

Trabajos

Modelado Molecular de péptidos que contienen secuencias repetitivas (FRA) de la proteína calpain cisteín peptidasa putativa de *Trypanosoma cruzi*

RECIBIDO: 4/7/07

ACEPTADO: 15/9/07

Peralta, J.M.¹ • Marcipar, I.² • Sferco, S.J.^{1,3}

1. Departamento de Física, Facultad de Bioquímica y Ciencias Biológicas, UNL, Paraje El Pozo, 3000 Santa Fe, Argentina.

2. Laboratorio de Tecnología Inmunológica, Facultad de Bioquímica y Ciencias Biológicas, UNL, Paraje El Pozo, 3000 Santa Fe, Argentina.

3. INTEC (CONICET-UNL), Güemes 3450, 3000 Santa Fe, Argentina.

RESUMEN: Se presentan predicciones de estructura secundaria, antigenicidad y modelos tridimensionales para 1, 2 y 6 repeticiones de 68 aminoácidos del péptido FRA de *Trypanosoma cruzi*. Los modelos tridimensionales fueron obtenidos mediante servidores en la web utilizando métodos de "fold-recognition". Nuestros resultados predicen para los tres péptidos una estructura básicamente formada por hélices alfa y espirales al azar. Se identifican los aminoácidos responsables del comportamiento antigénico de estos péptidos, y los modelos los ubican en un loop formado por 17 a 22 aminoácidos, con anclajes cargados y prácticamente formado por aminoácidos hidrofóbicos.

PALABRAS CLAVE: FRA, secuencias repetitivas, antígeno *Trypanosoma cruzi*

SUMMARY: Secondary Structure, antigenicity and tridimensional models predicted for 1, 2 and 6 repetitions of 68 aa of *Trypanosoma cruzi* FRA peptides are presented. The tridimensional models were obtained using web servers and using fold-recognition. For the three peptides, our results predicts a structure basically formed by alfa helices and coils. The antigenic aminoacids are identified as belonging to a 17 to 22 aminoacids large loop, with anchored charged but essentially formed by hydrophobic amino acids.

KEY WORDS: FRA, repetitive sequence, *Trypanosoma cruzi* antigen

I. Introducción

Se ha descrito que los genomas de ciertos protozoos poseen secuencias genómicas de baja complejidad. (1) Estas se expresan generalmente como regiones no globulares de proteínas o como secuencias repetitivas que separan dominios bien conservados, los cuales pueden ser alineados con proteínas homólogas de otros organismos. En el caso del *T. cruzi*, a comienzos de los años 90, cuando comenzaron a clonarse y secuenciarse las proteínas más antigénicas de este protozoo, se determinó que estas estaban compuestas esencialmente por este tipo de secuencias repetitivas. (2) A pesar del tiempo transcurrido desde esta observación, la razón que incrementa la antigenicidad de estas estructuras repetitivas no está aún aclarada.

Un caso particular, es el Antígeno Flagelar Repetitivo (Flagellar Repetitive Antigen, FRA) correspondiente al extremo amino terminal de la proteína calpaín cisteín peptidasa de *T. cruzi* de 1342 aminoácidos (aa). Esta proteína posee 6 repeticiones de 68 aa que ha resultado ser uno de los péptidos más antigénicos del *T. cruzi*, además de mostrar alta especificidad y sensibilidad (3) en su utilización en ensayos de ELISA para diagnóstico de la enfermedad. (4) Esta proteína se diferencia de las demás repetitivas por poseer repeticiones de una media seis veces mayor que las demás, las cuales presentan motivos repetidos que oscilan en tamaños de 5 a 14 aa. (2) Estas diferencias de tamaño podrían estar involucradas en la importante interacción antígeno-anticuerpo que presenta esta molécula. Dado que hasta la fecha no se han descrito estructuras tridimensionales de esta molécula, nos hemos propuesto analizar los posibles plegamientos de la misma para estimar la conformación tridimensional de su interacción con

los anticuerpos. Por lo tanto, en el presente trabajo se aborda el modelado tridimensional de los péptidos correspondientes a 1 y 2 repeticiones de 68 aa del extremo N terminal de FRA que permitan tener una primera aproximación del plegamiento de las repeticiones y evaluar si dicha conformación está involucrada en la alta antigenicidad de la molécula. En un trabajo posterior estos resultados serán contrastados con el comportamiento experimental de estas moléculas clonadas y expresadas en un sistema procarionota. También se aborda el modelado de los primeros 450 aa de la proteína tal como es derivada del gen que la codifica, a fin de entender cual es la correlación con las repeticiones clonadas.

En virtud del tamaño de las estructuras a modelar hemos decidido utilizar las técnicas pertenecientes al campo del modelado comparativo. En particular, la búsqueda de proteínas homólogas en el Protein Data Bank (PDB) (<http://www.rcsb.org/pdb>, consultado desde el 11-2006 al 03-2007), (5) es decir proteínas cuya estructura tridimensional sea conocida y guarden alguna relación evolutiva con la FRA, muestra resultados de baja identidad de secuencias, tal que hace inviable un modelado por los métodos tradicionales de homología. Por esta razón, se emplean en este trabajo los métodos basados en reconocimiento de plegamiento (Fold Recognition, FR), que en la actualidad combinan el análisis evolutivo (la secuencia incógnita se compara con bases de datos de secuencias, para generar alineamientos múltiples, que son a la vez usados como incógnitas para obtener secuencias moldes del PDB) y el análisis estructural (threading). (6,7) Los algoritmos modernos de FR han alcanzado un alto grado de sensibilidad y especificidad, en promedio puede afirmarse que aproximadamente el 50% de los mode-

los primarios generados por los servidores públicos encuentran el tipo de plegamiento correcto, y si bien poseen baja resolución, son de gran utilidad para el entendimiento de la relación estructura-función de las proteínas y la asignación de funciones a nivel genómico. Desde 1994 se lleva adelante el experimento colaborativo bianual "Critical Assessment of Protein Structure Prediction Methods" (CASP), (8) en el cual proteínas próximas a ser resueltas por cristalografía son puestas a disposición de la comunidad de grupos predictores bioinformáticos. Los resultados de la resolución cristalográfica final son exhaustivamente comparados con las predicciones y los métodos y algoritmos de predicción que son luego clasificados según su performance. Recientemente surgieron otros nuevos experimentos comunitarios para aclarar varias inquietudes de la comunidad científica a la hora de decidir entre qué métodos, algoritmos y servidores públicos elegir para llevar adelante un proceso de modelado. "Critical Assessment of Fully Automated Protein Prediction Methods" (CAFASP), (9) evalúa sólo los servidores que proveen procedimientos de trabajo enteramente automatizados; LiveBench (10) es una iniciativa para evaluar en lapsos semestrales la performance de los servidores públicos de modelado. En el presente trabajo, utilizaremos los servidores de "fold-recognition" mejor posicionados en los experimentos colaborativos mencionados. Por otra parte, el análisis de los modelos y el refina-

miento de los mismos es uno de los problemas más activos dentro de la bioinformática actual, la evaluación de la calidad de los modelos se lleva cabo con algoritmos que verifican el ambiente fisicoquímico de cada aminoácido a partir de las coordenadas espaciales generadas por el programa predictor. Los más utilizados en la actualidad son Verify3D, (11) ANOLEA, (12) PROSAIL, (13) PROCHECK. (14) Según la sofisticación del método presentan desde sólo valores discretos para cada aminoácido en función de la secuencia o reportes completos con información estereoquímica y datos de energía potencial de cada posición y del modelo total. En este trabajo usaremos sistemáticamente el método Verify3D.

Nuestro objetivo es obtener modelos tridimensionales para los péptidos formados por 1, 2 y 6 repeticiones de la proteína FRA de *Trypanosoma cruzi*, utilizando los métodos bioinformáticos de "fold-recognition" disponibles en la actualidad, y analizar la factibilidad de su comportamiento antigénico.

II. Sistemas estudiados y métodos empleados

Dado que los resultados aquí obtenidos se contrastarán en un trabajo posterior con estudios de antigenicidad experimental que se llevará a cabo con construcciones realizadas mediante ingeniería genética, se modelan aquí las secuencias tal como se obtendrán expresadas.

