



Casa abierta al tiempo

UNIVERSIDAD AUTONOMA
METROPOLITANA

UNIDAD IZTAPALAPA

DIVISION DE CIENCIAS BASICAS E INGENIERIA
DEPARTAMENTO DE QUIMICA

PREDICCIÓN DEL ARREGLO TRIDIMENSIONAL DEL
PROSEGMENTO DE PAPAÑA SIN EL EMPLEO DE
HOMOLOGOS ESTRUCTURALES. POSIBLES
ESCENARIOS DE LAS FUNCIONES QUE REALIZAN
LOS PROSEGMENTOS DE PROTEASAS
SULFHIDRILICAS.

T E S I S

QUE PRESENTA:

ALBERTA JAQUELINE PADILLA ZUÑIGA

PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

JULIO DE 2001



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

PREDICCIÓN DEL ARREGLO TRIDIMENSIONAL DEL PROSEGMENTO
DE PAPAÍNA SIN EL EMPLEO DE HOMÓLOGOS ESTRUCTURALES.
POSIBLES ESCENARIOS DE LAS FUNCIONES QUE REALIZAN LOS
PROSEGMENTOS DE PROTEASAS SULFHIDRÍLICAS.

TESIS QUE PRESENTA:

ALBERTA JAQUELINE PADILLA ZÚÑIGA

PARA OBTENER EL GRADO DE:

DOCTOR EN CIENCIAS

JULIO DE 2001

DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA
DEPARTAMENTO DE QUÍMICA
UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD IZTAPALAPA

ESTE TRABAJO SE REALIZÓ EN EL ÁREA DE BIOFISICOQUÍMICA DEL DEPARTAMENTO DE QUÍMICA, BAJO LA ASESORÍA DEL DR. ARTURO ROJO DOMÍNGUEZ Y CON EL APOYO FINANCIERO TANTO DE CONACyT, A TRAVÉS DE LA BECA NÚMERO 894455 Y DEL CONVENIO 400200-5-29124E, COMO DEL CONVENIO 98-252 DE LA THIRD WORLD ACADEMY OF SCIENCES.

AGRADECIMIENTOS

Deseo expresar mi profundo agradecimiento a las personas que hicieron posible la realización de este trabajo:

Agradezco al Dr. Ruy Pérez Montfort, a la Dra. Emma Saavedra Lira y a la Mtra. María Elena Chánez Cárdenas por su valiosa ayuda al lograr la expresión del prosegmento de papaína.

Agradezco muy sinceramente al Dr. Peter Reilly, al Dr. Andrés Hernández Arana y al Dr. Robin P. Sagar por sus acertadas observaciones y comentarios al respecto de la redacción del artículo original.

Asimismo agradezco la cooperación del Dr. Alfonso Valencia, del Dr. Hannes Floeckner, del Dr. Jong Park y del Dr. Yo Matsuo por su gentileza al aceptar someter la secuencia de propapaína a sus algoritmos de cómputo.

Al Dr. Arturo Rojo Domínguez le agradezco sinceramente el haber compartido sus conocimientos y experiencia para la conceptualización, la interpretación y cauce de este trabajo de tesis.

También agradezco a mis compañeros y amigos del Área de Biofísicoquímica y de la Universidad, por sus alentadoras palabras y su ayuda incondicional.

Agradezco a mi familia por su comprensión, su apoyo, su infinita paciencia y su cariño.

"Perhaps the most remarkable features of the molecule are its complexity and its lack of symmetry. The arrangement seems to be almost totally lacking of regularities which one instinctively anticipates, and it is more complicated than has been predicted by any theory of protein structure"

John Kendrew, 1958

Opinión respecto a la estructura tridimensional de la molécula de mioglobina, determinada por propio Dr. Kendrew en ese año.

ÍNDICE

CAPÍTULO 1

ANTECEDENTES	7
PREDICCIÓN ESTRUCTURAL DE PROTEÍNAS	7
PROTEASAS Y PROSEGMENTOS	9
FUNCIÓN DE ASISTENCIA AL PLEGAMIENTO EJERCIDA POR LOS PROSEGMENTOS	11
POSIBLES IMPLICACIONES DEL ESTUDIO DE PROSEGMENTOS DE PROTEASAS	11
OBJETIVO Y MARCO DE DESARROLLO DEL TRABAJO	12

CAPÍTULO 2

RESULTADOS Y DISCUSIÓN	15
INDICIOS DE LA CONFORMACIÓN PLEGADA DEL PROPÉPTIDO DE PAPAÍNA	15
PREDICCIÓN DEL ARREGLO TRIDIMENSIONAL DEL PROSEGMENTO DE PAPAÍNA	16
BÚSQUEDA DE SECUENCIAS HOMÓLOGAS AL PROSEGMENTO DE PAPAÍNA Y ALINEAMIENTO MÚLTIPLE	18
PREDICCIÓN DE ESTRUCTURA SECUNDARIA	22
PREDICCIÓN DE LOS RESIDUOS EXPUESTOS AL SOLVENTE	25
PREDICCIÓN DE LA REGIÓN DE CONTACTO ENZIMA-PROSEGMENTO	28
ENSAMBLE TRIDIMENSIONAL DE LOS ELEMENTOS ESTRUCTURALES PROPUESTOS EN EL TRABAJO DE PREDICCIÓN	31
PURIFICACIÓN Y CARACTERIZACIÓN DEL PROSEGMENTO DE PAPAÍNA	32
ESTUDIOS ESPECTROSCÓPICOS SOBRE EL PROSEGMENTO DE PAPAÍNA	35

CAPÍTULO 3

COMPARACIÓN ENTRE LAS CARACTERÍSTICAS ESTRUCTURALES PREDICHAS PARA LOS PROSEGMENTOS DE CADENA LARGA Y LAS GEOMETRÍAS CRISTALOGRÁFICAS DE HOMÓLOGOS CERCANOS	39
CAMBIO CONFORMACIONAL DE LOS PROSEGMENTOS LIBRES DE LA PORCIÓN MADURA	45
RELACIÓN ESTRUCTURA-FUNCIÓN DE LOS PROPÉPTIDOS DE PROTEASAS SULFHIDRÍLICAS	50
ASISTENCIA AL PLEGAMIENTO	50
AUTOPROCESAMIENTO	54

CAPÍTULO 4

CONCLUSIONES Y PERSPECTIVAS _____ **56**

CONCLUSIONES 56

PERSPECTIVAS 61

CAPÍTULO 5

MATERIALES Y MÉTODOS _____ **64**

ANÁLISIS DE SECUENCIAS Y DE ESTRUCTURAS TRIDIMENSIONALES 64

SIMULACIÓN MOLECULAR 68

PURIFICACIÓN Y CARACTERIZACIÓN DEL PROSEGMENTO DE PAPAÍNA 69

ESTUDIOS ESPECTROSCÓPICOS 70

BIBLIOGRAFÍA _____ **72**

CAPÍTULO I

ANTECEDENTES

PREDICCIÓN ESTRUCTURAL DE PROTEÍNAS

Actualmente uno de los mayores retos en biología estructural es la interpretación de las reglas que gobiernan el plegamiento de las proteínas y que las conducen hacia la adquisición de una estructura única entre la inmensa cantidad de conformaciones posibles [Levinthal, 1968]. El conocimiento de las bases moleculares de estas reglas permitiría la comprensión de los elementos estructurales que confieren a las proteínas tanto estabilidad como actividad biológica e imprimiría un gran impulso al diseño de macromoléculas con funciones nuevas, desarrolladas de manera racional. Además, haría posible anticipar la función y la estructura de proteínas en los casos en que la única información disponible fuese la estructura primaria.

El hecho de que existan cerca de 15,000 proteínas con estructura tridimensional conocida a través de métodos experimentales de difracción de rayos X y RMN, ha permitido estudiar y clasificar a estas macromoléculas de acuerdo con sus características conformacionales. Sin embargo, dado que la cantidad de secuencias de aminoácidos determinada hasta ahora es al menos un orden de magnitud mayor que el número de proteínas con estructura tridimensional caracterizada, y que esta diferencia tiende a ser aún mayor, el empleo de métodos alternos para identificar la arquitectura nativa o los rasgos estructurales globales de estas macromoléculas, se vuelve cada vez más necesario.

Para abordar el estudio estructural de proteínas existen actualmente diversos métodos alternativos a las aproximaciones experimentales. En estos métodos se llevan a cabo predicciones tridimensionales que son especialmente útiles cuando se carece de una muestra física de la macromolécula de interés o cuando el manejo de la misma resulta difícil. La predicción estructural en proteínas es una labor que ha venido desarrollándose desde hace varios años con resultados alentadores [Morea, 1998; Sánchez, 1997]. En la mayoría de los casos el éxito de

una predicción depende del tipo y cantidad de información que se encuentre disponible, así como de la intuición y experiencia del investigador. En la actualidad es posible predecir con detalle atómico la geometría de una cadena polipeptídica mediante el uso de métodos de homología en secuencia con proteínas ya caracterizadas. Incluso, es factible identificar estructuras que sean acordes con una secuencia de aminoácidos dada a través de técnicas de reconocimiento de patrones de plegamiento, las cuales realizan una búsqueda sobre bases de datos de información tridimensional.

De manera alterna a la predicción basada en la comparación directa con la estructura tridimensional de proteínas, existen métodos que se fundan en observaciones generales inferidas del estudio y clasificación de las características geométricas de grandes conjuntos de proteínas. Tales métodos, que no emplean homólogos estructurales, tienen como propósito el generar un arreglo espacial de la proteína de interés a partir de predicciones de estructura secundaria y terciaria que deriven del análisis de alineamientos múltiples de secuencias emparentadas con la molécula de estudio. El trabajo de predicción en estos casos tiene mejores oportunidades de ser exitoso si se dispone de información experimental con implicaciones estructurales de la propia proteína o de algún homólogo cercano; ya que de esta forma la construcción del modelo podría contar con útiles restricciones. Otros métodos, que tampoco hacen uso de estructuras tridimensionales, consisten en búsquedas conformacionales de la geometría del estado nativo, o de diferentes etapas del proceso de plegamiento de proteínas, mediante el empleo de péptidos simplificados. Los resultados que rinden estos métodos proveen modelos tridimensionales con distinto grado de detalle que han llegado a ser de gran utilidad cognoscitiva y que incluso, podrían aportar información no accesible por métodos experimentales [Karplus, 1995; Morea, 1998].

PROTEASAS Y PROSEGMENTOS

Las proteasas son enzimas que hidrolizan enlaces peptídicos en los sistemas biológicos. Estas enzimas se encuentran en todas las células y cumplen la función de degradación de proteínas no necesarias o que hayan sufrido algún daño. Además, permiten la digestión de los alimentos, regulan la función de otras proteínas y están involucradas en el desarrollo de ciertas enfermedades. Muchas proteasas han sido ampliamente estudiadas y son actualmente de considerable interés industrial [ACE; Albrecht; Deerland Corp.]. De acuerdo con los grupos químicos que constituyen sus sitios activos se conocen cuatro clases generales de enzimas proteolíticas: metaloproteasas, proteasas serínicas, proteasas ácidas o aspárticas y proteasas sulfhidrúlicas o cisteínicas.

Prácticamente todas las proteasas se sintetizan en las células en la forma de precursores inactivos o zimógenos [Erickson, 1989; Neurath, 1989]. Estos precursores están constituidos por una cadena de aminoácidos conocida como prosegmento, propéptido o prorregión, la cual se une covalentemente al extremo amino (N) terminal de la enzima a través de un enlace peptídico. La extensión de los prosegmentos puede variar ampliamente desde menos de 30 residuos, como en el caso de la tripsina, hasta más de 200 aminoácidos, como el propéptido de la catepsina C [Jerala, 1998].

La gran diversidad de funciones que realizan los prosegmentos hace muy atractivo su estudio ya que resultan ser indispensables en los procesos de expresión y de compartimentalización celular [Vernet, 1995; Chapman, 1997], así como de inhibición y adquisición de la estructura funcional de la proteína a la que se encuentran unidos [Fusek, 1991; Winther, 1991; Baker, 1992; Fox, 1992; Hu, 1994]. Aunque esta serie de funciones está bien documentada, los mecanismos moleculares de acción de los prosegmentos todavía no se han esclarecido completamente. Se sabe que cuando se genera alguna señal en el medio celular, los precursores son procesados para dar lugar a la proteína madura y completamente funcional. Como lo ha propuesto Neurath [1989], las precauciones que toma la célula en proveer a estas enzimas de un prosegmento parecen derivar de la necesidad de regular estrechamente la actividad proteolítica de las mismas.

El mecanismo de procesamiento mediante el cual se escinde el enlace covalente que une al prosegmento con la enzima madura es todavía incierto, pues para llevar a cabo esta reacción de lisis existen informes de la intervención de moléculas del zimógeno o de la propia enzima activa (llamados procesos *cis* inter e intramolecular, respectivamente) [Vernet, 1995], y también de proteasas exógenas (proceso *trans*) [Vernet, 1995; Jerala, 1998]. La existencia del proceso de activación intramolecular se ha demostrado con el hallazgo de actividad marginal en precursores de proteasas serínicas, aspárticas, metaloproteasas, [Kassell, 1973], y recientemente también en proteasas cisteínicas [Vernet, 1991; Mason, 1992; Ishido, 1995; Doran, 1999]. Aunque aún no se ha establecido si este mecanismo de activación es significativo en el medio celular, existe evidencia de su manifestación *in vitro* [Taylor, 1995; Mach, 1994].

En las proteasas sulfhidríticas se han hecho estudios comparativos sobre las secuencias de las regiones maduras y de los prosegmentos, y se ha propuesto la existencia de dos subfamilias genéticas que se distinguen por la extensión de su cadena polipeptídica y por presentar huellas o firmas características en su secuencias [Karrer, 1993]. Las dos subfamilias comprenden por una parte a las proteínas del tipo de la papaína, con propéptidos de más de 90 residuos, y por otra a las del tipo catepsina B, que presentan longitudes cercanas a los 60 residuos en su proregión (prosegmentos de cadena larga y corta, respectivamente).

Con relación a la acción inhibitoria de los prosegmentos de proteasas sulfhidríticas se sabe que las constantes de disociación de los complejos enzima cognada-prosegmento son del orden de 10^{-7} a 10^{-9} M [Taylor, 1995]. También se ha reportado que los valores de estas constantes presentan diferencias hasta de tres órdenes de magnitud cuando se comparan los procesos de unión de un mismo propéptido con distintas proteasas homólogas [Taylor, 1995; Fox, 1992]; evidencia de que los prosegmentos son inhibidores altamente específicos de sus proteasas correspondientes.

FUNCIÓN DE ASISTENCIA AL PLEGAMIENTO EJERCIDA POR LOS PROSEGMENTOS

Quizá la función más impresionante y menos comprendida de los prosegmentos sea la de facilitar o guiar el plegamiento, e incluso inducir el replegamiento, de la enzima que se sintetiza covalentemente unida a ellos. Mientras que en algunos trabajos sobre prosegmentos de proteasas se indica que los residuos cercanos al extremo carboxilo terminal de las proregiones resultan ser los más importantes en el plegamiento de la proteína [Peters, 1998], en otros casos se afirma que son los residuos intermedios en la cadena del propéptido quienes parecen conducir el correcto plegamiento de la enzima [Vernet, 1995]. Incluso, el prosegmento ha llegado a ser prescindible en el plegamiento de una proteasa sometida a mutaciones dirigidas [Bryan, 1992]. Por otra parte, respecto al modo de operación se ha propuesto que los prosegmentos funcionan como verdaderos chaperones moleculares [Winther, 1994], aunque también se ha establecido que en realidad favorecen el plegamiento de manera específica [Baker, 1993]. Las ideas contradictorias en estos reportes ponen de manifiesto la limitada comprensión que se tiene sobre la forma en que los prosegmentos promueven el plegamiento de la enzima madura.

POSIBLES IMPLICACIONES DEL ESTUDIO DE PROSEGMENTOS DE PROTEASAS

Debido a la importancia de las funciones que realizan los prosegmentos en el medio celular, cualquier contribución a la elucidación de los detalles estructurales de sus mecanismos de acción podría ser de utilidad en el bosquejo de las rutas de adquisición de la estructura nativa de proteínas asistidas por otras moléculas. Con ello tal vez fuese posible generar elementos para replantear el dogma central del plegamiento de proteínas, el cual establece que la estructura tridimensional de estas macromoléculas se encuentra codificada completamente en su secuencia de aminoácidos [Branden, 1991].

En el campo de las aplicaciones farmacéuticas, los avances en el estudio de las interacciones entre el prosegmento y la enzima podrían ser importantes para sentar las bases del desarrollo de inhibidores específicos con alta afinidad. Esto

permitiría regular la actividad enzimática de proteasas involucradas en la metástasis tumoral y en el desarrollo de enfermedades degenerativas como artritis, resorción de hueso, osteoporosis y Alzheimer [Tao, 1994]. Por otra parte en el campo industrial resultaría de gran importancia incidir sobre el diseño de proteínas que pudieran manipularse con facilidad y que toleraran tiempos prolongados de almacenamiento; a través del control de su autólisis. Igualmente sería deseable contribuir a la creación de polipéptidos que hicieran posible revertir, al menos en parte, el proceso de desplegamiento de las enzimas proteolíticas de uso más extendido.

OBJETIVO Y MARCO DE DESARROLLO DEL TRABAJO

El interés en el estudio conformacional del propéptido de papaína surge de trabajos previos realizados en el Área de Biofísicoquímica de la UAM-Iztapalapa sobre la estructura y estabilidad de diferentes miembros de la familia de las proteasas sulfhidríticas vegetales [Solís, 1992; 1993; 1998]. Las enzimas de esta familia están conformadas por una sola cadena polipeptídica de cerca de 200 residuos, su estructura corresponde a la clasificación alfa + beta, poseen dos dominios estructurales de aproximadamente 100 aminoácidos cada uno. Su sitio activo está integrado por los residuos histidina y asparagina, además de la cisteína esencial y se sabe que sufren un proceso de desplegamiento irreversible. La proteína de esta familia que más ampliamente se ha estudiado es la papaína, a la que se considera el arquetipo de las proteasas sulfhidríticas. La papaína madura cuenta con una cadena de 212 residuos que se pliega para dar lugar a una región con un contenido helicoidal relativamente alto (dominio 1) y a otra constituida principalmente por hojas beta (dominio 2). Adicionalmente a la región madura, el zimógeno de papaína comprende un propéptido de 107 residuos que, por su extensión, es equiparable con cualquiera de los dominios de la enzima funcional.

Dada la gran diversidad de funciones que se han asociado a los segmentos de aminoácidos previos a la porción madura de las distintas proteasas, el estudio de la interacción de la papaína con su propéptido resultaba de gran interés en nuestro grupo para esbozar interpretaciones estructurales de los mecanismos de

acción del prosegmento que pudieran generalizarse a otros miembros de la familia de las proteasas cisteínicas.

Un hecho trascendente al iniciar el estudio estructural del propéptido de papaína era que no se contaba con una muestra biológica que permitiera la caracterización de la estructura tridimensional por métodos experimentales. Sin embargo, se conocía la existencia de diversos métodos y bases de datos de proteínas, así como de algunas estrategias para la asignación de estructura secundaria y terciaria a partir del análisis de secuencias. De esta forma, el objetivo general del trabajo de tesis fue la predicción del arreglo tridimensional del prosegmento de papaína.

Una situación favorable que rodeaba a este trabajo era el anuncio de la liberación próxima de las coordenadas cristalográficas de un homólogo en secuencia al zimógeno de papaína, la procatepsina L humana, cuyo prosegmento comparte con el propéptido de papaína un 37% de identidad en estructura primaria. Este hecho establecía un marco semejante a las llamadas "pruebas ciegas" que resultan ser invaluable en la determinación de la confiabilidad de un método de predicción estructural, puesto que permiten estimar en corto plazo los aciertos y las sobreestimaciones en las decisiones tomadas durante el desarrollo del trabajo.

Cuando se tenían algunos resultados preliminares del modelado, fue posible obtener del laboratorio del Dr. Ruy Pérez Montfort (Instituto de Fisiología Celular, UNAM) una muestra recombinante del prosegmento de papaína. Esto abrió la posibilidad de realizar estudios espectroscópicos del prosegmento que fueran de utilidad para validar algunas de las premisas generadas durante el trabajo de predicción. Los resultados de estos experimentos se incorporaron al presente reporte escrito.

Antes de la liberación de la estructura tridimensional de procatepsina L fue publicada la estructura cristalográfica de un homólogo aún más cercano al precursor de papaína, la procaricaína [Groves, 1996], cuyo prosegmento mantiene un porcentaje de identidad en secuencia del 73% con el propéptido de la proteína bajo estudio. La relevancia de este hecho era que se tendría acceso a la

información tridimensional de dos proteínas pertenecientes a la subfamilia genética de la papaína. Todos estos elementos hicieron posible el complementar el trabajo teórico con un estudio comparativo entre las estructuras tridimensionales experimentales y las características geométricas propuestas para los prosegmentos de cadena larga. También de esta manera surgió la posibilidad de generar propuestas moleculares para explicar, al menos en parte, algunas de las funciones de estos prosegmentos.

Con la información reunida en las etapas de análisis teórico, de trabajo experimental y de revisión estructural de proteínas emparentadas, el esquema de la tesis se dividió en tres partes: la predicción del arreglo tridimensional del prosegmento de papaína, la comparación entre el modelo generado y las estructuras cristalográficas de los homólogos cercanos, y finalmente, la interpretación de la información originada en las dos etapas anteriores para proponer algunos detalles sobre las funciones que realizan los prosegmentos y precursores de las proteasas sulfhidríticas

CAPÍTULO 2

RESULTADOS Y DISCUSIÓN

INDICIOS DE LA CONFORMACIÓN PLEGADA DEL PROPÉPTIDO DE PAPAÍNA

Una de las inquietudes surgidas al inicio del trabajo de tesis era saber si el propéptido bajo estudio tendría realmente una conformación plegada. Cuando se revisaron algunas secuencias homólogas al prosegmento de papaína, se encontraron dos proteínas de ratón (con códigos CT2A_MOUSE y CT2B_MOUSE) las cuales se sintetizan en las células sin la zona correspondiente a la región enzimática de las proteasas sulfhidríticas. Ambas proteínas son potentes inhibidores de papaína [Delaria, 1994; Yamamoto, 1999] y, de acuerdo con un análisis de huellas funcionales sobre sus secuencias, poseen un sitio potencial de reconocimiento de nucleótidos. Estos hechos parecen indicar que las cadenas de tales proteínas consiguen un arreglo espacial que las habilita para realizar funciones biológicas, lo cual constituye una evidencia indirecta de que el propéptido podría tener una estructura terciaria particular. Por otro lado, se identificaron dos secuencias de homólogos a las enzimas proteolíticas de estudio, codificadas como: TES1_RAT y P34_SOYBN, donde la cisteína del sitio activo ha mutado e incluso donde, según reportes experimentales, la porción madura de TES1_RAT ha perdido la función proteolítica y su propéptido carece ya de propiedades inhibitorias [Cheng, 1993]. Dado que se sabe que estas dos proteínas existen como especies de 37 y 34 kDa, respectivamente [Cheng, 1993; Kalinski, 1992], es de suponer que poseen tres dominios funcionales en su molécula y que la prorregión se ha convertido en una parte integral del sistema molecular. Adicionalmente, de acuerdo con el reporte cristalográfico de Gallagher *et al.* [1995], se sabe que la región correspondiente al prosegmento de subtilisina posee estructura globular.

