hay unas miles de estructuras tridimensionales en el PDB (Protein Data Bank, <http://www.rcsb.org/>). La velocidad con la que aparecen nuevas secuencias es exponencial, así que ahora se está haciendo un esfuerzo por tratar de explicar las funciones de esas proteínas. Una herramienta importante para entender, describir y manipular a las proteínas es conociendo su estructura tridimensional (Sali y Blundell, 1993; Marti-Renom *et al.*, 2000; Bino y Sali, 2003). También a partir de la estructura secundaria de una proteína se puede entender mucho. Por ejemplo, las asas son sitios de interacción proteína-proteína y determinan la función específica de una proteína. Igualmente, en una familia con determinado plegamiento, la variabilidad estructural es el resultado de sustituciones, inserciones y deleciones en uno o más residuos entre los diferentes miembros de las familias. Dichos cambios frecuentemente corresponden a regiones expuestas que conectan elementos de la estructura secundaria en la proteína plegada (Fine *et al.*, 1986; Martin y Thornton, 1996; Oliva *et al.*, 1997; van Vlijmen y Karplus, 1997; Rapp y Friesner, 1999; Xiang *et al.*, 2002; Fiser y Sali, 2003). En el caso de que se sepa la función celular de una proteína, el conocimiento de su función

bioquímica puede ayudar a develar el mecanismo molecular en el proceso celular e incluso sugerir los ligandos a los cuales se puede unir la proteína (Zhang y Kim, 2003).

Debido a la dificultad que se presenta al resolver experimentalmente la estructura tridimensional de una proteína, el uso del modelaje puede proveer un modelo útil de una proteína (Sali y Blundell, 1993; Marti-Renom, 2000). Las técnicas que se utilizan pueden ser por medio de predicciones físicas o empíricas. Los métodos de predicción física están basados en las interacciones entre átomos incluyendo la dinámica molecular y la minimización de energía (Brooks III *et al.*, 1988; Sali y Blundell, 2003); mientras que los métodos empíricos dependen de la similitud que tenga la proteína problema con al menos una estructura protéica que ha sido determinada experimentalmente (Sali y Blundell, 2003), ya sea por espectroscopia de baja resolución como cristalografía de rayos X (Blundell y Johnson, 1976) o de media resolución como la resonancia magnética nuclear (NMR, por sus siglas en inglés) (Bax, 1989).

El modelo comparativo u homólogo utiliza proteínas determinadas experimentalmente para predecir la conformación de otras proteínas con secuencias de aminoácidos similares. Esto es posible porque un cambio pequeño en la secuencia usualmente resulta en un pequeño cambio en la estructura tridimensional (Lesk y Chothia, 1986; Hubbard y Blundell, 1987; Sali y Blundell, 1993). Así pues, el modelaje de una secuencia basada en una estructura conocida consiste en cuatro pasos: i) encontrar una estructura conocida relacionada con la secuencia que se quiere modelar (por ejemplo, un molde), ii) el alineamiento de la secuencia con el molde, iii) construir el modelo y, iv) validar el modelo (Marti-Renom *et al.*, 2000; Baker y Sali, 2003). No obstante, los errores

encontrados en el modelaje comparativo abarcan desde errores en el plegamiento y alineamiento de la proteína, que ocurren cerca de un 30%, así como distorsiones en las asas y en los núcleos de las proteínas, y un mal empaquetamiento de las cadenas (Eswar *et al.*, 2003). Aún así, la exactitud de la proteína modelada obtenida por un modelaje homólogo es favorable comparada con otros modelos calculados por otros métodos teóricos (Topham *et al.*, 1991; Sali y Blundell, 1993). Por supuesto, la precisión de la estructura tridimensional depende de la similitud que haya entre la proteína problema y el molde (Marti-Renom *et al.*, 2000; Baker y Sali, 2003). Un buen modelo se presenta cuando al menos hay un 50% de similitud entre el molde y la proteína blanco. Un buen molde, por ejemplo, son estructuras que han sido resueltas por cristalografía de rayos X y tienen una resolución menor o igual a 1.5Å, obteniéndose modelos con un RMS (Root Mean Square) significativamente bajo (Topham *et al.*, 1991; Sali y Blundell, 1993; Baker y Sali, 2003). Por ende, dos condiciones deben estar presentes para construir un buen modelo. Primero, la secuencia de la proteína problema y la secuencia de la proteína molde que se encuentra resuelta experimentalmente deben ser muy semejantes (mínimo 50%). Segundo, debe haber un alineamiento correcto entre la secuencia de la proteína problema y su molde correspondiente (Marti-Renom *et al.*, 2000).

En la actualidad existe una gran variedad de programas que cumplen con los requisitos básicos del modelaje por homología. Entre estos programas se encuentran el MODELLER y el SAM-T06.

MODELLER (Bino y Sali, 1993; <http://salilab.org/modeller>) es un programa computacional que permite el modelaje comparativo. Primero MODELLER realiza una búsqueda de secuencias similares a la secuencia

de la proteína objetivo mediante el alineamiento de esta secuencia con las proteínas registradas en la base de datos del Protein Data Bank (PDB, <http://www.rcsb.org/>). En el PDB se encuentran los datos y las coordenadas de las estructuras tridimensionales de proteínas resueltas experimentalmente ya sea por rayos-X o por NMR (Bino y Sali, 1993). Después, MODELLER parte de los alineamientos obtenidos para poder generar un modelo tridimensional de la estructura de la proteína al **"enhebrar" la secuencia objetivo sobre el modelo utilizando las coordenadas tridimensionales de los átomos del esqueleto de la proteína modelo para las posiciones de los aminoácidos que alinearon con la secuencia objetivo, y aplicando a los átomos correspondientes de los aminoácidos de la secuencia objetivo. Finalmente, MODELLER reconstruye las partes faltantes mediante métodos de mecánica y dinámica molecular.**

MODELLER está escrito en Fortran 90 y corre en una Pentium PC (Linux y Windows XP), Apple Macintosh (OS X) y estaciones de trabajo de Silicon Graphics (IRIX), Sun (Solaris), IBM (AIX) y DEC Alpha (OSF/1). El programa es utilizado con su propio idioma y no incluye interfaz gráfica, lo que le hace rápido y potente (Eswar *et al.*, 2003).

Otro servidor es SAM-T06 (Karplus *et al.*, 1998, 2001, 2003, 2005; http://www.soe.ucsc.edu/research/compbio/SAM_T06/T06-query.html), donde SAM es alineamiento y modelaje de secuencia (Sequence Alignment and Modeling). Dada una secuencia de una proteína problema, los algoritmos de SAM construyen un modelo de Hidden Markov (HMM) y un alineamiento múltiple que después buscan en subgrupos de base de datos utilizando el WU-BLAST en lugar de buscar en toda la base de datos no redundante (NR). Las secuencias de las posibles proteínas moldes son alineadas con la el modelo de HMM.

Teniendo los moldes, se construye el modelo de la proteína problema (Karplus *et al.*, 1998). Para el reconocimiento del plegamiento, el método SAM es extremadamente bueno con niveles aceptables de error, encontrando muchas más estructuras homólogas que los métodos tradicionales (Karplus *et al.*, 1998; 2005).

III. Hipótesis

La proteína NaTrxh es una tiorredoxina tipo *h* de *Nicotiana glauca* cuyo motivo **N β de la región N-terminal** tiene la información suficiente para que se secrete, por lo que es probable que dicha región esté expuesta al solvente en el polipéptido como un posible sitio de interacción con otras proteínas.

IV. **O**jetivos

IV.1. **O**jetivos Generales

- a)** Obtener un modelo tridimensional de la proteína NaTrxh para inferir si la región $N\beta$ se encuentra expuesta al solvente

- b)** Evaluar el modelo tridimensional de NaTrxh con diferentes simulaciones hasta llegar a un modelo con un estado de energía mínima

v. Materiales y Métodos

V.1. PROGRAMAS

Se utilizaron diferentes programas para el modelaje de la proteína, dinámica y minimización del modelo, su representación gráfica y finalmente validación de la estructura tridimensional de la NaTrxh.

V.1.1. MODELLER

Es un programa computacional que permite el modelaje comparativo, realizando búsqueda de secuencias de proteínas resueltas experimentalmente, con una similitud de por lo menos el 50% con la secuencia de la proteína objetivo, en este caso la NaTrxh. Después construye un modelo al que finalmente le hace una validación (Sali y Blundell, 1993). Para llegar al modelo final se utilizaron diferentes comandos, descritos a continuación.

V.1.1.1. Comando **profile.build()**

Éste comando busca las secuencias relacionadas de la estructura conocida (Tabla 1).

V.1.1.2. Comando **alignment.compare_structures()**

El **alignment.compare_structures()** se utilizó para valorar la estructura y la similitud de la secuencia de NaTrxh con los 21 posibles templados (Tabla 2).

V.1.1.3. Comando **align2d()**

Sirvió para alinear a NaTrxh con la 1ti3 (Tabla 3)

```

from modeller import *
log.verbose()
env = environ()
#-- Prepare the input files
#-- Read in the sequence database
sdb = sequence_db(env)
sdb.read(seq_database_file='pdb_95.pir', seq_database_format='PIR',
         chains_list='ALL', minmax_db_seq_len=(30, 4000),
         clean_sequences=True)
#-- Write the sequence database in binary form
sdb.write(seq_database_file='pdb_95.bin', seq_database_format='BINARY',
         chains_list='ALL')
#-- Now, read in the binary database
sdb.read(seq_database_file='pdb_95.bin', seq_database_format='BINARY',
         chains_list='ALL')
#-- Read in the target sequence/alignment
aln = alignment(env)
aln.append(file='Nalata_trxh.pir', alignment_format='PIR', align_codes='ALL')
#-- Convert the input sequence/alignment into
#   profile format
prf = aln.to_profile()
#-- Scan sequence database to pick up homologous sequences
prf.build(sdb, matrix_offset=-450, rr_file='${LIB}/blosum62.sim.mat',
         gap_penalties_ld=(-500, -50), n_prof_iterations=1,
         check_profile=False, max_aln_evalue=0.01)
#-- Write out the profile in text format
prf.write(file='build_profileTRXH.prf', profile_format='TEXT')
#-- Convert the profile back to alignment format
aln = prf.to_alignment()
#-- Write out the alignment file
aln.write(file='build_profileTRXH.ali', alignment_format='PIR')

```

Tabla 1. Comando del MODELLER para buscar alineamientos relacionados con NaTrxh. El nombre del archivo fue *'BasicModellinScript1.py'*.

```

from modeller import *

env = environ()
aln = alignment(env, file='pdball.pir',
               align_codes=(#'lerv',
                           ## la lerv la quite porque:
                           ##     File "BasicModellinScript2.py", line 7, in ?
                           ##     aln.malign3d()
                           ##     File "/Library/modeller-9v2/modlib/modeller/alignment.py", line
                           ##     296, in malign3d
                           ##     edit_file_ext)
                           ## _modeller.error: rdpdb__303E> No atoms were read from the
                           ## specified input PDB file, since the starting residue number and/or
                           ## chain id in MODEL_SEGMENT (or the alignment file header) was not
                           ## found; requested starting position: residue number " 1", chain " "

'1cqqA', '1dbyA', '1ep7A', '1faaA', '1fb6A', '1nswA', '1syrA',
'1ti3A', '1xwaA', '1m7tA', '1t00A', '1xflA'))
aln.malign()
aln.malign3d()
aln.compare_structures()
aln.id_table(matrix_file='family.mat')
env.dendrogram(matrix_file='family.mat', cluster_cut=-1.0)

```

Tabla 2. Comando del MODELLER **alignment.compare_structures()** para valorar la estructura y la similitud de la secuencia entre los posibles templados. El nombre del archivo fue *'BasicModellinScript2.py'*.

```

from modeller import *

env = environ()
aln = alignment(env)
mdl = model(env, file='1ti3', model_segment=('FIRST:A','LAST:A'))
aln.append_model(mdl, align_codes='1ti3A', atom_files='1ti3.pdb')
aln.append(file='build_profileTRXH.ali',
align_codes='Trxh,_N._alata_(DQ021448)')
aln.align2d()
aln.write(file='NATrxh-1ti3.ali', alignment_format='PIR')
aln.write(file='NATrxh-1ti3.pap', alignment_format='PAP')

```

Tabla 3. Comando del MODELLER **align2d()** que sirvió para alinear a la NaTrxh con la tiorredoxina de *Populus tremula* (1ti3). El nombre del archivo fue '*BasicModelinScript3.py*'.

V.1.1.4. Comando **automodel**

El programa MODELLER construyó varios modelos de la proteína NaTrxh utilizando el comando **automodel**. En este caso se le indicó que fueran 5 modelos diferentes (`a.ending_model = 5`) partiendo de un templado (`a.starting_model = 1`) (Tabla 4).

```

.
from modeller import *
from modeller.automodel import *

env = environ()
a = automodel(env, alnfile='NATrxh-1ti3.ali',
knowns='1ti3A', sequence='Trxh,_N._alata_(DQ021448)')
a.starting_model = 1
a.ending_model = 5
a.make()

```

Tabla 4. Comando del MODELLER **automodel**, el cual indica cuantos modelos se quieren obtener. *a. starting model* muestra que se parte de un molde (1ti3) y *a. ending_model* señala que se obtendrán 5 modelos diferentes. El número de modelos obtenidos fue una decisión aleatoria. El nombre del archivo fue '*BasicModelinScript4.py*'. Los modelos obtenidos se llamaron '*BasicModelinScript5_1.py* al *5.5.py*'

V.1.1.5. Comando **assess_dope**

Este comando se utilizó para elegir el modelo óptimo de los cinco obtenidos con MODELLER. Se utilizó el mismo comando para cada uno de los modelos (Tabla 5).

```

from modeller import *
from modeller.scripts import complete_pdb
log.verbose() # request verbose output
env = environ()
env.libs.topology.read(file='${LIB}/top_heav.lib') # read topology
env.libs.parameters.read(file='${LIB}/par.lib') # read parameters
# read model file
mdl = complete_pdb(env, 'Trxh,_N._alata_(DQ021448).B99990001.pdb')
# Assess with DOPE:
s = selection(mdl) # all atom selection
s.assess_dope(output='ENERGY_PROFILE NO_REPORT',
file='Trxh,_N._alata_(DQ021448).B99990001.profile',
normalize_profile=True, smoothing_window=15)

```

Tabla 5. Comando del MODELLER **assess dope**. Éste comando utiliza el algoritmo potencial DOPE (DOPE potential). El comando se utilizó para cada uno de los modelos, cambiando el nombre del archivo y obteniendo cinco archivos diferentes. En la tabla se muestra el archivo *'BasicModelinScript5_1.py'* generado para el NaTrxh_1a.

V.1.2. Swiss-PdbViewer

Debido a la presencia de un nudo que se generó en los modelos que se hicieron con el MODELLER hacía el extremo carboxilo, se utilizó el programa Swiss-PDB-Viewer (Guex y Peitsch, 1997; <http://expasy.org/spdbv/text/refs.htm>) que dentro de sus herramientas se encuentra el algoritmo Ramachandran que genera una interfaz gráfica basada en los ángulos phi y psi, en donde define las parejas de valores de phi y psi que están permitidos (Coutsias *et al.*, 2004). En este caso, el programa nos permitió modificar las conformaciones de aminoácidos individuales con sólo **“arrastrar”** los puntos del mapa de un sitio a otro, deshaciendo el nudo que se formó al final del carboxilo terminal.

V.1.3. HyperChem

El hecho de haber deshecho el nudo del carboxilo terminal del modelo de NaTrxh basados en la gráfica del Ramachandran, podría implicar que el modelo no era el adecuado. Se decidió entonces utilizar el programa HyperChemTM (HyperCube, Inc; <http://www.hyper.com/>), ya que dentro del programa se encuentra un algoritmo llamado **“recocimiento simulado”** (simulated annealing; SA). Este algoritmo se podría calificar como escalador estocástico. En

este caso, se utilizó una temperatura alta y una baja, de tal forma que el modelo se calentara pero sin romper los enlaces covalentes entre los aminoácidos. Después, la temperatura descendió poco a poco permitiendo que los átomos de los aminoácidos se acomodaran de la mejor manera posible. Todo esto se simuló en condiciones de vacío.

V.1.4. Visual Molecular Dynamics (VMD)

Todas las estructuras tridimensionales presentadas se analizaron y animaron con el programa Visual Molecular Dynamics (VMD) el cual presentó los resultados en tercera dimensión a partir de los archivos PDB obtenidos en el transcurso del presente trabajo de los diferentes modelos tridimensionales obtenidos de la NaTrxh (Humphrey *et al*, 1996; <http://www.ks.uiuc.edu/Research/vmd/>). La versión que se utilizó fue la 1.8.6, la cual soporta Microsoft Windows 2000/XP/Vista (32-bit) utilizando Open GL. También está disponible para Linux y MacOS.

Dentro del programa VMD existen incluidos diferentes analizadores bioinformáticos, de los cuales el más utilizado, en este caso, fue MultiSeq 2.0 (Roberts *et al.*, 2006; <http://www.scs.uiuc.edu/~schulten/multiseq>), que incluía algoritmos como Qres (Eargle *et al*, 2006) y STAMP (Russell y Barton, 1992). Tanto Qres como STAMP fueron utilizados para hacer las superposiciones entre los diferentes modelos tridimensionales obtenidos y hacer una comparación visual.

VI. Resultados

VI.1. MODELAJE DE LA PROTEÍNA NaTrxh con MODELLER

La proteína NaTrxh es la primera Trx de plantas que se demuestra que se secreta hacia el apoplasto y además datos de nuestro grupo demuestran que la región responsable de dicha secreción es la N-terminal, específicamente la región N β (Juárez-Díaz, 2006b; Cruz-García, datos no publicados). Se podría considerar entonces, que dicha región estaría expuesta facilitando la interacción con otras proteínas como parte del mecanismo de secreción. Una forma de obtener evidencia para apoyar esta propuesta es modelando la estructura tridimensional de la proteína, ya que aún no se encuentra cristalizada, ni se ha resuelto su estructura tridimensional.

Para realizar el modelaje se recurrió a una variedad de estrategias, apoyadas en el uso de diversos programas de cómputo. En todos los modelos obtenidos, la región correspondiente al extremo amino resultó con un grado importante de accesibilidad al solvente. El proceso de construcción de los modelos y de sus características se describen en el presente capítulo.

VI.1.1. Obtención de un primer modelo

Una primera aproximación fue el uso del programa MODELLER (Bino y Sali, 1993; <http://salilab.org/modeller>). Este programa que consta de varios módulos, realizó primeramente una búsqueda de secuencias similares a la de la secuencia de la NaTrxh mediante el alineamiento de la NaTrxh con proteínas resueltas experimentalmente ya sea por cristalografía de rayos-X o por NMR que se encuentran registradas en el PDB. A partir de los alineamientos obtenidos, se buscó la proteína modelo más adecuada para modelar a NaTrxh y después se modeló a la NaTrxh utilizando las coordenadas tridimensionales de los átomos del esqueleto de la proteína modelo

para las posiciones de los aminoácidos que alinearon con la NaTrxh. Finalmente, MODELLER reconstruyó las regiones faltantes, aquellas que no tuvieron similitud con el molde, mediante métodos de mecánica y dinámica molecular.

VI.1.1.1. Búsqueda de las estructuras relacionadas con NaTrxh

Para realizar la búsqueda, se empleó el módulo **profile.build()** (Tabla 1), que requiere la secuencia en formato PYR, con ciertos datos adicionales embebidos en el encabezado. El archivo introducido a MODELLER para la NaTrxh en el formato PYR (Sali y Blundell, 1993; Fiser y Sali, 2003; Eswar *et al.*, 2000; Marti-Renom *et al.*, 2000) se muestra en la Figura 7. Los datos en el archivo son: 1) el nombre de la secuencia (línea 1), antecedido por el código '>P1' que indica que se trata de una secuencia de aminoácidos; 2) Campos de datos sobre la secuencia referida, separados por el carácter ':' (línea 2); y 3) la secuencia de aminoácidos en el código de una letra y terminada con el carácter '*' (línea 3).