FRA-1:

**AMADIGSEFKKKLADRAFLEQKPEGVPLRELPLDDSDVFAMEQERRQLEKDPRRNAREIAALEESM
NARAQELAR**

FRA-2:

**AMADIGSEFKKKLADRAFLEDKPEGVPLRELPLDDSDVFAMEQERRQLEKDPRRNAKEIAALEESM
NARAQELAREKKLADRAFLDQKPEGVPLRELPLDDSDVFAMEQERRQLEKDPRRNAREIAALEESM
NARAQELAR**

Los residuos en **negrita** corresponden a los aa aportados por el vector de expresión pET32a que se utilizará para la expresión de la proteína.

FRA-450:

LLEKDPRRNAKEIAALEESMNARAQELAREKKLADRAFLDQKPEGVPLRELPLDDSDVFAMEQERR
QLEKDPRRNAREIAALEESMNARAQELARKLADRAFLDQKPEGVPLRELPLDDSDVFAMEQERRQ
LLEKDPRRNAREIAALEESMNARAQELAREKKLADRAFLDQKPEGVPLRELPLDDSDVFAMEQERR
QLEKDPRRNAKEIAALEESMNARAQELAREKKLADRAFLDQKPEGVPLRELPLDDSDVFAMEQER
RQLEKDPRRNVQKIADLEESMNARAQELAREKKLADRAFLDQKPEGVSLRELPLDDSDVFAMEQ
RRQLEKDPRRNVQIADLEESM**NARAQELAREKKLADRAFLDQKPEGVSLRELPLDDSDVFAMEQ**
ERRQLEKDPHRNAKEIAALEES**MNVCARNLAFDIRSRERDFLDDVVR**

Esta es la secuencia nativa de los primeros 450 aa de la FRA, utilizada para corroborar la coherencia entre las predicciones hechas para las porciones aisladas FRA-1 y FRA-2 y el extremo N terminal de la proteína.

Secuencia completa de FRA según GenBank XM_799896

LLEKDPRRNAKEIAALEESMNARAQELARE**KKLADRAFLDQKPEGVPLRELPLDDSDVFAMEQERR**
QLEKDPRRNAREIAALEESMNARAQELARKLADRAFLDQKPEGVPLRELPLDDSDVFAMEQERRQ
LLEKDPRRNAREIAALEESMNARAQELARE**KKLADRAFLDQKPEGVPLRELPLDDSDVFAMEQERR**
QLEKDPRRNAKEIAALEESMNARAQELARE**KKLADRAFLDQKPEGVPLRELPLDDSDVFAMEQER**
RQLEKDPRRNVQKIADLEESMNARAQELARE**KKLADRAFLDQKPEGVSLRELPLDDSDVFAMEQER**
RRQLEKDPRRNVQIADLEES**MNVCARNLAFDIRSRERDFLDDVVR**GIPLDALSLNDDNELCLLEAR
 RRELLKTSSAENSPELVELEKKIADRVDFLAVNFGHELLSFLDSKPEGISLSELELNGDLEFCNMERVLV
 ELMRARRQNAEAIKQQYAMNNRVHELAAQLLRSDREYLHPEPQGVPGQDLDDPVFHEMELQRR
 KLKDPERNAIKISELEKKLNDRADIALLRAKERAFLELEPEGIPIERLPLNEDPILHELETNYRRLKVT
 PRDKKAIRGIEEKIRSRVHELAVQQRGWQDEEFHESNKHMAEEWPRICELYPEGIRDVPEKTLPSQ
 VSSAPLELGYLAPFIAAMSRHPPLIDRLFDSKEHPVNGPYSFIFYDPSNPVRVEIDRVPVDANMEPKF
 TRVPKRSWYPLLEKAYAKFVGGYSRLDQCTPHETLRDLTGRPVTHIPFEDKRAEGIKMGDFRSAQFW
 REIHSDLAKGDIITAMSNKHVPDGIHPLCSYALFAVIETVKESNDPADIVIKLHNCYFDEPFYSGPLNRND
 GGWTTELMNACRYNPSEEEFLYLPQSVLNNFSSMQRCHINCGDRLTAIGEDWTKSCGGNPKFTTFR
 NNPIYLVENKSSRPVIRLAEALRHQAPVFDADSVGHYHQTGLALLQHGDGVSVLSGIITNSTHNFQKGI
 MLDTREVC SRMEIPTSTCILIPYMKRGCLGKFSVSIYPGDSSVNFMLPTPLSVTHGFCDDVDVILTPGSR

EGKRIEFVVGACDAHLLLRQNKITDPASIKKGDVLAEDDVMMLLYDEYMTRLASTGDATSAREHSLAL
 QLPSAGRYSVLLACPNKPVTVGNPCSLYIYTPKQIATRILPRPTNGTPQILPFLSLPQSSKGAARGNVKQ
 KVIGAGDVATGTGNRPVETQGMKLPNPPRNGKPKYHR

Están resaltados los primeros 450 residuos que contienen las 6 repeticiones de 68 aa, que es uno de los sistemas estudiados en este trabajo. Asimismo, en este grupo de 450 aa, están contenidas las dos primeras repeticiones (subrayadas) correspondientes a los otros dos sistemas estudiados.

Análisis de la estructura primaria

El análisis de la estructura primaria de la secuencia completa se llevó a cabo a través de la base de datos de alineamientos múltiples y Modelos Ocultos de Markov (HMM), Pfam. (15) Se identifican en este análisis los posibles motivos biológicos y/o estructurales presentes en nuestra secuencia incógnita. El análisis del índice de antigenicidad se llevó a cabo mediante el servidor del European Molecular Biology Laboratory (EMBL) a través de su programa EMBOSS (<http://liv.bmc.uu.se/cgi-bin/emboss/antigenic>, consultado el 26-02-2007) que implementa una modificación del algoritmo de Jameson-Wolf (16) propuesta por Kolaskar y Tongaonkar. (17)

Predicción de la estructura secundaria

Los métodos de análisis de la estructura secundaria han alcanzado un alto grado de precisión y exactitud en la actualidad. Todos los métodos modernos utilizan algoritmos de redes neuronales y la información evolutiva contenida en un alineamiento múltiple de la secuencia incógnita con una base de datos de secuencias, para aplicar algoritmos de desplazamiento por ventanas que recorren toda la secuencia asignado propensiones de estructura secundaria en tres

categorías: hélices α , filamentos β y espirales al azar. La mayoría de los métodos actuales alcanza un nivel de confiabilidad de 76-84 % (mayor en hélices α). (18,19) Para el presente trabajo se utilizaron los servidores de predicción Jpred (<http://www.compbio.dundee.ac.uk/~www-jpred/>, consultado el 15-02-2007) (20) y SPINE (<http://theory.med.buffalo.edu>, consultado el 21-02-2007). (21)

Predicción de la estructura terciaria

La predicción de la estructura terciaria se realizó vía web utilizando servidores públicos. De la gran cantidad de opciones que existen en la actualidad, se hizo una selección sobre la base de los análisis de las últimas rondas de CASP, CAFASP y LiveBench y dentro de los mejores posicionados se puso mayor énfasis en aquellos servidores que generan modelos completos ("full atom models"), generalmente mediante la utilización del software de modelado molecular MODELLER. (22) Resulta de gran utilidad el desarrollo de los metaservidores, que son servidores públicos que envían consultas a múltiples servidores primarios. Algunos de estos, poseen a su vez algoritmos de selección de los mejores modelos sobre la base de criterios estructurales y evolutivos, como por ejemplo PCONS. (23)

Tabla 1: Servidores utilizados para consultas remotas. (FR), Fold Recognition; (Hom), Modelado por Homología; (Ab), Ab Initio; (Me), metaservidor; (*) producen modelos tridimensionales completos minimizados utilizando MODELLER; (+) producen modelos crudos; (x) producen alineamientos entre la secuencia incógnita y un molde. (a) SwissModel posee su propio software de modelado por homología, PROMOD. (33)