Una de las evidencias experimentales más directamente relacionadas con la conformación del prosegmento de papaína es el reporte de Taylor *et al.* [1995], quienes encontraron un comportamiento cooperativo de la cadena del propéptido

durante su desplegamiento térmico. Por otra parte, un análisis de la secuencia del prosegmento realizado bajo el criterio de globularidad, desarrollado por Rost [1998], indicó una alta probabilidad de que el polipéptido libre de la enzima madura adquiriera una conformación semejante a la de un ovillo.

PREDICCIÓN DEL ARREGLO TRIDIMENSIONAL DEL PROSEGMENTO DE PAPAÍNA

Uno de los métodos que ha rendido mejores resultados en la predicción de la geometría de proteínas es el conocido como modelado por homología o por comparación. El fundamento de este método es que existe una clara relación entre la identidad en secuencia y la similitud estructural de las proteínas. De hecho, se ha visto que proteínas con una identidad en secuencia mayor al 50% poseen una diferencia promedio entre los átomos de sus cadenas superpuestas menor o igual a 1 Å [Morea, 1998]. De esta forma, mientras la identidad en secuencia sea suficientemente alta, una proteína de estructura conocida podría funcionar como buen templete o molde tridimensional para construir la molécula de interés [Mosimann, 1995]. Normalmente se considera que el límite inferior de identidad en secuencia para realizar un modelado por homología confiable es del 25% [von Rague-Schleyer, 1998].

En el caso que nos ocupa, el primer intento para asignar una estructura tridimensional al prosegmento de papaína fue el emprender un modelado por homología. Para ello se realizaron búsquedas masivas por identidad y similitud en secuencia sobre las estructuras depositadas hasta ese momento en el Banco de Datos de Proteínas, PDB [Bernstein, 1977]. Después de probar muchas combinaciones en los datos de entrada, que consistieron incluso en utilizar fracciones de la cadena polipeptídica del prosegmento, se obtuvieron como resultado diversas proteínas o segmentos de ellas que al ser comparadas entre sí mostraron elementos estructurales incompatibles, lo cual impidió la identificación de una geometría útil para representar al propéptido bajo estudio.

El siguiente intento para localizar una estructura tridimensional que sirviera como punto de partida para el modelado consistió en la identificación de patrones

de plegamiento compatibles con el prosegmento de papaína. Estos métodos tienen la ventaja de no requerir identidad en secuencia detectable entre proteínas, y se basan en la idea de que la gran mayoría de las estructuras de proteína que se conocen actualmente pueden ser agrupadas en un número limitado de arreglos geométricos característicos [Zvelebil, 1987], por lo cual, las nuevas proteínas secuenciadas muy probablemente poseerán patrones de plegamiento reportados con antelación. Al someter la secuencia del prosegmento de papaína a diversos programas de reconocimiento de patrones de plegamiento, solamente se identificaron pequeños fragmentos moleculares aislados con puntuación de baja significancia estadística, y nuevamente hubo casos donde al emplear programas independientes para analizar la misma zona de la cadena se obtuvieron predicciones contradictorias de estructura secundaria. Dos interpretaciones posibles de tales resultados son que los algoritmos empleados no poseían la sensibilidad necesaria para la localización de una geometría congruente con la secuencia de prueba, o bien, que no existía una estructura afín al prosegmento en las bases de datos correspondientes. Esto último podría deberse a su vez a que los criterios para seleccionar las proteínas en tales bases de datos soslayaban algunas estructuras tridimensionales, o quizá también, a que el patrón de plegamiento del prosegmento de papaína era inédito en el universo de proteínas caracterizadas.

Debido a la falta de una estructura tridimensional que pudiera emplearse como punto de partida en la asignación de las características conformacionales del propéptido de papaína, el modelado por homología y la asignación de un patrón de plegamiento resultaban inaplicables. De esta forma, la única ruta viable era la predicción estructural a partir de principios de construcción generales de proteínas, donde se hace uso tanto de la información experimental disponible como de aquella contenida en las secuencias de los prosegmentos homólogos a la proteína bajo estudio.

Un camino alternativo para alcanzar el objetivo propuesto en este trabajo era el considerar a la estructura tridimensional del propéptido de catepsina B de rata como templete adecuado para el prosegmento de papaína. De hecho, los detalles

atómicos de esta estructura eran accesibles a través del PDB [Cygler, 1996]. Sin embargo, al comparar la secuencia del prosegmento de papaína con la correspondiente a catepsina B resultó evidente el bajo porcentaje de identidad existente entre ellas (menor al 18%); la gran diferencia en longitud de las cadenas polipeptídicas (62 y 107 residuos en los prosegmentos de catepsina B y papaína, respectivamente); y la presencia de distintas huellas de aminoácidos conservados en los alineamientos múltiples reportados por Karrer *et al.* [1993]. Todas estas divergencias derivadas del hecho de que los precursores de papaína y catepsina B pertenecen a subfamilias genéticas diferentes. Adicionalmente, a partir de la inspección visual de la estructura cristalográfica del zimógeno de catepsina B, no es imaginable una conformación globular para su prosegmento libre.

Las etapas de desarrollo del trabajo de tesis durante la predicción del arreglo tridimensional del prosegmento de papaína fueron: a) búsqueda de las secuencias de homólogos al propéptido bajo estudio, b) construcción de un alineamiento múltiple de esas secuencias homóloga, c) predicción de las estructuras secundarias del propéptido de papaína, d) predicción de las posiciones expuestas al solvente y de la región de contacto con la enzima, e) ensamble tridimensional de la prorregión, y finalmente, f) estudios espectroscópicos sobre el prosegmento de papaína.

BÚSQUEDA DE SECUENCIAS HOMÓLOGAS AL PROSEGMENTO DE PAPAÍNA Y ALINEAMIENTO MÚLTIPLE

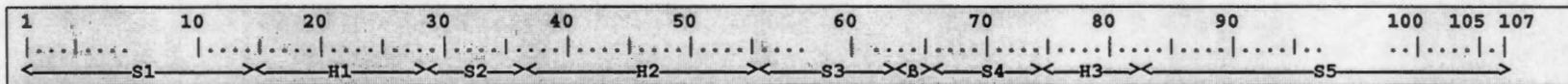
Cuando se llevó a cabo la búsqueda automática de proteínas similares al zimógeno de papaína se obtuvieron cerca de 500 secuencias distintas. Al realizar una revisión de estos resultados fue clara la presencia de representantes de las dos subfamilias genéticas, donde la mayor parte de las proteínas había sido localizada por su similitud con la región madura del zimógeno de papaína. Incluso algunas de las secuencias carecían del segmento correspondiente a la prorregión. Para conseguir resultados más concretos, en un siguiente intento se sometió al programa de búsqueda únicamente la porción de la cadena polipeptídica que corresponde al prosegmento de papaína (107 residuos). En esta ocasión se

encontraron un poco más de 150 secuencias, las cuales fueron revisadas en su significancia estadística y en su función para seleccionar los homólogos más cercanos al propéptido bajo estudio. El resultado de esta selección comprendió 141 secuencias que pertenecían en su totalidad a la subfamilia de prosegmentos largos; lo cual fue muy alentador, porque si bien debía enfrentarse un tipo de predicción estructural sin templete tridimensional, se tenía por otro lado una alta posibilidad de localizar las zonas de importancia estructural dado el gran número de proteínas homólogas disponibles.

Para continuar con el estudio de la información obtenida mediante la búsqueda de homólogos al propéptido de papaína, las 141 secuencias íntegras se sometieron al programa de alineamiento múltiple ClustalW [Thompson, 1994]. El alineamiento automático obtenido mostró desajustes en las zonas de mayor mutabilidad que dificultaban su interpretación. Con el objeto de refinar el acomodo de aminoácidos, las cadenas fueron analizadas por subconjuntos de proteínas emparentadas, los cuales permitieron identificar más fácilmente las regiones conservadas en cada posición de las secuencias. La información derivada de estos análisis se empleó como guía para realizar un ajuste manual al alineamiento de los 141 prosegmentos. En la figura 1 se muestra una parte representativa del alineamiento final, en el que sólo se han incluido las secuencias correspondientes a algunos de los prosegmentos considerados.

A pesar de que en el alineamiento múltiple existe una gran diversidad filogenética, que se extiende desde organismos unicelulares hasta mamíferos, en la figura 1 resultan evidentes varias posiciones que mantienen invariantes sus propiedades físicas y químicas (polaridad, tamaño, carga, hidrofobicidad). Tales posiciones conforman tres huellas de aminoácidos que poseen más de un 70% de conservación en las 141 secuencias analizadas. Sobre estas huellas se encuentran indicados en color gris los sitios ocupados por residuos hidrofóbicos según se muestra en la figura 1. La primera huella contiene tres sitios marcados y dos de ellos incluyen residuos aromáticos. La segunda huella contiene la firma reportada por Karrer *et al.* [1993], ERFNIN (en código de una letra), que es característica de los prosegmentos de cadena larga y que comprende en el

a)



PAP: CARPA DFSIVGYSQ----NDLTSTERLIQLFESWMLKHNKIYKNIDEKIYRFEIFKDNLKYIDETNKK---N-NSYWLGLNVFADMSNDEFKEKYTGSIAGNYTTTTELSY----EEVLNDGDVN----
 PAP3 CARPA DFSIVGYSQ----DDLTSIERLIQLFNSWMLNHNKIFYENVDEKLYRFEIFKDNLNYIDETNKK---N-NSYWLGLNEFADLSNDEFNEKYVGLIDATIEQSYD----EEFINEDTVN----
 gp|X97789| DFYTVGYSQ----DDLTSIERLIQLFDSWMLKHNKIYESIDEKIYRFEIFRDNLMYIDETNKK---N-NSYWLGLNGFADLSNDEFKFKYVGFVAEDFTGLEHFD----NEDFTYKHVTN----
 gp|X69877| DFSIVGYSQ----DDLTSIERLIQLFNSWMLNHNKIFYENVDEKLYRFEIFKDNLNYIDETNKK---N-NSYRLGLNEFADLSNDEFNEKYVGLIDATIEQSYD----EEFINEDIVN----
 gp|X78056| DFSIVGYSQ----DDLTSIERLIQLFNSWMLKHNKIYKNIDEKLYRFEIFKDNLKYIDERNKM---I-NGYWLGLNEFSDLSNDEFKEKYVGLSPEDYTNQPYD----EEFNVEDIVD----
 gp|U41902| DFSIISYDS----QDLIGDDAIMELYELWLAHQKAYNGLEKQKFFSVFKDNFLYIHQHNQO---N-PSYKGLGLNFADLSHEEFKAAAYLGTKLD----AKRLRSRSPRYQYVSGED----
 gp|U25970| AMSI-----INYSENEVMDYEEWLKHKRKYVNGLEKQKFFSVFKDNFLYIHQHNQO---N-NTYTLGLNKFADITNKEKYRAMYLGTTRDAKR---RVMK---TQNTGHRAYVNSGDQ---
 gp|X75749| DIQLP-----TGRSNDEVTMYEELVKHQKQVYNGLEKQKFFSVFKDNFLYIHQHNQO---N-YTYIVGLNKFADITNKEKYRAMYLGTTRDAKR---RVMK---TQNTGHRAYVNSGDQ---
 gp|X66061| DMCIIISYDKTHPKSTPRTNDQVLTMYEELVKHQKQVYNGLEKQKFFSVFKDNFLYIHQHNQO---N-LSFRLGLNRFADLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 gp|U44947| DMSS-----GRSNKEVMTMYEELVKHQKQVYNGLEKQKFFSVFKDNFLYIHQHNQO---N-HSYRVGLNRFADLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 ORYA_ORYSA DMSIVSYGERS-----EEEARLYAEWKAHEGKSYNAVGEERRYAAFRDNLRYIDEHNAADAGVHSFRLGLNRFADLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 ACTN_ACTCH AFNAKN-----LTQRTNDEVKAMYESWLIKYGKSYNSLGEWERREYRFEIFKELTRFIDEHNAAD---N-RSYKVLNRFADLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 CYS3 PHAVU VANSDFDHD-----KDLASEESLWDLYERWRSHHTVSRD---LGEKHKRFRVFKANLMHVHTNKM---D-KPYKLGKLNKFADITNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 gp|U32430| GLVAVPTDG-----ALEDSLMMERFHWMAKHKGSYAGVEEKLRRDFIFRNVFEIEAANRD---GRLSYTLGVNQFADLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 gp|U13940| LAAARSLQD-----ASMRHEEEMASYGKRVYKADINEKQKRYKIFEENVALIESNND---ANKPYKLVNQFADLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 gp|U37336| ITLSRPLDN-----ELIMQKRHIEWMTKHGRVYADVKEENRVRVFKNNVERIEHLNSIP---APRTFKLVNQFADLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 P34 SOYBN HRSILDLDL-----TKFTTQKQVSSLFQWKSEHGRVYHNHEEAKRLEIFKNNSNYIRDMMANRKS---HSHRLGLNKFADITNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 RD19 ARATH DLVIRQVVG-----GAEPQVLTSEDHFLFKRKFQKGYKADNEEHYRFSVFKANLRRARRHQL---D-PSATHGVTOFSDLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 gp|U45402| DPLIRQVVP-----GDDNDLELNAESHFLFKRKFQKGYKADNEEHYRFSVFKANLRRARRHQL---D-PSATHGVTOFSDLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 CYS3 HEMSP IAQSIPTFE-----KDLASEDSLWNLYEKWRTHHTVARD---LDEKRRFRVFKENVKFIEHFNQK---D-APYKLVNQFADLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 gp|U34747| AATAIDID-----KDLETEDSLWNLYEKWRTHHTVARD---LDEKRRFRVFKENVKFIEHFNQK---D-APYKLVNQFADLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 gp|U232795| TTVOIARLKLKLG---DNELLRTEKFKVFMENYGRSYSTEEYLRLRGIFAQNMVRAAEHQALD---P-TAVHGVTOFSDLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 CYS1 HORVU AAVELCSAIPMED---KDLESEALWDLYERWQSAHRVRRH---HAEKHRRFRTFKSNAHFHSHNKR---D-HPYRLHLNRFQDMQAEFRATFVGDRLRRDTP---A-KPPS---VPGFMYAALNVSD---
 CATH RAT ELTVNAIEK-----FHFTSWMKQKQKYS---SREYSHRLQVFNANWRKIQAHNRN---HTFKMGLNQFSDMSFAEIKHKYLWSEPN---CSATKSNYLRGTGPY-----
 gp|U07374| AVAQLHKDP-----TLDHHLWKKTHRRKQYNNKVDIISRLIWEKNLKYISIHNLASLGVHTYELAMNHLGDMTSEEVVQKMTGLKVP-----LSHRSNDTLYIPEWEGR-----
 CATK HUMAN SFALYPEE-----ILDTHWELWKKTHRRKQYNNKVDIISRLIWEKNLKYISIHNLASLGVHTYELAMNHLGDMTSEEVVQKMTGLKVP-----LSHRSNDTLYIPEWEGR-----
 CATL HUMAN SATLTFDHD-----SLEAQWTKWKAMHNLXG---MNEEGWRRAVWEKMMKIELHNQYREGKHSFTMAMNAGDMTSEEVVQKMTGLKVP-----LSHRSNDTLYIPEWEGR-----
 CT2B MOUSE SAAPSPDP-----SLDNEWKEWTKTFKAYS---LDEERHRLVWEENKKEIEAHNADYERKTSFYMGLNQFSDLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 CATK_RABIT SFALHPEE-----ILDTHWELWKKTHRRKQYNNKVDIISRLIWEKNLKYISIHNLASLGVHTYELAMNHLGDMTSEEVVQKMTGLKVP-----LSHRSNDTLYIPEWEGR-----
 TES1 RAT DSTAPTPDP-----SLDVEWNEWRTKHGKTYN---MNEERLKRRAVWEKMMKIELHNQYREGKHSFTMAMNAGDMTSEEVVQKMTGLKVP-----LSHRSNDTLYIPEWEGR-----
 CT2A MOUSE SAAPSPDP-----SLDNEWKEWTKTFKAYS---LDEERHRLVWEENKKEIEAHNADYERKTSFYMGLNQFSDLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 gp|L33772| VFASND-----LWHQWKRIYNKEYK---GADDDHRRNIWEQNVKHIQEHNLRLHDLGLVTKLGLNQFSDLTNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 LCPB LEIME ACAPARAIH-----VGTAAALFEFKRTYGRAYETLAEQORLANFERNLELMREHQR---N-PHAQFGITKFDLSEAEFAARYLNGAAY---FAAAKRHAHQHYRKRADLSA---
 gp|X77219| MFAFLSAGTS-----EKAFLAYMRETGNFTT---GDEYHFRGLYLANKRLVQEHNAAN---K-GFKLGLNLAHLTQSEYRSLGAKRLG---QKSNFFKCDAP-----
 gp|X77200| ATASSKLFQHE-----EKAFLDWMRSTNNMFV---GDEYHFRGLYLANKRLVQEHNAAN---K-GFKLGLNLAHLTQSEYRSLGAKRLG---QKSNFFKCDAP-----
 gp|X77218| MMQAHE-----QKSFGLWMTRETGNFTT---GDEYHFRGLYLANKRLVQEHNAAN---K-GFKLGLNLAHLTQSEYRSLGAKRLG---QKSNFFKCDAP-----
 gp|X80989| LAAANP-----SWEFFKGFGRKQYVDAEEDSYRRVIFEQNKYIEEFNKKYENGEVTFNLAMNKFQDMTLEEFNAVMKGNIPR-----RSAPVSVFYPKKETGP-----
 CYS2 HOMAM LAAASP-----SWEFFKGFGRKQYVDAEEDSYRRVIFEQNKYIEEFNKKYENGEVTFNLAMNKFQDMTLEEFNAVMKGNIPR-----RSAPVSVFYPKKETGP-----
 CYS3 HOMAM LATASP-----SWDFKTOYGRKYGDAAEELRYQRVFOQNEQLIEDFNKFFENGEVTFNVAMNKFQDMTLEEFNAVMKGNIPR-----RSAPVSVFYPKKETGP-----
 gp|L36204| LLVSYASAK-----QQFS---ELQYRNAFTNMQARQRTYS---SEEFNARYQIFKSNMDDYVHOWNKSG---GETVLGLNVFADITNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 gp|L36205| LLVSVATAK-----QQFS---ELQYRNAFTNMQARQRTYS---SEEFNARYQIFKSNMDDYVHOWNKSG---GETVLGLNVFADITNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 CYS1 DICDI FVSSRGIP-----EEQSQFLEFQDKFNKYS---HEEYLERFEIFKSNMDDYVHOWNKSG---GETVLGLNVFADITNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 CYS2 DICDI FSANVRPNG-----RRFS---ESQYRTAFTEWTLKFNQYS---SEEFNARYQIFKSNMDDYVHOWNKSG---GETVLGLNVFADITNKEEYRTRFLGTTRINPNR---RNRK---VNSQTNRYATRVGDK---
 CATV_NPVAC NSAAIDLLK-----APNYAEFVWLFHFNKDYGESEKLRRFIFQHNLNEIINKNQ---D-S-AYEINKFSDLSKDETIKYTGLSLP-----IQTQNFCKVIVLDQPPGK-----

b)

HUELLA 1

17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
*	LAVY	x	*	x	FWY	*	*	WFY	KMLR	x	*	HYF	GNQT	KRV	x	YFR
81%	75%		75%		98%	76%	76%	98%	84%		91%	91%	88%	93%		85%

HUELLA 2

38 E	39	40	41	42 R	43	44	45	46	47	48	49 N	50	51	52	53	54	55	56	57 N	58	59
ED	KEQH	x	x	RK	FLY	x	VI	FWY	*	*	N	LVMY	*	FKRH	IVA	*	x	x	NQA	x	*
97%	75%			98%	76%		84%	98%	82%	79%	93%	75%	74%	75%	93%	55%			96%		78%

HUELLA 3

64	65	66 G	67	68 N	69	70	71	72 D	73	74	75	76	77	78	79	80	81	82	
YFA	*	LFMV	GA	VILM	NT	*	FL	ASG	DA	IMI	ST	x	carga	ED	FY	RKN	x	x	YMF
90%	85%	79%	76%	96%	97%	83%	97%	93%	97%	93%	95%		76%	99%	82%	75%			81%

FIGURA 1. Alineamiento múltiple de secuencias homólogas al prosegmento de papaína. Las secuencias presentadas fueron obtenidas de las bases de datos Swiss-Prot y GenBank (gp). **a)** Se muestran 44 secuencias representativas de las 141 empleadas en este trabajo. La numeración de los residuos tiene como referencia la cadena de la proregión de papaína (PAPA_CARPA). Las primeras 23 secuencias corresponden a plantas y de ellas las primeras cinco a proteasas de papaya (incluyendo caricaína; PAP3_CARPA). Las siguientes secuencias provienen de mamíferos (8), protozoarios (5), otros animales (3), hongos (4) y virus (1). Se resaltan los códigos de identificación de las proteínas mencionadas en el texto: P34_SOYBN, CATL_HUMAN, TES1_RAT y CT2A_MOUSE. Con negritas se destacan las posiciones conservadas a lo largo del alineamiento múltiple. Se indica la posición de las estructuras secundarias y los segmentos que las bordean (**S1 a S5**). **b)** Huellas o patrones conservados en las secuencias. Cada cuadro representa una posición del alineamiento en donde se anota la frecuencia de aparición del conjunto de aminoácidos mostrado. Los residuos polares se señalan con (*), las posiciones sin conservación con (x) y los residuos hidrofóbicos conservados se resaltan en gris. También se señalan las posiciones de las firmas EF1FNIN y GNFD en las huellas 2 y 3, respectivamente.

presente análisis a los residuos 38 a 59 del prosegmento de papaína (numeración de referencia en todo este trabajo). Asimismo, fue posible encontrar otra huella que incluye los aminoácidos 63 a 82 y que contiene el arreglo GNFD, el cual parece estar ausente en los propéptidos del tipo catepsina B [Vernet, 1995]. En esas zonas existen posiciones que resultan estar ocupadas por un restringido número y tipo de residuos; tal es el caso de los aminoácidos: 22, 25, 28, 31, 38, 42, 46, 49, 53, 54, 57, 63, 67, 68, 70, 71, 72, 73, 74 y 77 que igualan o rebasan el 90% de conservación en el alineamiento de los 141 prosegmentos y donde se mantienen fundamentalmente las características hidrofóbicas o polares. Dado que se trata de un estudio sobre secuencias homólogas pertenecientes a especies muy diferentes, es altamente probable que los residuos conservados, aún en las proteínas más distantes, desempeñen papeles importantes en la función o en la estabilidad de los prosegmentos analizados [Benner, 1995]. De manera complementaria, en el propio alineamiento múltiple son evidentes las regiones de alta variabilidad que bordean a las huellas anteriormente mencionadas.