Los campos de datos en la línea 2 son:

- 1: Origen de los **datos de la secuencia**: "structureX (resuelta por rayos X), structureN (resuelta por NMR), structureM (estructura obtenida por modelado), sequence (no hay datos de estructura, **aplicable a la secuencia objetivo**)".
- 2: Código del PDB correspondiente. Debe ser único para las secuencias en un archivo. Si se usa este campo, el archivo estructural correspondiente debe estar disponible al programa.
- 3 a 6: Identificadores de la cadena polipeptídica y el intervalo de los aminoácidos a considerar (vgr: A: 25:A:75, del aminoácido 25 al 75 de la cadena A). Se pueden **usar claves como '@'** para referirse a la cadena, FIRST para el primer aminoácido de la cadena, LAST para el último, '+' para referirse a una posición relativa respecto a la anterior y '.' Para referirse a un identificador de cadena ausente, ya que un espacio puede ser un identificador de cadena válido).

- 7: Nombre de la proteína (opcional)
- 8: Organismo de origen de la proteína (opcional)
- 9: Resolución del cristal si existe (opcional)
- 10: Factor R del análisis cristalográfico, si aplica (opcional)

Sólo dos de estos campos son necesarios para las secuencias a modelar (campo 1 'sequence' y el nombre). En la figura 7 se muestra la secuencia objetivo, cuyo nombre es 'Trxh._N._alata_(DQ021448)'.

```
>P1;Trxh,_N._alata_(DQ021448)
sequence:Trxh,_N._alata_(DQ021448):.....:0.00: 0.00
MGSYLSSLLGGGAAEAAEAESGSSSEPSRVIAFHSSNRWQLHFNSSKQLNKLIVVDFAAATWCGPCKMM
EPVINAMSAKYTDVDFVKIDVDELSDVAQEFQVQAMPTFLLLKQGKEVERVVGAKKDELEKKILKHREA
PKFAASNYRTRKFHVQ*
```

Figura 7. Secuencia de aminoácidos de NaTrxh en formato PYR y los campos para MODELLER

El resultado del comando **profile.build()** (Tabla 1) utilizando la matriz BLOSUM62, fue una serie de números de identificación de estructuras del PDB, con las secuencias de aminoácidos para proteínas que tenían similitud a la NaTrxh y que eran posibles candidatos para ser las proteínas modelo (Figura 8).

VI.1.1.2. Selección de un molde

La Figura 8 muestra un total de 21 estructuras de proteínas relacionadas a la Trx de *N. alata*. Todas las Trx presentan el motivo del sitio activo característico de las Trx, la gran mayoría con la secuencia WCGPC (Holmgren, 1985). Ninguna de las secuencias encontradas tiene una extensión hacia el amino terminal ni hacia el

carboxilo terminal que coincide con la de *N. alata*. De las 21 Trx con estructura resuelta y relacionada con la NaTrxh hay 7 tipos diferentes que se presentan con sus códigos PDB: Trx varias (Trx: 2trx, 1r26, 1syr, 1thx, 1xwa, 1v98, 1t00, 1m7t), Trx tipo 1 (Trx1: 1oaz), Trx tipo *h* (Trx *h*: 1erv, 1ep7, 1ti3), Trx tipo *f* (Trx *f*: 1faa), Trx tipo *m* (Trx *m*: 1dby, 1fb6), proteínas semejantes a Trx (Trx like protein: 1gh2) y otras (1mek, proteína disulfuro isomerasa; 1xf1, c5a, peptidasas). Su origen biológico incluye Trx de *Drosophila melanogaster* (1xwa); de *Homo sapiens* (1erv, 1gh2, 1mek, 1nsk); de bacterias como *Escherichia coli* (2trx, 1oaz), de *Thermus hermophilus* (1v98), de *Streptomyces coelicolor* (1t00) y *Streptococcus pyogenes* (1xf1); hongos como *Chlamydomonas reinhardtii* (1dby, 1ep7); de plantas como espinaca (1faa, 1fb6, de cloroplasto); de parásitos como *Trypanosoma brucei brucei* (1r26), y *Plasmodium falciparum* (1syr); de algas como la *Anabeana* sp (1thx); y hasta una quimera recombinante de humano con *E. coli* (1m7t).

Por supuesto, las ideales a escoger serían las Trx tipo *h* ya que la NaTrxh es una Trx de este tipo (Juárez-Díaz *et al.*, 2006a y b). Además, se debe considerar la calidad de los alineamientos obtenidos por MODELLER y por supuesto el porcentaje de identidad, que debe ser por lo menos del 50% (Topham *et al.*, 1991; Sali y Blundell, 1993; Baker y Sali, 2003).

Se prosiguió entonces a hacer un escrutinio más exhaustivo para descartar las secuencias que no eran adecuadas para el modelaje de la NaTrxh, así que se utilizó el comando **alignment.compare_xstructures()** (Tabla 2), que valoró la estructura y la similitud de la secuencia entre los 13 posibles moldes, obteniéndose una comparación de semejanza y coincidencia entre las secuencias (Figura 9).

```

# Number of sequences:      21
# Length of profile :      152
# N_PROF ITERATIONS :      1
# GAP_PENALTIES_1D :      -500.0  -50.0
# MATRIX_OFFSET :          -450.0
# RR_FILE :                 $(LIB)/blosum62.sim.mat

 1 NaTrxh S    0  152    1  152    0    0    0    0.0    0.0    MGSYLSS
 2 lerv X     1  105   50  131   20  101   82  51.0    0.0    -----
 3 lcqgA X     1  105   50  131   20  101   82  50.0    0.0    -----
 4 ldbya X     1  107   53  122   22   92   70  41.0    0.0    -----
 5 lep7A X     1  112   30  137    4  112  108  43.0    0.0    -----
 6 2trxA X     1  108   53  129   23  101   77  37.0    0.74E-11 -----
 7 lfaaa X     1  121   50  131   34  116   82  41.0    0.0    -----
 8 1fb6A X     1  104   53  134   20  103   82  36.0    0.0    -----
 9 1gh2A X     1  107   50  136   21  107   87  33.0    0.16E-09 -----
10 1mek X     1  120   26  138    5  118  107  31.0    0.41E-09 -----
11 1nswA X     1  105   42  131    9  100   90  36.0    0.0    -----
12 1oazA X     1  115   53  129   23  108   71  33.0    0.92E-07 -----
13 1r26A X     1  113   52  131   28  107   80  35.0    0.49E-11 -----
14 1syrA X     1  103   40  134    8  102   95  41.0    0.0    -----
15 1thx X     1  108   53  127   23   99   75  32.0    0.13E-07 -----
16 1ti3A X     1  113   25  137    1  113  113  50.0    0.0    -----
17 1xwaA X     1  111   46  131   21  107   86  47.0    0.0    -----
18 1m7tA X     1  107   50  129   20  100   80  48.0    0.0    -----
19 1v98A X     1   92   52  138    4   87   83  36.0    0.37E-11 -----
20 1t00A X     1  112   50  123   23   97   74  39.0    0.0    -----
21 1xflA X     1  114   24  134    2  112  111  46.0    0.0    -----

 1 NaTrxh LLGGGAEEAAEAESGSSSEPSRVIAFHSSNRUQLHFNSSKQLNKLIVVDF AATWCGPCKMMEPVINAM
 2 lerv -----DKLVVVD FSATWCGPCKMIKPFHSL
 3 lcqgA -----DKLVVVD FSATWCGPCKMIKPFHSL
 4 ldbya -----VLVDFWAPWCGPCRIIAPVVDEI
 5 lep7A -----VIVIDSKAAWDAQLAKGKEEHKPIVVDF TATWCGPCKMIAPLFETL
 6 2trxA -----ILVDFWAEWCGPCKMIAPILDEI
 7 lfaaa -----DKPVVLD MF TQWCGPCKAMAPKYEKL
 8 1fb6A -----VMVDFWAPWCGPCKLIAPVIDEL
 9 1gh2A -----SRLAVVKFTMRGCGPCLRIAPAFSSM
10 1mek -----EEDHVLVLRKSN-----FAEALA AHKYL LVEFYAPWCGHCKALAPAAGKL
11 1nswA -----NFQQA I QGDGPVLVDFWAAWCGPCRMMAPVLEEF
12 1oazA -----ILVDFWAEWCGPCKMIAPILDEI
13 1r26A -----LTVAWFTAVWCGPCKTIERPMEKI
14 1syrA -----QAEFDS I ISQNELVIVDFFAEWCGPCKRIAPFYEEC
15 1thx -----VLVYFWASWCGPQQLMSPLINLA
16 1ti3A -----AEEGQVIACHTVDTWKEHFEKGGKSQKLI VVDF T ASWCPPCKMIAPIF AEL
17 1xwaA -----TKASGLVVDLDFATWCGPCKMISP KLVEL
18 1m7tA -----DKLVVVD FSATWCGPCKMIKPFHSL
19 1v98A -----LTLVDF F APWCGPCLRVSP ILEEL
20 1t00A -----DKPVLVDFWAAWCGPCRQIAP SLEAI
21 1xflA -----ASEEGQVIACHTVETWNEOLOKANESKTLVVVDFTASWCGPCRFIAPFFADL

 1 NaTrxh SAKYTDVDFVKIDVDELSDVAQEFVQAMP T F L L L KQKKEVERVVGAKKDELEKKILKHREAPKF AASNVRTK FHVQ
 2 lerv SEKYSNVIFLEVDVDDCQDVASECEVKSMP T F O F F KKGQKVG E F S GANKEKLEATI-----
 3 lcqgA SEKYSNVIFLEVDVDDAQDVASEAEVKATP T F O F F KKGQKVG E F S GANKEKLEATI-----
 4 ldbya AGEYKDLKCVKLN TDESPNVASEYGI RSIPTIMV F KGGKKE T I IGA-----
 5 lep7A SNDYAKVIFLKVVDVDAVA A VAE AAGITAMP T F HVYKDG V KADDLV GASQDKL KALVAKHAAA-----
 6 2trxA ADEYQKLVAKLNIDQNPGTAPKYGIRGIP T L L L F KNGEV AATKVGASKGOLKE-----
 7 lfaaa AEEYLDVIFLKLDCNENKTLAKELGIRVVP T F KILKENS VVGEV T GAKYDKLEAI-----
 8 1fb6A AKEYSKI AVYKLN TDEAPGIATQYNIRSIPTLVLFKNGERKESI I GAPKSTL TDSIEKY-----
 9 1gh2A SNKYPQAVFLEVDVHQCQGTAA TNNISATP T F O F F R N K V R I D Q Y Q G A D A V G L E E K I R Q H L E-----
10 1mek KAEGSEIRLAKVDATEESDLAQQGVGRGYP T I K F F R N G D T A E Y T A G R E A D D I - V N W L K K R T G P-----
11 1nswA AEAHADV T VAKLNVDENPETSQF G I M S I P T L I L F K G G R P V K Q L I G Y O K E Q L E A Q L-----
12 1oazA -----LTVAKLNIDQNPGTAPKYGIRGIP T L L L F KNGEV AATKVGASKGOLKE-----
13 1r26A AYEFPTVKFAKVDADNNSIVSKCRVLQLP T F I I A R S G K M L G H V I G A N P G M L R Q K L-----
14 1syrA SKTYTKMVF I K V D V D E V S E V T E K E N I T S M P T F K V Y R N G S S V D T L L G A N D S A L K Q L I E K Y-----
15 1thx ANTYSDLKVVKLEIDPNPTTVKKYKVEGVPALRLVKGEQILDSTEGVSKDKL-----
16 1ti3A AKKFPNV T F L K V D V D E L K A V A E E W N V E A M P T F I F L K D G K L V D K T V G A D K D G L P T L V A K H A T A-----
17 1xwaA STQFADVVVLKVDVDECEDIAMEYNISSMPT F V F L K N G V K V E E F A G A N A K R L E D V I-----
18 1m7tA SEKYSNVIFLEVDVDDAQDVAPKYGIRGIP T L L L F K N G E V A A T K V G A S K G O L K E-----
19 1v98A ARDHAGLKVVVWVNDHEPGLAARYGVRSVPTLVLFRRGAPVATWVGAS----PRRVLEERLRP-----
20 1t00A AA E Y G D I E I V K L N I D E N P G T A A K Y G V M S I P T L M V Y Q G G E V A K T I V G A K-----
21 1xflA AKKLPNVFLKVD T D E L K S V A S D W A I O A M P T F M F L K E G K I L D K V V G A K K D E L O S T I A K H-----

```

Figura 8. Resultado obtenido después de introducir el comando **build.profile()** en el MODELLER. Alineamiento de las secuencias de las posibles proteínas molde. La primera proteína es NaTrxh. Se observa que muy pocas proteínas tienen una extensión hacia la región amino terminal, e incluso ninguna abarca la región completa.

A)

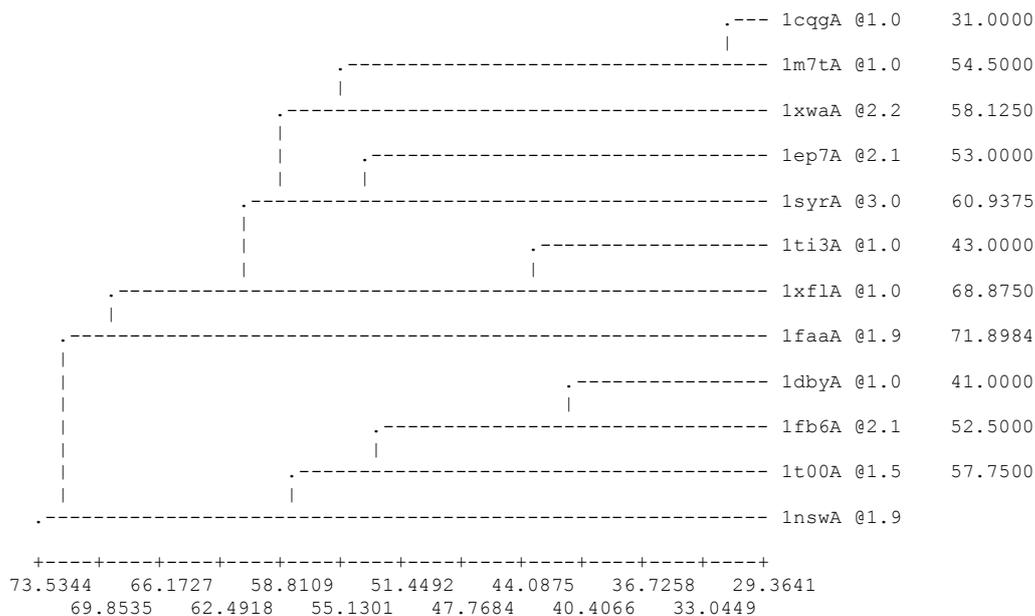
Comparación de secuencia de identidad (ID_TABLE):

Diagonal ... número de residuos;
 Triangulo superior ... número de residuos idénticos;
 Triangulo inferior ... % de identidad de secuencia, id/min(length).

	1cggA	@1ldbYA	@1lep7A	@21faaA	@11fb6A	@21nswA	@11syrA	@31ti3A	@11xwaA	@21m7tA	@11t00A	@11xf1A	@1
1cggA @1	105	27	48	38	30	30	38	39	49	72	23	42	
ldbYA @1	26	107	31	31	61	44	35	31	33	32	55	36	
lep7A @2	46	29	112	34	33	32	48	55	45	48	30	48	
1faaA @1	36	29	30	121	31	27	33	33	37	38	31	34	
1fb6A @2	29	59	32	30	104	39	37	32	39	38	46	34	
1nswA @1	29	42	30	26	38	105	35	27	34	34	47	27	
1syrA @3	37	34	47	32	36	34	103	38	44	40	35	36	
1ti3A @1	37	29	49	29	31	26	37	113	43	36	24	64	
1xwaA @2	47	31	41	33	38	32	43	39	111	47	27	40	
1m7tA @1	69	30	45	36	37	32	39	34	44	107	35	38	
1t00A @1	22	51	27	28	44	45	34	21	24	33	112	28	
1xf1A @1	40	34	43	30	33	26	35	57	36	36	25	114	

B)

Agrupamiento medio de pares en grupo basados sobre la distancia de la matriz:



Tiempo total de CPU [segundos]:6.83

Figura 9. Comparación de semejanza entre secuencias. **A)** Matriz de comparaciones de semejanza de secuencia entre el objetivo y las secuencias de proteínas del PDB que reportan un valor del estadístico E altamente significativo. **B)** Dendrograma de la matriz. La primera columna muestra el nombre del PDB de la secuencia que se está comparando en la base de datos del programa MODELLER. La segunda columna indica la resolución cristalográfica en ångströms. En el eje de las X se observa la identidad en porcentaje de las diferentes secuencias. La 1ti3:A es la secuencia que tiene mayor identidad con la NaTrxh, además de que es una Trx tipo *h* pero del subgrupo I, aunque tiene una extensión hacia el extremo amino y el carboxilo. A las proteínas resueltas mediante NMR se les asigna una resolución genérica de 1.0 Å.

A partir de la Figura 9 se eligió a 1ti3 (cadena A) como molde, con base en las siguientes consideraciones: a) es una Trx tipo *h*, al igual que NaTrxh; b) el organismo de origen es una planta superior, al igual que NaTrxh; c) tiene una mejor coincidencia de secuencia con NaTrxh (50%) que otras Trx *h* de plantas con estructura conocida (1xfl y 1faa); y, d) posee una extensión hacia el extremo amino, aunque no tan larga como la de NaTrxh y está también entre las que poseen una porción más extensa del extremo carboxilo, de nuevo, no tan extensa como la de NaTrxh (Figura 10).

Una vez que se escogió la proteína modelo que serviría como base para modelar por homología a la NaTrxh, lo siguiente fue alinear a la NaTrxh con 1ti3 que es una Trx tipo *h* de *Populus tremula*. Para esto se utilizó el comando **align2d()** (Tabla 3).

El comando **align2d()** (Tabla 3) está basado en un algoritmo programado dinámicamente, diferente de los métodos estándares de alineamiento de secuencia-secuencia ya que toma en cuenta la información estructural acerca de los moldes cuando se construye el alineamiento. Con estos métodos, los errores de alineamiento se reducen aproximadamente a un tercio con respecto a otros métodos (Sali y Blundell, 1993; Fiser y Sali, 2003; Eswar *et al.*, 2000; Marti-Renom *et al.*, 2000).