Nombre	URL (http://)	Referencias
I-TASSER	Zhang.bioinformatics.ku.edu/I-TASSER/ (consultado 05-2007)	FR-Ab; *;(24)
SPARKS (sp3 – sp4)	sparks.informatics.iupui.edu/ (consultado 02-2007)	FR; *;(25)
Genesilico	genesilico.pl/meta/ (consultado 03/04-2007)	FR-Me; *;(6, 26)
3D-Jury	www.bioinfo.pl/meta/ (consultado 04-2007)	FR-Me; +;(27)
@TOME	bioserv.cbs.cnrs.fr/HTML_BIO/frame_meta.html (consultado 04/05-2007)	FR-Me; *;(28)
LOOPP	cbsuapps.tc.cornell.edu/toopp.aspx (consultado 03-2007)	FR; *;(29)
PCONS	www.bioinfo.se/pcons/ (consultado 04/05-2007)	Me; *;(23)
FUGUE	www-cryst.bioc.cam.ac.uk/~fugue/ (consultado 02-2007)	FR; x (30)
3D-PSSM	www.sbg.bio.ic.ac.uk/3dpssm/html/ffrecog_simple.html (consultado 03-2007)	FR; +; (31)
mgenthreader	bioinf.cs.ucl.ac.uk/psipred/ (consultado 02-2007)	FR; x (32)
PredictProtein	www.predictprotein.org/ (consultado 02-2007)	FR-Me; x; (33)
SwissModel	swissmodel.expasy.org/SWISS-MODEL.html (consultado 11-2006, 02-2007 y 05-2007)	Hom; *; (34)a
PROTINFO	protinfo.compbio.washington.edu/ (consultado 05-2007)	FR-Ab; *;(35)
FFAS03	ffas.ijcf.edu (consultado 03-2007)	FR; x (36)
SAM-T02	www.soe.ucsc.edu/research/compbio/HMM (consultado 03-2007)	FR; x (37)
INBGU	www.cs.bgu.ac.il/~bioinbgu/form.html (consultado 04-2007)	FR; *;(38)
WURST	www.zbh.uni (consultado 03/2007)	FR; +; (39)
ORFEUS	grdb.bioinfo.pl (consultado 05-2007)	FR; x; (40)
Robetta	robeta.bakerlab.org/ (consultado 02/2007 resultados obtenidos 04/2007)	Ab-Me; *; (41)
LOMMETS	Zhang.bioinformatics.ku.edu/I-LOMMETS/ (consultado 06-2007)	FR-Ab; *;(42)
HHPRED	Toolkit.tuebingen.mpg.de/hhpred (consultado 02-2007)	Hom; *;(43)
INUB	inub.cse.buffalo.edu/ (consultado 03/04-2007)	FR; +; (44)
FORTE	www.cbrc.jp/forte/ (consultado 04/05-2007)	FR; *;(45)

Verificación de los modelos

Los modelos obtenidos fueron analizados con el algoritmo Verify3D, que actualmente es uno de los mecanismos de análisis más utilizados para evaluar los modelos primarios de los servidores, (6,7) en conjunto con diversas herramientas adicionales. A pesar de ser uno de los algoritmos más antiguos y sencillos, se ha visto excelentemente posicionado en las últimas rondas de evaluación de los métodos de análisis de modelos.

(46) El algoritmo se ejecutó en forma remota en el servidor del National Institute of Health (NIH) (http://nihserver.mbi.ucla.edu/Verify_3D/, consultado en 06/2007). Cada posición aminoacídica es evaluada con referencia a 18 categorías fijadas por los autores, según los ambientes fisicoquímicos estándar de los aminoácidos en las proteínas con estructura conocida. Para cada posición el algoritmo devuelve un valor que cuanto más cercano a cero es, menor calidad del modelo indica; y en caso de ser negativo, indica una zona con problemas que debiera ser refinada.

Los modelos seleccionados por Verify3D fueron posteriormente optimizados por minimización energética con el programa Swiss-PDB-Viewer. (47) Se utilizó una rutina de optimización aplicando el método de gradientes conjugados, de 3000 pasos en tres series de 1000 pasos con el campo de fuerzas GROMOS96. (48)

III. Resultados y Discusión

Análisis de la estructura primaria:

Los resultados del análisis de la estructura primaria se muestran en la Fig. 1. La secuencia completa posee cinco dominios asignables por alineamiento con los perfiles alojados en Pfam, dos de los cuales no poseen un nivel de probabilidad suficiente para ser considerados confiables. Como se aprecia en la Fig. 1 la proteína es una Cis-

teín Proteinasa Calcio Dependiente tipo Calpaína. El dominio catalítico se encuentra entre los residuos 720-995. Un Dominio Calpaína III adicional se encuentra entre 995-1140. Dos regiones a las que no se puede asignar dominios con significación estadística suficiente ocupan las porciones de 428-720 y 1140-1342 aminoácidos. Finalmente la porción en estudio posee similitud con un dominio conservado de ATPasa de reparación del ADN.

La actividad enzimática de la FRA no ha sido experimentalmente comprobada hasta el momento, pero del análisis bioinformático de la secuencia primaria se infiere, con una gran probabilidad, que actúe como una calpaína. Las calpaínas de otros organismos estudiadas hasta el momento poseen extensiones variables que van de los 345 aa hasta los 968 aa en calpaínas multidominio. (49) Es una característica general de ciertas proteínas presentes en protozoos que infectan a los seres humanos la presencia de inserciones aminoacídicas de hasta 900 aa. En algunos casos corresponden a la presencia de varios dominios funcionales de las cisteína proteasas que actúan en la regulación enzimática, pero es común observar, como ocurre en el caso de FRA, la existencia de zonas de baja complejidad de información y secuencias repetitivas. (1) La presencia de estas grandes inserciones supone un incremento en el gasto energético considerable para estos organismos unicelulares, en orden de replicar, transcribir y reparar ADN y para traducirlo en proteínas. Este gasto diferencial debe estar plenamente justificado desde el punto de vista evolutivo, ya que la vida como sistema termodinámico es sumamente conservadora en cuanto al gasto energético. Por ello estas inserciones, que en el caso de FRA están representadas por los primeros 720 aa y los últimos 202, de-

Figura 1: Análisis de la estructura primaria de la proteína FRA completa utilizando la base de datos Pfam. Se muestran esquemáticamente los dominios encontrados y aquellos donde la propia base de datos no puede encontrar dominios asignables.

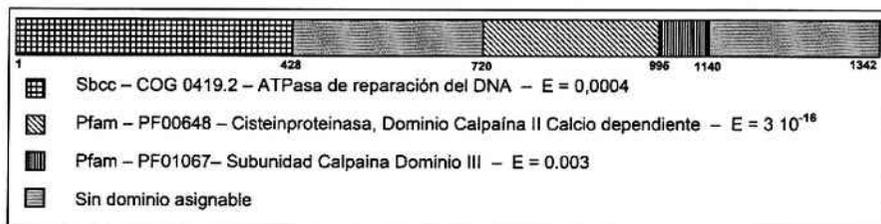
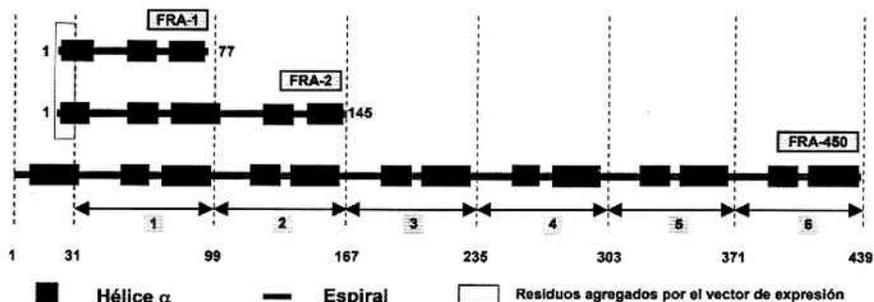


Figura 2: Esquema de la estructura secundaria predicha por consenso entre los algoritmos Jpred y SPINE. Se observa esquemáticamente la repetición de un arreglo de estructura secundaria que se conserva aún en las porciones separadas analizadas en forma aislada. 1, 2, 3, 4, 5 y 6 indican las repeticiones.



ben tener una función biológica justificada. Para los primeros 428 aa, la hipótesis actual es considerar estas secuencias repetitivas como señuelos antigénicos que evitan el desarrollo de anticuerpos por parte del huésped hacia los epitopes de la proteína presentes en la porción con actividad catalítica, así pues la proteína desarrolla su actividad catalítica aun en presencia de respuesta de anticuerpos por parte del huésped. (1)