PREDICCIÓN DE ESTRUCTURA SECUNDARIA

Cuando se intenta asignar características tridimensionales a una proteína, es útil contar con la ubicación tentativa de los elementos de estructura secundaria sobre la cadena polipeptídica. En este trabajo se emplearon los resultados generados por diversos algoritmos de predicción de estructura secundaria que permiten asignar un tipo de conformación específico (hélice alfa, cadena beta, estructura irregular y en algunos casos giro) a cada uno de los aminoácidos de una proteína, ya sea individualmente o dentro de un alineamiento múltiple de secuencias. Los resultados de la predicción sobre el prosegmento de papaína se muestran en la figura 2a donde, a pesar de la variabilidad en los atributos de conformación secundaria, es posible reconocer zonas amplias con predominante tendencia helicoidal y regiones más localizadas que son propensas a poseer conformación beta.

Un problema común en la predicción de estructuras secundarias es la determinación precisa de los extremos inicial y final de los elementos individuales

involucrados. A fin de establecer límites de confiabilidad ponderable Benner *et al.* [1991] desarrollaron un algoritmo que se funda en el análisis de alineamientos múltiples de grupos de secuencias homólogas. La idea básica de este método es que la conservación de una posición en las secuencias es más valiosa en información cuanto más divergentes sean las proteínas estudiadas; mientras que las diferencias locales son más significativas cuando éstas se verifican sobre polipéptidos semejantes. Así, dado que los prosegmentos de proteasas sulfhidrúlicas considerados en este trabajo comparten una función biológica, es de esperar que en general, los elementos de estructura secundaria se mantengan conservados aún entre las proteínas más alejadas entre sí, y que los límites de estos elementos puedan ser determinados a partir de las regiones de mayor variabilidad o de mayor número de inserciones en los alineamientos múltiples. Las inserciones en un alineamiento de homólogos indican generalmente la presencia de giros o asas estructurales, las cuales desfavorecen la propagación de los elementos de estructura secundaria. De acuerdo con Benner [1995], la existencia de residuos de prolina o glicina adyacentes a una posición variable en un alineamiento múltiple de homólogos cercanos sugiere una interrupción de los elementos de estructura secundaria. Estas ideas y la tendencia reportada para ciertos aminoácidos a encontrarse en los extremos de elementos helicoidales [Parker, 1997; Aurora, 1998] fueron empleadas en el presente trabajo para deslindar las zonas de estructura secundaria predichas. Durante esta etapa de la predicción se emplearon nuevamente los subconjuntos de proteínas homólogas mencionados en la construcción del alineamiento múltiple y los resultados obtenidos se presentan en la figura 2a. Puede observarse que el uso de la información que delimita a las estructuras secundarias permite identificar, sobre el propéptido de papaína, tres zonas helicoidales y una región beta denotadas por H1, H2, H3 y β , respectivamente. Las señales de terminación débiles que aparecen sobre la segunda hélice del propéptido podrían sugerir una ligera flexión en la parte central de este elemento. Los segmentos que conectan a las estructuras secundarias predichas también se muestran en la figura 2a y se representan con los símbolos S1 a S5.

PREDICCIÓN DE LOS RESIDUOS EXPUESTOS AL SOLVENTE

Para estimar la localización espacial de los residuos de una proteína pueden generarse predicciones de la exposición al solvente de las cadenas de estas macromoléculas. Ello se consigue con el uso de algoritmos disponibles a través de servidores remotos conectados en red, o bien con la utilización de métodos de revisión visual sobre conjuntos de proteínas homólogas alineadas.

En este trabajo se utilizaron dos algoritmos independientes para predecir la exposición al solvente de la cadena lateral de cada aminoácido: PHDacc [Sander, 1991] y el método de Benner *et al.* [1994]. El primero de éstos realiza su predicción con un proceso automático basado en redes neuronales mientras que el segundo propone un análisis manual fundado en la identificación de zonas conservadas o de alta variabilidad de residuos; pero la estrategia en ambos casos es el análisis de conjuntos de secuencias homólogas. La información obtenida con estos métodos no sólo representa una valiosa herramienta para predecir características tridimensionales de las proteínas de interés, sino que además, anticipa la hidrofobicidad de cada sitio en la cadena polipeptídica. La asignación de exposición al solvente para las cadenas laterales del prosegmento de papaína se presenta en la figura 2b, donde es posible observar la sorprendente semejanza entre los resultados de ambos métodos.

Al analizar la disposición de los residuos predichos como ocultos al solvente parece razonable esperar, de acuerdo con las observaciones de Lim [1974], que los elementos de estructura secundaria correspondientes a estas zonas posean naturaleza anfipática, por lo que probablemente se encuentren localizados en la superficie de la proteína. Además, dada la coincidencia de los sitios ocultos con las posiciones hidrofóbicas conservadas en el alineamiento múltiple (señaladas con una barra vertical en la figura 2b), parece viable la existencia de un núcleo hidrofóbico común a los elementos de estructura secundaria. Este núcleo podría estar conformado por los residuos Leu18, Phe22, Trp25 del primer segmento helicoidal; Phe43, Ile45, Phe46, Leu50, Ile53 de la segunda hélice; Tyr63, Leu65 del segmento beta; y finalmente Leu67, Phe70, Met73, Phe78 y Tyr82 que

pertenecen al segmento S4 y a la tercera región helicoidal propuesta (ver figura 1b).

Es importante señalar que además de las asignaciones de exposición al solvente que reportan los programas mencionados en párrafos precedentes, es posible identificar, por inspección de las diferentes secuencias de aminoácidos, algunos sitios de glicosilación potenciales. En las proteínas caracterizadas tridimensionalmente, estos sitios de glicosilación se localizan siempre en la superficie de las moléculas y de ahí resulta su importancia para la asignación de regiones expuestas en el prosegmento de papaína. Las huellas específicas para la unión de los carbohidratos a las proteínas son ampliamente conocidas [Bairoch, 1993]. Una de estas huellas puede describirse con los símbolos en código de una letra; N-{P}-[ST] que indica la presencia en forma consecutiva de una asparagina, un aminoácido diferente a prolina y una serina o treonina [Hofmann, 1999]. Las posiciones de glicosilación para la proteína de estudio en este trabajo se detectaron experimentalmente sobre los aminoácidos 60 y 89 del prosegmento. La posición 54, si bien corresponde a un aspartato en la secuencia de la prorregión de papaína, ha sido reportada como asparagina por Ramjee *et al.* [1996], lo que genera un sitio de glicosilación potencial sobre el propéptido y a ello se debe su inclusión en la predicción como sitio expuesto al solvente. Un hecho interesante es que a lo largo de las 141 secuencias del alineamiento múltiple también fue posible detectar varios sitios de glicosilación potenciales, y éstos se localizaron principalmente en la región que corresponde al segmento S5. Del análisis que dio lugar a la figura 3, donde se presentan las posiciones de mayor variabilidad sobre las secuencias en el alineamiento múltiple de prosegmentos, fue patente la alta incidencia de inserciones y pérdidas de aminoácidos en los segmentos S1, S3 y S5. Dado que se conoce que los mayores cambios entre secuencias homólogas se presentan fundamentalmente en regiones que no corresponden a estructuras secundarias, pero que las conectan entre sí, y que estas regiones se hallan normalmente en la superficie molecular [Branden, 1991], puede esperarse que justamente en el propéptido de papaína tales segmentos se encuentren expuestos al solvente, y que posiblemente constituyan giros o asas estructurales.

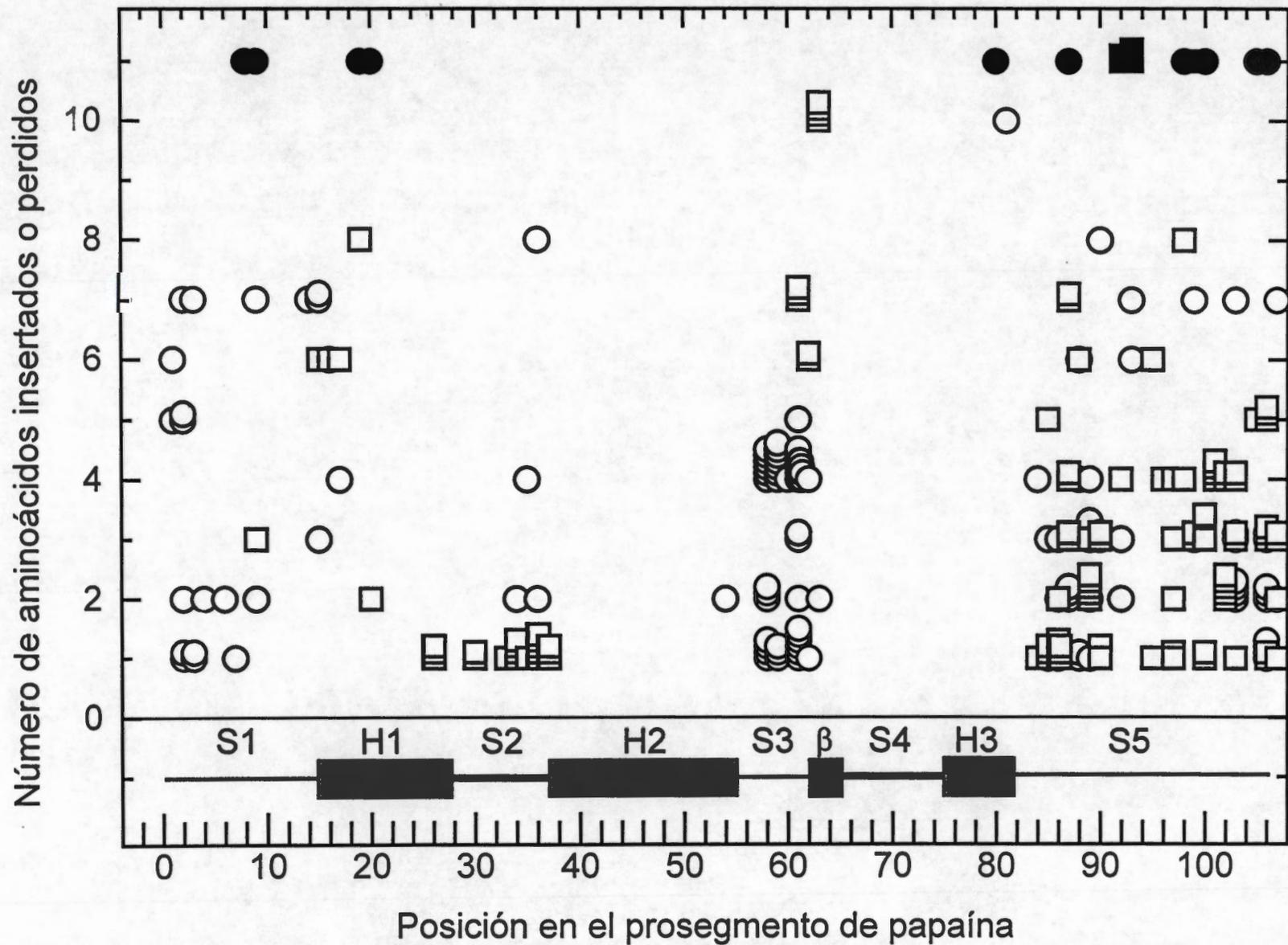


FIGURA 3. Inserciones y pérdidas de aminoácidos en el alineamiento múltiple. Los círculos y cuadros vacíos representan respectivamente las posiciones de inserción o pérdida de aminoácidos con relación a la secuencia del prosegmento de papaína. El eje vertical señala la longitud de dichos segmentos. Los símbolos llenos corresponden a segmentos mayores de 10 residuos. En la parte inferior del gráfico se muestran con rectángulos las posiciones de los elementos de estructura secundaria predichas para el propéptido.

Una forma alterna de localizar residuos expuestos al solvente en una proteína [Benner, 1995] es la identificación, a través del análisis de alineamientos de secuencias emparentadas, de residuos triptofano (Trp) que han mutado por arginina (Arg) o viceversa a lo largo de la evolución. El argumento para justificar la predicción terciaria de estos sitios es la semejanza en los codones que codifican a tales residuos y el hecho de que el cambio de un residuo hidrofóbico y extenso como Trp por un grupo con carga eléctrica (o viceversa) causaría un serio conflicto en el interior de una proteína. Después de una revisión sobre los subconjuntos de propéptidos homólogos, originados en la etapa de alineamiento, la ocurrencia de este tipo de mutaciones se observó en los residuos: 19, 39, 41, 56, 58, 64, 75, 84, 86 y 106. Tales posiciones se consideraron expuestas al solvente y se anexaron a las predicciones descritas en los párrafos anteriores para la construcción del patrón tridimensional que nos ocupa en este trabajo.

PREDICCIÓN DE LA REGIÓN DE CONTACTO ENZIMA-PROSEGMENTO

Debido a que los prosegmentos son fuertes inhibidores de sus enzimas cognadas, era de esperar que los propéptidos se unieran en las vecindades del sitio activo de tales moléculas. Al realizar un examen sobre las secuencias alineadas y las estructuras cristalográficas de las proteasas sulfhidríticas maduras, se hallaron algunas evidencias que ayudaron a situar espacialmente la interacción enzima-prosegmento. Se encontró por ejemplo, que existe una notable ausencia de posiciones de glicosilación potenciales cercanas al sitio activo de las regiones maduras de las proteasas cisteínicas y, así también, pudo determinarse que la accesibilidad al solvente de algunos residuos que se encuentran reunidos alrededor del sitio activo resulta ser mayor que aquélla predicha a partir de sus secuencias. Igualmente pudo observarse que las cistatinas, que son potentes inhibidores de proteasas sulfhidríticas, se fijan a los sitios activos de estas enzimas. De hecho estas interacciones se caracterizan por poseer altas constantes de unión [Stubbs, 1990; Turk, 1991]; lo cual también se ha reportado para la formación de los complejos enzima-prosegmento [Fox, 1992; Taylor, 1995; Pol, 1995].

Así como sucede con los sitios de glicosilación, las regiones de alta variabilidad en secuencia en el alineamiento múltiple podrían indicar no solamente zonas superficiales de las proteínas, sino también, lugares alejados de la región de unión entre éstas y la enzima madura. A partir de tal interpretación parecía ser claro que los segmentos S1, S3 y S5 no estaban involucrados en la interacción con la enzima madura, mientras que por su parte, los segmentos S2 y S4 que muestran una marcada conservación en secuencia y en longitud de la cadena (ver figura 3); podrían jugar un papel importante en el reconocimiento molecular. Como información adicional se conocía que tanto la estefina humana como la cistatina de pollo conforman el complejo con la papaína a través de dos asas estructurales [Stubbs, 1990; Auerswald, 1995].

Respecto a las condiciones que favorecen la generación de la enzima funcional era conocido el hecho de que la activación del zimógeno, en todas las proteasas cisteínicas estudiadas, se desencadena al llevar el valor del pH del medio hasta valores cercanos a 4 [Vernet, 1990 y 1991; Mason, 1992; Rowan 1992; Jerala 1998]. Esto sugería que en el proceso de activación estaba involucrado al menos un grupo carboxilato. Evidencia experimental realizada por Carmona *et al.* [1996] sobre procathepsina L señalaba que la pérdida de un fragmento que contiene a tres grupos con carga eléctrica, entre los que se encuentra el equivalente al residuo Glu38 del prosegmento de papaína, aumentaba el valor de la constante de inhibición 33,000 veces. A lo largo del alineamiento múltiple de las prorregiones existen tres residuos ácidos que están fuertemente conservados y que corresponden a los aminoácidos Glu38, Asp72 y Glu77. Tales grupos además, se hallan en las cercanías o dentro de los segmentos S2 y S4 (ver figuras 1 y 4a). Estos hechos hicieron suponer que los residuos mencionados podrían formar parte de la zona de contacto enzima-prosegmento en los precursores de las proteasas sulfhidríticas y que la carga eléctrica correspondiente estaría compensada por la enzima madura.

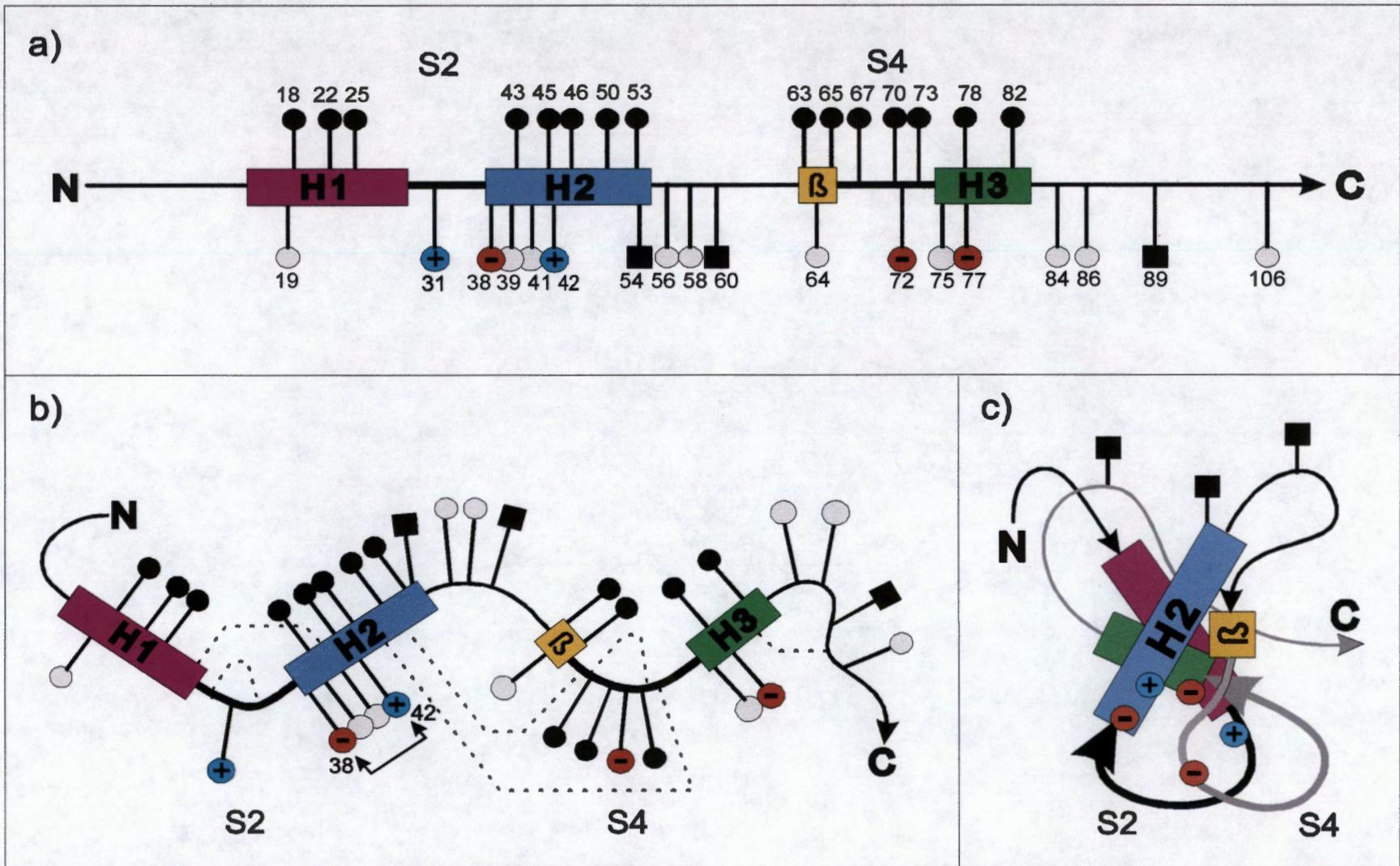


FIGURA 4. Ensamble de los elementos estructurales predichos para el prosegmento. (a) Representación de los residuos relevantes para la predicción estructural. Se marcan los aminoácidos conservados: hidrofóbicos (●), ácidos (-) y básicos (+); así como los sitios de glicosilación (■) y los de reemplazo de arginina por triptofano (○). Se muestran los segmentos S2 y S4 con línea gruesa. Con rectángulos se representan los elementos de estructura secundaria. (b) Restricciones espaciales predichas. Los segmentos S2 y S4 apuntan hacia la misma dirección. Las líneas punteadas indican pares de sitios con mutaciones correlacionadas y se representa un probable puente salino entre las posiciones 38 y 42. (c) Esquema del patrón tridimensional globular predicho para el prosegmento de papaína, donde las estructuras secundarias se orientan alrededor de un núcleo hidrofóbico común.