VI.1.1.3. Alineamiento de NaTrxh con la proteína molde (1ti3)

En la Figura 10 se observa el alineamiento de la NaTrxh con 1ti3 (PtTrxh1) donde hay más de 50 posiciones conservadas entre ambas secuencias. También se observa que de los 28 aminoácidos que corresponden a la región N-terminal de NaTrxh, PtTrxh1 presenta tres aminoácidos de esta región. Aunque aparentemente sean pocos aminoácidos, a diferencia del resto de las secuencias con las que se comparó en el momento de escoger el molde (Figura 8), PtTrxh1

tiene una extensión hacia el N-terminal que resultó conveniente para el modelaje de la NaTrxh.

```

_aln.pos      10      20      30      40      50
1ti3A      -----AEEGQVIACHTVDTWKEHFEEKGKSQK
NaTrxh      MGSYLSSLLGGGAEEAAEAESGSSSEPSRVIAFHSSNRWQLHFNSSKQLNK
_consrvd      *   * * * *   *   * *   *   *
              60      70      80      90      100
1ti3A      LIVVDFTASWCPPCKMIAPIFAELAKKFPNVTFLKVDVDELKAVAEAWNVE
NaTrxh      LIVVDFAATWCGPCKMMEPVINAMSAKYTDVDFVKIDVDELSDVAQEFVQ
_consrvd      ***** * * * * * *   * * * * * * * * *
              110     120     130     140     150
1ti3A      AMPTFIFLKDGKLVDKTVGADKDGLEPTLVAKHATA-----
NaTrxh      AMPTFLLLKQGKEVERVVGAKKDELEKKILKHREAPKFAASNYRTKHFVQ
_consrvd      ***** ** * * *   * * * * *   * * *

```

Figura 10. Alineamiento de NaTrxh con la Trx tipo *h* del *P. tremula* (PtTrxh1). Hay varias regiones que se encuentran conservadas (regiones indicadas por asterisco), además de la presencia de 24 aminoácidos más hacia el amino terminal (rojo) y 15 aminoácidos más hacia el extremo carboxilo terminal (verde) por parte de la NaTrxh a diferencia de 1ti3 que sólo muestra 3 aminoácidos hacia el N-terminal (azul).

Ya se tenía entonces el molde que se alineó con la NaTrxh con 1ti3 (Figura 11), por lo que se prosiguió a construir el modelo de la Trx de *N. alata*. El programa MODELLER tiene la capacidad de construir automáticamente el modelo, y eso fue lo que se le solicitó, utilizando el comando de **automodel** (Tabla 4). Se creó un objeto **automodel** llamado 'a' y una serie de parámetros para guiar el modelo en su construcción. El archivo de datos indicados en el parámetro **alnfile** (*Trxh._N._alata_(DQO21448.ali)*) contiene el alineamiento del molde 1ti3 y el objetivo en formato PYR. El archivo indicado por el parámetro **knowns** lista los archivos PDB de las estructuras descritas en el archivo **alnfile**. El parámetro **sequence** definió el nombre del objetivo en **alnfile**. El parámetro **assess_methods** indica los criterios a utilizar para la calificación de los modelos resultantes. El parámetro **starting_model** y **ending_model** definieron el número de serie de los modelos a generar y por lo tanto, el número de

modelos solicitados, en este caso 5. El comando **make** ordenó generar los modelos de la NaTrxh (Sali y Blundell, 1993; Eswar *et al.*, 2000; Fiser y Sali, 2003; Marti-Renom *et al.*, 2000).

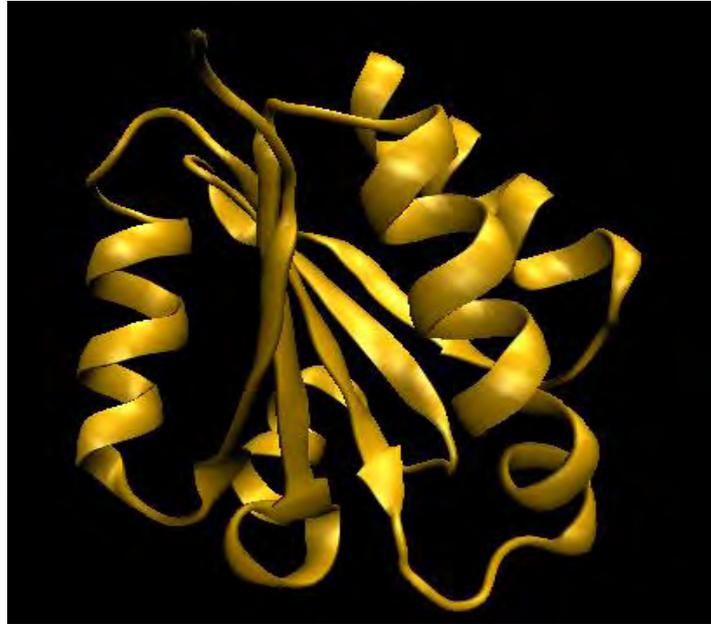
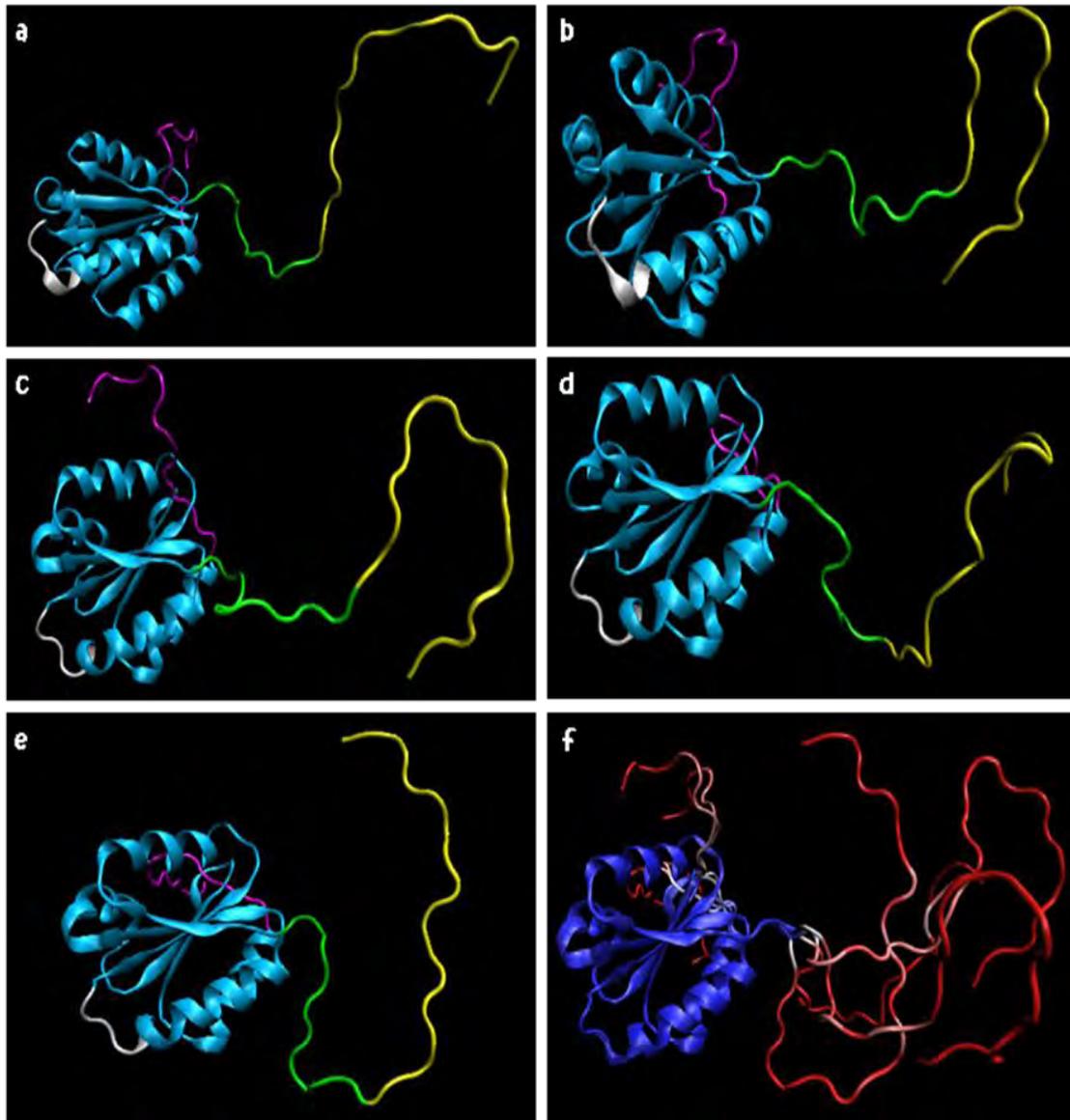


Figura 11. Modelo de la estructura tridimensional de 1ti3 (PtTrxh1, Trx de *P. tremula*). PtTrxh1 se tomó como proteína modelo debido a que tiene una similitud muy alta con la secuencia de NaTrxh y además es una Trx *h* pero del subtipo I; aún así, ésta presenta una pequeña extensión hacia el N-terminal.

VI.1.1.4. Construyendo los modelos de NaTrxh en base a 1ti3

En la Figura 12 de los paneles **a** al **e** se muestran los 5 modelos generados por **automodel**. Claramente se observa que tanto la región N-terminal como la C-terminal tienen una estructura poco compacta y que varía mucho entre los modelos; no obstante, la región que abarca el núcleo de la Trx resultó muy semejante en los 5 modelos. En la Figura 12 panel f, se muestra la sobreposición de los 5 modelos de la NaTrxh y la PtTrxh1 como molde.

A)



B)

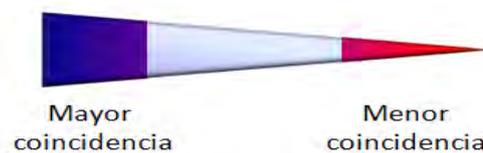


Figura 12. Comparaciones de los modelos tridimensionales de NaTrxh. **A.** (a) NaTrxh_1a, (b) NaTrxh_1b, (c) NaTrxh_1c, (d) NaTrxh_1d, (e) NaTrxh_1e, (f) superposición de los 5 modelos y de PtTrxh1 utilizando el algoritmo STAMP y Qres. **B.** Esquematización de la relación entre la coincidencia de los algoritmos Qres y STAMP. A mayor coincidencia el color es azul oscuro, a menor coincidencia el color es rojo. En **a**, **b**, **c**, **d** y **e** se observa de color azul claro el núcleo de la NaTrxh, en amarillo la región N α , en verde la región N β , en blanco el sitio activo (WCGPC) y en magenta la secuencia de aminoácidos del carboxilo terminal (138-152 aminoácidos)

La superposición se realizó empleando los algoritmos STAMP (Russell y Barton, 1992) y Qres (Eargle *et al.*, 2006) del módulo Multiseq (Roberts *et al.*, 2006), incluidos en el programa Visual Molecular Dynamics (VMD) (Humphrey *et al.*, 1996) a partir de la versión 1.8.5. El algoritmo STAMP (Russell y Barton, 1992) minimiza **la distancia del Ca entre los residuos alineados en cada molécula** a través de las transformaciones de cuerpo-rígido globalmente óptimas (Mizuguchi *et al.*, 1998). El algoritmo STAMP es un módulo iterativo que genera un alineamiento estructural y se emplea para calificar el alineamiento. Qres determina el grado de similitud estructural por residuo, sólo para los alineamientos del alineamiento estructural múltiple (Eargle *et al.*, 2006). Las regiones coloreadas en azul poseen alta semejanza entre los residuos estructuralmente equivalentes (Figura 12B). Es decir, los 5 modelos de la NaTrxh poseen el mismo plegamiento que el de la PtTrxh1. Este resultado es de esperarse, ya que PtTrxh1 y NaTrxh se alinearon de manera continua (sin huecos) en toda esa región.

VI.1.1.5. Evaluación de los modelos para la NaTrxh

Con los 5 modelos diferentes, se procedió a calificar la geometría para elegir el mejor modelo (Figura 13). Dentro del programa MODELLER se encuentra el comando **assess_dope** (Tabla 5), el cual utiliza un potencial estadístico denominado DOPE (Discrete 'Optimized' Protein Energy) **utilizado para valorar modelos generados** por homología en la predicción de estructuras de proteína. MODELLER usa este módulo para calcular la energía del modelo de la NaTrxh evaluando todas las interacciones que se establecen en cada modelo. Dichas interacciones son calculadas para cada residuo y se genera un índice por residuo a lo largo del modelo (Sali y Blundell, 1993; Fiser y Sali, 2003; Eswar *et al.*, 2000; Marti-Renom *et al.*, 2000).

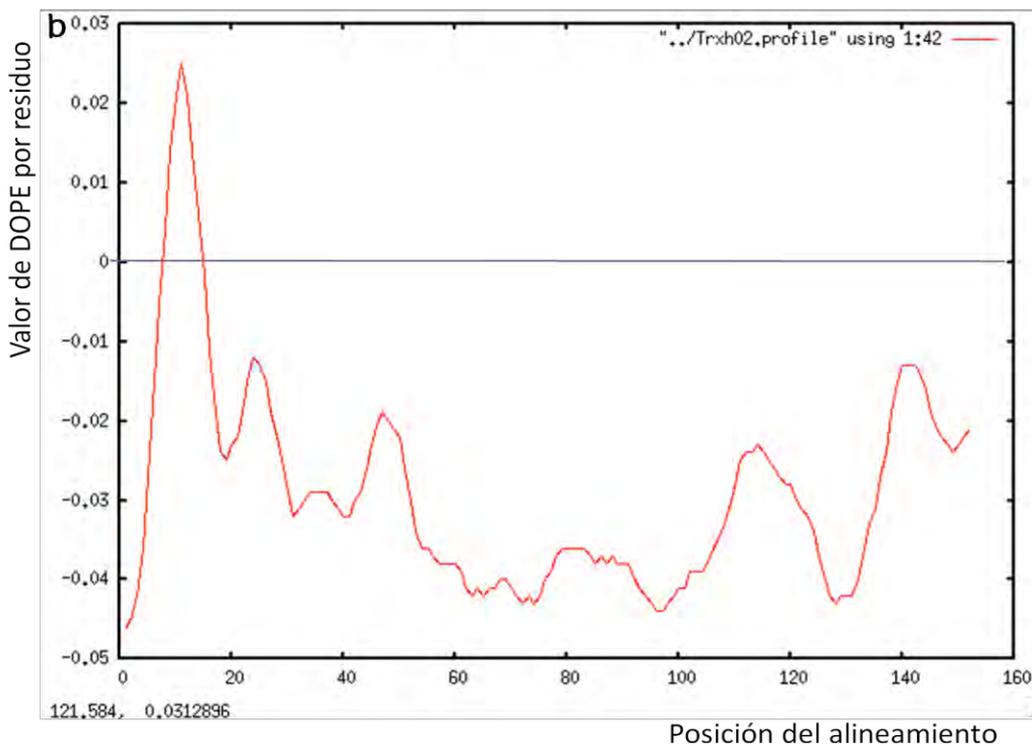
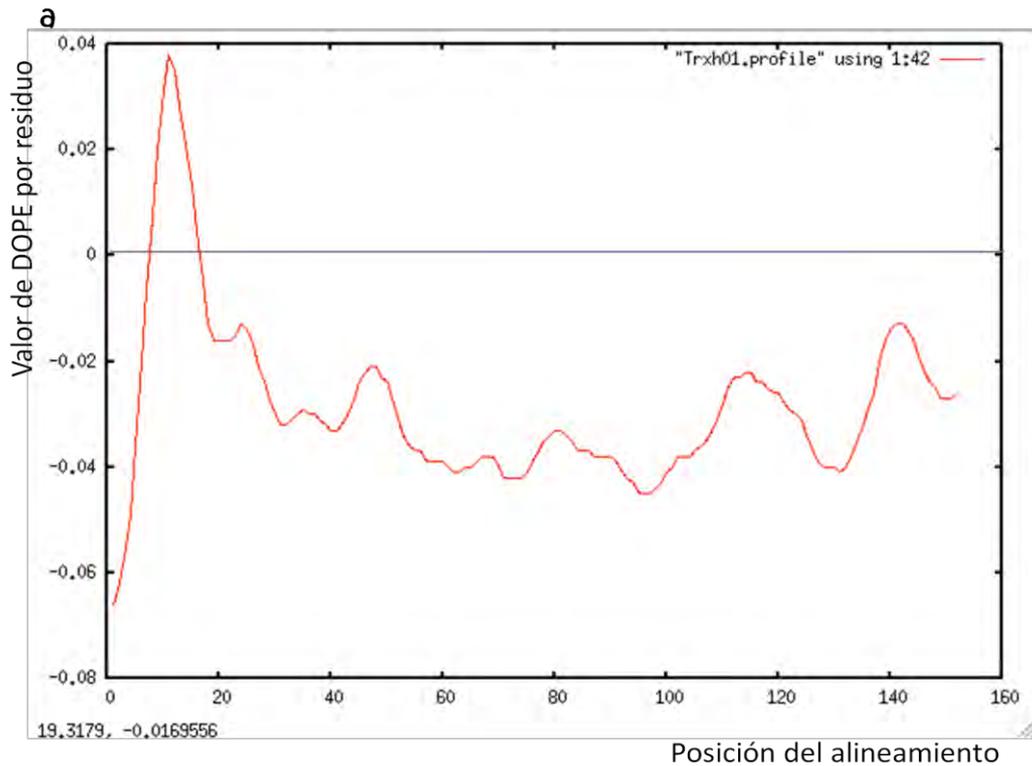


Figura 13. Selección del modelo más apropiado. Gráficas del parámetro de coincidencia estructural generado por `assess_dope()` para los modelos. El modelo `NaTrxh_1d` (**d**) es el que tiene un valor más cercano a cero, por lo cual es el que se tomó para continuar otras simulaciones. En el eje Y se muestra el valor por residuo de DOPE y en el eje X la posición en el alineamiento.

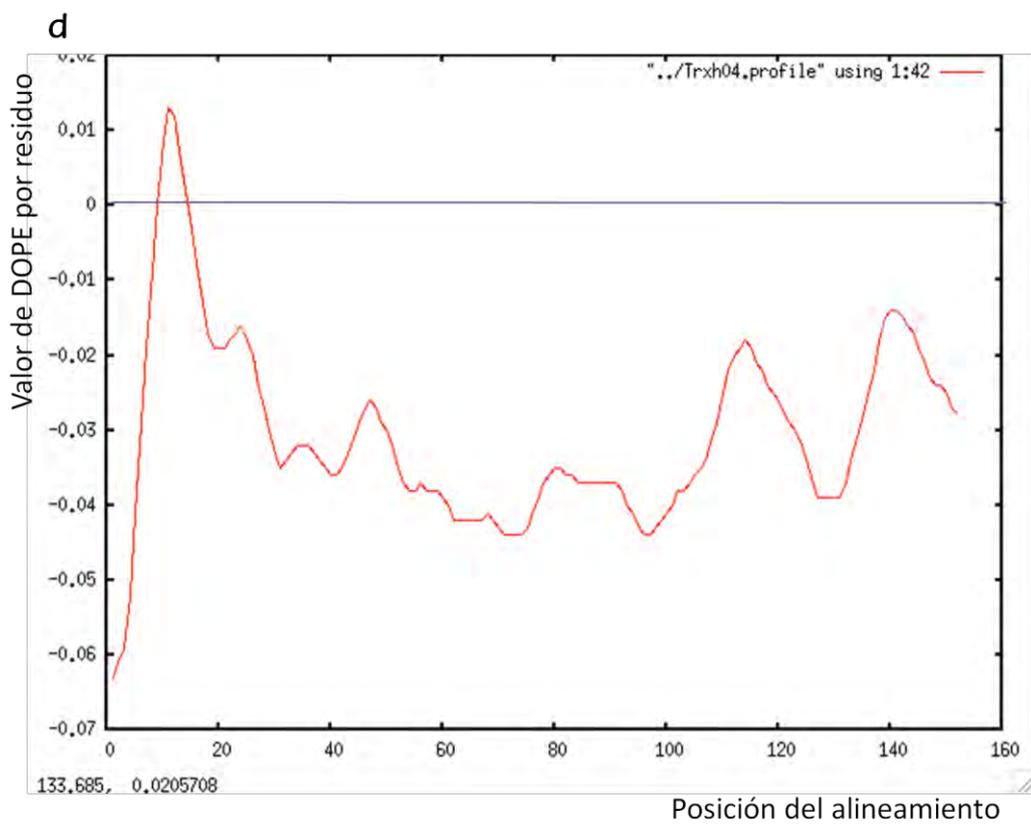
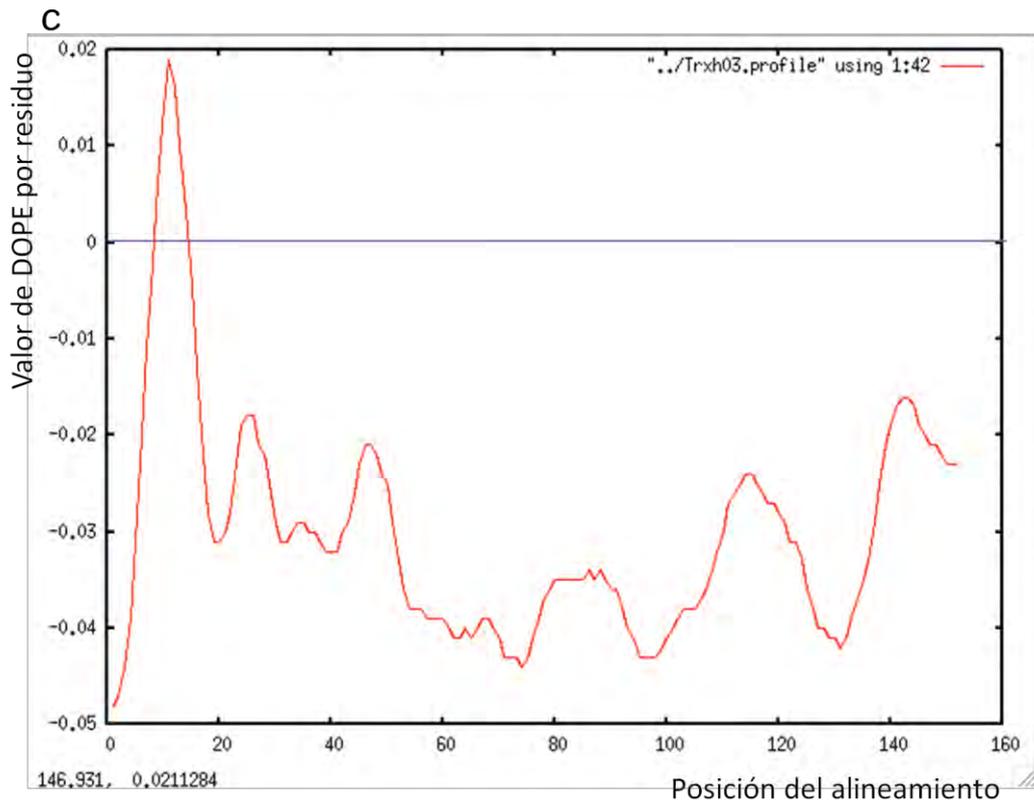


Figura 13. Continuación.