Predicción de estructuras secundarias

La predicción de la estructura secundaria se resume en la Fig. 2 para las tres secuencias analizadas y, en detalle, en las Figs. 3, 8 y 13, para la FRA-1, FRA-2 y FRA-450, respectivamente. La concordancia entre los resultados de los dos métodos de predicción de estructura secundaria utilizados es prácticamente total. La Fig. 2 incluye un consenso de las dos predicciones, donde puede observarse un arreglo regular de estructura secundaria (formado por dos hélices con un pequeño loop entre ellas), que se unen al

motivo de la repetición siguiente a través de un loop manifiestamente más grande que el anterior, que se mantiene incluso en las secuencias aisladas FRA-1 y FRA-2 alineadas con respecto a la secuencia nativa FRA-450. Se realizó, además, un análisis de la presencia de hélices transmembrana llevado a cabo con los algoritmos de predicción primarios consultados por el meta-servidor BPROMPT (www.jenner.ac.uk/BPROMPT, consultado el 16-02-2007). (50) No se detectó presencia de hélices transmembrana en ninguna de las secuencias en estudio, ni en ningún otra parte de la secuencia proteica completa (resultados no mostrados). Esto indicaría que FRA no es una proteína de membrana, en acuerdo con la actividad de las calpaínas, que es principalmente citosólica. (49)

Predicción de antigenicidad

Los resultados de la predicción de la antigenicidad se encuentran resumidos en las Figs. 3, 8 y 13. Es de destacar que estos resultados son coherentes con las predicciones de estructura secundaria, y los elementos de mayor potencial antigénico se encuentran en las zonas predichas como espirales al azar, las cuales forman los loops de conexión entre hélices, y son por lo general sus partes más expuestas al solvente. Adquiere además relevancia el hecho que el mayor índice de antigenicidad se encuentre alojado en un tipo de loop en particular, de 16 a 22 aa de extensión separando dos hélices, una corriente abajo de 19-22 aa y otra corriente arriba, de 13-15 aa. Estas zonas, además, cuando se estudian en conjunto con la secuencia primaria muestran presencia de residuos ácidos y básicos en los extremos y no polares en el centro. Es de resaltar que, patrones similares pueden ser encontrados en las regiones determinantes

de la complementariedad de los anticuerpos (RDC). (52) Se predice además que el segmento de mayor antigenicidad, contenido en el loop de tipo largo posee entre 9 y 12 aa, longitud congruente con el tipo de péptidos presentados por las moléculas del complejo mayor de histocompatibilidad tipo II (CPHII).

Esta coherencia entre los resultados marca una pauta importante y nos han llevado a tomar como criterio el de utilizar la congruencia de los resultados de predicción de estructura secundaria en conjunto con los de predicción de la antigenicidad para evaluar los modelos tridimensionales.

Predicción de estructura terciaria

Los modelos tridimensionales presentados en este trabajo corresponden a los obtenidos directamente por los servidores, a los que sólo se les ha efectuado una minimización de energías utilizando el campo de fuerzas GROMOS96. Son, por lo tanto, modelos preliminares que requerirán refinamientos por técnicas de Dinámica Molecular o Monte Carlo. Sin embargo, estos modelos preliminares, ya permiten orientar la validación experimental por medio de hidrólisis selectiva, mutagénesis dirigida o mediciones de difracción circular. (6)

Péptido FRA-1

De los servidores públicos incluidos en la Tabla 1, se obtuvieron 221 resultados que se clasifican en la Tabla 2:

Tabla 2: Tipos de resultados obtenidos de los servidores públicos para FRA-1

Modelos completos		Modelos incompletos		Sólo alineamientos		Total
Servidor	Nº de resultados	Servidor	Nº de resultados	Servidor	Nº de resultados	
@TOME	4	INUB	11	3DPSSM	20	
Genesilico	3	Wurst	5	FFAS	20	
I-TASSER	5	3D-Jury	10	FUGUE	10	
LOOPP	50			HHPRED	20	
PCONS	16			PHYRE	20	
PROTINFO	4			PsiPred	8	
Robetta	10					
SPARKS	5					
	97		26		98	221

Sólo se procedió al análisis detallado de los 97 "full-atom models". Los restantes resultados requerían de un procesamiento posterior con el software MODELLER que hubiera sido extremadamente intensivo desde el punto de vista computacional. La simple inspección visual de los modelos, y el hecho de que varios de ellos no cubrían la secuencia completa fue el primer criterio de selección. Muchos modelos mostraban claras faltas de congruencia con los supuestos básicos acerca de la estructura secundaria y terciaria de las proteínas. De esta inspección se seleccionaron 46 modelos para ser evaluados con el algoritmo Verify3D. Los resultados de esta evaluación, fueron graficados para todos los modelos de forma tal de poder inferir la calidad de los mismos a lo largo de la secuencia (datos no mostrados). La suma de los valores puntuales de Verify3D para cada aa se utilizó como un parámetro escalar único con el cual caracterizar la calidad global de un modelo en particular. Luego de esta segunda ronda de análisis quedaron sólo 12 modelos seleccionados. A continuación, se compararon los elemen-

tos de estructura secundaria predichos por los servidores de estructura terciaria para estos 12 modelos, con aquellos predichos por los métodos de predicción de estructura secundaria. Aquellos que se alejaban notoriamente (más del 15 % de error) de las pautas predichas por Jpred y SPINE fueron descartados. Sólo cuatro modelos llegaron a la última etapa del análisis que es la minimización energética por relajamiento de las cadenas laterales y algunos ángulos diedros del esqueleto carbonado. Finalmente se seleccionaron sólo dos modelos para el péptido FRA-1, que poseían la mejor conjunción de calidad según Verify3D, acuerdo con la estructura secundaria y menor energía luego de la minimización: estos fueron los modelos I-TASSER-5 y @TOME-2. La Fig. 3 resume la estructura secundaria obtenida por cada uno de estos modelos para el péptido FRA-1, junto con las predicciones de estructura secundaria y de antigenicidad.

Figura 3: Resultados alineados de la predicción de estructura secundaria por Jpred y SPINE; de antigenicidad por EMBOSS y de la estructura secundaria asignada en los modelos tridimensionales finales @TOME-2 e I-TASSER-5 para FRA-1. Puede verse el acuerdo entre los resultados de cada método de predicción y la presencia de antigenicidad máxima en las zonas de mayor flexibilidad y accesibilidad al solvente. H, hélice α , C, espiral al azar. Secuencia aminoacídica en código de una letra.

Posición	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	
Secuencia aminoacídica	A	M	A	D	I	G	S	E	F	K	K	K	L	A	D	R	A	F	L	E	Q	K	P	E	G	V	P	L	R	E	L	P	L	D	
Predicción est. 2° SPINE	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
Predicción est. Sec. JPRED	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
ANTIGENICIDAD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Modelo @TOME-2	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
Modelo ITASSER-5	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C

Posición	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68
Secuencia aminoacídica	D	D	S	D	F	V	A	M	E	Q	E	R	R	Q	L	L	E	K	D	P	R	H	A	R	E	I	A	A	L	E	S	M		
Predicción est. 2° SPINE	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C	H	H	H	H	H	H	H	H	H	H
Predicción est. Sec. JPRED	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C	H	H	H	H	H	H	H	H	H	H
ANTIGENICIDAD	0	0	1.02	1.02	1.02	1.02	1.02	0	0	0	0	1.00	1.00	1.00	1.00	1.00	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Modelo @TOME-2	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C	H	H	H	H	H	H	H	H	H	H	H
Modelo ITASSER-5	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H

Posición	69	70	71	72	73	74	75	76	77
Secuencia aminoacídica	N	A	R	A	Q	E	L	A	R
Predicción est. 2° SPINE	H	H	H	H	H	H	H	C	C
Predicción est. Sec. JPRED	H	H	H	H	H	H	H	C	C
ANTIGENICIDAD	0	0	0	0	0	0	0	0	0
Modelo @TOME-2	H	H	H	H	H	H	H	C	C
Modelo ITASSER-5	H	H	H	H	H	H	H	C	C

Los resultados de la calidad de los modelos en función de la posición según Verify3D son presentados en la Fig. 4. Ambos modelos poseen todos valores positivos para las 77 posiciones de FRA-1, lo que indica con alta probabilidad, que la calidad de la predicción es buena. Cada modelo, además, posee mejor calidad en ciertas porciones del péptido, así I-TASSER-5 es más adecuado para explicar la porción 0-41 de FRA-1, mientras que @TOME-2 posee mayor calidad en la porción 42-77.