ENSAMBLE TRIDIMENSIONAL DE LOS ELEMENTOS ESTRUCTURALES PROPUESTOS EN EL TRABAJO DE PREDICCIÓN

Con la información obtenida en las etapas previas de predicción y de asignación espacial se creó un esquema del arreglo tridimensional del propéptido de papaína. La construcción de este esquema se presenta en la figura 4 donde se resumen además las distintas fases de ensamble. Los elementos de estructura secundaria se dispusieron de forma que permitieran el contacto a los grupos hidrofóbicos conservados en el alineamiento múltiple. Los sitios de glicosilación y las posiciones donde se detectaron mutaciones de tipo Trp/Arg se orientaron para exponerlas al solvente. Los grupos ácidos conservados se colocaron cercanos entre sí para conformar el sitio de reconocimiento con la enzima y los segmentos S2 y S4 se orientaron de forma que apuntaran hacia la misma dirección relativa. Con todo ello se respetaban las premisas derivadas tanto del estudio de predicción como de la información experimental disponible. Por su conservación en el alineamiento múltiple y por su posible ubicación sobre una misma cara de la segunda hélice predicha, era de esperar que los residuos Glu38 y Arg42 se encontraran interaccionando electrostáticamente en la estructura tridimensional funcional.

Durante el ensamblado estructural la secuencia del propéptido de papaína fue sometida a dos tipos de predicción de interacciones terciarias: mutaciones correlacionadas [Pazos, 1997] y detección de contactos beta [Hubbard, 1995]. Este último análisis sugirió la ausencia de láminas beta en el prosegmento de papaína; lo cual estaba de acuerdo con las expectativas. Por su parte, el estudio de mutaciones correlacionadas rindió resultados con baja significancia estadística, tal vez debido a que el método empleado no fue suficientemente sensible para detectar señales útiles en el amplio y variable conjunto de secuencias analizado en este trabajo. Los contactos tridimensionales de mayor confiabilidad que fueron predichos por el programa de mutaciones correlacionadas se muestran en la figura 4b, y contribuyeron como guía secundaria en la construcción del arreglo espacial que se presenta en la figura 4c.

Finalmente, con el propósito de verificar que las restricciones geométricas derivadas del trabajo de predicción fueran estructuralmente compatibles entre sí y con la secuencia del prosegmento, se analizó la integridad del conjunto de premisas de construcción durante una simulación del movimiento molecular. Tales premisas incluyeron: la longitud de los elementos de estructura secundaria; el ensamblado tridimensional de estos mismos elementos mediante el establecimiento de interacciones hidrofóbicas particulares; la cercana colocación en el espacio de los segmentos S2 y S4; y la ubicación en el exterior de la molécula tanto de los sitios de glicosilación, como de las posiciones de mutación W/R y de los segmentos S1 a S5. Con base en este conjunto de restricciones geométricas se construyó un modelo con detalle atómico que fue sometido a diversas etapas de minimización de energía y posteriormente a una simulación de dinámica molecular de 200 ps a una temperatura de 300 K. Al final de este último proceso pudo observarse que el modelo alcanzó y mantuvo una geometría estable que cumplía prácticamente con todas las restricciones originales. Este resultado puede ser interpretado como una comprobación de la viabilidad molecular simultánea del grupo de predicciones realizadas sobre el prosegmento.

PURIFICACIÓN Y CARACTERIZACIÓN DEL PROSEGMENTO DE PAPAÍNA

Durante la realización del trabajo de predicción fue posible obtener una muestra física del prosegmento de papaína que dio oportunidad a la realización de estudios con implicaciones estructurales. Esto, como ya se había mencionado, era de suma importancia puesto que presentaba una ocasión para apuntalar o modificar las premisas generadas durante la labor de predicción, así como durante el análisis de los datos experimentales preexistentes de homólogos al prosegmento bajo estudio.

Después de su expresión en *E. coli*, el prosegmento de papaína fue colectado como cuerpos de inclusión y disuelto en hidrocloreuro de guanidina 8 M. La muestra fue dializada extensivamente contra un regulador de fosfatos 0.05 M a pH 6.5 y la solución resultante fue filtrada y sometida a purificación por cromatografía de intercambio aniónico. El programa de elución incluyó un

gradiente salino de NaCl según se indica en la figura 5a. A continuación se realizó una segunda cromatografía donde se obtuvo una sola señal intensa como se aprecia en la misma figura. La fracción correspondiente fue colectada para su caracterización posterior.

A fin de estimar el tamaño molecular y determinar la homogeneidad de la sustancia aislada en la cromatografía líquida, se realizó una electroforesis en gel de poliacrilamida en condiciones desnaturizantes. El carril de la muestra presentó una sola banda y, de acuerdo a la migración de marcadores estándar, su masa molecular correspondió a (12 ± 1) kDa (figura 5b). Posteriormente se identificó el punto isoeléctrico de la proteína bajo estudio, el cual tuvo un valor de 4.37 ± 0.10 (figura 5c). Tanto la masa molecular como el punto isoeléctrico determinados en los experimentos anteriores resultaron ser compatibles con los valores calculados teóricamente a partir de la secuencia del prosegmento de papaína. Estos valores fueron respectivamente, 12.7 kDa y 4.40.

Los resultados anteriores permitieron tener confianza en que el péptido aislado correspondía a la prorregión de papaína, sin embargo, para lograr una caracterización más completa de la muestra de trabajo, era necesario demostrar que ésta poseía capacidad de reconocimiento de la proteína madura. Dado que se conoce que el prosegmento bajo estudio es un potente inhibidor de papaína, se probó el efecto de la prorregión sobre la actividad proteolítica de esta enzima. La figura 5d muestra el desarrollo de la absorbancia a 400 nm generada por la acción de la papaína sobre el sustrato sintético α -N-benzoil-DL-arginina-p-nitroanilida en función del tiempo. Con el propósito de contrastar el comportamiento del sistema, se registró la evolución de la señal de aparición del mismo producto de hidrólisis en un segundo experimento donde se habían incubado previamente la papaína y el prosegmento. Los resultados de este ensayo se muestran también en la figura 5d. Al comparar la actividad de la enzima en ambos casos se observa una disminución significativa de la señal causada por la formación del complejo papaína-prorregión, lo cual representa una evidencia de la funcionalidad molecular de la muestra recombinante purificada.

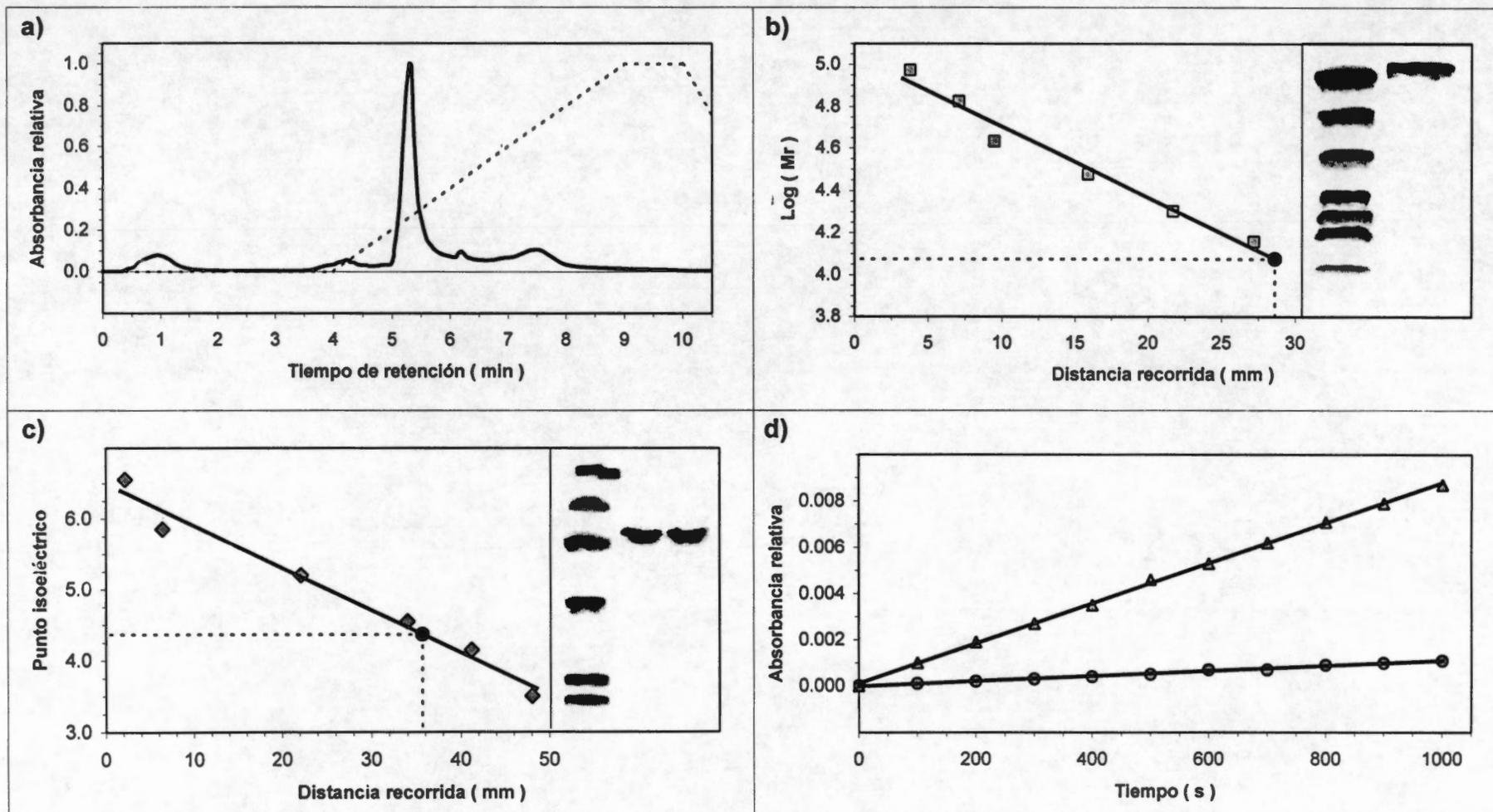


FIGURA 5. Purificación y caracterización del prosegmento de papaína. a) Cromatograma de la proregión de papaína. La línea punteada marca el cambio gradual en la concentración de NaCl en la fase móvil desde 0 hasta 1 M. Se muestran las curvas de calibración y los gels correspondientes a las determinaciones de masa molecular **b)** y punto isoelectrico **c)** del propéptido purificado. Los valores estimados fueron de 12 kDa y 4.4, respectivamente. **d)** Cinética de la hidrólisis de BAPNA catalizada por papaína en ausencia (▲) y presencia (●) de una cantidad equimolar del prosegmento de papaína.

ESTUDIOS ESPECTROSCÓPICOS SOBRE EL PROSEGMENTO DE PAPAÍNA

Se llevaron a cabo dos tipos de estudios experimentales sobre el propéptido de papaína: dicroísmo circular en la región del ultravioleta lejano y espectroscopía de fluorescencia.

El espectro de dicroísmo circular del prosegmento de papaína nativa, en el intervalo de 250 a 190 nm, se presenta en la figura 6. En esta figura se muestran también los ajustes a los datos experimentales logrados por dos programas de evaluación del contenido de estructura secundaria. La primera interpretación que puede hacerse del espectro es que parece ser representativo de una proteína helicoidal; dado que los mínimos de la señal se encuentran aproximadamente en 210 y 220 nm. El análisis del espectro de dicroísmo circular realizado con el método de Hennessey y Johnson [1981] indicó un $39 \pm 5\%$ de residuos en conformación helicoidal y un $20 \pm 7\%$ de residuos en conformación beta. El método k2d [Andrade, 1993] reportó valores de 31% y 11%, respectivamente para la proporción de aminoácidos en conformaciones helicoidal y beta, con un error global estimado del 8%. Los resultados de ambos métodos fueron compatibles con las predicciones de estructura secundaria del prosegmento a partir de su secuencia, las cuales arrojaron un 38% y 3% de residuos con estructuras helicoidal y beta, respectivamente. Cabe señalar que esta predicción se realizó sólo en el intervalo de residuos 12 a 82 debido a la baja conservación de los segmentos S1 y S5 en el alineamiento múltiple de prosegmentos.

Un estudio sobre el progreso de la señal de dicroísmo circular a 220 nm reveló un perfil sigmoïdal para el desplegamiento térmico de la proteína (ver recuadro en la figura 6). Este hecho es acorde con la cooperatividad observada por Taylor *et al.* [1995] y sugiere cierto grado de interrelación de la cadena polipeptídica del prosegmento, al menos bajo las condiciones de trabajo empleadas.

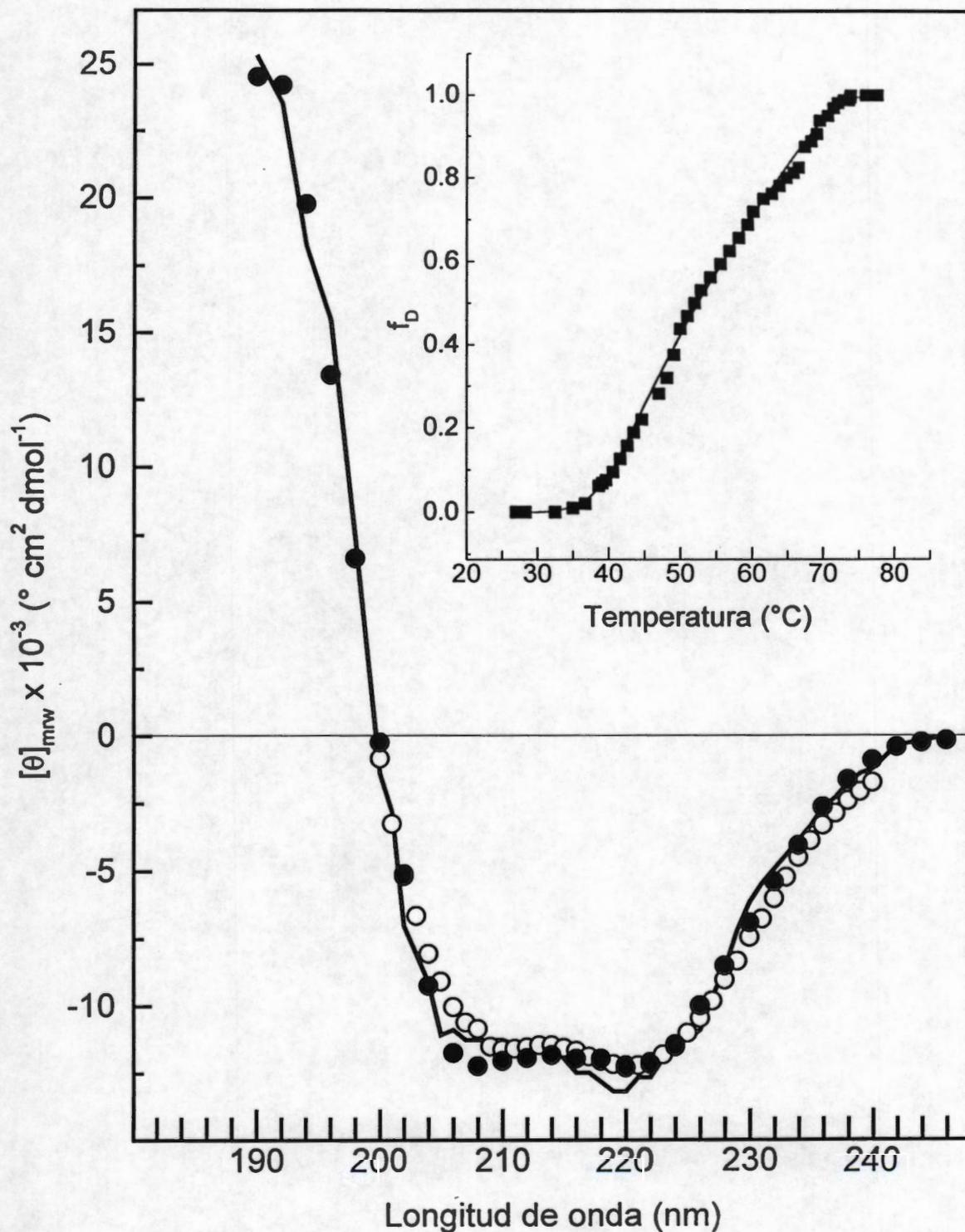


FIGURA 6. Dicroísmo circular en el UV-lejano. El trazo continuo corresponde al espectro experimental de dicroísmo circular a pH 6.5 para el prosegmento de papaína libre. Se muestran los ajustes al espectro realizados por los métodos de Hennessey y Johnson (círculos llenos) y k2d (círculos vacíos). En el recuadro se ilustra la dependencia de la señal a 220 nm con la temperatura, expresada como fracción desplegada.

El segundo tipo de ensayos espectroscópicos involucró el análisis tanto de la fluorescencia intrínseca del propéptido como de la radiación fluorescente emitida por la molécula de prueba ANS (ácido 8-anilino-naftalen 1-sulfónico) en presencia de la prorregión de papaína. El examen del espectro de fluorescencia intrínseca del prosegmento, al cual contribuyeron fundamentalmente los residuos de triptofano de la proteína [Permyakov, 1993], parece revelar la existencia de entornos diferentes para cada uno de los dos residuos de este tipo que están presentes de la cadena del prosegmento. Ello significa que uno de los residuos de Trp se encuentra oculto y el otro completamente expuesto al solvente, en las condiciones de trabajo establecidas (ver figura 7). Este resultado experimental concuerda con la predicción de que la cadena lateral del Trp 25, que se encuentra altamente conservada en el alineamiento múltiple, está dirigida hacia el interior de la proteína; mientras que el grupo Trp 64, altamente variable (incluso cambia algunas veces por arginina), se asume en contacto con el solvente en la región de la cadena beta anfipática. Por su parte, la fluorescencia del ANS reveló un cambio conformacional en el propéptido que, a juzgar por los resultados experimentales, sucede en el intervalo de pH de 5.7 a 3.0 (recuadro en la figura 7); lo cual habla de un posible estado distendido del polipéptido en estas condiciones de trabajo. De manera correspondiente, podría esperarse una estructura mejor empacada para el prosegmento de papaína en la zona de pH comprendida entre 5.7 y 6.7.

Con la información antes expuesta resulta evidente que el incremento en la concentración de los protones del medio altera la conformación de la molécula del prosegmento. Este hecho podría ser muy significativo en la interpretación del mecanismo molecular de activación del precursor de papaína, y de otras proteasas homólogas, cuando se trabaja condiciones de pH bajo.

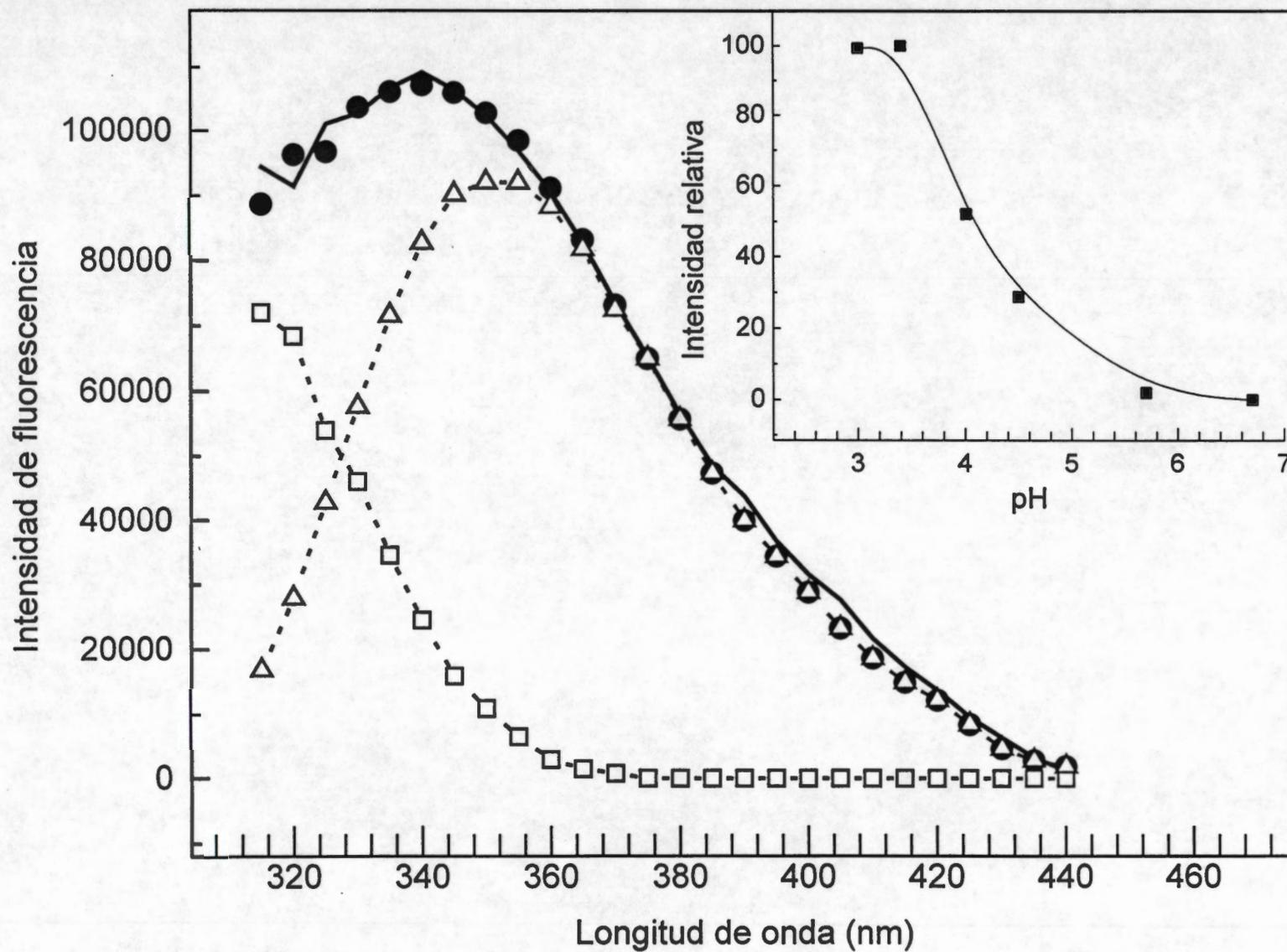


FIGURA 7. Fluorescencia. Espectro experimental de la fluorescencia intrínseca del prosegmento de papaína libre (línea continua). Las curvas con símbolos vacíos corresponden a las contribuciones de las cadenas laterales de un triptofano oculto al solvente (cuadros) y de otro expuesto (triángulos). La suma de ambas contribuciones se muestra con círculos llenos. El recuadro muestra el efecto del pH en la fluorescencia del ANS a 488 nm en presencia de la proregión de papaína.