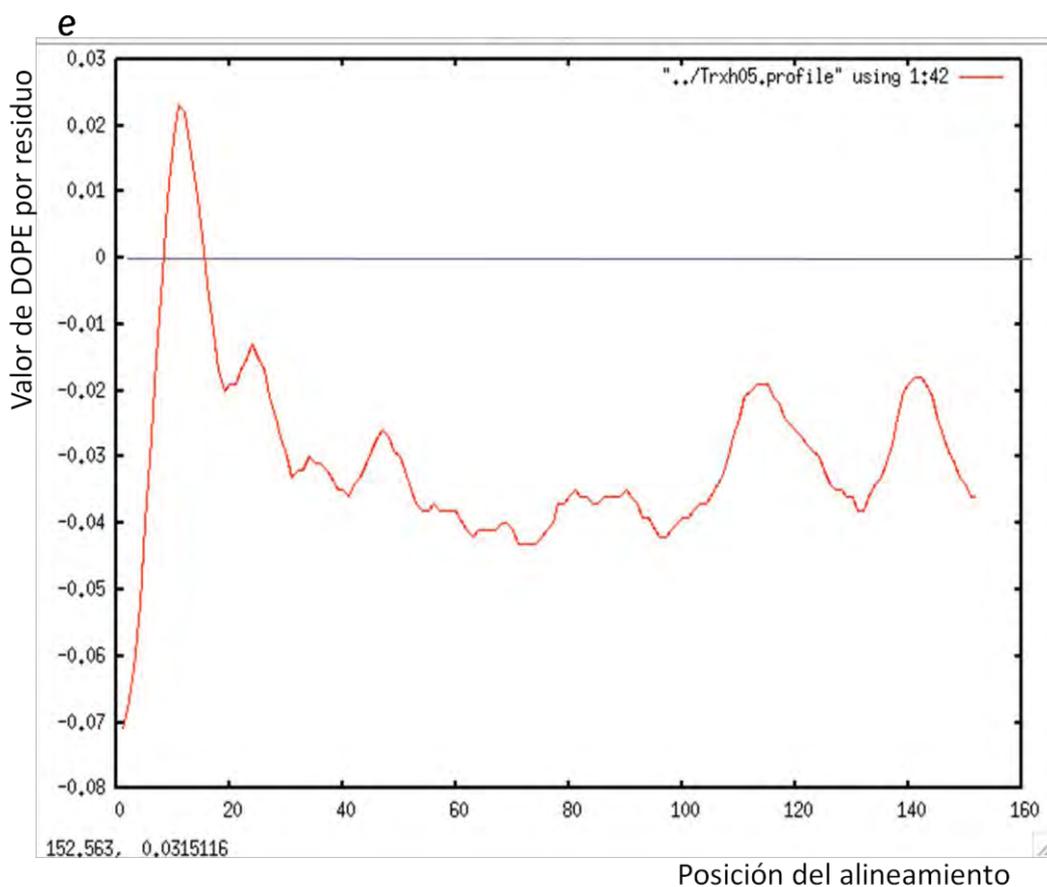


Figura 13. Continuación.

De los 5 modelos, el mejor construido por MODELLER fue el de NaTrxh_1d (Figura 13d), debido a que en la gráfica muestra un valor cercano a 0.

VI.1.2. Presencia de un "nudo" en los modelos de MODELLER

El modelo NaTrxh_1d se escogió como punto de partida, puesto que el modelo sólo cubría la porción central de la secuencia, el modelaje de los extremos amino y carboxilo resultó incierto y debía mejorarse. Dado que estas regiones son más cortas, es posible mejorar este modelo mediante simulaciones de dinámica molecular.

Además, se observó que los 5 modelos presentaban un nudo al final del carboxilo. Tales estructuras anudadas no han sido observadas en las estructuras de las proteínas naturales. Estas estructuras se encuentran tan constreñidas que los métodos de dinámica molecular son inadecuados para corregirlas.

El nudo se eliminó manualmente del modelo_1d por medio del programa Swiss PDB Viewer (Figura 14) (Guex y Peitsch, 1997); <http://www.expasy.org/spdbv/>), empleado un algoritmo asociado al mapa de Ramachandran (Coutsias *et al.*, 2004), que permite modificar las conformaciones de aminoácidos individuales con simplemente "arrastrar" los puntos del mapa de un sitio a otro. El mapa de Ramachandran es una representación gráfica de los ángulos dihedros de los enlaces entre el carbono alfa y el nitrógeno amídico y el carbono alfa y el carbonilo de cada residuo de aminoácido en la estructura de la proteína. Debido a la planaridad y rigidez del enlace peptídico, dichos ángulos son los que influyen en mayor medida en la dirección que sigue la cadena polipeptídica en el espacio. Al modificarlos, se puede modificar la conformación local de aminoácidos en la representación tridimensional de una proteína y cambiar "manualmente" la estructura de dicho modelo.

En la Figura 14a y 14c se muestra el nudo formado entre los aminoácidos lisina, glutamina, glicina, lisina y ácido glutámico (111-115) de color rojo y los del carboxilo terminal (138-144) de color magenta. En la figura 14b y 14d se muestra la estructura de NaTrxh_2 sin el nudo.

Para confirmar que la conformación general del núcleo modelado por MODELLER no se hubiera modificado por el ajuste manual del modelo, se realizó una superposición de ambas estructuras. En la Figura 15 se muestra el resultado y se observa que no hubo cambios en el plegamiento del núcleo de la Trx. El color azul indica que hay mucha similitud entre las dos estructuras y el rojo que hay poca. Sólo en el carboxilo terminal se observan regiones con color rojo. La

superposición del modelo de NaTrxh_2 con la estructura de la Trx de *P. tremula* (molde) (Figura 15 c y d) muestra también la coincidencia de plegamientos en las regiones en las que las secuencias coinciden.

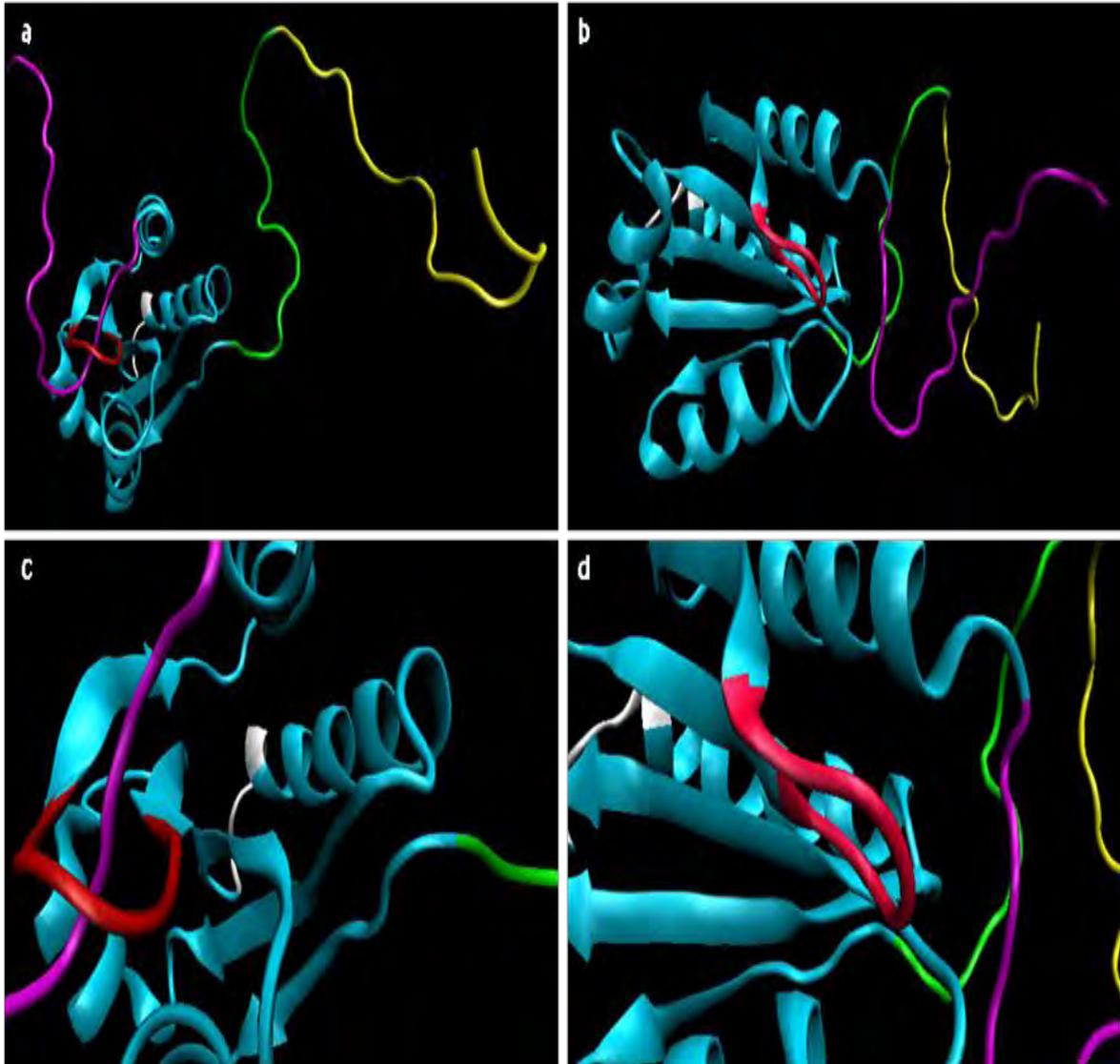


Figura 14. Presencia de un nudo al final del carboxilo terminal en el modelo de la NaTrxh_1d. (a) Modelo NaTrxh_1d de MODELLER en dónde se presenta un nudo al final del carboxilo terminal, (b) Modelo NaTrxh_2 de MODELLER sin el nudo el cual se eliminó manualmente con "Swiss PDB Viewer", (c) Acercamiento de la formación del nudo entre los aminoácidos 112-115 y los del carboxilo terminal y, (d) Acercamiento de los aminoácidos 112-115 y los del carboxilo terminal sin el nudo. En color azul claro se observa el núcleo de la NaTrxh, en amarillo la región N α , en verde la región N β , en blanco el sitio activo (WCGPC), en magenta la secuencia de aminoácidos del carboxilo terminal (138-152 aminoácidos) y en rojo la secuencia de aminoácidos remanentes en la zona que contenía inicialmente el nudo (112-115) de la secuencia completa de NaTrxh.

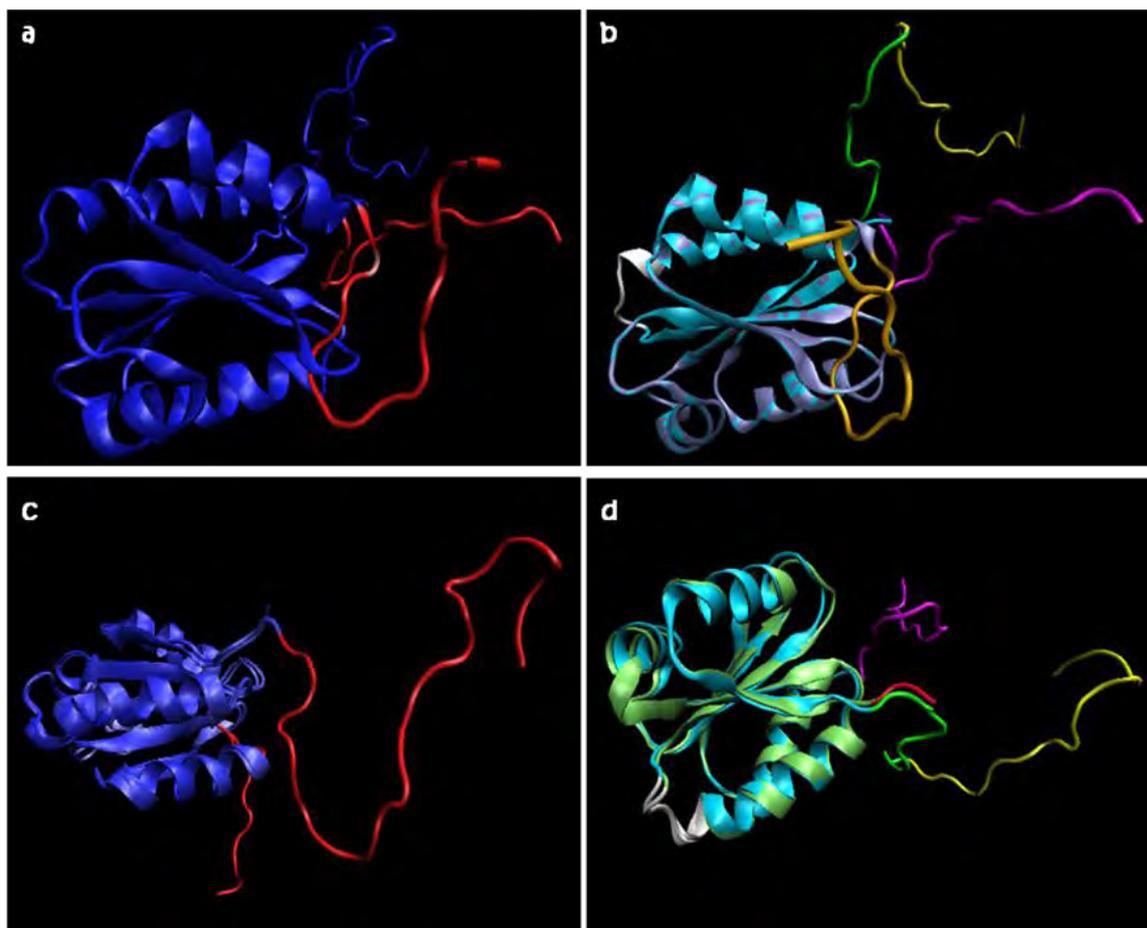


Figura 15. Superposiciones de la NaTrxh_1d con otros modelos. (a-b) Superposición de NaTrxh_1d con NaTrxh_2 donde no hay cambios en el núcleo ni en la región N-terminal y, (c-d) Superposición de NaTrxh_2 con la proteína molde 1ti3. Se observa de color rojo la región N-terminal de NaTrxh_2 que 1ti3 no posee. En a y c se utilizaron los algoritmos Qres y STAMP. En b el núcleo de NaTrxh_2 es de color azul claro y el carboxilo terminal de color magenta; el núcleo de NaTrxh_1d es de color lila y la C-terminal de color naranja. En ambos casos la región N α en amarillo, N β en verde y el sitio activo en blanco. En c el núcleo de NaTrxh_2 es igual al de b, en cambio el núcleo de 1ti3 es de color verde claro y sus primeros cinco aminoácidos de su región N-terminal son de color rojo.

VI.1.3. Recocimiento simulado del modelo tridimensional de la NaTrxh desanudado

Se sometió el modelo de NaTrxh_2a a un recocimiento simulado (SA) (Nilges *et al.*, 1988) mediante el programa HyperChem™ (HyperCube, Inc). Dicho método consiste en realizar una simulación de dinámica molecular elevando notoria y rápidamente la temperatura (de 198 a 500 K en 5 ns en este caso), mantener el

sistema en esta condición por un periodo corto (10 ns) y permitir entonces un enfriamiento gradual del sistema hasta una temperatura que produzca un estado de baja energía (de 500 a 100 K en 75 ns). Debido a que la región más incierta eran los extremos amino y carboxilo, la simulación se realizó fijando la posición de los átomos del esqueleto en la región central del modelo (del modelo 31 al 139). Adicionalmente, para acelerar el colapso de estas regiones la simulación se realizó en el vacío. La figura 16 muestra el modelo obtenido (NaTrxh_3) al final de dicha simulación.

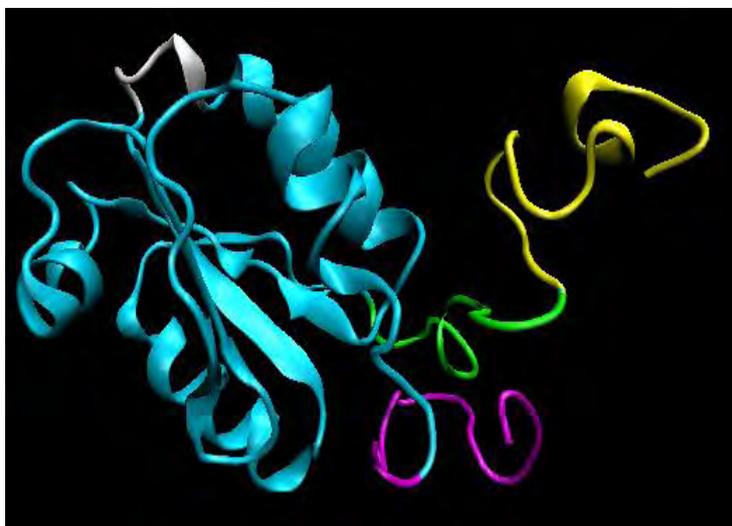


Figura 16. NaTrxh_3 luego de un recocido simulado en condiciones de vacío, con los átomos del esqueleto fijos para los aminoácidos (31-139). En color azul claro se observa el núcleo de la NaTrxh, en amarillo la región N α , en verde la región N β , en blanco el sitio activo (WCGPC), en magenta la secuencia de aminoácidos del carboxilo terminal (138-152 aminoácidos) y en rojo la secuencia de aminoácidos que formaron el nudo (112-115) de la secuencia completa de NaTrxh. La simulación consistió en calentamiento de 298 a 500 K en 5 ns, seguida de 10 ns a 500 K y un enfriamiento gradual de 500 a 100 K en 75 ns, con saltos de 50 K. El incremento de integración fue de 0.002 ps y el campo de fuerza empleado fue Amber95.