La forma tridimensional de los dos modelos seleccionados se presenta en las Figs. 5, 6 y 7. En la Fig. 5 se muestra una superposición de ambos modelos para apreciar sus similitudes y diferencias. Se observa claramente que el lazo largo está expuesto al solvente y de esta manera sería accesible para interactuar con la región determinante de la complementariedad del anticuerpo (RDC).

Ambos modelos además, presentan una estructura compacta, con un núcleo hidro-

fóbico y un exterior hidrofílico. Todas estas características hacen plausibles las estructuras de los modelos desde el punto de vista biológico y permiten explicar de manera coherente su actividad antigénica. En la Fig. 6 se muestran dos vistas del modelo obtenido por el servidor I-TASSER en las que se aprecian el empaquetamiento compacto de las tres hélices α que se predicen para la estructura secundaria y la exposición del loop en donde se predice el mayor índice de antigenicidad.

La Fig. 7 presenta los mismos resultados, pero ahora obtenidos por el servidor @TOME para la FRA-1, con el agregado de la orientación de las cadenas laterales en el loop largo del modelo. Puede verse que las cadenas laterales se exponen al solvente a pesar de ser hidrofóbicas, lo que indicaría que a pesar de incrementar la energía, se adopta una estructura tal como la sugerida, que puede interactuar con posibles zonas hidrofóbicas en las RDC de los anticuerpos. Esto explicaría la respuesta antigénica de esta secuencia aminoacídica en términos de estructura terciaria tridimensional.

Figura 4: Gráfica de resultados de Verify3D en función de la secuencia de aa para los modelos de FRA-1 I-TASSER-5 y @TOME-2. Se observa una mejor calidad en el modelo I-TASSER-5 en la porción inicial del péptido y mejor calidad en el modelo @TOME-2 para la segunda porción de FRA-1. Valores cercanos a cero indican zonas con probables problemas en el modelado; valores negativos indican zonas probablemente erróneas en el modelo. Notar que todos los valores son positivos para ambos modelos, indicando la buena calidad de los mismos.

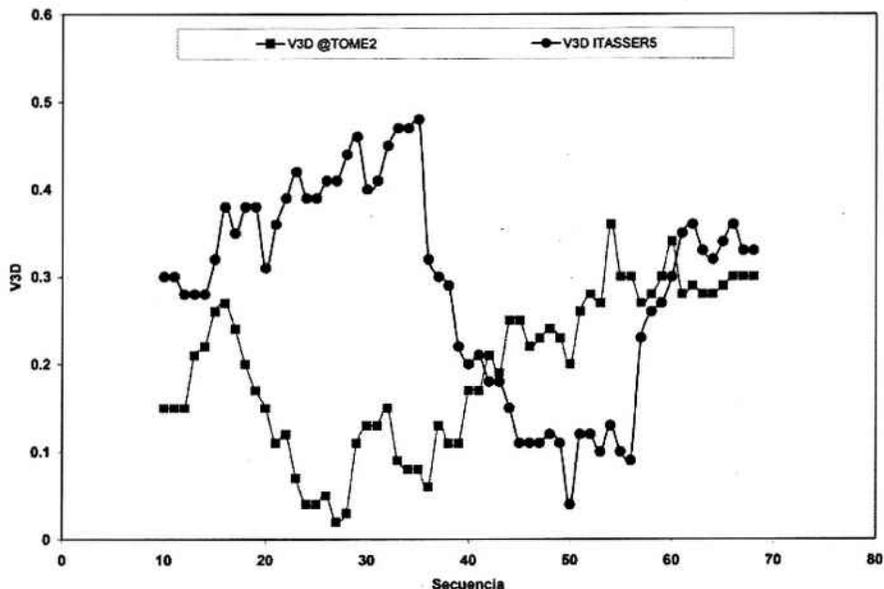


Figura 5: Modelos superpuestos de FRA-1. I-TASSER-5 (negro) y @TOME-2 (gris). Vista en perspectiva de los loops principales expuestos al solvente. Imagen generada con Swiss-PDB-Viewer. (47)



Figura 6: Vistas del modelo I-TASSER-5 para FRA-1, luego de la minimización energética. (a) Detalle del empaquetamiento de las hélices α . (b) Perspectiva del modelo mostrando la forma en que extrude del empaquetamiento de hélices α el loop donde se encuentra predicho el mayor índice de antigenicidad. Imagen generada con Swiss-PDB-Viewer. (47)

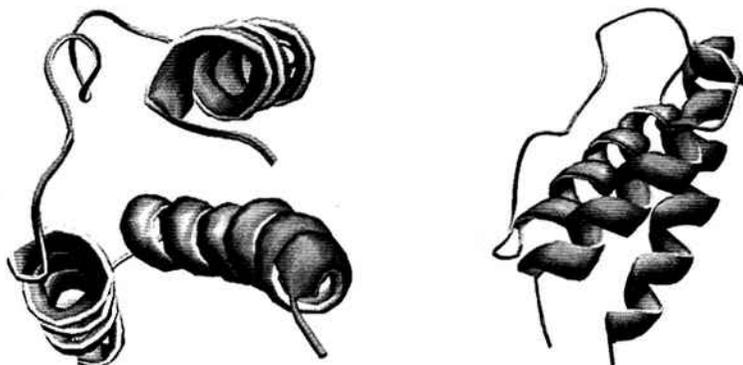
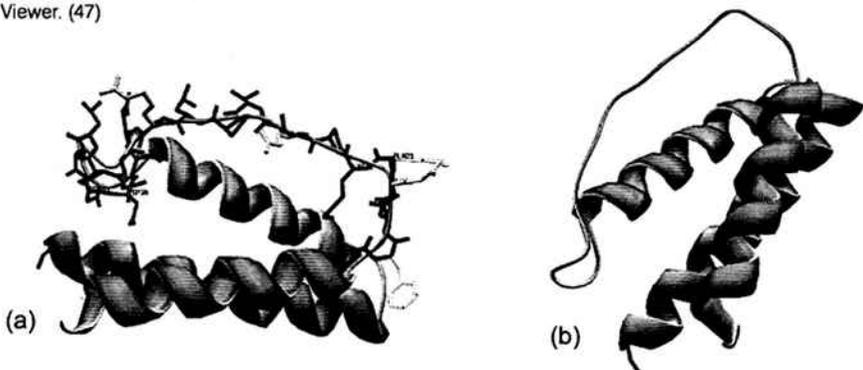


Figura 7: Vistas del modelo @TOME-2 para FRA-1 luego de minimización energética. (a) Detalle de la orientación de las cadenas laterales del loop principal. (b) Vista del modelo mostrando el loop donde se encuentra predicho el mayor índice de antigenicidad. Imagen generada con Swiss-PDB-Viewer. (47)



Péptido FRA-2

Siguiendo el mismo esquema de trabajo que para FRA-1, para el péptido con dos repeticiones, FRA-2, se obtuvieron 194 resultados de los servidores, según se detalla en la Tabla 4.

Tabla 4: Tipos de resultados obtenidos de los servidores públicos para FRA-2

Modelos completos		Modelos incompletos		Sólo alineamientos		Total
Servidor	Nº de resultados	Servidor	Nº de resultados	Servidor	Nº de resultados	
@TOME	3	INUB	10	3DPSSM	20	
Genesilico	3	Wurst	5	FUGUE	10	
I-TASSER	5	3D-Jury	10	HHPRED	10	
LOOPP	50			PHYRE	10	
PCONS	20			PsiPred	10	
PROTINFO	2					
LOMMETS	11					
Robetta	10					
SPARKS	5					
	109		25		60	194

Luego de la inspección inicial se procedió al análisis sistemático por Verify3D a sólo 47 de los 109 full-atoms models. La aplicación del algoritmo de evaluación sólo dejó 7 modelos como posibles candidatos para un ulterior análisis según la congruencia con la estructura secundaria predicha y la predicción de antigenicidad. Los 7 modelos fueron evaluados por minimización energética. Finalmente el mejor modelo según la conjunción de calidad según Verify3D, coherencia en la predicción de la estructura secundaria y menor energía fue forte-5, que fue seleccionado a su vez, por el metaservidor PCONS. La Fig. 8 resume la estructura secundaria obtenida para este modelo, com-

parándola con las obtenidas utilizando los predictores de estructura secundaria y la predicción de antigenicidad.