CAPÍTULO 3

COMPARACIÓN ENTRE LAS CARACTERÍSTICAS ESTRUCTURALES PREDICHAS PARA LOS PROSEGMENTOS DE CADENA LARGA Y LAS GEOMETRÍAS CRISTALOGRAFICAS DE HOMÓLOGOS CERCANOS.

Durante la última etapa del modelado se publicó la estructura cristalográfica de un homólogo cercano al zimógeno de papaína, la procaricaína [Groves, 1996], cuyo prosegmento mantiene una identidad en secuencia del 73% con el de papaína. Poco después se reportó también la estructura resuelta por difracción de rayos X de otro zimógeno homólogo, la procatepsina L [Coulombe, 1996] (37% de identidad entre prosegmentos). Ambas estructuras, junto con la de procatepsina B (15% de identidad topológica con el prosegmento de caricaína), fueron empleadas para evaluar el impacto de las decisiones tomadas y la confiabilidad de los resultados del trabajo de predicción. Los análisis realizados se basaron en el hecho experimental de que proteínas que realizan la misma función y mantienen homología en secuencia poseen características tridimensionales comunes [Morea, 1998].

Al comparar las predicciones de exposición al solvente en el prosegmento de papaína con la estructura cristalográfica de procaricaína (que presenta la mayor homología en secuencia con propapaína), se encontró que los métodos de Benner *et al.* [1994] y PHDacc resultaron con una exactitud cercana al 80%, ver figura 2b. Como se advierte en la figura 8 y en la tabla I, también fue acertada la predicción de las cadenas expuestas al solvente con el criterio de los sitios de glicosilación, y con el de la aparición alternada de los residuos Trp y Arg en una misma posición a lo largo de los alineamientos de secuencias homólogas (excepto para los residuos Thr 56, Gly 84 y Leu 86). Por su parte, la identificación de los elementos de estructura secundaria también fue acertada y alcanzó un 81% de exactitud en términos de la asignación de carácter helicoidal, beta e irregular para cada aminoácido de la cadena (ver figura 2a).

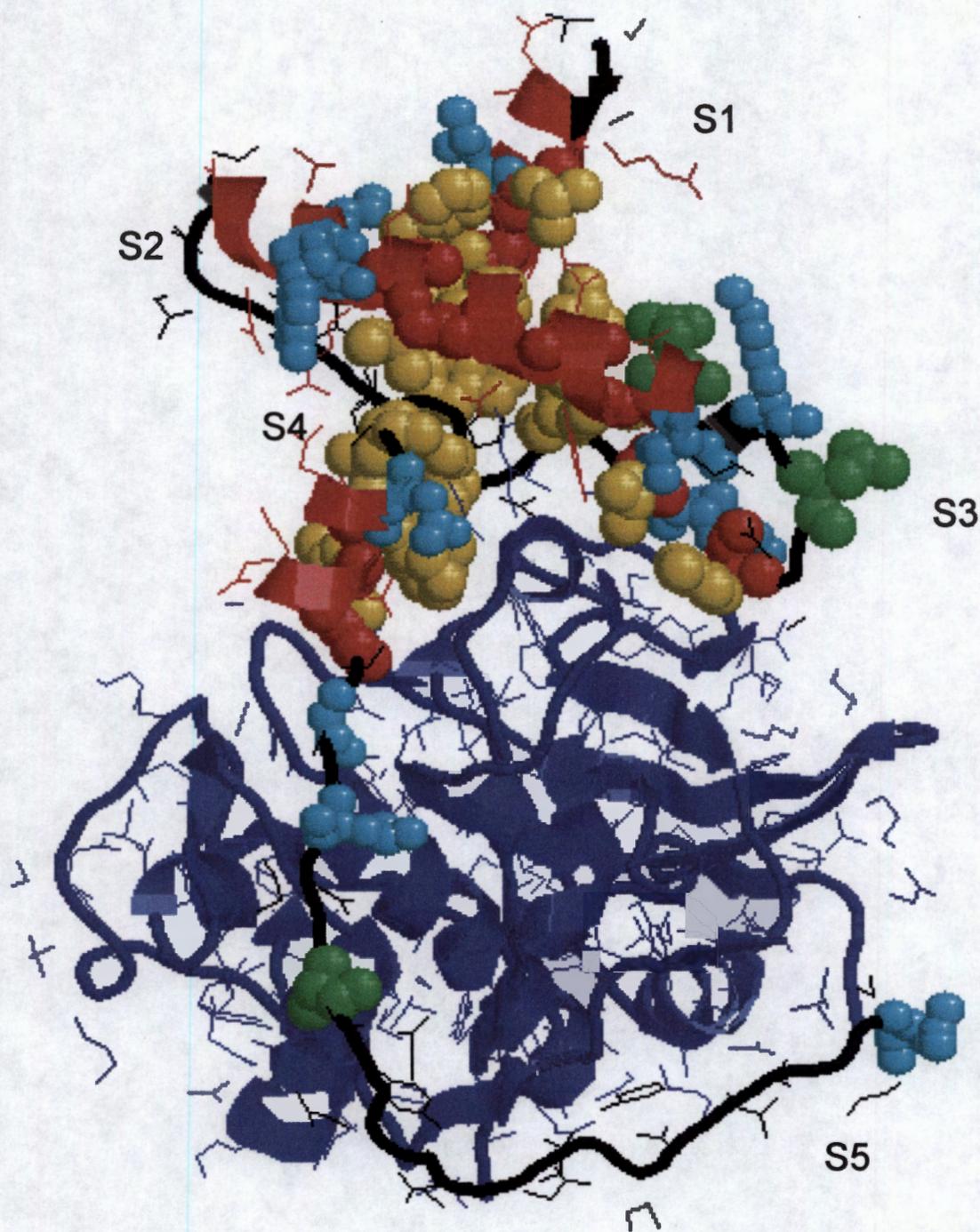


FIGURA 8. Estructura cristalográfica de procaricaína. La prorregión se muestra en color rojo y la porción madura en azul marino. Se presenta la ubicación de los segmentos S1 a S5 y se han resaltado las cadenas laterales de los residuos predichos ocultos al solvente en el núcleo hidrofóbico ■, y expuestos tanto por presentar mutaciones Trp/Arg ■ como por ser sitios de glicosilación ■. Se observa la disposición en cruz de las hélices constituyentes de la prorregión.

<i>RESIDUO</i>	<i>EXPOSICIÓN ABSOLUTA (Å²)</i>	<i>EXPOSICIÓN RELATIVA (%)</i>
<i>LEU 18</i>	<i>32.78</i>	<i>23.3</i>
<i>PHE 22</i>	<i>0.2</i>	<i>0.0</i>
<i>TRP 25</i>	<i>9.02</i>	<i>4.3</i>
<i>PHE 43</i>	<i>30.48</i>	<i>18.6</i>
<i>ILE 45</i>	<i>26.65</i>	<i>19.3</i>
<i>PHE 46</i>	<i>0.67</i>	<i>0.4</i>
<i>LEU 50</i>	<i>39.07</i>	<i>27.8</i>
<i>ILE 53</i>	<i>8.39</i>	<i>6.1</i>
<i>TYR 63</i>	<i>6.76</i>	<i>3.8</i>
<i>LEU 65</i>	<i>1.46</i>	<i>1.0</i>
<i>LEU 67</i>	<i>31.21</i>	<i>22.2</i>
<i>PHE 70</i>	<i>4.30</i>	<i>2.6</i>
<i>LEU 73</i>	<i>17.03</i>	<i>12.1</i>
<i>PHE 78</i>	<i>0.00</i>	<i>0.0</i>
<i>TYR 82</i>	<i>31.00</i>	<i>17.6</i>
<i>ILE 19</i>	<i>73.62</i>	<i>53.4</i>
<i>LYS 39</i>	<i>68.13</i>	<i>41.9</i>
<i>TYR 41</i>	<i>119.02</i>	<i>67.4</i>
<i>THR 56</i>	<i>18.39</i>	<i>18.1</i>
<i>LYS 58</i>	<i>142.25</i>	<i>87.5</i>
<i>TRP 64</i>	<i>129.13</i>	<i>61.2</i>
<i>ASN 75</i>	<i>37.80</i>	<i>35.6</i>
<i>GLY 84</i>	<i>5.11</i>	<i>15.8</i>
<i>LEU 86</i>	<i>14.85</i>	<i>10.6</i>
<i>ASN 106</i>	<i>104.65</i>	<i>98.5</i>
<i>ASP 54</i>	<i>61.31</i>	<i>59.8</i>
<i>ASN 60</i>	<i>116.21</i>	<i>109.4</i>
<i>ALA 89</i>	<i>29.95</i>	<i>43.3</i>

TABLA I. Exposiciones absoluta y relativa de las cadenas laterales del propéptido de caricaína calculadas con el programa NACCESS. Se presentan los valores para los residuos predichos ocultos en un núcleo hidrofóbico (amarillo) y expuestos al solvente por ubicarse tanto en sitios de mutación Trp/Arg (azul) como en sitios de glicosilación (verde).

Como puede advertirse en la figura 8, los residuos no polares conservados se observan ciertamente ocultos al solvente en la estructura cristalográfica, y forman parte de núcleos hidrofóbicos. La suposición de que los segmentos S2 y S4 se encontraban orientados en la misma dirección relativa también resultó ser consistente con lo observado, pero de hecho, el único segmento que interacciona con la enzima madura es el S4. El segmento S5, por su parte, adopta una conformación irregular, como lo sugería la alta variabilidad en secuencia de esa zona (figuras 1 y 3) y la presencia de una multitud de señales de terminación de estructura secundaria asignadas por el método de Benner *et al.* (figura 2a). Los residuos carboxilato conservados se encuentran en efecto cercanos en el espacio, aunque no están interaccionando con la enzima. El puente salino predicho para los residuos Glu38 y Arg42 (localizados en vueltas consecutivas de la hélice 2) se establece realmente a juzgar por los resultados del análisis geométrico de las cadenas laterales respectivas, realizado con programas de representación molecular. Adicionalmente se analizó el acomodo de las cadenas ocultas al solvente del prosegmento de caricaína con la construcción de poliedros de Voronoi [Pontius, 1996]. Los resultados indicaron que los átomos del interior del propeptido se encuentran empacados de una manera similar a la observada en proteínas nativas; y ello resulta ser consistente con la visión de una molécula que en medio acuoso no favorece la generación de radiación fluorescente del ANS.

En resumen puede afirmarse que la exposición y la colocación relativa de la mayor parte de los elementos de estructura secundaria, así como la extensión de éstos y su conectividad a través de los segmentos S2 a S4, se encontraron bien representados en el modelo tridimensional construido en la etapa de predicción estructural.

Por otra parte puede afirmarse que los resultados de baja significancia reportados por los métodos de reconocimiento de patrones de plegamiento se debieron fundamentalmente a la ausencia de estructuras representativas del arreglo espacial de los prosegmentos en las bases de datos correspondientes. De hecho, los elementos helicodales en los propeptidos de cadena larga se disponen en un patrón de cruz que no había sido reportado para otras proteínas (figura 8).

Recientemente estos mismos métodos de reconocimiento, con información actualizada, detectan fácilmente las estructuras conocidas de los zimógenos de caricaína y catepsina L como compatibles con la secuencia del precursor de papaína.

Una de las mayores divergencias entre el esquema tridimensional generado en este trabajo y las estructuras cristalográficas de homólogos empleados en el estudio comparativo, fue la colocación de la tercera hélice del prosegmento. Esta hélice en realidad se encuentra en la zona de contacto con la proteína madura en los tres zimógenos con estructura conocida. Además, en ellos resulta evidente que el prosegmento se une al resto de la proteína en sentido inverso al observado tanto para la unión enzima-cistatina, [Stubbs, 1990; Turk, 1991] como para la interacción enzima-sustrato [Baker, 1987]. Este último punto significa que la proregión no emplea dos asas estructurales para el reconocimiento de la enzima (lo cual era una de las premisas del trabajo de predicción) sino que en realidad presenta una unión del tipo prosegmento-catepsina B, esto es, una cadena simple de aminoácidos consecutivos en dirección opuesta a la de un sustrato. La procatepsina B y la procaricaína, que son miembros de subfamilias genéticas distintas, constituyen un ejemplo de polipéptidos con muy baja identidad en secuencia que no obstante, mantienen una alta conservación en estructura tridimensional. La baja homología entre los prosegmentos de estas proteínas dificultaba la revisión comparativa de las secuencias, y fue por esta razón que resultó necesario realizar el alineamiento topológico de las dos cadenas polipeptídicas que se muestra en la figura 9. Al analizar las posiciones conservadas entre los miembros de cada una de las subfamilias genéticas, con base en este alineamiento topológico, se encontró que cerca de 10 de los residuos equivalentes mantuvieron relativamente invariantes sus propiedades fisicoquímicas de manera simultánea en ambos grupos de proteínas; lo cual sugiere un origen común para ellas. La experiencia más útil de este sorprendente resultado es que la conservación en la disposición atómica espacial puede prevalecer aún entre proteínas que han sufrido grandes cambios en su secuencia a lo largo de la evolución.

Por otra parte, los residuos apolares de las huellas 1 a 3, que en el trabajo de predicción se habían supuesto como integrantes de un núcleo hidrofóbico común, fueron encontrados ciertamente en sitios ocultos al solvente. De hecho los residuos hidrofóbicos conservados que constituyen las hélices 1 y 2 del prosegmento se integran para formar un núcleo común. Sin embargo, en la estructura cristalográfica de procaricaína la mayor parte de los residuos apolares de la huella 3 se localizan en la región interfacial enzima-proregión. Este resultado muestra que cuando se detectan grupos apolares conservados en un alineamiento múltiple de secuencias existe una alta probabilidad de encontrar a tales grupos resguardados del solvente; pero al tratarse de polipéptidos que en su lapso de vida útil deban unirse a otras proteínas o a otros segmentos de la propia cadena, los residuos hidrofóbicos conservados tienen la oportunidad de ser hallados en la superficie de reconocimiento molecular.

CAMBIO CONFORMACIONAL DE LOS PROSEGMENTOS LIBRES DE LA PORCIÓN MADURA

La labor de predicción de la estructura tridimensional del propéptido de papaína, sin el empleo de coordenadas cristalográficas de base, condujo a reunir una gran cantidad de información experimental sobre homólogos en secuencia a la proteína de interés. Esta información, proveniente de fuentes diversas, refiere características importantes del comportamiento de precursores y de sus correspondientes prosegmentos en distintas condiciones del medio. Por ejemplo, el reporte de Jerala *et al.* [1998] menciona la proclividad a la agregación del propéptido de catepsina L en medios con pH superiores al de activación del precursor. Carmona *et al.* [1996] manifiestan el uso de mezclas agua-acetonitrilo en el manejo de la proregión de catepsina L libre de la proteína madura. Durante los experimentos hechos en el presente trabajo sobre el prosegmento de papaína se observó que la muestra disminuía su solubilidad al eliminar el cloruro de sodio, adquirido en la etapa de purificación. En el manuscrito de Vernet *et al.* [1995] se reporta un incremento del pH óptimo de activación del precursor de papaína al introducir la mutación Phe70His en el propéptido. Existen además los estudios ya

mencionados sobre la activación de los precursores en valores de pH cercanos a 4 para ambas subfamilias genéticas; la cita de Carmona *et al.* [1996] sobre el decremento de las constantes de unión enzima-prosegmento en medio ácido, así como también, el informe de Ishido *et al.* [1995] respecto a la activación *in vitro* de precursores de proteasas sulfhidríticas que se produce espontáneamente en presencia de polianiones. Finalmente en el recuadro de la figura 7 puede apreciarse que la fluorescencia de la molécula de prueba ANS, en presencia del prosegmento de papaína, es sensible a las variaciones de pH del medio.

Toda la evidencia experimental citada en el párrafo precedente invitaba a plantear una interpretación molecular fundada en la existencia de un factor electrostático responsable de la modificación geométrica y por tanto funcional de los prosegmentos estudiados. Para explorar las posibilidades de esta propuesta se procedió a la revisión de algunos datos generados durante el trabajo de predicción, así como al estudio de las estructuras de los precursores caracterizados tridimensionalmente.

Al analizar las particularidades de los residuos con carga eléctrica a pH fisiológico, Lys31; Glu38; Arg42; Asp72 y Glu77, los cuales pueden observarse en el alineamiento múltiple de prosegmentos (figura 1), resulta evidente que ellos no sólo están altamente conservados en el curso de la evolución sino que también se encuentran expuestos al solvente en las estructuras cristalográficas de procaricaína y procatepsina L; lo cual sugiere su participación en la función molecular de los prosegmentos en las diferentes especies consideradas. De hecho estos residuos cargados confluyen en las cercanías de los segmentos S2 y S4 donde constituyen un entramado electrostático local, como se aprecia en la figura 9a para procaricaína. Otro resultado importante, aún cuando no está relacionado directamente con el factor electrostático, es que la superficie expuesta al solvente en el prosegmento cristalográfico de caricaína, en ausencia de la región madura, muestra carácter apolar en un poco más de la mitad de su área total (59%) y que la zona interfacial con la proteína madura presenta aún mayor proporción de área hidrofóbica (64%).

Con la información reunida hasta este punto del trabajo se intentó probar la propuesta del factor electrostático a través de una simulación molecular sobre la porción estructurada del prosegmento de caricaína cristalográfica (figura 10a). Esta simulación inició con la creación de dos ambientes electrostáticos distintos, en uno de ellos se trabajó con los residuos ácidos completamente protonados para hacer la representación de un medio con pH bajo. En el segundo caso se mantuvieron intactos los carboxilatos para simular las condiciones previas al cambio en la señal fluorescente del ANS. El prosegmento fue posteriormente sometido a una serie de etapas de dinámica molecular bajo las dos condiciones señaladas, a fin de observar el efecto de las cargas negativas en la conformación global. Durante el ensayo con los carboxilatos protonados se hizo evidente, desde etapas tempranas del cálculo computacional, un movimiento de la hélice 3 dirigido hacia las superficies hidrofóbicas que presentan las otras dos hélices y la cadena beta (figura 10c). En el segundo cálculo, con las cargas intactas, la hélice 3 realizó un desplazamiento en dirección al cúmulo de grupos cargados (ver figura 10b). La marcada diferencia en la posición de la hélice 3, entre ambas condiciones de estudio mencionadas, permitió proponer una interpretación al comportamiento experimental del propéptido libre de la porción madura. La presencia del cúmulo electrostático, integrado por los residuos cargados conservados, parece favorecer la formación de puentes salinos múltiples en medios con pH superior a 6. Ello propiciaría la colocación de la hélice 3 de tal manera que se expusiera al solvente una cantidad relativamente alta de área hidrofóbica, proveniente de algunos residuos de la cadena beta, y la propia hélice 3; que en conjunto constituyen la región de contacto con la enzima madura. Como consecuencia, en condiciones que mantuvieran inalteradas las interacciones del cúmulo de cargas, el prosegmento sería poco soluble y presentaría tendencia a la agregación si estuviese aislado (lo cual explicaría la observación de Jerala *et al.* [1998] y el empleo de acetonitrilo por Carmona *et al.* [1996]), pero también en esas condiciones sería altamente afin a la enzima madura y podría inactivarla poderosamente; hecho que es compatible con lo reportado experimentalmente por Taylor *et al.* [1995] y que asimismo resulta consistente con la geometría

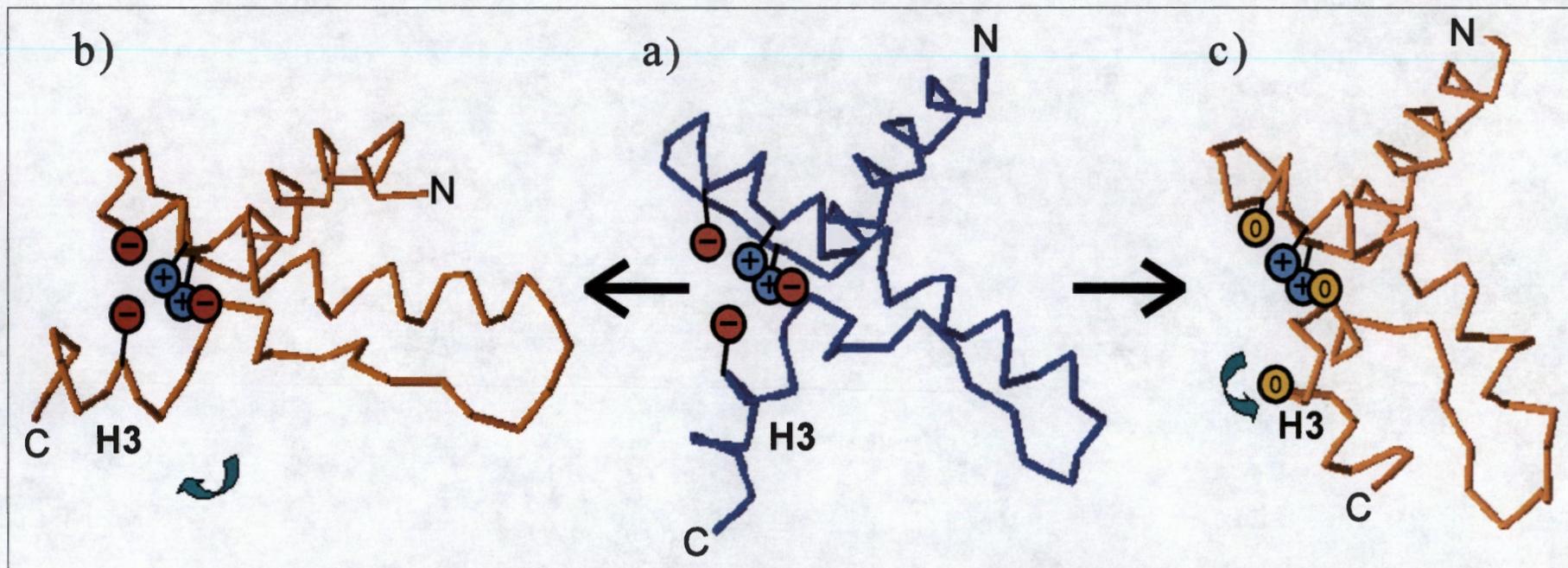


FIGURA 10. Resultados de dos simulaciones de dinámica molecular. Cada cálculo involucró trayectorias de 200 ps sobre la porción globular del prosegmento de caricaina. En ambas simulaciones se utilizaron las coordenadas cristalográficas de la molécula como punto de partida (a). Se muestran esquemas de las conformaciones finales en las dinámicas con las cargas intactas (b) y con los grupos carboxilato sin carga eléctrica (c).

cristalográfica reportada para procaricaína. En contraste, con las interacciones electrostáticas desestabilizadas (valores bajos de pH, condiciones de fuerza iónica alta, o incluso presencia de especies polianiónicas), la hélice 3 podría desplazarse hacia el resto de los elementos de estructura secundaria para ocultar área hidrofóbica; con un consecuente incremento en la solubilidad de la molécula libre y una disminución de interacciones potenciales con la enzima madura.