En la Figura 17 se muestra la superposición de la NaTrxh_3 con la NaTrxh_2. Se observa que hay un cambio estructural en las regiones del extremo amino como en las del carboxilo. Aún no hay una estructura definida.

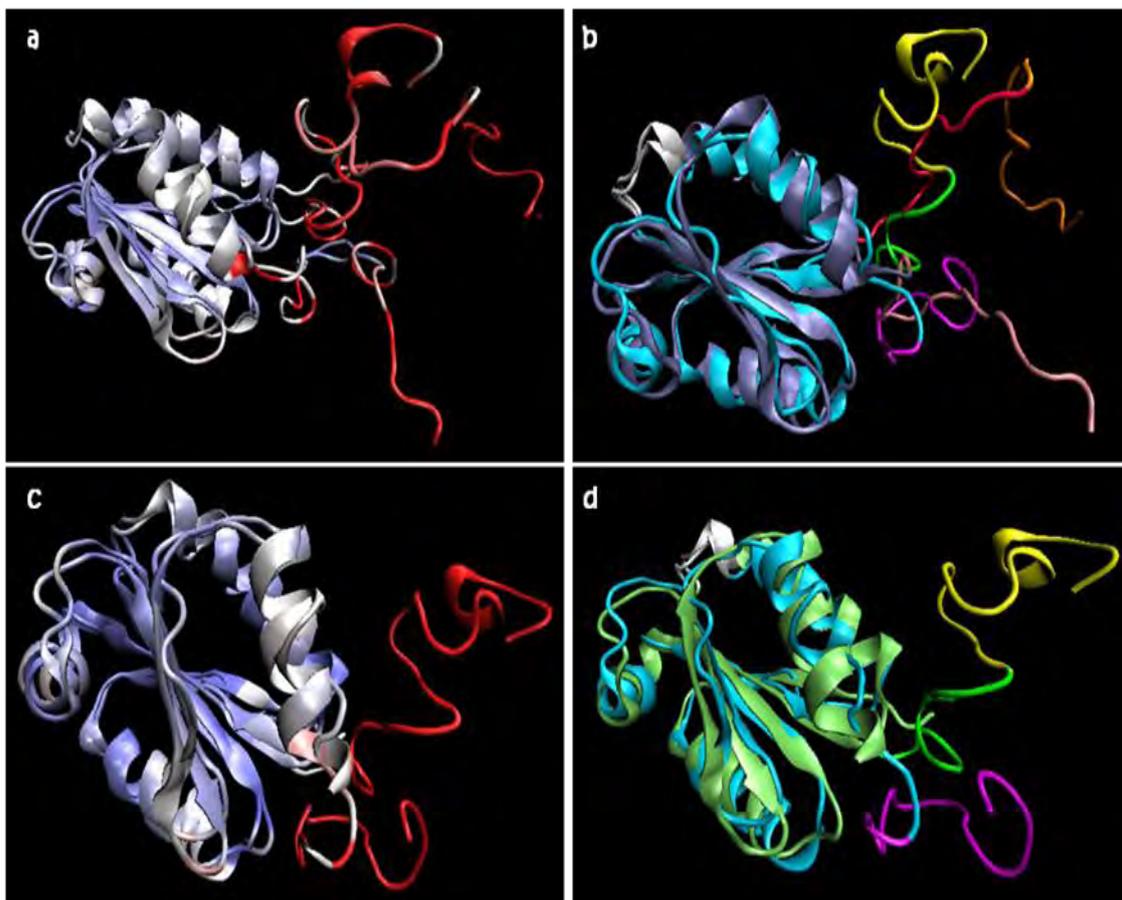


Figura 17. Superposición de NaTrxh_3 con NaTrxh_2 donde se observa que hay un cambio en la estructura del amino terminal de la NaTrxh_3 (a-b), c-d NaTrxh_3 superpuesta con 1ti3 de *P. tremula*. a y c muestran diferencias en el núcleo de la proteína pero aún con similitud entre los aminoácidos, lo contrario en el N-terminal que es completamente rojo en donde no hay semejanza entre ellos. De color azul claro se observa el núcleo de la NaTrxh_3, en lila el núcleo de la NaTrxh_2 y en verde claro la de 1ti3. La región $N\alpha$ de NaTrxh_2 es naranja y la de NaTrxh_3 es amarilla. En verde la región $N\beta$ de NaTrxh_3 y en rojo la de NaTrxh_2; en blanco el sitio activo (WCGPC) y, en magenta la secuencia de aminoácidos del carboxilo terminal (138-152 aminoácidos) de NaTrxh_3 y en rosa la de NaTrxh_2.

VI.1.4. Simulación dinámica molecular de NaTrxh_3

Debido a que la simulación en el vacío es poco realista, y considerando además que el colapso de la región amino era incompleto, se sometió el modelo a una simulación de dinámica molecular más larga a 313 K en medio acuoso, con NaCl 0.15 M. Debido a las limitaciones del equipo y del software, dicha simulación no se pudo hacer en un tiempo razonable en HyperChem por lo que

se realizó en GROMACS 3.3, usando un campo de fuerzas GROMAS 53a6 y los modelos de solvente SPC216.

El intervalo de integración fue de 0.002 ps y la simulación se dejó correr 35 ns (17.5 millones de pasos), colectando coordenadas del sistema cada 2.5 ps y datos de energías cada 0.125 ps. En esta simulación, los extremos alcanzaron un estado compacto luego de aproximadamente 27 ns y se mantuvieron razonablemente estables hasta el final de la simulación (Figura 18)

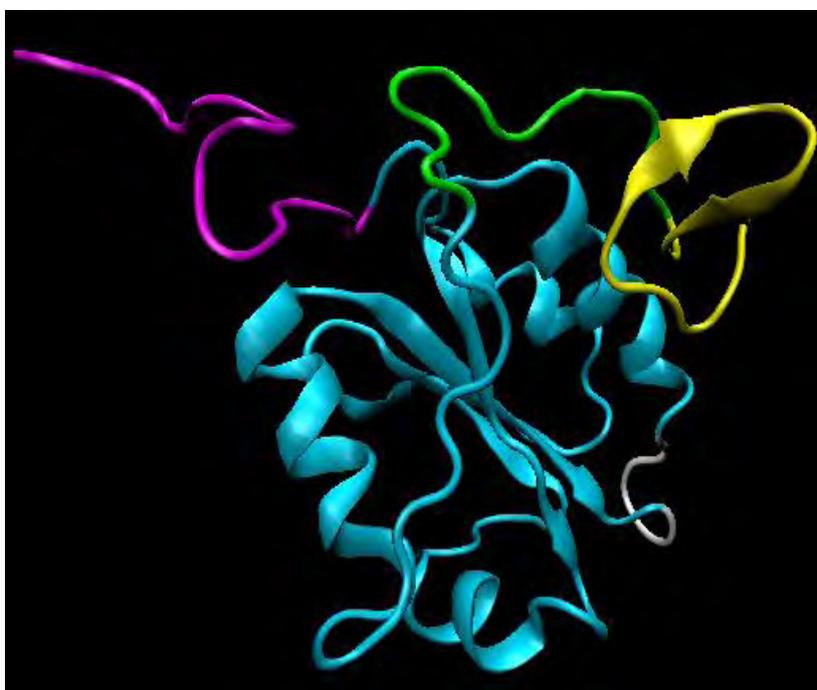


Figura 18. Modelo de NaTrxh_5 mostrando dos hebras betas en la región N α y los aminoácidos expuestos al solvente en la región N β que podría permitir la interacción con alguna proteína. De color azul claro se observa el núcleo de la NaTrxh, en amarillo la región N α , en verde la región N β , en blanco el sitio activo (WCGPC), en magenta la secuencia de aminoácidos del carboxilo terminal (138-152 aminoácidos).

Sin embargo, al hacer la superposición entre la NaTrxh_5 y la Trx de *P. tremula*, se observa que en el núcleo clásico de las Trx no se conserva (Figura 19) (Holmgren, 1985; Powis y Montfort, 2001). No obstante, NaTrxh_5, es el modelo más estable que se tenía hasta ese momento.

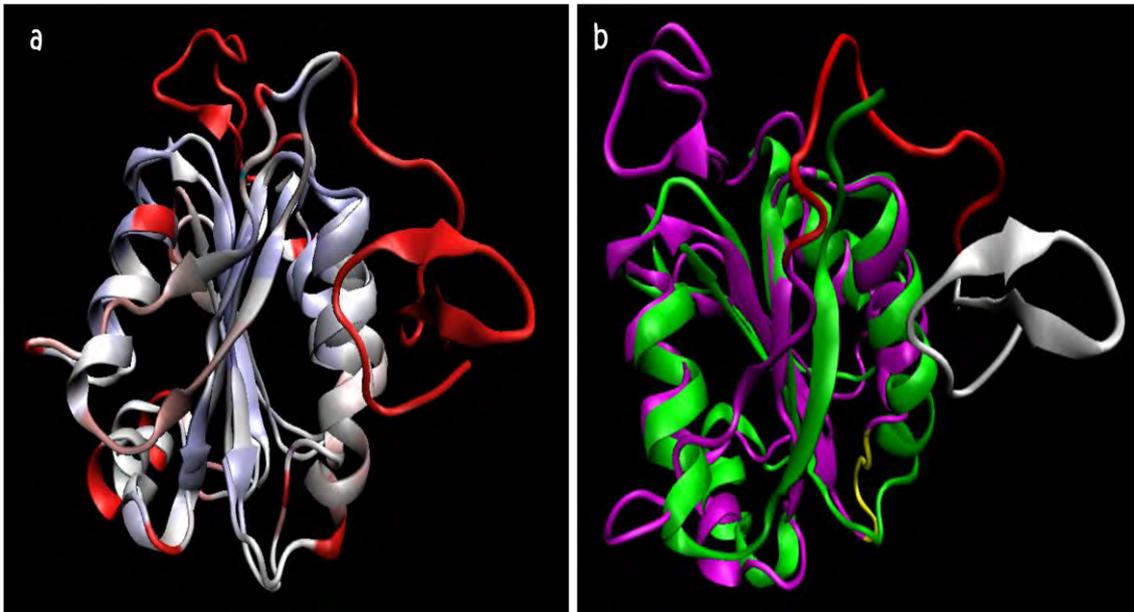


Figura 19. Superposición de NaTrxh_5 con 1ti3. Se observa, que casi no se conservan los elementos estructurales secundarios principales del núcleo de la proteína (color blanco-azul). Por supuesto, la estructura del N-terminal es muy diferente con respecto a la Trx de álamo (color rojo) que no presenta una extensión tan amplia de aminoácidos hacia ese extremo. **(b)** se muestra de color magenta a NaTrxh_5 con la región N α de color blanco y N β en rojo. La proteína de álamo se muestra en color verde. De color magenta se observa el núcleo de la NaTrxh, en blanco la región N α , en rojo la región N β , y en amarillo el sitio activo (WCGPC).

VI.2. MODELAJE DE LA SECUENCIA DE LA NaTrxh CON SAM-T06

A lo largo de todo este proceso de modelado, se observó que los extremos amino y carboxilo de la proteína permanecían bastante expuestos, lo que permite proponer que dichas regiones de la proteína son relativamente hidrofílicas y no interaccionan con regiones internas, es decir, aun si el modelo es un tanto defectuoso, dichas regiones están posiblemente expuestas.

El problema del plegamiento de una proteína a partir de la estructura primaria no es un problema resuelto. De hecho, aun cuando el modelaje por homología es el que presenta mayor nivel de éxito, existe un grado importante de incertidumbre en los modelos generados. Por esta razón, se generaron otros modelos empleando

servidores de predicción de estructura disponibles en internet. Entre los servidores con mayor éxito se encuentran I-Tasser, SAM-T06 y Robetta. De ellos, sólo SAM-T06 (actualmente la versión 08 es la activa) puede generar un modelo en pocas horas. Dichos modelos suelen ser certeros en plegamiento general, pero tiene defectos en la conectividad de algunos enlaces peptídicos, por lo que es necesario minimizar sus geometrías con métodos de mecánica molecular. En este caso, se envió la secuencia de la NaTrxh para generar un modelo con SAM-T06 (Figura 20).

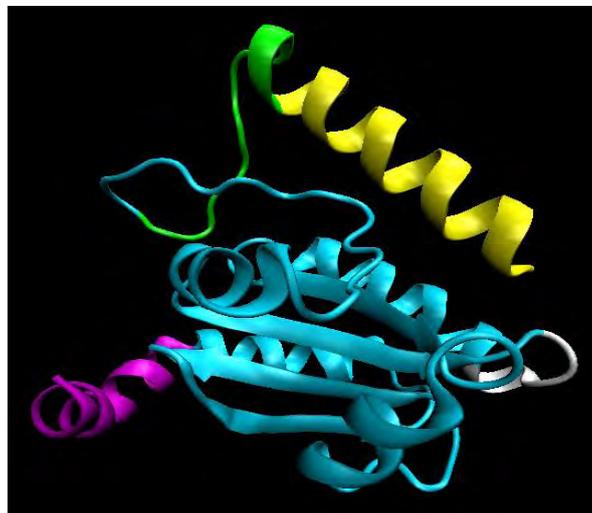


Figura 20. Modelo de la secuencia de NaTrxh_6 propuesta por el servidor SAM-T06. De color azul claro se observa el núcleo de la NaTrxh, en amarillo la región $N\alpha$, en verde la región $N\beta$, en blanco el sitio activo (WCGPC), en magenta la secuencia de aminoácidos del carboxilo terminal (138-152 aminoácidos).

Finalmente, todo modelo debe ser estable, por lo que es conveniente someterlos a simulaciones de dinámica molecular para analizar su estabilidad y permitir la relajación de la estructura a un mínimo de energía. Sin embargo, al momento de minimizar las geometrías de NaTrxh_6, la estructura creada para la región amino y carboxilo se deshizo. No obstante, el modelo muestra una región amino expuesta.

VI.3. HvTrxh2 COMO MOLDE PARA GENERAR UN NUEVO MODELO CON MODELLER

Debido a que los proyectos genómicos producen mucho más secuencias en poco tiempo que aquel en que se puede resolver su estructura tridimensional, ya sea por cristalografía de rayos X o por NMR, se ha optado por resolver su estructura por homología comparativa, que es relativamente rápido, dependiendo del método de selección para obtener el modelo tridimensional de la proteína que se requiera. Por supuesto, la calidad de los modelos de las proteínas difiere, por lo que es necesario someter el modelo a una evaluación (Rodríguez *et al.*, 1998).

En este caso se sometió el modelo NaTrxh_5 a una evaluación con el programa Rosetta-HMMer (Rodríguez-Sotres y Martínez-Castilla; sin publicar); sin embargo, el programa mostró que la proteína NaTrxh no se parecía a ninguna Trx, por lo que se decidió someter la secuencia de NaTrxh en búsqueda de una proteína que estuviera relacionada con nuestra Trx y además estuviese resuelta su estructura. Como resultado, se obtuvo a HvTrxh2, que es una Trx de *Hordeum vulgare* (cebada) y además tiene una extensión hacia el amino terminal (Maeda *et al.*, 2003a; Maeda *et al.*, 2006) y muestra un 51% de similitud con la NaTrxh al hacer un alineamiento múltiple con Clustal X (Higgins *et al.*, 1988; Higgins *et al.*, 1989; Thompson *et al.*, 1997) y un 50% de similitud con PtTrxh (Figura 21). En la Figura 21 se marcan con asteriscos las posiciones que son idénticas; las posiciones conservadas están indicadas con dos puntos y las poco conservadas se indican con un solo punto (Higgins *et al.*, 1988; Higgins *et al.*, 1989; Thompson *et al.*, 1997). Cabe destacar, que el número de aminoácidos en el extremo amino y en el carboxilo está mayoritariamente abarcado por HvTrxh2 en lugar de PtTrxh; **asimismo, el número de 'espacios' se ve limitado a las regiones**

extremas de la secuencia y no en el núcleo de la proteína. Por ende, HvTrxh2 es un mejor molde para la NaTrxh que PtTrxh.

```

      ....|....| ....|....| ....|....| ....|....| ....|....|
          5         15         25         35         45         55
1|NaTrxh  MGSYLSSLLG  GGAAEAAEAE  SGSSSEPSRV  IAFHSSNRWQ  LHFNSSKQLN  KLIVVDFEAT
2|HvTrxh2  -GSHMA---A  SATAAAVAAE  -----V  ISVHSLEQWT  MQIEEANTAK  KLVVIDFTAS
3|PtTrxh   -----  -----AEEGQ  -----V  IACHTVDTWK  EHFEEKGKSQ  KLIVVDFEAT
Clustal Co          *  .:          * *: *: : *  :::::  : **:*:**:*:

      ....|....| ....|....| ....|....| ....|....| ....|....|
          65         75         85         95         105        115
1|NaTrxh  WCGPCKMMEP  VINAMSAKYT  DVDFVKIDVD  ELSDVAQEFQ  VQAMPFTLLL  KQGKEVERV
2|HvTrxh2  WCGPCRIMAP  VFADLAKKFP  NAVFLKVDVD  ELKPIAEQFS  VEAMPFTLFM  KEGDVKDRV
3|PtTrxh   WCPPCKMIAP  IFAELAKKFP  NVTFLKVDVD  ELKAVAEAWN  VEAMPFTFIFL  KDGKLVDKT
Clustal Co  ** *.:.: *  ::  :: *.:  .:  *:*:* ** .:*.:. *:*:*:*: *:*  :.:*

      ....|....| ....|....| ....|....| ..
          125        135        145
1|NaTrxh  GAKKDELEKK  ILKHREAPKF  AASNYRTKFK  VQ
2|HvTrxh2  GAIKEELTAK  VGLHAAAQ--  -----  --
3|PtTrxh   GADKDGLEPTL  VAKHATA---  -----  --
Clustal Co  ** *: *  :  *  *

```

Figura 21. Alineamiento de NaTrxh, HvTrxh2 y PtTrxh con el programa ClustalX. Se observa que hay más aminoácidos hacia el extremo amino y carboxilo de HvTrxh2 con respecto a NaTrxh que PtTrxh. Los asteriscos ('*') indican las posiciones que son idénticas, los dos puntos (':'), regiones conservadas y un punto ('.'), regiones poco conservadas.

Sin embargo, el PDB de HvTrxh2 (2wit) muestra a la proteína formando un complejo con una proteína sustrato denominada α -amilasa/inhibidor de subtilisina (BASI) (Figura 22a) (Maeda *et al.*, 2006), por lo que fue necesario hacer un archivo PDB que sólo contuviera la información de la HvTrxh2 utilizando el programa VMD (Figura 22b). Este monómero mantiene la estructura altamente conservada en varias especies, conformada de 5 hebras beta y 4 hélices alfa (Holmgren, 1985).