El modelo reproduce las pautas predichas para FRA-1 por I-TASSER y @TOME, es decir, la presencia de un empaquetamiento compacto de las tres hélices α que conforman una repetición de 68 aa en la secuencia primaria con un loop extenso entre dos hélices que se expone al solvente. Este esquema aparece repetido dos veces en el modelo tridimensional, tal como ocurre con la predicción de estructura secundaria. La posibilidad de interacción con las RDC de los anticuerpos es también altamente probable para una estructura de este tipo.

procedimiento debe ser realizado formalmente mediante una simulación por docking de proteínas, esta comparación de tamaños relativos sugiere que tal docking es posible, tanto para la FRA-1 como para la FRA-2.

Figura 9: Gráfica de resultados de Verify3D en función de la secuencia de aa para el modelo forte-5. Pueden observarse zonas con valores negativos, que indican zonas a refinar. Las zonas en la que se predice una alta antigenicidad poseen un alto valor y por tanto son de buena calidad.

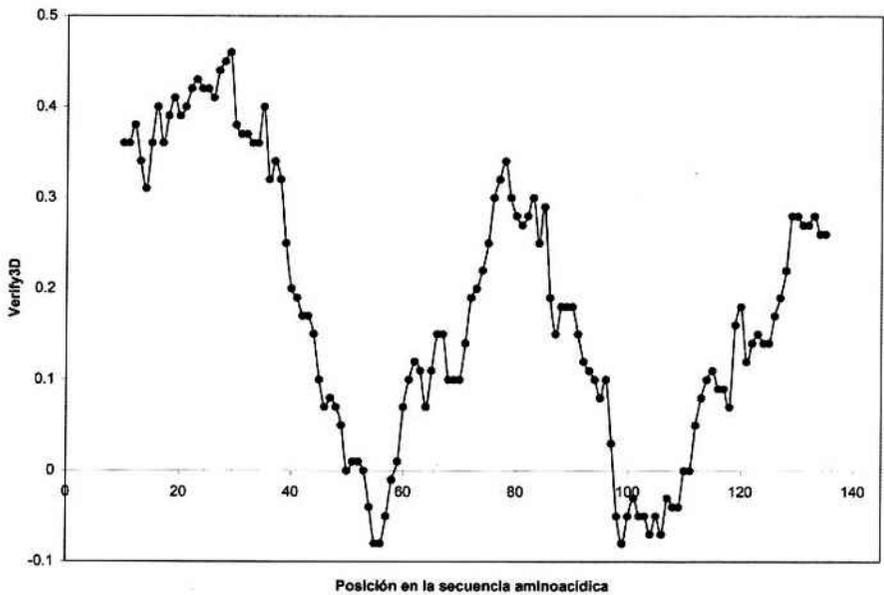


Figura 10: Modelo de FRA-2 generado por el servidor primario FORTE y seleccionado por el meta-servidor PCONS. (a) Vista del modelo luego de la minimización energética. (b) Detalles de las cadenas laterales de los dos loops principales. Imagen generada con Swiss-PDB-Viewer. (47)

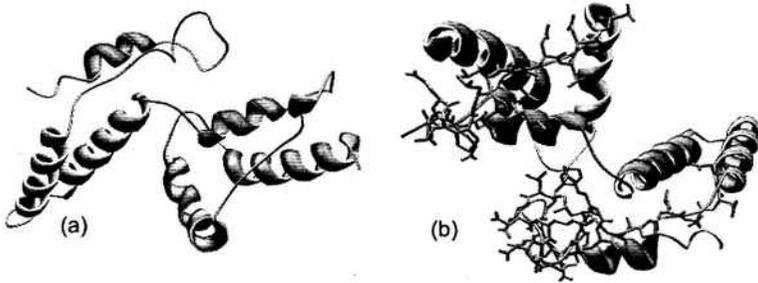


Figura 11: (a) Relaciones de escala entre las estructuras predichas para FRA-1 y un modelo de un anticuerpo IgG1 teórico tomado del PDB. (b) @TOME-2 con detalles de las cadenas laterales. Imagen generada con Swiss-PDB-Viewer. (47)

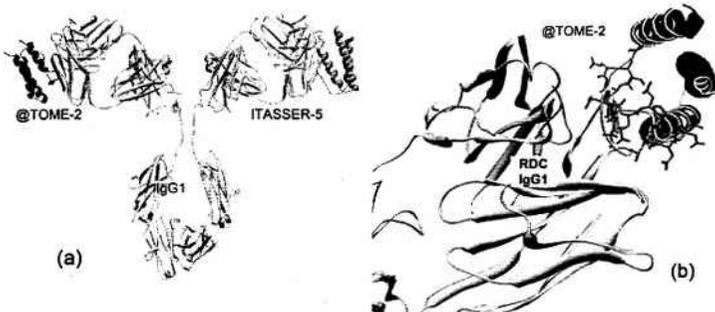
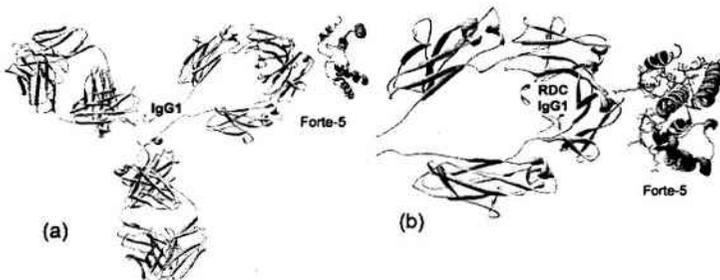


Figura 12: (a) Relación de escala entre un modelo de IgG1 y el modelo forte-5 para FRA-2. (b) Posible arreglo intermolecular del modelo de FRA-2 y el modelo de IgG1, donde se muestra un detalle de las cadenas laterales que podrían entrar en contacto con las RDC del anticuerpo. Imagen generada con Swiss-PDB-Viewer. (47)



Péptido FRA-450:

De los servidores públicos se obtuvieron 135 resultados que se clasifican en la Tabla 6:

Tabla 6: Tipos de resultados obtenidos de los servidores públicos para FRA-2

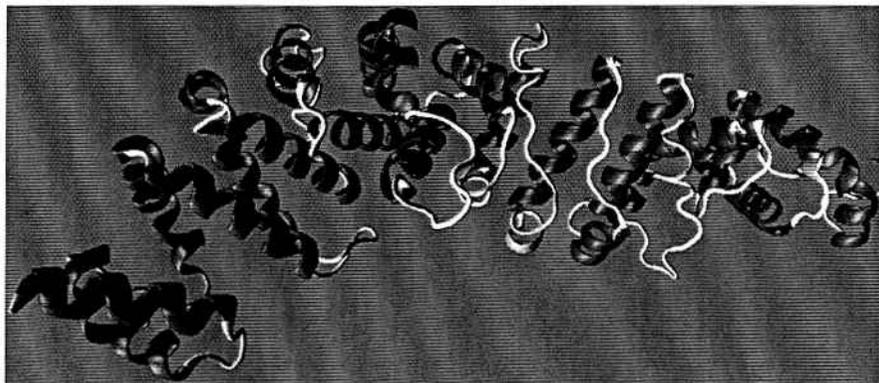
Modelos completos		Modelos incompletos		Sólo alineamientos		Total
Servidor	Nº de resultados	Servidor	Nº de resultados	Servidor	Nº de resultados	
Genesilico	3	INUB	10	@TOME	16	
ITASSER	5	3D-Jury	10	SAM-T02	30	
LOOPP	50					
HHPRED	1					
SPARKS	10					
	69		20		46	135

Luego de la inspección inicial de los modelos se seleccionaron sólo 15 para ser evaluados por Verify3D, la comparación de la congruencia de los resultados de la predicción de la estructura secundaria, la antigenicidad y los resultados mismos de Verify3D indicaron que sólo 3 modelos eran adecuados para encarar un minimización energética. La Fig. 13 resume el análisis de predicción de estructura secundaria, de antigenicidad y estructura secundaria asignada por el modelo I-TASSER-5 para FRA-450. Puede verse el acuerdo entre los resultados de cada método de predicción y la presencia de antigenicidad máxima en las zonas de mayor flexibilidad y accesibilidad al solvente. En la Fig. 14 se muestra un esquema de la estructura tridimensional de los 450 aa iniciales de FRA.