El movimiento de la tercera hélice del prosegmento, inducido por cambios en las condiciones del medio, podría ser de suma importancia para explicar tanto las primeras etapas del mecanismo de activación de los zimógenos de las proteasas cisteínicas, como la inhibición de las catepsinas L y B por sus prosegmentos respectivos. En catepsina L se ha reportado que la constante de unión enzima-prosegmento decrece conforme el pH se torna ácido [Carmona, 1996]. Los autores han demostrado que este hecho se debe a la disminución en la constante de velocidad para la formación del complejo; dado que la constante de velocidad de la reacción de disociación se mantuvo inalterada en el intervalo de valores de pH estudiado. La existencia de una conformación globular para el prosegmento libre en medio ácido (del tipo mostrado en la figura 10c) podría explicar su baja afinidad por la enzima madura; lo cual haría decrecer el valor de la constante de velocidad de asociación sin afectar el del proceso inverso. Bajo esta interpretación es comprensible ahora que la eliminación de una porción que contiene a los residuos conservados Lys 31, Glu 38 y Arg 42, realizada por estos mismos autores sobre el propéptido de catepsina L, resulte en un decremento considerable en el valor de la constante de unión. En el caso de la catepsina B, que es miembro de la otra subfamilia genética, se ha reportado una respuesta análoga de la constante de unión enzima-prosegmento frente a cambios de pH. Este hecho resulta ser un tanto inesperado ya que, aunque en las proteínas tipo catepsina B existe un conjunto de grupos cargados conservados, éste es diferente al que aparece en la otra subfamilia genética. De acuerdo con un trabajo de Fox *et al.* [1992], el valor de la constante de unión para catepsina B y su propéptido decrece con valores bajos de pH debido a un incremento en la constante de velocidad para la reacción de disociación y dado que la constante de velocidad para la formación del

complejo permanece inalterada en el intervalo de valores de pH estudiado. Al realizar un análisis sobre la estructura cristalográfica del zimógeno de catepsina B pudieron identificarse dos grupos carboxilato en la enzima madura (Glu 122 y Asp 124, numeración de catepsina B) que establecen puentes salinos con un grupo positivo del prosegmento (Lys 40, equivalente a la posición 80 del prosegmento de caricaína, figura 9b). Es probable que al protonar los grupos con carga negativa, debido a la disminución del pH del medio, se generara un cambio conformacional que promoviera la disociación del complejo, con el consiguiente incremento en la constante de velocidad de disociación.

Como ya se había mencionado, una de las divergencias más notables entre el esquema tridimensional propuesto en este trabajo y la información cristalográfica resultó ser la colocación de la tercera hélice, la cual fue encontrada en la región de contacto con la enzima madura y no como integrante del núcleo hidrofóbico en el prosegmento. Sin embargo, del conjunto de resultados descrito hasta aquí, surge la posibilidad de que la predicción propuesta inicialmente acerca de una estructura globular para el propéptido sea cierta al menos en medio ácido; aunque de acuerdo con el resultado de la fluorescencia de ANS en este estudio, el grado de empaquetamiento estructural que se espera encontrar en esas condiciones es bajo.

RELACIÓN ESTRUCTURA-FUNCIÓN DE LOS PROPÉPTIDOS DE PROTEASAS SULFHIDRÍLICAS

ASISTENCIA AL PLEGAMIENTO

El análisis de las estructuras cristalográficas de los zimógenos conocidos hasta ahora nos permite especular sobre los detalles moleculares de la función del prosegmento en la promoción del plegamiento de la enzima madura. De acuerdo con ese análisis, en la zona de unión entre la enzima y el propéptido parecen ocultarse y anclarse varios residuos aromáticos. Esta reducción en la movilidad podría explicar el gran incremento en la señal de dicroísmo circular en el ultravioleta cercano observado por Carmona *et al.* [1996] durante la formación del complejo catepsina L-proorregión. Es posible que el anclaje de tales grupos

contribuyera de manera importante a la afinidad entre el prosegmento y la enzima; lo cual resultaría ser consistente con la hidrofobicidad relativamente alta encontrada en este trabajo para la zona interfacial. Las estructuras cristalográficas también revelaron que los prosegmentos interaccionan principalmente con el segundo dominio de las enzimas maduras. Este dominio está constituido por un barril beta flanqueado por pequeñas hélices alfa. El propéptido establece múltiples contactos tanto con el extremo superior del barril como con una cara de la hélice inferior, y además parece incorporar los primeros residuos de la cadena madura al barril beta (figura 11). En todos estos casos es posible apreciar que la presencia de la prorregión genera una gran cantidad de puentes de hidrógeno que extienden la red de este tipo de interacciones en el segundo dominio enzimático.

Por otra parte, un rasgo sobresaliente de las proteasas sulfhidríticas es que los últimos residuos de su cadena presentan una topología de pseudonudo. En esta conformación los residuos del extremo C terminal parecen estar atrapados entre los dos dominios de la enzima y el segmento que une esas dos porciones de la cadena (figuras 9 y 11). A pesar de que las topologías de este tipo son poco usuales en las estructuras de proteínas [Creighton, 1993], es posible observarlas en las proteasas del tipo papaína y también del tipo catepsina B. Este hallazgo parece indicar que el plegamiento de la cadena madura sucede de manera no secuencial y que, por tanto, los últimos residuos de esta cadena establecen sus interacciones terciarias antes de que puedan crearse los contactos interdominio de la enzima. Bajo esta suposición pueden plantearse dos rutas o escenarios probables de adquisición de la estructura nativa para la proteína madura en las cuales el prosegmento ya estructurado asiste el plegamiento del resto de la cadena. En uno de esos escenarios, el primer dominio constituido principalmente por elementos helicoidales se plegaría de forma prácticamente independiente del resto de la cadena; puesto que se ha observado que las hélices mantienen información de corto alcance en la secuencia y que son estructuras autoestables [Brown, 1971; Dobson, 1994; Aurora, 1997; Sheinerman, 1998]. En este esquema el segundo dominio podría requerir de una fuerza inductora, proporcionada por el prosegmento, para armar su barril beta. Una vez formados, es posible que el

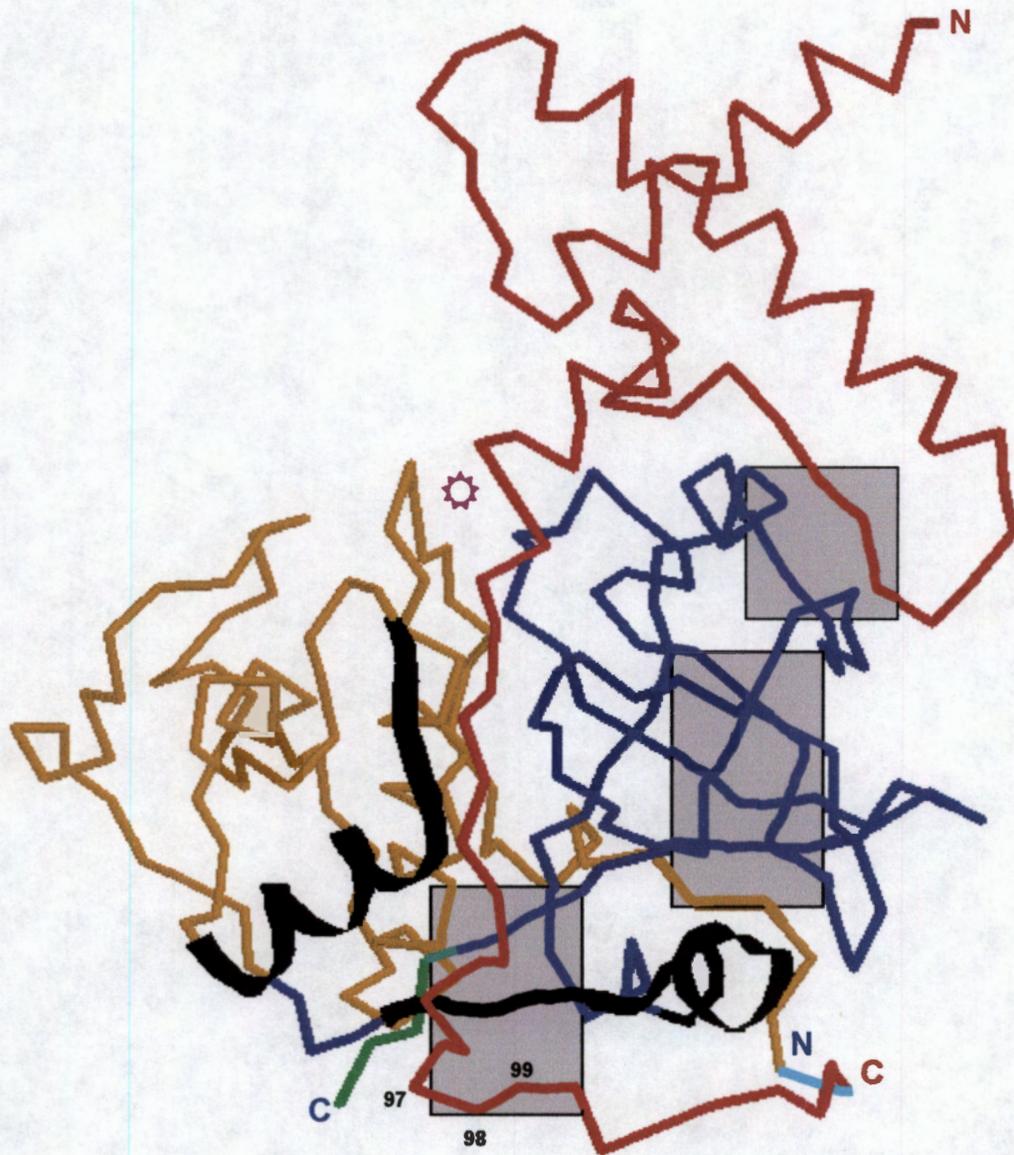


FIGURA 11. Ubicación del prosegmento en el zimógeno cristalográfico de caricaina. Con colores oro y azul se destacan los dominios helicoidal y beta de la enzima, respectivamente; la cadena del propéptido se resalta con color rojo. Las unidades de plegamiento independiente inducidas por el prosegmento se indican con listones negros, mientras que las regiones de interacción por puente de hidrógeno generados por el prosegmento se enmarcan en cuadros color gris. Se muestra la colocación de los residuos 97 a 99 de la proregión, así como la ubicación del sitio activo (magenta); de los últimos residuos de la cadena enzimática (verde); de los extremos N y C terminales del propéptido (rojo) y de la enzima madura (azul marino). El enlace peptídico entre el prosegmento y la enzima se presenta en color azul claro.

propéptido acercara y orientara a los dos dominios estructurales de la enzima y colaborara al establecimiento de las interacciones específicas entre ellos. La otra ruta factible es la que contempla como etapa inicial el plegamiento del segundo dominio auxiliado por el prosegmento. Este dominio, una vez organizado, podría servir como base estructural para el resto de la cadena posibilitando así la formación de los elementos helicoidales y del resto de las interacciones que conforman al primer dominio. Uno de tales elementos, la hélice central marcada en azul en la figura 9a, es fundamental para la actividad de la enzima madura y además parece contener baja información local sobre su propia conformación. De aquí resulta factible suponer que tal segmento estructural requiera ayuda para conseguir su plegamiento funcional.

Una forma de abordar el estudio del proceso de plegamiento en proteínas, es a través de la identificación de las zonas llamadas "unidades de plegamiento independiente" propuestas por Moult [1991] y Pedersen [1995]. Estas unidades se asumen como segmentos de polipéptido que se pliegan de manera autónoma y que sirven como sitios de nucleación durante la formación de la estructura funcional de una proteína. Las unidades de plegamiento independiente, según los autores, pueden ser localizadas sobre las conformaciones tridimensionales de las proteínas nativas como zonas que ocultan una proporción significativa de área hidrofóbica. Al evaluar la exposición relativa en las estructuras cristalográficas de caricaína madura y de su zimógeno, se encontró que la prorregión inducía la formación de dos de estas unidades en la superficie de contacto enzima-propéptido. Una de ellas se ubica en la zona interdominio de la enzima madura y la otra en la base del barril beta del segundo dominio de la enzima (figura 11). Cabe señalar que la existencia de esta última unidad de plegamiento independiente parece promover la interacción indirecta entre el extremo C terminal de la enzima y la prorregión. Así, los residuos 97 a 99 del prosegmento podrían inducir la primera etapa de la formación del pseudonudo a través de la construcción de una hoja beta de tres cadenas (figura 11). Por otra parte, los experimentos de desplegamiento, realizados en nuestro laboratorio sobre papaína madura, han mostrado que el reestablecimiento de las condiciones iniciales del

medio no induce la recuperación de estructura secundaria en la molécula. Este resultado representa una fuerte evidencia en contra de la primera de las rutas propuestas en el párrafo anterior, por lo que es posible que el plegamiento del dominio helicoidal requiera de la conformación nativa del dominio beta; mientras que este último, a su vez, quizá necesitara del prosegmento para lograr su estructura tridimensional final. Aparentemente, la función del propéptido no radica sólo en la estabilización del segundo dominio enzimático, ya que también induce la formación de una unidad de plegamiento independiente en el dominio helicoidal. De esta manera y para continuar con el esquema planteado, la prorregión podría favorecer el acercamiento de la porción desplegada de la cadena (que da origen al primer dominio) hacia el resto de la proteína ya estructurada y ello propiciaría la formación de la zona interfacial final.

AUTOPROCESAMIENTO

La ruptura del enlace que une al prosegmento con la enzima madura es un evento llamado indistintamente procesamiento o activación del zimógeno, y el mecanismo molecular involucrado permanece sin ser caracterizado completamente. Como ya se ha mencionado, existe evidencia experimental acerca de que los precursores de estas enzimas poseen cierto grado de actividad catalítica y capacidad de autopprocesamiento (mecanismo *cis* intramolecular) en ausencia de alguna otra enzima proteolítica activa [Kassell, 1973; Vernet 1991; Ishidoh, 1995], lo que implica que el sitio activo es al menos parcialmente funcional aún en el zimógeno. Adicionalmente se sabe que los precursores son susceptibles a adquirir actividad biológica a través de un ataque realizado por otras proteasas que coexistan en el medio (mecanismo *trans*) [Holwerda, 1990; Carmona, 1996]. En el contexto del autopprocesamiento del zimógeno en una acción *cis* intramolecular se llevó a cabo una modificación conformacional de un segmento N terminal de la caricaina para averiguar si le era posible alcanzar el sitio catalítico de la enzima. Se observó que el movimiento simple de los primeros 12 residuos de la cadena permite al extremo amino tener acceso al sitio activo en

el sentido adecuado para la hidrólisis. Un movimiento similar ha sido propuesto para el autoprocesamiento del precursor de subtilisina [Gallagher, 1995].

Es posible que, en las condiciones ácidas de activación, los prosegmentos de cadena larga alteren su conformación (como se ha planteado en la sección precedente) y se alejen de la porción madura de forma tal que el inicio de la cadena madura se coloque suficientemente cerca del sitio activo y en la orientación de un sustrato natural para favorecer así la lisis intramolecular. Esta acción podría ocurrir sobre un enlace alterno al que une al propéptido con la enzima y, en tal caso, alguna molécula activa del medio tendría oportunidad de refinar el corte dando lugar a la enzima madura y funcional que se conoce experimentalmente. Es importante mencionar que el análisis anterior no excluye la posibilidad de un mecanismo simultáneo o incluso alterno de procesamiento *cis* intermolecular en el medio celular.

El hecho de que los zimógenos de ambas subfamilias genéticas se activen en condiciones ambientales similares (valores bajos de pH o presencia de polianiones), sugiere semejanzas en su mecanismo de adquisición de la función biológica. Dado que los prosegmentos del tipo catepsina B tienen una longitud aproximadamente 40% menor a los del tipo papaína y que sus estructuras no conservan el cúmulo de cargas previamente discutido para los propéptidos largos, la porción madura podría ser responsable de la alteración electrostática que induce la activación. Esto se logra posiblemente a través de un segmento de 18 residuos conocido como "asa ocluyente" (figura 9b), que está ausente en las demás proteasas sulfhidrúlicas y que soporta dos residuos con carga eléctrica conservados en este grupo de enzimas. Recientemente se ha reportado información experimental que parece apoyar la participación de este elemento estructural en el proceso de activación del zimógeno de catepsina B inducido por cambios en el pH del medio [Quraishi, 1999]. De esta forma, la activación y el procesamiento parecen ser dos procesos distintos, pero relacionados, que ocurren en forma sucesiva para lograr la conversión de los precursores de las proteasas sulfhidrúlicas en enzimas completamente funcionales.

CAPÍTULO 4

CONCLUSIONES Y PERSPECTIVAS

CONCLUSIONES

En este reporte se describe el curso seguido para predecir características tridimensionales en el prosegmento de papaína. Durante el desarrollo del trabajo se utilizó información experimental del zimógeno, de su prorregión y de los homólogos cercanos. También a través de la labor de predicción se manejaron secuencias de aminoácidos de un gran número de precursores de proteasas cisteínicas y se emplearon diversos programas de cómputo locales y servidores de uso público por red. El diagrama tridimensional de la prorregión de papaína, generado en la primera etapa del trabajo, fue contrastado principalmente con la estructura cristalográfica de la procaricaína, que es la proteína más cercana en secuencia al zimógeno de papaína. A partir de estas comparaciones fue posible plantear la existencia de un cambio conformacional en el propéptido que explicara satisfactoriamente la evidencia experimental disponible sobre los precursores de cadena larga. Finalmente se analizó el proceso de inhibición de la enzima por su prosegmento y se esbozaron vías factibles de activación y autoprocésamiento del zimógeno, así como del plegamiento de la porción madura asistido por el propéptido.

Es importante mencionar que el libre acceso a datos, programas y el enlace a servidores en red se vuelve indispensable cuando se aborda un caso de predicción de proteínas sin homólogos estructurales. En estas condiciones la búsqueda y el examen de las secuencias correspondientes resultan más expeditos, y se hace factible el empleo de una gran variedad de herramientas que difícilmente serían desarrolladas por un solo usuario.

La predicción de un arreglo espacial para el prosegmento de papaína representó un importante desafío debido tanto a la ausencia de una estructura tridimensional que permitiera su modelado por homología, como al tipo y cantidad de información que tuvo que ser empleada durante el desarrollo de cada etapa del proceso. La diversidad en las fuentes de origen y en la extensión de las

secuencias de aminoácidos, así como el considerable número de éstas, requirió de una gran inversión de tiempo en su edición, manejo y análisis final de los resultados. Aunado a ello, la exigencia de algunos formatos particulares para el acceso a programas automáticos de predicción y alineamiento, dificultó el avance y la revisión de la información que fuera significativa para los propósitos de este trabajo.

A pesar de los inconvenientes manifestados en el párrafo precedente, una de las ventajas de la labor de predicción en este trabajo fue precisamente el gran número de homólogos encontrados para el propeptido de papaína. La riqueza de información que se obtiene en estos casos es de gran utilidad para definir características geométricas de una proteína, ya que tanto la identificación de los residuos primordiales como la probable exposición al solvente de los grupos constituyentes de una molécula, sólo resultan evidentes del análisis conjunto de varias cadenas polipeptídicas relacionadas entre sí. A partir de la comparación con las estructuras tridimensionales, puede afirmarse que los datos obtenidos del análisis del alineamiento múltiple permitieron plantear predicciones confiables sobre el carácter anfipático, la longitud y la ubicación de los elementos de estructura secundaria, así como de las posiciones internas y externas en los prosegmentos del tipo papaína. Por ello, el modelo generado con base en estas predicciones describió adecuadamente el acomodo relativo de la primera y segunda hélices, la disposición de esta segunda hélice y la región beta, así como la identificación de la mayor parte de los grupos que establecen interacciones hidrofóbicas en estos elementos geométricos. Es útil mencionar que tanto los programas PHDacc, PHDsec, SSP, SSPred y PREDATOR, como los algoritmos de Benner *et al.* y el criterio de Lim [1974], resultaron ser herramientas de asignación estructural muy confiables en el trabajo de modelado del prosegmento de papaína. Por su parte las mutaciones correlacionadas, que se emplearon como guía auxiliar en el ensamble de los elementos de estructura secundaria, fueron de utilidad, pues a falta de restricciones tridimensionales experimentales ellos permitieron la construcción de un arreglo inicial de la cadena del prosegmento.

Un hecho que se volvió sustancial durante el desarrollo del presente trabajo fue el que las bases de datos representativas en los programas empleados no incluían la información tridimensional del prosegmento de la cathepsina B. Al parecer los criterios de selección en tales programas evitaban reemplazar las proteínas originales por los archivos más completos de homólogos que se publicaban posteriormente. Ello ocasionó que tanto los algoritmos de predicción de estructura secundaria como los programas de reconocimiento de patrones de plegamiento vieran decrecido su potencial para identificar características geométricas en los prosegmentos estudiados. Afortunadamente ya está habiendo cambios favorables en este sentido y algunos programas son ahora más efectivos a partir del uso de bases de datos ampliadas. Como ejemplo de estos pueden citarse los servidores desarrollados por Cuff y Barton [1999], y por Eyrich y Rost [2000].