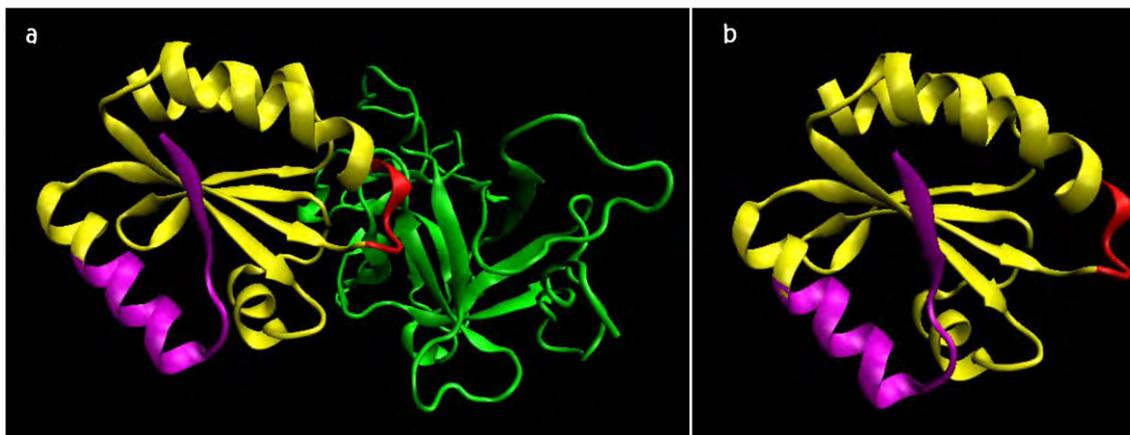


Figura 22. Archivo PDB de HvTrxh2 (2wit). **(A)** Complejo formado entre HvTrxh2 y la proteína α -amilasa/inhibidor de subtilina (BASI). **(B)** Proteína HvTrxh2, donde se muestra el extremo N-terminal que abarca 16 de los 28 aminoácidos de NaTrxh. En color magenta se muestra el N-terminal; en amarillo el núcleo de HvTrxh2; en rojo el sitio activo (WCGPC); y en verde la proteína BASI.

Un dato importante es que de las 21 secuencias arrojadas por el comando **profile.build()** (Tabla 1) de MODELLER (Figura 9), no se encontró a la secuencia de 2wit como una secuencia de aminoácidos que se asemejaba a la NaTrxh.

Se procedió a modelar a la NaTrxh tomando como molde a la HvTrxh2 en el programa MODELLER. Antes de modelar con MODELLER, se le especificó al programa Rosetta-HMMer que hiciera un archivo en el cual tomara el núcleo de HvTrxh2 como molde y utilizara como molde el extremo amino y carboxilo de la NaTrxh_5. Se hicieron 15 modelos y se seleccionó el más apropiado denominado NaTrxh_7 (Figura 23).

El modelo NaTrxh_7 se minimizó con AMBER 95 en Hyperchem, empleando un esquema progresivo que minimiza las cadenas laterales con el esqueleto fijo por unos pocos pasos, seguido de otros pocos pasos con las cadenas laterales fijas y el esqueleto libre. El proceso se repitió hasta reducir el gradiente a $0.25 \text{ Kcal/mol } \text{\AA}^2$. Finalmente, la estructura se liberó por completo y la minimización se prolongó hasta alcanzar un gradiente de $0.1 \text{ Kcal/mol } \text{\AA}^2$ (NaTrxh_8)

(Figura 24), obteniéndose dos hebras beta en la extensión amino y una en el carboxilo terminal.

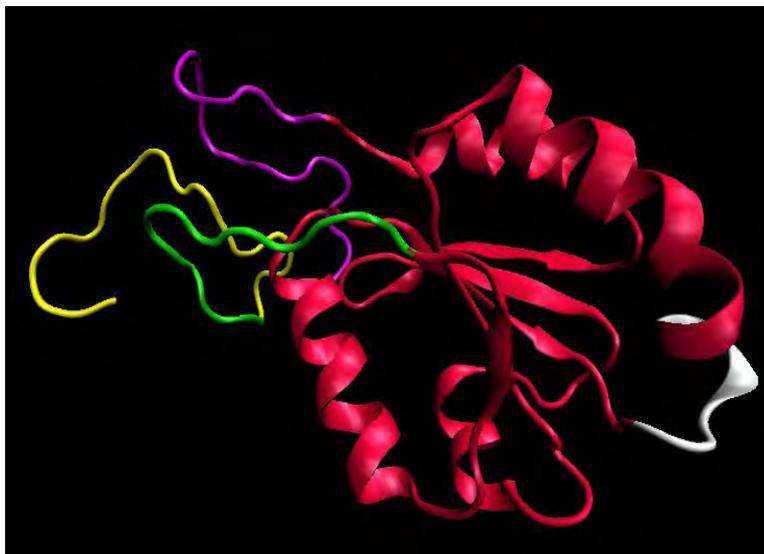


Figura 23. Modelo NaTrxh_7 obtenido con el MODELLER utilizando como molde para el núcleo a HvTrxh2 y como molde al extremo amino y carboxilo de NaTrxh_5. En color rosa oscuro se observa el núcleo de la NaTrxh, en amarillo la región N α , en verde la región N β , en blanco el sitio activo (WCGPC) y en magenta la secuencia de aminoácidos del carboxilo terminal (138-152 aminoácidos).



Figura 24. Modelo NaTrxh_8 después de la minimización con AMBER 95 en HyperChem, el cual se repitió hasta reducir el gradiente a 0.25 Kcal/mol \AA^2 , hasta que se liberó la estructura por completo y la minimización se continuó hasta alcanzar un gradiente de 0.1 Kcal/mol \AA^2 . Además muestra 3 hebras beta, 2 en el amino y una en el C-terminal. De color rosa oscuro se observa el núcleo de la NaTrxh, en amarillo la región N α , en verde la región N β , en blanco el sitio activo (WCGPC) y en magenta la secuencia de aminoácidos del carboxilo terminal (138-152 aminoácidos).

NaTrxh_8 se sometió a simulación de dinámica molecular por 35 ns para obtener el modelo final. Este modelo se extrajo de la dinámica eligiendo la estructura más representativa mediante el algoritmo g_cluster implementado en GROMACS. Se calificó a NaTrxh_8 con el programa Rosetta-HMMer, y dio la secuencia de la NaTrxh con un valor de E de 3.4×10^{-3} (Figura 25) lo que sugiere fuertemente que el modelo NaTrxh_8 realmente ocurre en la estructura natural de la NaTrxh.

```

hmmsearch - search a sequence database with a profile HMM
HMMER 2.3.2 (Oct 2003)
Copyright (C) 1992-2003 HHMI/Washington University School of Medicine
Freely distributed under the GNU General Public License (GPL)
-----
HMM file: thxh_2_01_min_a-0.hmm [thxh_2_01_min_a-0]
Sequence database: /usr/home/rogelio/xtra/models/nathx/thxh_2/thx2/thx2_a-
O.xids
per-sequence score cutoff: [none]
per-domain score cutoff: [none]
per-sequence Eval cutoff: <= 10
per-domain Eval cutoff: [none]
-----

Query HMM: thxh_2_01_min_a-0
Accession: [none]
Description: [none]
[HMM has been calibrated; E-values are empirical estimates]

Scores for complete sequences (score includes all domains):
Sequence Description Score E-value N
-----
gi|66096352|gb|AAY42864.1| thioredoxin H [N 4.3 0.0034 1
gi|147821566|emb|CAN70031.1| hypothetical pro -20.2 1.1 1
gi|33621080|gb|AAQ23133.1| thioredoxin H2 [ -21.7 1.5 1
gi|125536296|gb|EAY82784.1| hypothetical pro -25.7 3.9 1
gi|125579022|gb|EAZ20168.1| hypothetical pro -25.7 3.9 1
gi|115488174|ref|NP_001066574.1| Os12g0281300 [Or -25.7 3.9 1
gi|125569177|gb|EAZ10692.1| hypothetical pro -25.7 3.9 1
gi|168013664|ref|XP_001759411.1| predicted protei -25.7 4 1
gi|78042550|ref|NP_001030273.1| glutaredoxin 3 [ -26.1 4.3 1
gi|7767672|gb|AAF69169.1|AC007915_21 F27F5.21 [Arabid -26.3 4.5 1
gi|53748519|emb|CAH59452.1| thioredoxin 3 [P -28.8 8.1 1
gi|30684711|ref|NP_188415.2| ATTDX (TETRATICO -28.8 8.1 1
gi|21536657|gb|AAM60989.1| tetratricoredoxi -28.8 8.1 1
gi|145332619|ref|NP_001078175.1| ATTDX (TETRATICO -28.8 8.1 1
gi|167525094|ref|XP_001746882.1| predicted protei -29.1 8.7 1

```

Figura 25. Calificación de NaTrxh con el programa Rosetta-HMMer. La calificación dio a la NaTrxh (gi|66096352|gb|AAY42864.1|) como el mayor modelo para la NaTrxh.

La estructura final (NaTrxh_9) (Figura 26) se minimizó en GROMACS para recuperar una estructura de equilibrio, ya que los intermediarios de la dinámica molecular son estados transitorios, no completamente relajados.

En la Figura 26 (a y b) se observa que el modelo NaTrxh_9 **mantiene expuesta la región N β** , ampliando la posibilidad de que sea un sitio de interacción con otra proteína y además tiene cambios importantes en la organización del núcleo de la proteína en comparación con los modelos hechos anteriormente, así como una reestructuración de los extremos amino y carboxilo. De igual forma, se vuelven a detectar las tres hebras beta antiparalelas que se encontraban en el modelo NaTrxh_8; una se encuentra en la región N α (N β 1), **la segunda en la región N β (N β 2)** y la tercera en el C-terminal (**C β 3**). En la Figura 26c se muestran los puentes de hidrógeno (color verde limón) y se puede apreciar que las tres hebras beta están interaccionando entre ellas (Figura 26c y d), estableciendo que la estabilidad estructural de dichas regiones depende una de la otra.

Por otro lado, no se presentaron cambios en el núcleo de la Trx (Figura 26e y f), manteniéndose la lámina beta característica de las Trx con sus tres hebras betas paralelas (color amarillo) y sus dos hebras antiparalelas (color rojo); así como las cuatro hélices alfa rodeando a la lámina beta (color morado).

Finalmente, después de la dinámica molecular que se le aplicó a la NaTrxh_8 sólo hubo un reacomodo en el esqueleto de la molécula aparentemente poco significativo (Figura 27a y b), y en realidad no se puede discernir entre los modelos NaTrxh_8 (color magenta) y el NaTrxh_9 (color blanco), ya que las dos formas de las moléculas pueden estar presentes naturalmente.

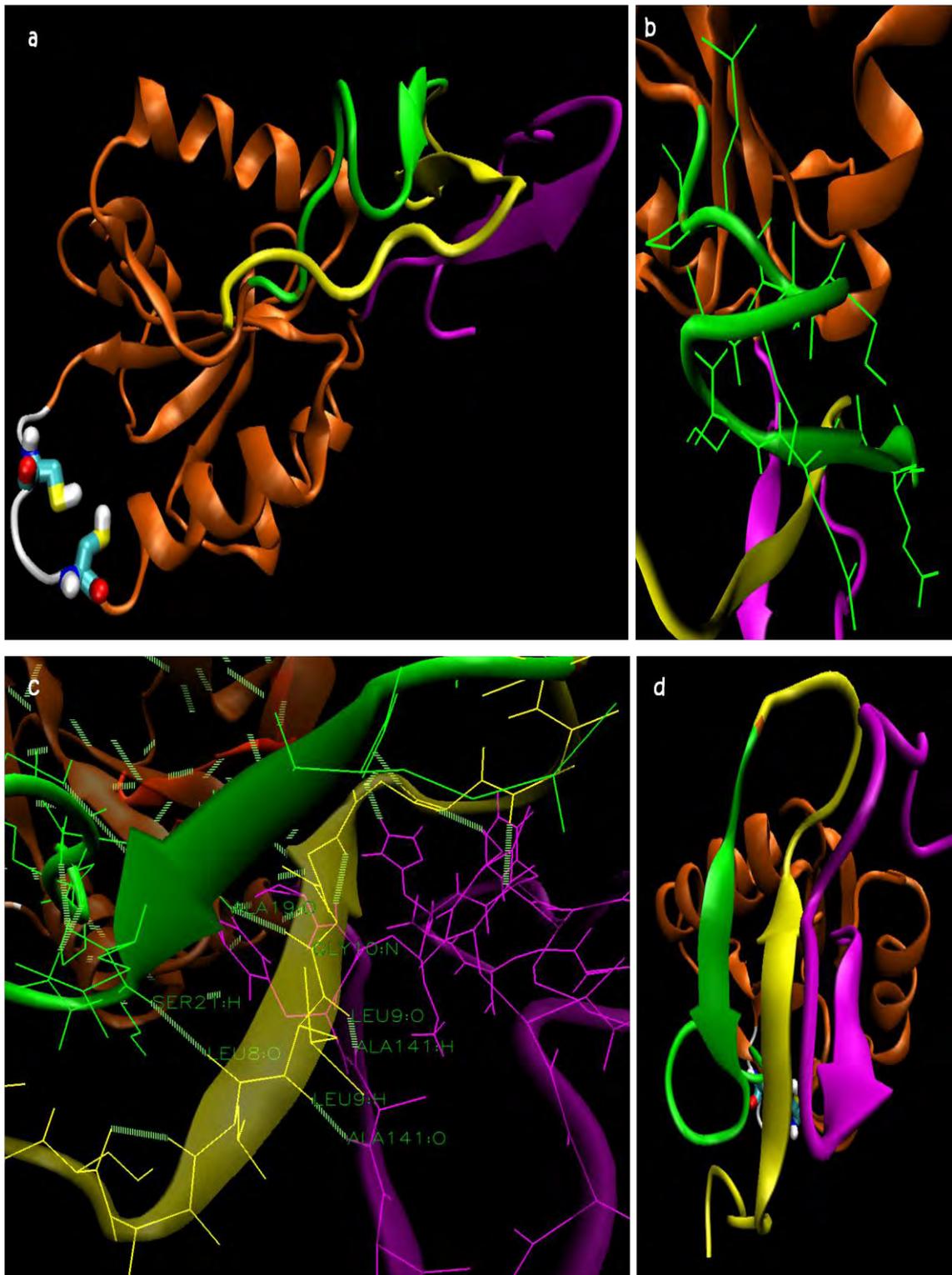


Figura 26. Modelo final y estable de la NaTrxh (NaTrxh_9). **(a)** NaTrxh_9. **(b)** Región N β expuesta al solvente. **(c)** Representación de las 3 hebras beta que dependen una de la otra para su estabilidad estructural, unidas por puentes de hidrógeno (color verde limón). **(d)** Otra representación de las tres hebras betas antiparalelas. **(e)** Núcleo de la NaTrxh. En rojo las hebras antiparalelas y en amarillo las tres betas paralelas. Las hélices alfa (morado) rodean a las hebras beta. **(a-c)** En color naranja se observa el núcleo de la NaTrxh, en amarillo la región N α , en verde la región N β , en color blanco el sitio activo (WCGPC) y en magenta la secuencia de aminoácidos del carboxilo terminal (138-152 aminoácidos).

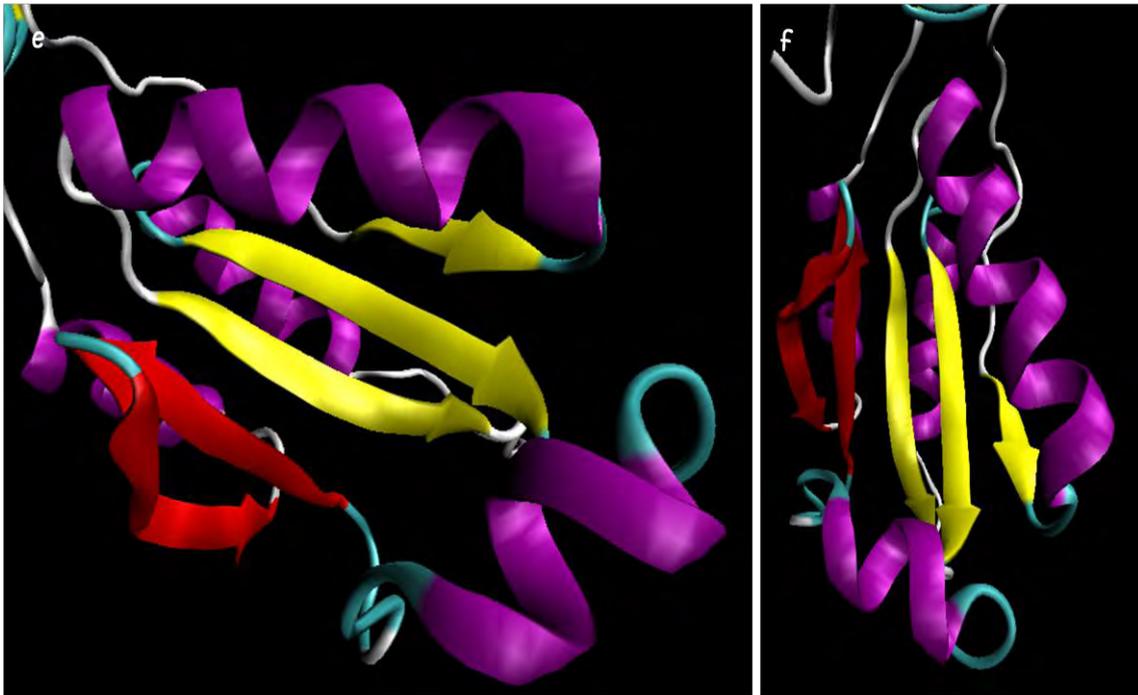


Figura 25. Continuación.

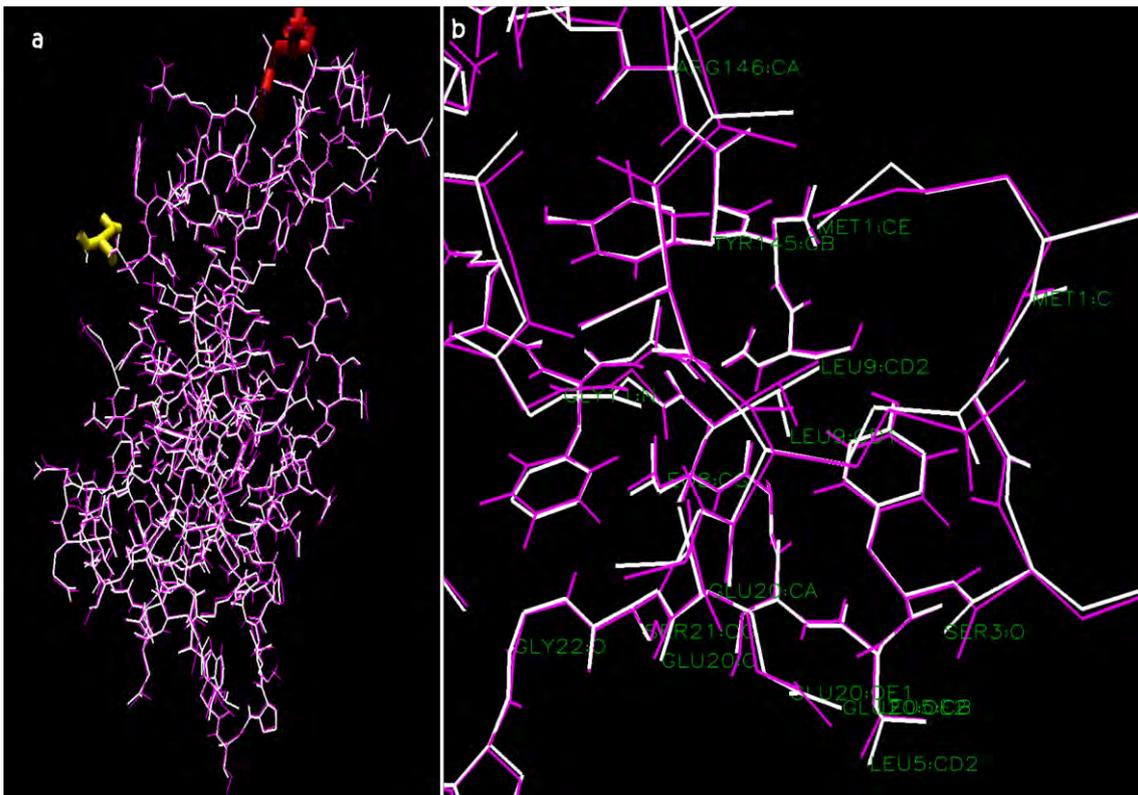


Figura 27. Superposición entre los modelos NaTrxh_8 y NaTrxh_9. **(a)** No se observa un cambio significativo en el esqueleto de los dos modelos. En rojo el aminoácido Met:1 y en amarillo el aminoácido Gln:152. **(b)** Acercamiento de la región amino terminal donde hay un cambio de conformación mínima en el esqueleto de la molécula. **(a-b)** En color magenta se muestra a la NaTrxh_8 y en blanco a la NaTrxh_9.