Figura 13 (pág. siguiente): Resultados alineados de la predicción de estructura secundaria por Jpred; de antigenicidad por EMBOSS y de la estructura secundaria asignada en el modelo tridimensional final I-TASSER-5 de FRA-450. H, hélice α , C, espirales al azar. Secuencia aminoacídica en código de una letra.

Figura 14: Modelo ITASSER-5 para FRA-450.

Imagen generada con Visual Molecular Dynamics-VMD (54)



La calidad de predicción de la estructura terciaria se muestra en función de la posición según Verify3D, en la Fig. 15. Existen zonas con claros problemas de asignación de estructura, indicados por valores negativos de Verify3D. La minimización energética no fue suficiente para relajar estas distorsiones y seguramente es necesario un método de refinamiento más sofisticado. Comparando los resultados de Verify3D para FRA-1, FRA-2 y FRA-450, podemos apreciar que la calidad del modelo disminuye con la cantidad de repeticiones. En la Fig. 16 pueden apreciarse dos representaciones esquemáticas del modelo sugerido para FRA-450, esta comparación de escala permite apreciar que la porción antigénica es incapaz por su tamaño de interactuar con las dos RDC de la molécula de inmunoglobulina, por lo que es poco probable que se pueda unir a ésta de manera divalente.

IV. Conclusiones

En este trabajo hemos presentado modelos tridimensionales para los péptidos repetitivos FRA de la proteína calpain cisteína peptidasa putativa de *Trypanosoma cruzi*, consistentes en 1 y 2 repeticiones de 68 aa y los primeros 450 aa de la proteína. Los modelos fueron seleccionados utilizando varios criterios; por un lado, a través de análisis exhaustivos de los modelos provistos por los servidores de "fold-recognition", y por otra parte, exigiendo que dichos modelos tengan alta similitud de estructura secundaria con aquella predicha, por sus predictores específicos, en forma independiente. De esta forma, se predice para los tres péptidos estudiados, una estructura secundaria formada por dos hélices alfa para cada repetición, conectadas por un loop largo de 17 a 22 aa. Dicho loop formado mayoritariamente por residuos hidrofóbicos posee anclajes cargados. Las predicciones de antigenicidad ubican precisamente en este loop el comportamiento antigénico, lo que permite comprender la presencia de tales residuos hidrofóbicos en una zona tan expuesta al medio. Los modelos tridi-

Figura 15: Gráfica de resultados de Verify3D en función de la secuencia de aa para el modelo I-TASSER-5 de FRA-450. Pueden observarse zonas con valores negativos, las cuales es necesario refinar.

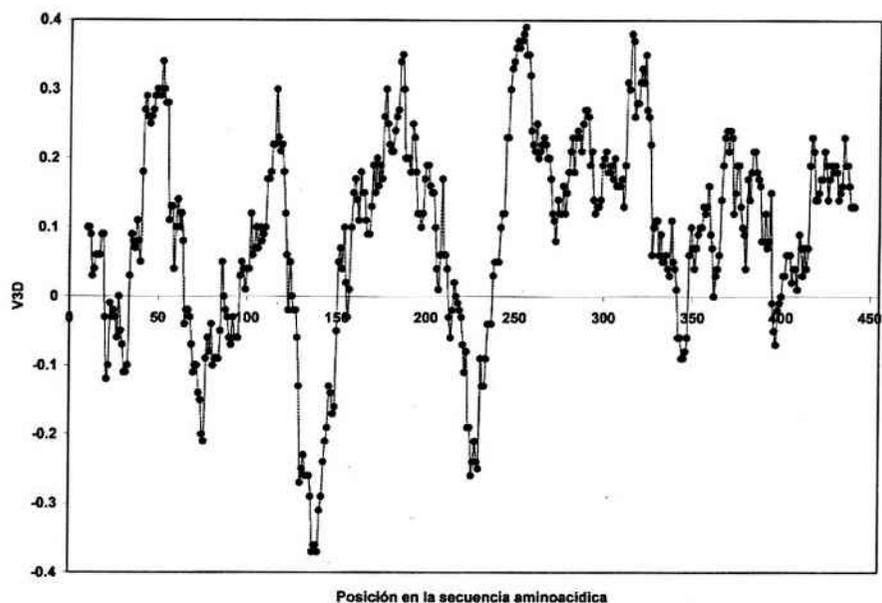
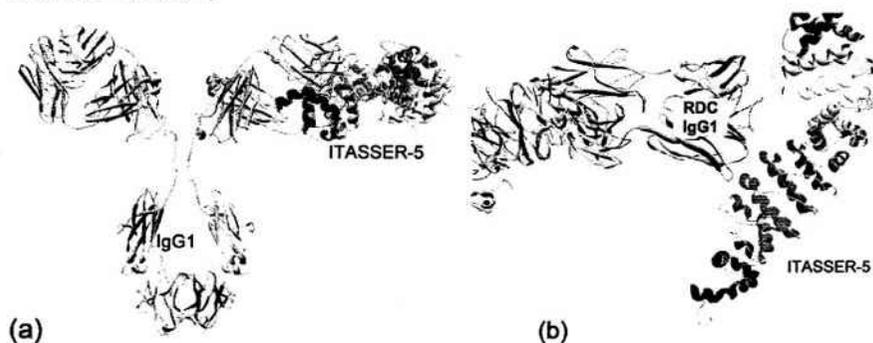


Figura 16: (a) Relación de escala entre un modelo de IgG1 y el modelo I-TASSER-5 para FRA-450. (b) Posible arreglo intermolecular del modelo de FRA-2 y el modelo de IgG1. Imagen generada con Swiss-PDB-Viewer. (47)



mensionales obtenidos para FRA-1 presentan un péptido formado por tres hélices alfa que forman una estructura compacta, mientras que el loop que une dos de ellas está expuesto al solvente. El modelo predicho para FRA-2, consiste básicamente en la repetición del modelo de FRA-1 (a menos de algunos cambios de orientación de las estructuras de alfa hélices), presentando dos loops hacia la misma zona expuesta al solvente. Al sólo efecto de efectuar una comparación de escala, se han enfrentado las estructuras modelos de FRA-1 y FRA-2 con la porción RDC de una inmunoglobulina tomada del PDB. Esta comparación sugiere que las superficies de contacto presentadas por FRA-1 y FRA-2 no tienen una diferencia de tamaño que permita predecir una interacción divalente o polivalente con el anticuerpo. Por eso, nuestros modelos no prevén encontrar una mayor antigenicidad de FRA-1 con respecto a FRA-2. Finalmente, el modelo obtenido para la FRA-450, correspondiente al número de repeticiones en la molécula nativa, sugiere una estructura repetitiva de los péptidos anteriores, pero que tampoco sería lo suficientemente grande como para actuar en forma divalente o polivalente con la Inmunoglobulina.

Por lo tanto, las conclusiones de nuestros modelos para FRA-1, FRA-2 y FRA-450 indican que, todas ellas, deberían presentar un nivel similar de antigenicidad. Esta predicción deberá ser contrastada con los resultados de la determinación experimental de la antigenicidad según el número de repeticiones, que es un tema de investigación aún abierto, no habiendo resultados concluyentes a la fecha. Las predicciones hechas en el presente trabajo pueden ser refinadas y, en particular, se implementarán simulaciones de dinámica molecular y Monte Carlo para verificar la estabilidad de estas estructuras.

Agradecimientos:

El presente trabajo fue realizado con fondos del Proyecto CAI+D 3/20/2006, de la Universidad Nacional del Litoral.