La realización del trabajo de tesis permitió formar un criterio sobre algunos aspectos de la predicción estructural en secuencias de aminoácidos. Respecto a la identificación de la zona de contacto proteína-proteína es de gran utilidad el conocer la función biológica que desempeñan las moléculas bajo estudio, así como el poseer información experimental que imponga restricciones tridimensionales a los modelos que se forjen durante el trabajo de predicción. Por otra parte, debe tenerse presente que cuando en el análisis de un alineamiento múltiple de secuencias homólogas se ha identificado un grupo de residuos conservados, ellos tienen alta probabilidad de ser esenciales para la estabilidad o para la función de la proteína de interés. Si tales grupos son apolares, ellos podrían conformar el núcleo hidrofóbico de un dominio o bien un sitio de reconocimiento molecular. En otro contexto, cuando una gran cantidad de programas de identificación del patrón de plegamiento reporte resultados de baja significancia estadística es muy probable que la secuencia de la proteína de interés no sea compatible con el universo de arreglos estructurales contenidos en las bases de datos correspondientes. Es importante considerar que las características terciarias de una proteína se encuentran mejor conservadas a lo largo la evolución que su composición o su secuencia de aminoácidos, sin

embargo, aquellas moléculas que comparten una función biológica y poseen alta identidad en secuencia, mantendrán muy seguramente un arreglo tridimensional semejante. Puede concluirse asimismo, que la globularidad de la cadena en una proteína es un requisito común pero no indispensable para que ésta sea funcional. También es útil considerar que la identificación de posiciones que presentan mutaciones correlacionadas a lo largo de un alineamiento de secuencias tiene mayor probabilidad de éxito si el análisis respectivo se verificara sobre conjuntos de proteínas que guarden alta homología relativa y, a fin de abordar los casos donde la naturaleza permite un reacomodo múltiple de cadenas, resultaría muy conveniente desarrollar estrategias de búsqueda de tres o más mutaciones simultáneas. En lo referente al empleo de los sitios de glicosilación como guía para localizar residuos expuestos al solvente, o zonas alejadas del reconocimiento entre proteínas, puede afirmarse que sus resultados son ampliamente confiables y que tal criterio debe ser incorporado a la predicción de características tridimensionales en estas biomoléculas. Finalmente, debido al rápido incremento en las capacidades de cómputo y en la versatilidad de los algoritmos de programación, se espera que los procesos de búsqueda conformacional, identificación de patrones estructurales e incluso detección de unidades de plegamiento independiente, se vuelvan más accesibles y eficientes a la vuelta de algunos años.

Una situación que se manifestó claramente en este trabajo, fue que el análisis conjunto de los resultados de la predicción tridimensional y de los detalles estructurales que brinda la cristalografía resultó tener más información de lo que podrían haber aportado ambas fuentes individualmente. Por ejemplo, al analizar la estructura tridimensional de la procaricaína es posible advertir que el segmento S5 se encuentra interaccionando con el subsitio 2 de la enzima [Baker, 1987]. Ello permite suponer que tal segmento tiene participación en la especificidad del prosegmento al ejercer su acción como inhibidor; lo cual explicaría la marcada variabilidad en la composición y en la extensión de este segmento a lo largo del alineamiento múltiple de proregiones de proteasas sulfhidríticas. En consecuencia, al realizar mutaciones en esta región de la cadena es posible que

se modificara la afinidad de los prosegmentos y que se lograra así mayor versatilidad en la inhibición de las proteasas sulfhidríticas. De manera análoga es probable que la función de los segmentos S2 y S4 en el propéptido, sea la de mantener la distancia y orientación adecuadas en la constelación de los grupos cargados conservados, y quizá por ello la extensión de estos segmentos se haya mantenido prácticamente invariante a lo largo de la evolución. El análisis realizado sobre el alineamiento múltiple de prosegmentos permitió identificar un conjunto de residuos ácidos y básicos conservados en las secuencias analizadas y este resultado, aunado a la localización externa de los mismos grupos en los precursores con estructura tridimensional caracterizada, hizo posible el planteamiento de una de las contribuciones más importantes de este trabajo de tesis que se refiere a una probable modificación conformacional del prosegmento, inducida por cambios en el medio que alteren las interacciones electrostáticas. Esta modificación conformacional es capaz de explicar satisfactoriamente las observaciones experimentales reportadas para diferentes prosegmentos y precursores de proteasas sulfhidríticas y sugiere que, en ausencia de la enzima madura, los prosegmentos de cadena larga pueden adoptar una estructura globular. Tal posibilidad representa una nueva vía en el estudio de la estabilidad estructural de un polipéptido que desempeña diversas funciones de elevada importancia biológica.

A partir de las estructuras cristalográficas reportadas para las enzimas maduras de proteasas sulfhidríticas resulta evidente que la cadena polipeptídica en esas proteínas sufre un plegamiento no secuencial. Además, de acuerdo con un estudio geométrico hecho en computadora sobre caricaína, es en principio posible que el procesamiento del precursor se lleve a cabo por el propio sitio activo de la enzima madura que constituye el zimógeno. También parece probable que en este grupo de proteínas, tanto la activación como el procesamiento sean eventos distintos pero que ocurren en forma consecutiva para rendir la enzima madura a partir del precursor.

De manera general puede afirmarse que un trabajo de predicción sin homólogos estructurales, como el que se reporta en este escrito, produce una

serie de datos que pueden aplicarse a la descripción de características tridimensionales del conjunto de proteínas empleadas en el mismo estudio. En estos casos los métodos de simulación por computadora se convierten en útiles herramientas, complementarias a los estudios experimentales, que permiten tener una visión alterna de las moléculas bajo estudio.

PERSPECTIVAS

Dada la gran cantidad de fenómenos por explicar que presentan los prosegmentos de proteasas sulfhidrúlicas, puede hacerse una clasificación de los factores por considerar en el estudio futuro de esos polipéptidos y de los precursores correspondientes.

Respecto a la estabilidad intrínseca y al cambio conformacional propuesto para los prosegmentos de cadena larga, podría estudiarse el efecto del cambio en las condiciones del medio (fuerza iónica, pH y concentración de cosolutos) sobre diferentes observables tales como la señal de dicroísmo circular en el ultravioleta lejano y cercano; la radiación fluorescente emitida por los cromóforos del prosegmento de papaína; e incluso, la fluorescencia del ANS en presencia del propéptido. Asimismo podrían mutarse los grupos cargados del propéptido de papaína para evaluar la contribución de cada uno de ellos a la estabilidad molecular. Del mismo modo, sería muy interesante observar el comportamiento de la cadena polipeptídica al mutar el residuo 71 del prosegmento, el cual se mantiene conservado a lo largo de la evolución como un grupo pequeño en la huella 3 sin que exista una razón clara de su función; aunque de acuerdo con su ubicación en el segmento S4, este residuo quizá sea imprescindible para la movilidad de la hélice 3. Una modificación química que se puede intentar sobre el prosegmento de papaína es el entrecruzamiento de grupos que inmovilicen a la hélice 3 y que permitan validar la hipótesis planteada a raíz de la simulación por computadora. Por su parte, también sería relevante estudiar el desplegamiento térmico del propéptido de papaína en diferentes condiciones del medio.

Un aspecto muy amplio por estudiar es el reconocimiento de la enzima madura por el prosegmento en los procesos de inhibición, plegamiento y replegamiento de la cadena. En este sentido, podría estudiarse el cambio en la afinidad enzima-prosegmento a partir de mutaciones sobre los grupos aromáticos que se anclan en la región interfacial o sobre los grupos del segmento S5 que interactúan con el subsitio S2 catalítico. El proceso de replegamiento de la cadena madura por el prosegmento, observado para otras familias de proteasas, tendría oportunidad de investigarse en la familia de la papaína bajo diferentes condiciones de trabajo donde se podría incluso mantener a la proteína parcialmente desplegada al momento de ser reconocida por el prosegmento, o bien, donde se prescindiera completamente del propéptido. Sería factible también trabajar con fragmentos de la cadena del prosegmento de papaína y estudiar la afinidad que desarrollarían por la enzima madura y asimismo podría analizarse el efecto de la pérdida de los últimos 25 residuos de la cadena del prosegmento sobre la afinidad molecular. De ser posible, resultaría muy interesante separar los dominios estructurales de la porción madura de papaína y revisar tanto su estabilidad térmica relativa como el reconocimiento que experimentarían por parte del propéptido correspondiente.

Por otra parte, el proceso de autoactivación podría ser abordado de forma que indicara la viabilidad del mecanismo *cis* intramolecular al disminuir el contacto entre los precursores en solución, ya sea por su dilución, por el incremento en la viscosidad del medio, o al impedir la migración de los primeros 12 residuos de la cadena madura. También sería interesante determinar la capacidad del precursor para hidrolizar su propia cadena en sentido inverso al de un sustrato natural.

En lo que respecta a la caracterización experimental de la estructura del prosegmento, es posible proyectar estudios de difracción de rayos X y de resonancia magnética nuclear de la cadena. El primer método permitiría determinar la estructura del prosegmento con detalle atómico y el segundo haría posible conocer además el efecto de la variación de las condiciones del medio de trabajo sobre la geometría molecular del propéptido en solución.

Finalmente es posible diseñar estudios por computadora para la simulación tanto del desplegamiento de la cadena madura de papaína, como del efecto del pH sobre la conformación del prosegmento de caricaína en presencia de moléculas de solvente.

CAPÍTULO 5

MATERIALES Y MÉTODOS

ANÁLISIS DE SECUENCIAS Y DE ESTRUCTURAS TRIDIMENSIONALES

Las secuencias homólogas utilizadas en este trabajo fueron identificadas por el programa BLAST [Altschul, 1990] sobre una base de datos de proteínas no redundante que incluyó al banco de secuencias de proteínas Swiss-Prot [Bairoch, 1998], al Protein Identification Resource [George, 1997] y al GenBank [Benson, 1998] (secuencias cuyo código inicia con gp en el alineamiento múltiple de la figura 1). Las estructuras primarias de propapaína, procaricaína, procatepsina L humana y procatepsina B de rata que se utilizaron a lo largo de este trabajo fueron tomadas de la base de datos Swiss-Prot. Los códigos de identificación correspondientes a estas proteínas, que se emplean en el alineamiento múltiple, son: PAPA_CARPA, PAP3_CARPA, CATL_HUMAN y CATB_RAT, respectivamente.

La predicción de globularidad a partir de la secuencia de aminoácidos del propéptido de papaína se llevó a cabo con el programa GLOBE [Rost, 1998]; el cual realiza una predicción de accesibilidad sobre la proteína de interés y la compara con la correlación experimental entre el área expuesta y la masa molecular de proteínas globulares caracterizadas tridimensionalmente.

Todas las coordenadas cristalográficas empleadas en este trabajo fueron obtenidas del Protein Data Bank [Bernstein, 1977], con los siguientes códigos de identificación: 1PCI, procaricaína; 1MIR, procatepsina B de rata; 1CJL, procatepsina L humana; 1STF, complejo papaína-estefina; 9PAP, papaína madura y 1PPO, caricaína madura. El banco de datos PDB fue revisado con el programa BLAST en busca de una estructura de base para el modelado por comparación que se menciona en el texto.

Para el reconocimiento del patrón de plegamiento se emplearon diversos programas, tales como: Hidden Markov Model [Krogh, 1994; Eddy, 1997], TOPITS [Rost, 1995], THREADER [Jones, 1992], Protein Fold Identification Tool (PROFIT) [Sippl, 1996], MAP [Russell, 1996], DOE [Fisher, 1996] y el método de Nishikawa y

Matsuo [Nishikawa, 1993]. Estos programas tienen el objetivo de identificar estructuras terciarias compatibles con la secuencia de la proteína problema a través de la generación de perfiles, consensos y predicciones de estructura secundaria y de exposición al solvente, así como de la evaluación de potenciales de contacto interresiduo.

Los alineamientos múltiples de secuencias de aminoácidos se realizaron con los programas Clustal W [Thompson, 1994] y MaxHom (del servidor Predict Protein). Ambos algoritmos emplean alineamientos locales y similitud de aminoácidos para construir los grupos crecientes de secuencias alineadas. La clasificación en subconjuntos de proteínas, empleados en este trabajo, hizo uso del servidor Darwin [Gonnet, 1992] y del programa Sequence Space [Casari, 1995], los cuales agrupan de manera progresiva las secuencias que mantienen la mayor homología o cantidad de rasgos invariantes.

Para la predicción de la estructura secundaria del prosegmento de papaína se empleó el grupo de métodos SOPM [Self Optimized Prediction Method] que incluye seis diferentes algoritmos: Gibrat *et al.* [1987], Levin *et al.* [1986], Double Prediction Method, DPM, [Deleage, 1987]; Self-optimized Method by Consensus Prediction from Multiple Alignment SOPMA, [Geourjon, 1995], Prediction at Heidelberg, PHDsec, [Rost, 1994 a] y Neural Network Prediction, NNPred, [Kneller, 1990]. Adicionalmente se utilizaron los métodos: SSPREDICT, SSPre, [Mehta, 1995], Quadratic Logistic, QL, [Di Francesco, 1995]; ZPREDICTION, Zpred, [Zvelebil, 1987]; Discrete State-Space, DSM, [Stultz, 1993; White, 1994]; PREDATOR, PRED, [Frishman, 1997] y los algoritmos SSP, Nearest-Neighbor Secondary Structure Prediction, NNSSP, y Secondary Structure Prediction from Alignments, SSPAL [Solovyev, 1994; Salamov, 1995]. Estos programas de predicción se basan en una gran variedad de métodos (redes neuronales, potenciales estadísticos, homología parcial con estructuras cristalográficas, etc.) y manejan como datos de entrada tanto secuencias individuales como alineamientos múltiples de proteínas homólogas. Además, en este trabajo se empleó el método de predicción manual de límites de estructuras secundarias desarrollado por Benner *et al.* [1991] en el que se analizan alineamientos múltiples de grupos de

proteínas cercanamente relacionadas. Un límite altamente probable se establece cuando en un alineamiento de proteínas homólogas se presenta indefectiblemente un residuo de prolina o de glicina adyacente a una posición variable en todas las secuencias analizadas. Un límite posible se presenta cuando sólo algunos miembros del conjunto estudiado muestran tal variación.

Para evaluar las predicciones realizadas sobre el prosegmento de papaína se empleó fundamentalmente la estructura cristalográfica de procaricaína, por ser ésta la proteína con mayor identidad en secuencia respecto al zimógeno de papaína. La evaluación de la exactitud en la predicción de estructura secundaria sobre el prosegmento de estudio se basó en el número de residuos correctamente asignados en conformación hélice alfa, hoja beta e irregular, sobre un segmento de 95 residuos (aminoácidos 12 al 107) del propéptido de papaína. Los primeros 11 residuos no fueron considerados en esta evaluación debido a que están ausentes en el archivo cristalográfico de referencia.

Las predicciones de la accesibilidad al solvente hechas sobre la secuencia de la porción madura y del propéptido de papaína fueron realizadas por el algoritmo de redes neuronales del programa PHDacc [Rost, 1994 b] y por el método manual propuesto por Benner *et al.* [1994], en el que se buscan residuos que conserven características hidrofóbicas (AILMYF, en código de una letra) o hidrofílicas (KREND) en una misma posición a lo largo de un alineamiento múltiple de secuencias con alta homología relativa.

Con el fin de evaluar la exactitud de las predicciones sobre la exposición de residuos en el prosegmento y para analizar la zona de contacto con la enzima madura, se llevaron a cabo cálculos de accesibilidad al solvente sobre las estructuras cristalográficas de las porciones maduras de papaína y caricaína. Estos cálculos emplean una versión del algoritmo de Lee y Richards [1971] que está instrumentado en el programa NACCESS [S. Hubbard, University College, London, UK]. Los parámetros empleados en este programa para la determinación del área expuesta fueron: una esfera de prueba con radio de 1.4 Å y un ancho de corte de 0.05 Å. Los resultados se reportan como área total y como área relativa a la máxima exposición posible; la cual fue estimada como aquella que exhibe la

cadena lateral del aminoácido central en un tripéptido Ala-X-Ala completamente extendido. Las cadenas laterales se consideraron ocultas cuando presentaron áreas relativas menores a 35%. La superficie de los átomos de carbono y azufre en cada aminoácido se consideró apolar, mientras que el área polar se estimó a partir de la suma de contribuciones de átomos de nitrógeno y oxígeno. El cálculo del área expuesta por el prosegmento de caricaína en la zona interfacial con la enzima se determinó como la diferencia entre las áreas del prosegmento aislado de la enzima madura y la correspondiente al prosegmento en el zimógeno. A fin de identificar los probables sitios de contacto con la prorrégion, se contrastó la predicción de la exposición al solvente, realizada en el trabajo, con los correspondientes valores obtenidos del programa NACCESS sobre la estructura tridimensional de papaína. Por su parte, las mutaciones correlacionadas dentro de las secuencias del alineamiento múltiple fueron estimadas con el algoritmo de Göbel *et al.* [1994] a través de la identificación de parejas de posiciones que presentaran variación de manera simultánea. La búsqueda de contactos beta en la cadena del prosegmento fue realizada con el método estadístico de Hubbard [1995] mediante la construcción de un mapa de interacciones entre pares de residuos. Adicionalmente, la identificación de firmas o huellas características de aminoácidos sobre las secuencias del alineamiento múltiple se llevó a cabo con el programa y base de datos PROSITE [Bairoch, 1993].

Las unidades de plegamiento independiente fueron detectadas con un programa de confección propia que se basó en un algoritmo previamente publicado [Moult, 1991; Pedersen, 1995]. Este programa realizó un análisis de la exposición de área apolar al solvente en ventanas de 12 aminoácidos. De acuerdo con los autores, un segmento cuya conformación nativa oculte en promedio más de 50 Å² de área apolar por residuo, respecto a la cadena desplegada, puede considerarse como una unidad de plegamiento independiente, y por tanto, como un probable sitio de nucleación durante el proceso de adquisición de la estructura final.

SIMULACIÓN MOLECULAR

La visualización y el modelado molecular se realizaron en un sistema Indigo² de Silicon Graphics con el paquete de cómputo BIOGRAF [Molecular Simulations]. Las minimizaciones de energía y las simulaciones de dinámica molecular emplearon el campo de fuerzas Dreiding II [Mayo, 1990]. Las etapas de cálculo en las regularizaciones de la geometría se efectuaron hasta alcanzar fuerzas menores a $4 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. Las simulaciones de dinámica molecular fueron realizadas por integración numérica de las ecuaciones del movimiento de Newton con el empleo del algoritmo de Verlet [1967]. Se empleó un incremento de tiempo de 1 fs con una temperatura de 300 K, en condiciones isotérmicas, y las velocidades se actualizaron cada 0.1 ps. Solamente se consideraron las interacciones no enlazadas a distancias menores a 0.90 nm y, a fin de evitar discontinuidades en las funciones de energía potencial, se empleó una función decreciente desde 0.80 nm hasta 0.85 nm. Durante el proceso de evaluación geométrica del modelo tridimensional del propéptido de papaína, fueron incluidos explícitamente todos los átomos de hidrógeno y, por economía computacional, los cálculos fueron realizados en vacío. Para simular los efectos de atenuación de carga que realiza el solvente, durante los cálculos de energía molecular se utilizó una constante dieléctrica dependiente de la distancia.

El cambio conformacional del segmento N terminal de papaína consistió exclusivamente en el movimiento de los 12 primeros residuos de la estructura cristalográfica de la porción madura. Esta modificación se completó después de varios ciclos alternados de regularización local de la estructura y de simulaciones de dinámica molecular con restricciones armónicas de distancia entre átomos.

Las dos simulaciones efectuadas sobre la conformación cristalográfica de procaricaína involucraron la región globular del prosegmento (residuos 15 a 82). Las condiciones ácida y fisiológica del medio fueron modeladas bajo la asignación de cargas netas en los grupos carboxilato con valores de 0 y -1, respectivamente. Ambas simulaciones de dinámica molecular se realizaron después de varias etapas de regularización geométrica y evolucionaron por 200 ps a 300 K.

La asignación de elementos de estructura secundaria al archivo cristalográfico de procaricaína se llevó a cabo con el programa PROCHECK [Morris, 1992; Laskowski, 1993]; mientras que el análisis del empaquetamiento molecular del prosegmento de esta proteína se realizó a través de la construcción de poliedros de Voronoi, los cuales permiten calcular el volumen de los átomos internos en una estructura tridimensional [Pontius, 1996].

El alineamiento topológico que se presenta en la figura 9 fue realizado por el servidor DALI [Holm, 1993] a partir de las coordenadas cristalográficas de los prosegmentos de caricaína y de catepsina B. El conjunto de programas contenidos en este servidor realiza comparaciones para identificar las similitudes espaciales existentes entre estructuras tridimensionales de proteínas.

PURIFICACIÓN Y CARACTERIZACIÓN DEL PROSEGMENTO DE PAPAÍNA

El prosegmento de papaína, empleado para fines experimentales, consistió en una cadena polipeptídica de 107 residuos cuya secuencia de aminoácidos se muestra en la figura 1 (primer renglón del alineamiento múltiple) y fue expresado en el laboratorio del Dr. Ruy Pérez Montfort (Instituto de Fisiología Celular, UNAM). El material insoluble obtenido fue disuelto y filtrado antes de ser sometido a separación cromatográfica en un equipo HPLC Varian 5000 con columna de DEAE para intercambio aniónico. La detección se realizó de manera continua por absorción a 280 nm y la fracción colectada fue sometida nuevamente a cromatografía, obteniéndose el gráfico que se muestra en la figura 5a.

Tanto la electroforesis como la determinación del punto isoeléctrico de la muestra fueron realizadas en un equipo Phast System (Pharmacia) en geles de poliacrilamida tipo homogéneo 20 e IEF 4-6.5, respectivamente. En ambas determinaciones se emplearon los marcadores estándar recomendados por el fabricante del equipo.

El coeficiente de extinción, empleado para estimar la concentración de las soluciones del prosegmento, tuvo un valor de 1.7 a 280 nm y 1 mg/mL y fue calculado a partir de la composición de aminoácidos de acuerdo con el método de Gill y von Hippel [1989].

Los ensayos de actividad proteolítica de papaína, a una concentración de 0.9 μM , se realizaron en regulador de fosfatos 0.05 M y pH 6.5 con incubación de la enzima por 10 minutos en una solución activadora que contenía 10 mM de cisteína y 1 mM de EDTA. Esta mezcla se colocó en un espectrofotómetro Shimadzu 160U y posteriormente se le adicionaron 400 μL del sustrato (α -N-benzoil-DL-arginina-p-nitroanilida, BAPNA) para lograr una concentración final de 0.5 mM. La generación del producto de hidrólisis fue seguida durante 15 minutos a 400 nm. El gráfico obtenido en estas condiciones mostró una tendencia lineal y se consideró como el total de la actividad de la muestra (100%). Para determinar la capacidad inhibitoria del prosegmento, éste fue añadido a la mezcla de incubación, previo a la adición del sustrato. La solución contenía una mezcla equimolar de enzima y prosegmento (0.9 μM).