VII. Discusión

VII.1. NaTrxh es una proteína importante en la autoincompatibilidad gametofítica

Muchas angiospermas emplean el sistema de autoincompatibilidad (AI) para prevenir la autopolinización (de Nettancourt, 1977; Haring *et al.*, 1990). Estudios genéticos establecen que el auto/no-auto reconocimiento en las especies es controlado por un locus multialélico (locus S) y la inhibición del polen ocurre cuando la misma especificidad del haplotipo-S es expresado tanto por el polen como por el pistilo (Franklin-Tong y Franklin, 2003; Takayama e Isogai, 2005). El locus S incluye al menos dos unidades transcripcionales organizados. Uno funcionando como determinante femenina y el otro como determinante masculino (Takayama e Isoagi, 2005). En el caso específico de la AI gametofítica la determinante femenina es la S-RNasa y la masculina es SLF. Por supuesto, en el sistema hay otros factores estilares no ligados al locus S, conocidos como genes modificadores, que son necesarios para modular la actividad biológica de la S-RNasa en el rechazo del polen (Anderson y de Winton, 1931; Mather, 1943; Martin, 1968; Murfett *et al.*, 1996; Tsukamoto *et al.*, 1999; McClure *et al.*, 2000).

En la búsqueda de genes modificadores, en nuestro laboratorio se aisló un cDNA que condifica la proteína denominada NaTrxh, una Trx extracelular, cuyos niveles de RNAm y proteína son más altos en especies incompatibles de *Nicotiana* (Cruz-García, sin publicar). Además, la NaTrxh reduce a la S-RNasa bajo condiciones *in vitro* (Juárez-Díaz *et al.*, 2006a). Lo anterior ubica a la NaTrxh como un gen modificador potencial.

La proteína NaTrxh tiene una extensión hacia el N-terminal que no es peculiar en un gran número de Trx vegetales y animales. Un análisis filogenético utilizando el método de *neighbor-joining* (Kumar

et al., 2001) indica que es una Trx tipo *h* del subgrupo II. Estudios posteriores demostraron que la secreción de la NaTrxh promovida hacia el citoplasma, depende de un N-terminal (Figura 5) (Juárez-Díaz *et al.*, 2006a y b). En esta región un motivo denominado **N β** parece ser suficiente para promover la secreción de NaTrxh (Figura 6) (Juárez-Díaz, 2006b). Basándose en estos datos, una parte importante de esta investigación fue tratar de obtener un modelo tridimensional de NaTrxh en donde se pudiera observar si la secuencia de aminoácidos del N-terminal estaba o no expuesta al solvente, para dilucidar si la NaTrxh es capaz de unirse a alguna proteína para su posterior secreción hacia el apoplasto siguiendo algún mecanismo de los propuestos por Juárez-Díaz (2006b) (Figura 6).

VII.2. Primera estructura tridimensional de la NaTrxh modelada por homología

Actualmente hay muchas secuencias protéicas, incluyendo la de la NaTrxh, pero sólo se han resuelto experimentalmente la estructura terciaria de unas cuantas (Petsko, 2006). Por ende, las funciones de estas proteínas deben ser descritas, entendidas y manipuladas a partir de su estructura tridimensional (Sali y Blundell, 1993; Marti-Renom *et al.*, 2000; Bino y Sali, 2003). En el caso particular de la NaTrxh, se conoce un poco de su función en las células del tejido de transmisión y con la estructura tridimensional se podría develar algo más acerca de su mecanismo molecular, o incluso sugerir si se puede unir a cofactores y/o sustratos para su posterior secreción (Zhang y Kim, 2003).

Debido a la facilidad que presenta hacer un modelo mediante predicciones empíricas utilizando un método comparativo que depende de la relación de la secuencia con al menos una estructura protéica que se determine experimentalmente (Sali y Blundell, 1993),

se decidió proceder por este camino para predecir la estructura tridimensional de la NaTrxh.

El programa que se escogió inicialmente fue MODELLER (Bino y Sali, 1993). Básicamente es un programa de modelaje por homología, que en este caso, sólo fue necesario buscar el molde, de acuerdo a ciertos parámetros y restricciones que tiene el programa, para modelar a la NaTrxh (Fiser y Sali, 2003; Eswar *et al.*, 2003).

Por medio de comandos incluidos en el MODELLER se buscaron secuencias homólogas que tuvieran su estructura tridimensional resuelta experimentalmente. De la búsqueda que se hizo se obtuvieron 21 secuencias (Figura 14), todas ellas Trx. Las secuencias eran de humano, una quimera recombinante de humano con *E. coli*, hongos, bacterias, parásitos y plantas. Por supuesto, se encontraron Trx tipo *h* específicamente del subgrupo I como las de *A. thaliana*, *Spinaceae loleraceae* y *P. tremula*. Lo ideal hubiera sido escoger inmediatamente una de las Trx del tipo *h* ya que la NaTrxh es una Trx tipo *h* (Juárez-Díaz *et al.*, 2006a); sin embargo, se debía tomar en cuenta la similitud de las secuencias de las Trx con la NaTrxh.

El programa comparó la identidad de secuencia de las Trx que habían tenido un valor de E igual a cero con la secuencia de NaTrxh, obteniéndose con un 50% de similitud en su secuencia a 1ti3.

1ti3 es una Trx *h* del subgrupo I de *P. tremula* (PtTrx*h*1) (Coudeville *et al.*, 2005), a diferencia de la NaTrxh que es una proteína del subgrupo II (Juárez-Díaz *et al.*, 2006a). La PtTrx*h*1 se encuentra en los tubos cribosos del floema, se sintetiza en las células acompañantes y se transfiere a través de los plasmodesmos de los tubos cribosos del floema sin núcleo (Ishiwatari *et al.*, 1998). Su región amino terminal es necesaria para la transferencia de célula a célula. Sin embargo, su sitio activo difiere en la secuencia del clásico (WCGPC), ya que en lugar de una glicina hay una prolina (WCPPC). Ésta mutación no genera un cambio conformacional importante, pero

si le da mayor rigidez a la estructura (Coudevylle *et al.*, 2005). En cambio, la NaTrxh que se encuentra en el subgrupo II tiene como característica principal la presencia de una extensión hacia el amino terminal (Gelhaye *et al.*, 2004) que en el caso particular, esta **extensión y principalmente la región N β** , es de suma importancia ya que contiene la información suficiente para dirigir la secreción de NaTrxh hacia el citoplasma sin que sea un péptido señal clásico (Juárez-Díaz, 2006b). Además tiene el sitio activo clásico de las Trx (WCGPC) (Juárez-Díaz, 2006b). Las diferencias entre ambas Trx se hicieron notar en el modelo final arrojado por MODELLER, de lo cual se va a discutir más adelante. Sin embargo, el hecho de que la PtTrxh1 tenga una extensión hacia el amino terminal fue determinante para escogerla como proteína molde.

Una vez que se alineó a la NaTrxh con PtTrxh1, se le solicitó a MODELLER la construcción de 5 modelos. Después se evaluaron y se observó que el modelo NaTrxh_1d (Figura 12d) era el más óptimo de los modelos, no obstante, no tenía una estructura en las regiones amino y carboxilo terminal, aunque presentó el núcleo clásico de las Trx (Holmgren, 1985; Powis y Montfort, 2001).

En el modelo NaTrxh_1d se encontró un nudo formado entre los aminoácidos lisina, glutamina, glicina, lisina y ácido glutámico (111-115) y los del carboxilo terminal (138-144). En general hay fallas en el modelaje que incluyen errores dentro de la cadena de empaquetamiento, distorsiones y cambios en los segmentos del núcleo que no siempre se pueden arreglar (Pieper *et al.*, 2002). En este caso hubo la posibilidad de deshacer manualmente el nudo por medio del programa Swiss PDB Viewer (Guex y Peitsch, 1997), empleando un algoritmo asociado al mapa de Ramachandran, que permitió tomar y modificar un punto en la gráfica donde se encontraban los aminoácidos que formaban el nudo al final del carboxilo terminal, deshaciendo el nudo manualmente.

Por supuesto cabía la posibilidad de que el núcleo de la Trx se viese afectado incluyendo la región N-terminal, por lo que fue necesario hacer una superposición entre la NaTrxh con el nudo (NaTrxh_1d) y la NaTrxh sin el nudo (NaTrxh_2). En la figura 15a se muestra claramente que no hay un cambio en la estructura, lo cual implicó que el haber deshecho manualmente el nudo no afectó al núcleo. Incluso la región N-terminal se superpone perfectamente en los dos modelos.

La región de interés en modelar (NH_3^+ -terminal) no tenía una estructura definida en la NaTrxh_2, por lo que fue necesario mejorar esa región sometiendo al modelo NaTrxh_2 a simulaciones de dinámica molecular. Primero se aplicó un recocimiento simulado (Nilges *et al.*, 1988) en condiciones de vacío para obtener a la NaTrxh_3 (Figura 16). Superponiendo a la NaTrxh_3 con la NaTrxh_2 (Figura 17a-b) se observa que hay un reacomodo de los aminoácidos en general, incluyendo las regiones del amino y del carboxilo pero no presentan una estructura secundaria. Además ya no hay una similitud entre el núcleo de la proteína molde (PtTrxh1) y la de la NaTrxh_3 (Figura 17c-d). Esto puede ser debido a que la identidad de secuencia entre NaTrxh y PtTrxh1 desde el inicio sólo fue del 50%, lo que indica que hubo un reacomodo en los aminoácidos después de la simulación de dinámica molecular.

Debido a la simulación poco realista (condiciones de vacío) a la que fue sometida NaTrxh_3, se prosiguió a aplicar una simulación más larga en medio acuoso con NaCl. Para la simulación se utilizó el programa GROMACS 3.3, usando un campo de fuerzas GROMAS 53a6 y los modelos de solvente SPC216, dando como resultado un estado más compacto del modelo en los extremo amino y carboxilo, principalmente (NaTrxh_5) (Figura 18) que comparada con PtTrxh1 se observan estos cambios de la región N-terminal, cambiando

también la semejanza entre los dos núcleos de las Trx (Figura 19c y d).

VII.3. Otras estructuras predichas por servidores basadas en modelaje comparativo

Aunque NaTrxh_5 fue el modelo más estable que se tenía hasta ese momento, se generaron otros modelos utilizando otros servidores. En el presente trabajo sólo se muestran los resultados del servidor SAM-T06 (Figura 20); el cual mostró que la región amino terminal se encontraba expuesta. Sin embargo, al momento de hacerle una simulación de dinámica molecular para analizar su estabilidad, la estructura creada por SAM-T06 para la región amino y carboxilo se deshizo. Desafortunadamente, ocurrió lo mismo para el resto de los modelos predichos con I-tasser y HMMSTer-Rosetta.

VII.4. Un modelo estable para la NaTrxh

Uno de los principales problemas a los cuales se enfrenta el modelaje tridimensional de una proteína es saber qué tan confiable es el modelo obtenido. En la actualidad hay muchos programas que son capaces de predecir la estructura; pero los parámetros que utilizan cada uno difieren entre ellos, por lo que es necesario calificar dichos modelos.

En nuestro caso utilizamos el programa Rosetta-HMMer para calificar el modelo NaTrxh_5 (Figura 18), que era el modelo más estable que se tenía hasta ese momento. Sin embargo, Rosetta-HMMer indicó que la NaTrxh no se parecía a ninguna Trx. A su vez, se sometió de nuevo la secuencia de NaTrxh en búsqueda de una proteína que estuviera relacionada con la proteína de interés y además que su estructura estuviera resuelta experimentalmente. HvTrxh2 fue la Trx que Rosetta-HMMer encontró con la mayor

similitud con la NaTrxh. Cuando se hizo el alineamiento con el programa Clustal X (Higgins *et al.*, 1988; Higgins *et al.*, 1989; Thompson *et al.*, 1997), se obtuvo una similitud del 51%, un porcentaje mayor que aquel que tenía con la PtTrxh. Además, la HvTrxh2 tiene una extensión hacía el amino terminal que abarca 16 de los 28 aminoácidos que conforman el N-terminal de NaTrxh y 11 aminoácidos más que PtTrxh en dicha región. Esto fue un punto muy favorable para considerar volver a modelar a la NaTrxh pero ahora tomando como molde a HvTrxh2.

El modelaje de la NaTrxh con la nueva secuencia consistió en tomar a la HvTrxh2 como molde únicamente para el núcleo, es decir, que sólo considerara las 5 hebras beta y las cuatro alfa hélices (Holmgren, 1985) y los extremos amino y carboxilos del modelo NaTrxh_5 (Figura 18). Se modeló con MODELLER y se obtuvieron 15 modelos, de los cuales al ser evaluados se seleccionó al más adecuado, denominándolo NaTrxh_7 (Figura 23).

NaTrxh_7, se minimizó con AMBER 95 en HyperChem empleando un esquema progresivo. Este esquema se repitió hasta alcanzar un gradiente tal que se obtuviera el modelo NaTrxh_8 (Figura 24), en donde se obtuvo una estructura de tres hebras beta antiparalelas al final de los extremos amino y carboxilo de la secuencia, las cuales, siguieron presentándose al final de una simulación de dinámica molecular de 35 ns (NaTrxh_9)(Figura 26).

El modelo de NaTrxh_9 **muestra la región N β expuesta al** solvente que podría ser un sitio de unión proteína-proteína. Además mantiene la estructura clásica de las Trx, es decir, presenta los elementos estructurales secundarios que incluyen más del 75% de los residuos aminoacídicos de la NaTrxh_9, que son 5 hebras beta, tres paralelas y dos antiparalelas que forman una región hidrofóbica rodeada de cuatro hélices alfa sobre la superficie externa de la NaTrxh_9 (Figura 26e y f).

A su vez, los residuos de aminoácidos de los extremos amino y carboxilo influyeron a la formación de las tres hebras beta. La hebra de la región $N\alpha$ (**$N\beta 1$**) **conformada de los aminoácidos 7 al 13** (Ser-Leu-Leu-Gly-Gly-Gly) es la primera en aparecer durante la minimización con AMBER 95 para la obtención del modelo NaTrxh_8. Esto pudo **deberse a que la Ser y Leu que conforman la $N\beta 1$ tienen predilección por la conformación de hebra beta**, en cambio la Gly preferencialmente la tiene a la formación de un asa, seguida de una hebra beta (Petsko y Ringe, 2004).

La formación de la $N\beta 2$ conformada por los aminoácidos 17 al 21 (Ala-Glu-Ala-Glu-Ser) configura prioritariamente una hebra beta. En cambio los aminoácidos de la **$C\beta 3$** (140-143; Lys-Phe-Ala) modelan preferencialmente hélices alfa (Lys y Ala) (Petsko y Ringe, 2004). Sin embargo, la presencia de los dos puentes de hidrógeno **entre los aminoácidos leucina 9 (Leu9) de la $N\beta 1$ y la alanina 141 (Ala141) de la $C\beta 3$** , pudieron haber influido en la formación de la **hebra beta $C\beta 3$ (Figura 26c)**. Cabe destacar que los dos puentes de hidrógeno que se presentaron entre la Leu9 y la Ala141, se formaron entre el oxígeno de la Leu9 (Leu9:O) y el hidrógeno de la Ala 141 (Ala141:H), y el segundo entre el hidrógeno de la Leu9 (Leu9:H) y el oxígeno de la Ala141 (Ala141:O) (Figura 26c).

Posteriormente se hizo la superposición entre los modelos NaTrxh_8 y NaTrxh_9 (Figura 27). El esqueleto de la NaTrxh en los dos modelos se conserva con cambios poco significativos, lo que indica que ambos modelos pueden presentarse biológicamente y más aún, da lugar a inferir que el modelo final que se presenta en este trabajo es muy confiable. Además, la propuesta de una estructura que no se ha reportado en las extensiones amino y carboxilo de una Trx de plantas.

VIII. Conclusión

De todas las estructuras obtenidas por modelaje comparativo de la proteína NaTrxh, la más estable podría ser NaTrxh_9, sin descartar la posibilidad de que la conformación de la NaTrxh_8 también pudiese estar presente biológicamente. Después de haber sido sometido a diferentes dinámicas moleculares y calificaciones con el programa Rosetta-HMMer, se obtuvo un modelo estable de tres hebras beta: una en la región N α , la segunda en la región N β , que también tiene una región expuesta al soluto, y la tercera en el extremo carboxilo terminal. Las hebras antiparalelas se encuentran interaccionando entre ellas y por ende, estabilizando la estructura. La lámina beta formada entre las extensiones amino y carboxilo terminal de la NaTrxh_9, es una estructura que no ha sido reportada hasta la fecha en una Trx de plantas.

Se concluye entonces que la región amino terminal, **específicamente la región N β , se encuentra expuesta al solvente**, por lo que la NaTrxh puede estar interaccionando en esta zona con alguna proteína que promueva su secreción hacia el apoplasto.

IX. Bibliografía

A

- ▣ Anderson, M. A., E. C. Cornish., I. Goeffrey., R. Bernatzky., A. Altkinson., T. Orpin., H. Dedman., G. Tregear., R. Fernley., y A. E. Clarke. 1989. Sequence variability of three alleles of the self-incompatibility gene of *Nicotiana alata*. *Plant Cell* **1**: 483-491.
- ▣ Anderson, E., y D. de Winton. 1931. The genetic analysis of an unusual relationship between self-sterility and self-fertility in *Nicotiana*. *Annals of the Missouri Botanical Garden* **18**: 97-116.
- ▣ Arnér, E. S. J., y A. Homgren. 2000. Physiological functions of thioredoxin and thioredoxin reductase. *European Journal of Biochemistry* **267**: 6102-6109.

B

- ▣ Balmer, Y., y B. B. Buchanan. 2002. Yet another plant thioredoxin. *Trends in Plant Science* **7**: 191-193.
- ▣ Baumann, U., y J. Juttner. 2002. Plant thioredoxins: the multiplicity conundrum. *Cellular and Molecular Life Sciences* **59**: 1042-1057.
- ▣ Baker, D., y A. Sali. 2001. Protein structure prediction and structural genomics. *Science* **294**: 93-96.
- ▣ Bax, A. 1989. Two-dimensional NMR and protein structure. *Annual Review of Biochemistry* **58**: 223-256.
- ▣ Bendtsen, J. D., H. Nielsen., G. von Heijne., y S. Brunak. 2004a. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**: 783-790.
- ▣ Bendtsen, J. D., L. J. Jensen., N. Blom., G. von Heijne., y S. Brunak. 2004b. Feature-based prediction of non classical and

- leaderless protein secretion. *Protein Engineering, Design and Selection* **17**: 349-356.
- ▣ Bino, J., y A. Sali. 2003. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Research* **31**: 3982-3992.
- ▣ Blundell, T. L., y L. N. Johnson. 1976. The protein data bank: a computer-based archival file for macromolecular structures. *Journal of Molecular Biology* **112**: 535-542.
- ▣ Bréhélin, C., N. Mouaheb., L. Verdoucq., J. M. Lancelin., e Y. Meyer. 2000. Characterization of determinants for the specificity of *Arabidopsis* thioredoxins h in yeast complementation. *Journal of Biological Chemistry* **275**: 31641-31647.
- ▣ Brooks III, C. L., M. Karplus., y B. M. Pettit. 1988. Proteins: A theoretical perspective of dynamics structure and thermodynamics. *John Wiley & Sons. USA*.
- ▣ Bower, M. S., D. D. Matias., E. Fernandes-Carvalho., M. Gu., S. J. Rothstein., y D. R. Goring. 1996. Two members of the thioredoxin-*h* family interact with the kinase domain of a *Brassica* S locus receptor kinase. *Plant Cell* **8**: 1641-1650.