Bibliografía

1. Pizzi E, Frontali C. (2001). Low-complexity regions in *Plasmodium falciparum* proteins. *Genome Res* 11:218-229
2. da Silveira, J. F.; Umezawa, E. S.; Ostermayer Luquetti, A. (2001) Chagas disease: recombinant *Trypanosoma cruzi* antigens for serological diagnosis. *Trends in Parasitology*, 17: 286-291.
3. Krieger, M. A., E. Almeida, W. Oelemann, J. J. Lafaille, J. Borges-Pereira, H. Krieger, M. R. Carvalho, and S. Goldenberg. (1992). Use of recombinant antigens for the accurate immunodiagnosis of Chagas' disease. *Am. J. Trop. Med. Hyg.* 46:427-434.
4. Umezawa, E.S.; Bastos, S.F.; Camargo, M.E.; Yamauchi, L.M.; Santos, M.R.; Gonzales, A., Zingales, B.; Levin, M.; Sousa, O.; Rangel-Aldao, R.; Franco Da Silveira, J. (1999). Evaluation of recombinant antigens for serodiagnosis of Chagas disease in South and Central America. *Journal of Clinical Microbiology*, 37 (5): 1554-1560.
5. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein data Bank. *Nucleic Acids Res.* 28, 235-242.
6. Bujnicki, J.M (Ed.). (2004). "Practical Bioinformatics", Springer-Verlag Berlin Heidelberg
7. Ginalski K., Grishin N.V., Godzik A., Rychlewski L. (2005). Practical lessons from protein structure prediction. *Nucleic Acids Res* 33: 6. 1874-1891
8. Moult, J., Hubbard, T., Fidelis, K., and Pedersen, J.T. (1999). Critical assessment of methods of protein structure prediction (CASP):

Round III. *Proteins* **37** (S3): 2–6.

9. Fischer D, Rychlewski L, Dunbrack RL Jr, Ortiz AR, Elofsson A. (2003). CAFASP3: the third critical assessment of fully automated structure prediction methods. *Proteins*, **53**(S6):503-516.

10. Rychlewski L, Fischer D. (2005). LiveBench-8: the large-scale, continuous assessment of automated protein structure prediction. *Protein Sci.*

11. Luthy R, Bowie JU, Eisenberg D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*; **356**:83–85.

12. A Melo, F., Devos, D., Depiereux, E. and Feytmans, E. (1997) "ANOLEA: a www server to assess protein structures." *Intelligent Systems for Molecular Biology* **97**, 110-113.

13. Sippl, M.J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins* **17**, 355-362.

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

15. Marshall KL, M, Sonnhammer EL. (2002). The Pfam protein families database. *Nucleic Acids Res* **30**:276–280.

16. Jameson B.A. Wolf H. (1988). The antigenic index: a novel algorithm for predicting antigenic determinants. *CABIOS*. **4**:1.181-186.

17. Kolaskar, AS Tongaonkar, PC. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters* **276**: 172-174.

18. Rost B. (2001). Protein Secondary Structure Prediction Continues to Rise. *J. Struc, Biol.* **134**:204-218.

19. Heringa J. (2000). Computational Methods for Protein Secondary Structure Prediction Using Multiple Sequence Alignments. *Curr. Prot. Pep. Sci.* **1**:3.273-301.

20. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**:892–893.

21. Dor O., Zhou Y. (2007). Achieving 80% Ten-fold Cross-validated Accuracy for Secondary Structure Prediction by Large-scale Training. *Proteins* **66**:838–845.

22. Sanchez R, Sali A. (2000). Comparative protein structure modeling. Introduction and practical examples with MODELLER. *Methods Mol Biol* **143**:97–129

23. Wallner B., Elofsson A. (2005). Pcons5: combining consensus, structural evaluation and fold recognition scores. *Bioinformatics*. **21**:23. 4248-4254.

24. Zhang Y., Arakaki A.K., Skolnick J. (2005). TASSER: An Automated Method for the Prediction of Protein Tertiary Structures in CASP6. *Protein (S7)* 91-98.

25. Zhou H., Zhou Y. (2005). SPARKS 2 and SP3 Servers in CASP6. *Protein (S7)* 152-156.

26. Kosinski J., Cymerman I.A., Feder M., Kurowski M.A., Sasin J.M., Bujnicki J.M. (2003). A "Frankenstein's Monster" Approach to Comparative Modeling: Merging the Finest Fragments of Fold-Recognition Models and Iterative Model Refinement Aided by 3D Structure Evaluation. *Protein*, **53**: 369-379.

27. Ginalski K., Elofsson A., Fischer D., Rychlewski L. (2003). 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* **18**:8. 1015-1018.

28. Douguet D, Labesse G. (2001). Easier threading through web-based comparisons and cross-validations. *Bioinformatics* **17**:752–753

29. Octavian Teodorescu, Tamara Galor, Jaroslaw Pillardy, and Ron Elber. *Bioinformatics*, **54**:41-48 (2004). "Enriching the sequence substitution matrix by structural information", *Proteins: Structure, Function and*

30. Shi J, Blundell TL, Mizuguchi K. (2001). Fugue: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol* **310**:243–257.

31. Kelley LA, McCallum CM, Sternberg MJ.

- (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* **299**:501–522.
32. Jones DT. (1999). GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol* **287**:797–815
33. Rost B., Yachdav G., Liu J. (2004). The Predict Protein server. *Nucleic Acid Res.* **32**: 321-326.
34. Arnold A., Bordoli L., Kopp J., Schwede T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics* **22**:2. 195-201.
35. Samudrala R, Levitt M. (2002). A comprehensive analysis of 40 blind protein structure predictions. *BMC Struct Biol* 2:3
36. Rychlewski L, Jaroszewski L, Li W, Godzik A. (2000). Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci* **9**:232–241
37. Karplus K., Barrett C., Cline M., Diekhans M., Grate L., Hughey R. (1999). Predicting protein structure using only sequence information. *Proteins (S3)*:121–125
38. Fischer D. (2000). Hybrid fold recognition: combining sequence derived properties with evolutionary information. *Pac Symp Biocomput*, pp 119–130
39. Torda A.E., Procter J.B., Huber T. (2004). Wurst: a protein threading server with a structural scoring function, sequence profiles and optimized substitution matrices. *Nucleic Acid Res.* **32**: 532-535
40. Ginalski K., Pas J., Wyrwicz L.S., von Grothuss M., Bujnicki J.M., Rychlewski L. (2003). ORFeus: detection of distant homology using sequence profiles and predicted secondary structure. *Nucleic Acid Res.* **31**:13. 3804-3807.
41. Kim D.E., Chivian D., Baker D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acid Res.* **32** 526-531.
42. Sitao Wu, Yang Zhang. (2007). LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Research*, vol **35**, 3375-3382.
43. Biegert A., Mayer C., Remmert M., Söding J., Lupas A.N. (2006). The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acid Res.* **34** 335-339.
43. Daniel Fischer. 2003. 3D-SHOTGUN: A Novel, Cooperative, Fold-Recognition Meta-Predictor. *Proteins* **51**:434-441.
45. Tomii K., Akiyama Y. (2004). FORTE: a profile–profile comparison tool for protein fold recognition. *Bioinformatics* **20**:4 594-595.
46. Susana Cristobal, Adam Zemla, Daniel Fischer, Leszek Rychlewski Arne Elofsson. (2001). A study of quality measures for protein threading models. *Bioinformatics* **2**:5
47. Guex N, Peitsch MC. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**:2714–2723.
48. Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Kruger P, van Gunsteren WF. (1999). The GROMOS biomolecular simulation program package. *J Phys Chem* **103**:3596–3607.
49. Khorchid A., Mitsuhiu I. (2002). How calpain is activated by calcium. *Nature Struc. Biol.* **9**:4 239-241
50. Taylor PD, Attwood TK, Flower DR (2003) BPROPMT: a consensus server for membrane protein prediction. *Nucleic Acids Res* **31**:3698-3700
51. Marcipar I., comunicación personal.
52. Searle J.S., Pedersen J., Henry A.H., Webster D.D., and Rees A.R., (1995). "Antibody structure and function" In "Antibody engineering". Ed Borrebaeck C. Second Edition. Oxford University Press. New York.
53. Clark M. (1997). IgG Effector Mechanisms. *Chemical Immunology* **65**, 88-110.
54. Humphrey, W., Dalke, A. and Schulten, K. (1996). "VMD - Visual Molecular Dynamics", *J. Molec. Graphics.* **14**: 33-38