C

- ▣ Cabrillac, D., J. M. Cock., C. Dumas., y T. Gaude. 2001. The S-locus receptor kinase is inhibited by thioredoxins and activated by pollen coat proteins. *Nature* **410**: 220-223.
- ▣ Campbell, I. D. 2002. Timeline: the march of structural biology. *Nature Reviews Molecular Cell Biology* **3**: 377-381.
- ▣ Cleves, A. E., D. N. W. Cooper., S. H. Barondes., y R. B. Nelly. 1996. A new pathway for protein export in *Saccharomyces cerevisiae*. *The Journal of Cell Biology* **133**: 1017-1026.
- ▣ Coudeville, N., A. Thyreau., C. Hemmerlin., E. Gelhaye., J. P. Jacquot., y M. T. Cung. 2005. Solution structure of a natural CPPC

active site variant, the reduced form of Thioredoxin *h1* form Poplar. *Biochemistry* **44**:2001-2008.

- ▣ Coutsiás, E. A., C. Seok., y K. A. Dill. 2004. Using quaternions to calculate RMSD. *Journal of Computational Chemistry* **25**: 1849-1857.

D

- ▣ de Nettancourt, D. 1977. Incompatibility in angiosperms. Monographs on theoretical and applied Genetics 3. *Springer-Verlag*, E. U. A.

E

- ▣ Eargle, E., D. Wright., y Z. Luthey-Schulten. 2006. Multiple Alignment of protein structures and sequences for VMD. *Structural Bioinformatics* **22**:504-506.
- ▣ Eswar, N., M. A. Marti-Renom., B. Webb., M. S. Madhusudhan., D. Eramian., M. Shen., U. Pieper., y A. Sali. Comparative Protein Structure Modeling With MODELLER. 2003. *Current Protocols in Bioinformatics, John Wiley & Sons, Inc.* **15**: 5.6.1-5.6.30.

F

- ▣ Fine, R. M., H. Wang., P. S. Shenkin., D. L. Yarmush., y C. Levinthal. 1986. Predicting antibody hypervariable loop conformations. II: minimization and molecular dynamics studies of MCPC603 from many randomly generated loop conformations. *Proteins* **1**: 342-362.
- ▣ Fiser, A., y A. Sali. 2003. ModLoop: automated modeling of loops in protein structures. *Bioinformatics Applications Note* **19**: 2500-2501.

- ▣ Fomenko, D. E., y V. N. Gladyshev. 2002. CxxS: fold-independent redox motif revealed by genome-wide searches for thiol/disulfide oxidoreductase function. *Protein Science* **11**: 2285–2296.
- ▣ Franklin-Tong, V. E., y F. C. H. Franklin. 2003. Gametophytic self-incompatibility inhibits pollen tube growth using different mechanisms. *TRENDS in Plant Science* **12**: 598 – 605.

G

- ▣ Gelhaye, E., N. Rouhier., P. Laurent., P. E. Sautiere., F. Martin., y J. P. Jacquot. 2002. Isolation and characterization of an extended thioredoxin *h* from poplar. *Plant Physiology* **114**: 165–171.
- ▣ Gelhaye, E., N. Rouhier., y J. P. Jacquot. 2004a. The thioredoxin *h* system of higher plants. *Plant Physiology Biochemistry* **42**: 265-271.
- ▣ Gelhaye, E., N. Rouhier., J. Gérard., Y. Jolivet., J. Gualberto., N. Navrot., P-I. Ohlsson., G. Wingsle., M. Hirasawa., D. B. Knaff., H. Wang., P. Dizengremel., Y. Meyer., y J. P. Jacquot. 2004b. A specific form of thioredoxin *h* occurs in plant mitochondria and regulates the alternative oxidase. *The Proceedings of the National Academy of Sciences* **101**: 14545-14550.
- ▣ Goldraij, A., K. Kondo., C. B. Lee., C. N. Hancock., M. Sivaguru., S. Vázquez-Santana., S. Kim., T. E. Phillips., F. Cruz-García., y B-McClure. 2006. Compartmentalization of S-RNase and HT-B degradation in self-incompatible *Nicotiana*. *Nature* **439**: 805-810.
- ▣ Guex, N., y M. C. Peitsch. 1997. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis* **18**: 2714-2723.

H

- ▣ Haring, V., J. E. Gray., B. A. McClure., M. A. Anderson., y A. E. Clarke. 1990. Self-incompatibility: a self-recognition system in plants. *Science* **250**: 937-941.
- ▣ Higgins, D. G., y P. M. Sharp. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**: 237-271.
- ▣ Higgins, D. G., y P. M. Sharp. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. *CABIOS*. **5**: 151-153.
- ▣ Holmgren, A. 1985. Thioredoxin. *Annals Review of Biochemistry*. **54**: 237-271.
- ▣ Hubbard, T. J. P., y T. L. Blundell. 1987. Comparison of solvent inaccessible cores of homologous proteins: definitions useful for protein modeling. *Protein Engineering, Design and Selection* **1**: 159-171.
- ▣ Humphrey, W., D alke., y Schulten, K. 1996. VMD – Visual Molecular Dynamics. J. Molecular Graphics. *Journal of Molecular Graphics and Modelling* **14**: 33-38.
- ▣ HyperChem(TM). Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA

I

- ▣ Ishiwatari, Y., C. Honda., I. Kawashiima., S. Nakamura., H. Hirano., S. Mori., T. Fujiwara., H. Hayashi., y M. Chino. 1995. Thioredoxin *h* is one of the major proteins in rice phloem sap. *Planta* **195**: 456-463.
- ▣ Ishiwatari, Y., T. Fujiwara., K. C. McFarland., K. Nemoto., H. Hayashi., M. Chino., y W. J. Lucas. 1998. Rice phloem

thioredoxin *h* has the capacity to mediate its own cell-to-cell transport through plasmodesmata. *Planta* **205**: 12–22.

J

- ▣ Juárez-Díaz, J. A., B. McClure., S. Vázquez-Santana., A. Guevara-García., P. León-Mejía., J. Márquez-Guzmán., y F. Cruz-García. 2006a. A novel Thioredoxin *h* is secreted in *Nicotiana alata* and reduces S-RNase *in vitro*. *Journal of Biological Biochemistry* **281**: 3418-3424.
- ▣ Juárez-Díaz, J. A. 2006b. NaTrxh es una tiorredoxina *h* de secreción que reduce *in vitro* a la S-RNasa de *Nicotiana alata*. Tesis de Doctorado en Ciencias Bioquímicas, Facultad de Química, UNAM.

K

- ▣ Karplus, K., S. Katzman., G. Shackleford., M. Koeva., J. Draper., B. Barnes., M. Soriano., y R. Hughey. 2005. SAM-T04: what's new in protein-structure prediction for CASP6. *Structure, Function, and Bioinformatics* **61**: 135-142.
- ▣ Karplus, K., R. Karchin., J. Draper., J. Casper., Y. Mandel-Gutfreund., M. Diekhans., y R. Hughey. 2003. Combining local-structure, fold-recognition, and new-fold methods for protein structure prediction. *Proteins: Structure, Function and Genetics* **6**: 491-496.
- ▣ Karplus, K., R. Karchin., C. Barrett., S. Tu., M. Cline., M. Diekhans., L. Grate., J. Casper., y R. Hughey. 2001. What is the value added by human intervention in protein structure prediction? *Proteins: Structure, Function and Bioinformatics* **45**: 86-91.

- ▣ Karplus, K., C. Barrett., y R. Hughey. 1998. Hidden Marcov models for detecting remote protein homologies. *Bioinformatics* **14**: 846-856.
- ▣ Krause, G., J. Lundström., J. López-Barrera., C. Pueyo-de la Cuesta., y A. Holmgren. 1991. Mimicking the active site of protein disulfide-isomerase by substitution of proline 34 in *Escherichia coli* thioredoxin. *Journal of Biological Chemistry* **266**: 9494-9500.
- ▣ Kumar, S., K. Tamura., I. B. Jakobsen., y M. Nei. 2001. MEGA2: molecular evolutionary genetics analisis software. *Bioinformatics* **17**: 1244-1245.

L

- ▣ Lack, A. J., y D. E. Evans. 2001. Instant Notes. Plant Biology. *Bios y Springer*. E. U. A.
- ▣ Laloi, C., N. Rayapuram., Y. Chartier., J. M. Grienenberger., G. Bonnard., e Y. Meyer. 2001. Identification and characterization of a mitochondrcial thioredoxin system in plants. *The Proceedings of the National Academy of Sciences* **98**: 14144-14149.
- ▣ Lesk, A. M., y C. H. Chothia. 1986. The response of protein structures to amino-acid sequence changes. *Philosophical Transactions of the Royal Society* **317**: 345-356.

M

- ▣ Maeda, K., C. Finnie., y B. Svensson. 2003a. Cy5 maleimide-labelling for sensitive detection of free thiols in native protein extracts: identification of seed proteins targeted by barley thioredoxin *h* isoforms. *Biochemical Journal* **24**: 497-507.
- ▣ Maeda, K., C. Finnie., O. Østergaard., y B. Svensson. 2003b. Identification, cloning and characterization of two thioredoxin *h*

- isoforms, HvTrxh1 and HvTrxh2, from the barley sered proteome. *European Journal of Biochemistry* **270**: 2633-2643.
- ▣ Maeda, K., P. Häggglund., C. Finnie., B. Svensson., y A. Henriksen. 2006. Structural basis for target protein recognition by the protein disulfide reductase thioredoxin. *Structure* **14**: 1701-1710.
- ▣ Marchand, C., P. L. Maréchal., Y. Meyer., M. Miginiac-Maslow., E. Issakidis-Bourguet., y P. Decottignies. 2004. New targets of Arabidopsis thioredoxins revealed by proteomic analysis. *Proteomics* **4**: 2696-2706.
- ▣ Martí-Renom, M. A., A. C. Stuart., A. Fiser., R. Sánchez., F. Melo., y A. Sali. 2000. Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure* **29**: 291-325.
- ▣ Martin, A. C., y J. M. Thornton. 1996. Structural families in loops of homologous proteins: automatic classification, modeling and application to antibodies. *Journal of Molecular Biology* **263**: 800-815.
- ▣ Martin, F. W. 1968. The behavior of *Lycopersicon* incompatibility alleles in an alien by proteomic analysis. *Proteomics* **4**: 2696-2706.
- ▣ Mather, K. 1943. Specific differences in *Petunia* I. Incompatibility. *Journal of Genetics* **45**: 215-235.
- ▣ McClure, B. A., F. Cruz-García., B. Beecher., y W. Sulaman. 2000. Factor affecting inter- and intra-specific Pollen Rejection in *Nicotiana*. *Annals of Botany* **85**: 113-123.
- ▣ Mestres-Ortega, D., e Y. Meyer. 1999. The *A. thaliana* genome encodes at least four thioredoxins *m* and a new prokaryotic-like thioredoxin. *Gene* **240**: 307-316.
- ▣ Meyer, Y. L., L. Verdoucq., y F. Vignols. 1999. Plant thioredoxins and glutaredoxins: identity and putative roles. *Trends in Plant Science* **4**: 388-394.

- ▣ Meyer, Y., F. Vignols., y J. P. Reichheld. 2002. Classification of plant thioredoxins by sequence similarity and intron position. *Methods in Enzymology* **347**: 394-402.
- ▣ Mizuguchi, K., C. M. Deane., T. L. Blundell., y J. P. Overington. 1998. HOMSTRAD: A database of protein structure alignments for homologous families. *Protein Science* **7**: 2469–2471.
- ▣ Murfett, J., T. J. Strabala., D. M. Zurek., B. Mou., B. Beecher., y B. A. McClure. 1996. S-RNase and interspecific pollen rejection in the genus *Nicotiana*: multiple pollen-rejection pathways contribute to unilateral incompatibility between self-incompatibility and self-compatible species. *Plant Cell* **8**: 943-958.
- ▣ Mustacich, D., y G. Powis. 2000. Thioredoxin reductase. *Biochemical Journal* **346**: 1-8.

N

- ▣ Nilges, M., G. M. Clore., y A. M. Gronenborn. 1988. Determination of three-dimensional structures of proteins from inter-proton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Letters* **229**: 317-324.
- ▣ Nombela, C., C. Gil., y W. L. Chaffin. 2006. Non-conventional protein secretion in yeast. *Trends in Microbiology* **14**: 15-21.

P

- ▣ Petsko, G. A. 2006. An introduction to Modeling Structure from Sequence. *In* Baxevanis, A. D., G. A. Petsko, L. D. Stein y J. R. Yates III (Eds.) Current Protocols in Bioinformatics. *John Wiley & Sons, Inc. USA*.
- ▣ Petsko, G., y D. Ringe. 2004. Protein Structure and Function. *Sinauer Associates, Inc. Publishers*.

- ▣ Pieper, U., N. Eswar., A. C. Stuart., V. A. Ilyin y A. Sali. 2002. MODBASE, a database of annotated comparative protein structure models. *Nucleic Acids Research* **30**: 255-259.
- ▣ Powis, G., y W. R. Montfort. 2001. Propierties and biological activities of thioredoxins. *Annual Review of Pharmacology and Toxicology* **41**: 261-295.

R

- ▣ Rapp, C. S., y R. A. Friesner. 1999. Prediction of loop geometries using a generalized born model of salvation effects. *Proteins* **35**: 173-183.
- ▣ Reichheld, J. P., D. Mestres-Ortega., C. Laloi., e Y. Meyer. 2002. The multigenic family of thioredoxin *h* in *Arabidopsis thaliana*: specific expression and stress response. *Plant Physiology and Biochemistry* **40**:685–690.
- ▣ Rivera-Madrid, R., D. Mestres., P. Marinho., J. P. Jacquot., P. Decottignies., M. Muginiac-Maslow., e Y. Meyer. 1995. Evidence for five divergent thioredoxin *h* sequences in *Arabidopsis thaliana*. *The Proceedings of the National Academy of Sciences* **92**: 6511-6514.
- ▣ Roberts, E., J. Eargle., D. Wright., y Z. Luthey-Schulten. 2006. Multiseq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* **7-382**: 1471-2105.
- ▣ Russell, R. B., y G. J. Barton. 1992. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins* **14**:309–323.

S

- ▣ Sali, A. 1998. 100,000 protein structures for the biologist. *Nature Structural Biology* **5**: 1029-1032.

- ▣ Sali, A., y Blundell, T. L. 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* **234**:779-815.
- ▣ Santandrea, G., Y. Guo., T. O'Connell., y R. D. Thompson. 2002. Post-phloem protein trafficking in the maize caryopsis: zmTRXh1a thioredoxin specifically expressed in the pedicel parenchyma of *Zea mays* L., is found predominantly in the placentochalaza. *Plant Molecular Biology* **50**: 743-756.
- ▣ Schürmann, P., y J. P. Jacquot. 2000. Plant thioredoxin systems revisited. *Annual Review of Plant Physiology and Plant Molecular Biology* **51**: 371-400.
- ▣ Serrato, A. J., J. M. Perez-Ruiz., y F. J. Cejudo. 2002. Cloning of thioredoxin *h* reductase and characterization of the thioredoxin reductasethioredoxin *h* system from wheat. *Biochemical Journal* **367**: 491-497.
- ▣ Sheen, J., H. Seongbin., Y. Niwa., H. Kobayashi., y D. W. Galbraith. 1995. Green-fluorescent protein as a new vital marker in plant cells. *The Plant Journal* **8**: 777-784.

T

- ▣ Takayama, S., y A. Isogai. 2005. Self-Incompatibility in Plants. *Annual Review of Plant Biology* **56**: 467-489.
- ▣ Tanudji, M., S. Hevi., y S. Cchuck. 2003. The nonclassic secretion of thioredoxin is not sensitive to redox state. *American Journal of Physiology* **284**: C1272-C1279.
- ▣ Thompson, J. D., T. J. Gibson., F. Plewniak., F. Jeanmougin., y D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*. **24**:4876-4882.
- ▣ Topham, C .M., P. Thomas., J. P. Overington., M. S. Johnson., F. Eisenmenger., y T. L. Blundell. 1991. An assessment of

COMPOSER: a rule-based approach to modeling protein structure.

Biochemical Society Symposium **57**: 1-9.

- ▣ Tsukamoto, T., T. Ando., H. Kokubun., H. Watanabe., M. Masada., X. Zhu., E. Marchesi., y T. H. Kao. 1999. Breakdown of self-incompatibility in a natural population of *Petunia axillaris* (Solanaceae) in Uruguay containing both self-incompatible and self-compatible plants. *Sexual Plant Reproduction* **12**: 6-13.

V

- ▣ van Vlijmen, H. W., y M. Karplus. 1997. PDB-based protein loop prediction: parameters for selection and methods for optimization. *Journal of Molecular Biology* **267**: 975-1001.

W

- ▣ Williams Jr., C. H., L. D. Arscott., S. Müller., B. W. Iennon., M. L. Ludwing., P. F. Wang., D. M. Veine., K. Becker., y R. H. Schirmer. 2000. Thioredoxin reductase. Two models of catalysis have evolved. *European Journal of Biochemistry* **267**: 6110-6117.
- ▣ Wollman, E. E., L. d'Auriol., L. Rimsky., A. Shaw., J. P. Jacquot., P. Winfield., P. Graber., F. Dessarps., P. Robin., F. Galibert., J. Bertoglio., y D. Fradelizi. 1988. Cloning and expression of a cDNA for human thioredoxin. *Journal of Biological Chemistry* **263**: 15506-15512.
- ▣ Wong, J. H., Y. B. Kim., P. H. Ren., N. Cai., M. J. Cho., P. Hedden., P. G. Lemaux., y B. B. Buchanan. 2002. Transgenic barley grain overexpressing thioredoxin shows evidence that the starchy endosperm communicates with the embryo and the aleurone. *The Proceedings of the National Academy of Sciences* **99**: 16325-16330.

-
- ▣ Wu, H. M., E. Wong., J. Ogdahl., y A. Y. Cheung. 2000. A pollen tube growth promoting arabinogalactan protein from *Nicotiana glauca* is similar to the tobacco TTS protein. *The Plant Journal* **22**: 165-176.

X

- ▣ Xiang, Z., C. S. Soto., y B. Hong. 2002. Evaluating conformational free energies: the colony energy and its application to the problem of loop prediction. *Proc. Natl. Acad. Sci. USA*. **99**: 7432-7437.

Z

- ▣ Zarembinski, T. I., L. W. Hung., H. J. Mueller-Dieckmann., K. K. Kim., H. Yokota., R. Kim., y S. H. Kim. 1998. Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structure genomics. *The Proceedings of the National Academy of Sciences* **95**: 15189-15193.
- ▣ Zhang, C., y S. H. Kim. 2003. Overview of structural genomics: from structure to function. *Current Opinion in Chemical Biology* **7**: 28-32.
- ▣ Zhu, H., M. Bilgin., R. Bangham., D. Hall, A. Casamayor., P. Bertone., N. Lan., R. Jansen., S. Bidlingmaier., T. Houfek., T. Mitchell., P. Miller., R. A. Dean., M. Gerstein., y M. Snyder. 2001. Global analysis of protein activities using proteome chips. *Science* **293**: 2101-2105.