ESTUDIOS ESPECTROSCÓPICOS

El espectro de dicroísmo circular del prosegmento de papaína fue registrado a 25 °C en un espectropolarímetro JASCO J-500 A calibrado con ácido [+]10-canforsulfónico [Hennessey, 1982]. Las mediciones se realizaron en la región del espectro electromagnético comprendida entre 190 y 270 nm, con soluciones de 0.1 mg/ml del prosegmento en regulador de fosfatos 0.05 M a pH 6.5 y en una celda de 0.05 cm de recorrido óptico. La determinación del contenido de estructura secundaria sobre el espectro de dicroísmo circular se realizó de acuerdo con el método de Hennessey y Johnson [Hennessey, 1981] y con el empleo del programa k2d [Andrade, 1993]. En ambos casos se consideró una masa molar promedio de 118.5 Da para el cálculo de la elipticidad media por residuo. En los experimentos de desnaturalización térmica, la muestra fue colocada en un portaceldas con recirculación de agua y la temperatura de la solución fue determinada con un termistor. Para el análisis de los datos se restó un espectro control obtenido al registrar la señal proveniente de una solución reguladora contenida en la misma celda experimental.

Los datos de emisión de la fluorescencia intrínseca del prosegmento de papaína fueron registrados en el intervalo de 320 a 440 nm en un

espectrofluorómetro ISS PC1. La muestra fue excitada a 295 nm con un ancho de banda de 5 nm. La concentración del prosegmento en la muestra fue de 10 µg/mL en una solución reguladora de fosfatos pH 6.5. Este mismo regulador fue empleado como testigo. El espectro de fluorescencia intrínseca fue descompuesto en una combinación lineal de los espectros reportados por Permyakov [1993] para triptofanos en ambientes hidrofóbico y acuoso ($\lambda_{\text{máx}} = 317$ y 352 nm, respectivamente). Los ensayos de fluorescencia del ácido 8-anilino-naftalen-1-sulfónico (ANS), se realizaron manteniendo la concentración de esta molécula de prueba en un exceso 20 molar respecto al prosegmento de papaína. El espectro de emisión se registró en el intervalo de 420 a 620 nm con una longitud de onda de excitación de 380 nm [Semisotnov, 1991] en un intervalo de valores de pH de 6.7 a 3. Un espectro blanco fue sustraído para obtener la información de la radiación fluorescente neta debida a la interacción del ANS con la proteína.

BIBLIOGRAFÍA

- ACE. Agriculture Commercialization and Enterprise. Agribusiness Information Centre (AIC) Market (<http://www.agroindia.org/world/ace/papain.htm>).
- Albrecht, K., University of Wisconsin - Eau Claire, Department of Chemistry (<http://www.chem.uwec.edu/Chem406/Webpages/KAREN/facts.html>).
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. y Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, **215**: 403-410.
- Andrade, M.A., Chacón, P., Merelo, J.J. y Morán, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng.*, **6**: 383-390.
- Auerswald, E.A., Nagler, D.K., Assfalg-Machleidt, I., Stubbs, M.T., Machleidt, W. y Fritz, H. (1995). Hairpin loop mutations of chicken cystatin have different effects on the inhibition of cathepsin B, cathepsin L and papain. *FEBS Lett.*, **361**: 179-184.
- Aurora, R., Creamer, T.P., Srinivasan, R. Y Rose, G.D. (1997). Local interactions in protein folding: lessons from the alpha-helix. *J. Biol. Chem.*, **272**: 1413-1416.
- Aurora, R. y Rose, G.D. (1998). Helix capping. *Protein Sci.*, **7**: 21-38.
- Bairoch, A. (1993). The PROSITE dictionary of sites and patterns in proteins, its current status. *Nucleic Acids Res.*, **21**: 3097-3103.
- Bairoch, A. y Apweiler, R. (1998). The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998. *Nucleic Acids Res.*, **26**: 38-42.
- Baker, D., Silen, J.L. y Agard, D.A. (1992). Protease pro region required for folding is a potent inhibitor of the mature enzyme. *Proteins*, **356**: 263-265.
- Baker, D., Shiau, A.K. y Agard, D.A. (1993). The role of regions in protein folding. *Curr. Op. Cell Biol.*, **5**: 966-970.
- Baker, E.N. y Drenth, J. (1987). (En) Biological Macromolecules and Assemblies John Wiley & Sons, N.Y., Vol. 3, 313-368.
- Benner, S.A. y Gerloff, D. (1991). Patterns of divergence in homologous proteins as indicators of secondary and tertiary structure: a prediction of the structure of catalytic domain of protein kinases, *Adv. Enzyme Regul.*, **31**: 121-181.

- Benner, S.A., Badcoe, I., Cohen, M.A. y Gerloff, D.L. (1994). *Bona fide* prediction of aspects of protein conformation. Assigning interior and surface residues from patterns of variation and conservation in homologous protein sequences. *J. Mol. Biol.*, **235**: 926-958.
- Benner, S.A. (1995). Predicting the conformation of proteins from sequences, progress and future progress, *J. Mol. Rec.*, **8**: 9-28.
- Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J. y Ouellette B.F. (1998). GenBank. *Nucleic Acids Res.*, **26**: 1-7.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.E. Jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. y Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **112**: 535-542.
- Branden, C. y Tooze, J. (1991). Introduction to protein structure. Garland Publishing, Inc., Nueva York y Londres, 1a. edición.
- Brown, J.E. y Werner, A.K. (1971). Helix-coil transition of the isolated amino terminus of ribonuclease. *Biochemistry*, **10**: 470-476.
- Bryan. P., Alexander, P., Strausberg, S., Schwartz, F., Lan, W., Gilliland, G. y Gallagher, D.T. (1992). Energetics of folding subtilisin BPN'. *Biochemistry*, **31**: 4937-4945.
- Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J.S. y Ménard, R. (1996). Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. *Biochemistry*, **35**: 8149-8157.
- Casari, G., Sander, C. y Valencia, A. (1995). A method to predict functional residues in proteins. *Nat. Struct. Biol.*, **2**: 171-178.
- Chapman, R.L., Kane, S.E. y Erickson, A.H. (1997). Abnormal glycosilation of procathepsin L due to N-terminal point mutations correlates with failure to sort to lysosomes. *J. Biol. Chem.*, **272**: 8808-8816.
- Cheng, C.Y., Morris, I. y Bardin, C.W. (1993). Testins are structurally related to the mouse cysteine proteinase precursor, but devoid of any protease/antiprotease activity. *Biochem. Biophys. Res. Commun.*, **191**: 224-231.
- Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J.S. y Cygler, M. (1996). Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J.*, **15**: 5492-5503.

- Creighton, T.E. (1993). *Proteins. Structures and Molecular Properties*, Freeman, 2ª edición, ISBN 0-7167-7030-X.
- Cuff, J.A. y Barton, G.J. (1999). Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins*, **34**: 508-519.
- Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A.C. y Mort, J.S. (1996). Structure of rat procathepsin B: Model for inhibition of cysteine protease activity by the proregion. *Structure*, **4**: 405-416.
- Deerland Corporation (<http://www.deerland-enzymes.com/>).
- Delaria, K., Fiorentino, L., Wallace, L., Tamburini, P., Brownell, E. y Muller, D. (1994). Inhibition of cathepsin L-like cysteine proteases by cytotoxic T-lymphocyte antigen-2 beta. *J. Biol. Chem.*, **269**: 25172-25177.
- Deleage, G. y Roux, B. (1987). An algorithm for protein secondary structure prediction based on class prediction. *Protein Eng.*, **1**: 289-294.
- Di Francesco, V., Munson, P.J. y Garnier, J. (1995). Use of multiple alignments in protein secondary structure prediction. 28th Hawaii International Conference on System Sciences, **5**: 285-291. IEEE, Los Alamitos, CA.
- Dobson, C.M., Evans, P.A. y Radford, S.E. (1994). Understanding how proteins fold: the lysozyme story so far. *Trends Biochem Sci.*, **1**: 31-37.
- Doran, J.D., Nomizu, M., Takebe, S., Ménard, R., Griffith, D. y Ziomek, E. (1999). Autocatalytic processing of the streptococcal cysteine protease zymogen. Processing mechanism and characterization of the autoproteolytic cleavage sites. *Eur. J. Biochem.*, **263**: 145-151.
- Eddy, S.R. (1997). Hidden Markov models. *Curr. Opin. Struct. Biol.*, **6**: 361-365.
- Erickson, A.H. (1989). Biosynthesis of lysosomal endopeptidases. *J. Cell. Biochem.*, **40**: 31-41.
- Eyrich, V. y Rost, B. (2000). META server (http://cubic.bioc.columbia.edu/predictprotein/doc/meta_intro.html)
- Fischer, D. y Eisenberg, D. (1996). Fold recognition using sequence-derived predictions. *Protein Sci.*, **5**: 947-955.
- Fox, T., de Miguel, E., Mort, J.S. y Storer, A.C. (1992). Potent slow-binding inhibition of cathepsin B by its propeptide. *Biochemistry*, **31**: 12571-12576.

- Frishman, D. y Argos, P. (1997). Seventy-five percent accuracy in protein secondary structure prediction. *Proteins*, **27**: 329-335.
- Fusek, M., Mares, M., Vagner, J., Voburka, Z. y Baudys, M. (1991). Inhibition of aspartic proteinases by propart peptides of human procathepsin D and chicken pepsinogen. *FEBS Lett.*, **287**: 160-162.
- Gallagher, T., Gilliland, G., Wang, L y Bryan, P. (1995). The prosegment-subtilisin BPN' complex: crystal structure of a specific 'foldase'. *Structure*, **3**: 907-914.
- George, D.G., *et al.* y Barker W.C. (1997). The Protein Information Resource (PIR) and the PIR-International Protein Sequence Database. *Nucleic Acids Res.*, **25**: 24-28.
- Geourjon, C. y Deléage, G. (1995). SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.*, **11**: 681-684.
- Gibrat, J.F., Garnier, J. y Robson, B. (1987). Further developments of protein secondary structure prediction using information theory. New parameters and consideration of residue pairs. *J. Mol. Biol.*, **198**: 425-443.
- Gill, S.C. y von Hippel, P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.*, **182**: 319-326.
- Göbel, U., Sander, C., Schneider, R. y Valencia, A. (1994). Correlated mutations and residue contacts in proteins. *Proteins*, **18**: 309-317.
- Gonnet, G.H., Cohen, M.A. y Benner, S.A. (1992). Exhaustive matching of the entire protein sequence database. *Science*, **256**: 1443-1445.
- Groves, M.R., Taylor, M.A.J., Scott, M., Cummings, N.J., Pickersgill, R.W. y Jenkins, J.A. (1996). The prosequence of procaricain forms an α -helical domain that prevents access to the substrate-binding cleft. *Structure*, **4**: 1193-1203.
- Hennessey, J.P. Jr. y Johnson, W.C. Jr. (1981). Information content in the circular dichroism of proteins. *Biochemistry*, **20**: 1085-1094.
- Hennessey, J.P., Jr. y Johnson, W.C., Jr. (1982). Experimental errors and their effect on analyzing circular dichroism spectra of proteins. *Anal. Biochem.*, **125**: 177-188.
- Hofmann, K., Bucher, P., Falquet, L. y Bairoch, A. (1999). The PROSITE database, its status in 1999. *Nucleic Acids Res.*, **27**: 215-219.

- Holm, L. y Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**: 123-138.
- Holwerda, B.C., Padgett, H.S. y Rogers, J.C. (1990). Proaleurain vacuolar targeting is mediated by short contiguous peptide interactions. *Plant Cell.*, **4**: 307-318.
- Hu, Z, Zhu, X., Jordan, F. e Inouye, M. (1994). A covalently trapped folding intermediate of subtilisin E: spontaneous dimerization of a prosubtilisin E Ser49Cys mutant *in vivo* and its autoprocessing *in vitro*. *Biochemistry*, **33**: 562-569.
- Hubbard, T.J. y Park, J. (1995). Fold recognition and ab initio structure predictions using hidden Markov models and beta-strand pair potentials. *Proteins*, **3**: 398-402.
- Ishido, K. y Kominami, E. (1995). Procathepsin L degrades extracellular matrix proteins in the presence of glycosaminoglycans *in vitro*. *Biochem. Biophys. Res. Commun.*, **217**: 624-631.
- Jerala, R., Zerovnik, E., Kidric, J. y Turk, V. (1998). pH-induced conformational transitions of the propeptide of human cathepsin L. *J. Biol. Chem.*, **273**: 11498-11504.
- Jones, D.T., Taylor, W.R. y Thornton, J.M. (1992). A new approach to protein fold recognition. *Nature*, **358**: 86-89.
- Kalinski, A., Melroy, D.L., Dwivedi, R.S. y Herman, E.M. (1992). A soybean vacuolar protein (P34) related to thiol proteinases is synthesized as a glycoprotein precursor during seed maturation. *J. Biol. Chem.*, **267**: 12068-12076.
- Karplus, M. y Sali, A. (1995). Theoretical studies of protein folding and unfolding. *Curr. Op. Struct. Biol.*, **5**: 58-73.
- Karrer, K.M., Pfeiffer, S.L. y DiTomas, M.E. (1993). Two distinct gene subfamilies within the family of cysteine protease genes. *Proc. Natl. Acad. Sci. USA*, **90**: 3063-3067.
- Kassell, B. y Kay, J. (1973). Zymogens of proteolytic enzymes. *Science*, **180**: 1022-1027.
- Kneller, D.G., Cohen, F.E. y Langridge, R. (1990). Improvements in protein secondary structure prediction by an enhanced neural network. *J. Mol. Biol.*, **214**: 171-182.

- Krogh, A., Brown, M., Mian, I.S., Sjölander, K. y Haussler, D. (1994). Hidden Markov models in computational biology. Applications to protein modeling. *J. Mol. Biol.*, **235**: 1501-1531.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. y Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**: 283-291.
- Lee, B. y Richards, F.M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**: 379-400.
- Levin, J.M., Robson, B. y Garnier, J. (1986). An algorithm for secondary structure determination in proteins based on sequence similarity. *FEBS Lett.*, **205**: 303-308.
- Levinthal, C. (1968). Are there pathways in protein folding? *J. Chim. Phys.*, **65**: 44-45.
- Lim, V. (1974). Algorithms for prediction of alpha-helical and beta-structural regions in globular proteins. *J. Mol. Biol.*, **88**: 873-894.
- Mach, L., Mort, J.S. y Glössl, J. (1994). Maturation of human procathepsin B. Proenzyme activation and proteolytic processing of the precursor to the mature proteinase, *in vitro*, are primarily unimolecular processes. *J. Biol. Chem.*, **269**: 13030-13035.
- Mason, R.W. y Massey, S.D. (1992). Surface activation of pro-cathepsin L. *Biochem. Biophys. Res. Commun.*, **189**: 1659-1666.
- Mayo, S.L., Olafson, B.D. y Goddard, W.A. III. (1990). Dreiding: A generic Forcefield for molecular simulations. *J. Phys. Chem.*, **94**: 8897-8909.
- Mehta, P.K., Heringa, J. y Argos, P. (1995). A simple and fast approach to prediction of protein secondary structure from multiply aligned sequences with accuracy above 70%. *Protein Sci.*, **4**: 2517-2525.
- Morea, V., Leplae, R. y Tramontano, A. (1998). Protein structure prediction and design. *Biotech. Ann. Rev.*, **4**: 177-214.
- Morris, A.L., MacArthur, M.W., Hutchinson, E.G. y Thornton, J.M. (1992). Stereochemical quality of protein structure coordinates. *Proteins*, **12**: 345-364.
- Mosimann, S., Meleshko, R. y James, M.N.G. (1995). A critical assesment of comparative molecular modeling of tertiary structures of proteins. *Protein Struct. Funct. Genet.*, **23**: 301-317.

- Moult, J. y Unger, R. (1991). An analysis of protein folding pathways. *Biochemistry*, **30**: 3816-3824.
- Neurath, H. (1989). Proteolytic processing and physiological regulation. *Trends Biochem. Sci.*, **14**: 268-271.
- Nishikawa, K. y Matsuo, Y. (1993). Development of pseudoenergy potentials for assessing protein 3-D-1-D compatibility and detecting weak homologies. *Protein Eng.*, **6**: 811-820.
- Parker, M.H. y Hefford, M.A. (1997). A consensus residue analysis of loop and helix-capping residues in four-alpha-helical-bundle proteins. *Protein Eng.*, **10**: 487-496.
- Pazos, F., Helmer-Citterich, M., Ausiello, G. y Valencia, A. (1997). Correlated mutations contain information about protein-protein interaction. *J. Mol. Biol.*, **271**: 511-523.
- Pedersen, J.T. y Moult, J. (1995). *Ab initio* structure prediction for small polypeptides and protein fragments using genetic algorithms. *Proteins*, **23**: 454-460.
- Permyakov, E.A. (1993). *Luminescent Spectroscopy of Proteins*. CRC Press, Boca Raton, Florida.
- Peters, R.J., Shiau, A.K., Sohl, J.L., Anderson, D.E., Tang, G., Silen, J.L. y Agard, D.A. (1998). Pro region C-terminus: protease active site interactions are critical in catalyzing the folding of α -lytic protease. *Biochemistry*, **37**: 12058-12067.
- Pol, E., Olsson, S.L., Estrada, S., Prasthofer, T.W. y Björk, I. (1995). Characterization by spectroscopic, kinetic and equilibrium methods of the interaction between recombinant human cystatin A (stefin A) and cysteine proteinases. *Biochem. J.*, **311**: 275-282.
- Pontius, J., Richelle, J. y Wodak, S. (1996). Deviations from standard atomic volumes as a quality measure for protein folds. *Nucleic Acids Res.*, **25**: 231-234.
- Quraishi, O., Nagler, D.K., Fox, T., Sivaraman, J., Cygler, M., Mort, J.S. y Storer, A.C. (1999). The occluding loop in cathepsin B defines the pH dependence of inhibition by its propeptide. *Biochemistry*, **38**: 5017-5023.
- Ramjee, M.K., Petithory, J.R., McElver, J., Weber, S.C. y Kirsch, J.F. (1996). A novel yeast expression/secretion system for the recombinant plant thiol endopeptidase propain. *Protein Eng.*, **9**: 1055-1061.

- Rost, B. y Sander, C. (1994 a). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins*, **19**: 55-72.
- Rost, B. y Sander, C. (1994 b). Conservation and prediction of solvent accessibility in protein families. *Proteins*, **20**: 216-226.
- Rost, B. (1995). TOPITS: Threading one-dimensional predictions into three-dimensional structures. In *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology*. (Rawlings, C.J., Clark, D., Altman, R., Hunter, L., Lengauer, T. y Wodak, S., eds.), pp. 314-321, AAAI Press, Menlo Park, CA, Cambridge, UK.
- Rost, B. (1998). Short yeast ORFs: expressed protein or not? Preprint. <http://dodo.cpmc.columbia.edu/~rost/Papers/98globe.html>.
- Rowan, A.D., Mason, P., Mach, L. y Mort, J.S. (1992). Rat procathepsin B. Proteolytic processing to the mature form *in vitro*. *J. Biol. Chem.*, **267**: 15993-15999.
- Russell, R.B., Copley, R.R. y Barton, G.J. (1996). Protein fold recognition by mapping predicted secondary structures. *J. Mol. Biol.*, **259**: 349-365.
- Salamov, A.A. y Solovyev, V.V. (1995). Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiple sequence alignments. *J. Mol. Biol.*, **247**: 11-15.
- Sánchez, R. y Sali, A. (1997). Advances in comparative protein-structure modeling. *Curr. Opin. Struct. Biol.*, **7**: 206-214.
- Sander, C. y Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*, **9**: 56-68.
- Sarath, G., De la Motte, R.S. y Wagner, F.W. (1993). Protease assay methods. In *Proteolytic Enzymes, a practical approach*. (Beynon, R.J. y Bond, J.S., eds.), pp 25-55, IRL Press, Oxford.
- Semisotnov, G.V., Rodionova, N.A., Razgulyaev, O.L., Uversky, V.N., Gripas, A.F. y Gilmanshin, R.I. (1991). Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers*, **31**: 119-128.
- Sheinerman, F.B. y Brooks, C.L. III. (1998). Molecular picture of folding of a small α/β protein. *Proc. Natl. Acad. Sci. USA.*, **95**: 1562-1567.

- Sippl, M.J. y Floeckner, H. (1996). Threading thrills and threats. *Structure*, **4**: 15-19.
- Solís-Mendiola, S., Arroyo-Reyna, A. y Hernández-Arana, A. (1992). Circular dichroism of cysteine proteinases from papaya latex. Evidence of differences in the folding of their polypeptide chains. *Biochim. Biophys. Acta*, **1118**: 288-292.
- Solís-Mendiola, S., Rojo-Domínguez, A. y Hernández-Arana, A. (1993). Cooperativity in the unfolding transitions of cysteine proteinases. Calorimetric study of the heat denaturation of chymopapain and papain. *Biochim. Biophys. Acta*, **1203**: 121-125.
- Solís-Mendiola, S., Gutiérrez-González L.H., Arroyo-Reyna, A., Padilla-Zúñiga, J., Rojo-Domínguez, A., y Hernández-Arana, A. (1998). pH dependence of the activation parameters for chymopapain unfolding: influence of ion pairs on the kinetic stability of proteins. *Biochim. Biophys. Acta*, **1388**: 363-72.
- Solovyev, V.V. y Salamov, A.A. (1994). Predicting alpha-helix and beta-strand segments of globular proteins. *Comput. Appl. Biosci.*, **10**: 661-669.
- Stubbs, M.T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. y Turk, V. (1990). The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. *EMBO J.*, **9**: 1939-1947.
- Stultz, C.M., White, J.V. y Smith, T.F. (1993). Structural analysis based on state-space modeling. *Protein Sci.*, **2**: 305-314.
- Tao, K., Stearns, N.A., Dong, J., Wu, Q.L. y Sahagian, G.G. (1994). The proregion of cathepsin L is required for proper folding stability, and ER exit. *Arch. Biochem. Biophys.*, **311**: 19-27.
- Taylor, M.A.J., Baker, K.C., Briggs, G.S., Connerton, I.F., Cummings, N.J., Pratt, K.A., Revell, D.F., Freedman, R.B. y Goodenough, P.W. (1995). Recombinant pro-regions from papain and papaya proteinase IV are selective high affinity inhibitors of the mature papaya enzymes. *Protein. Eng.*, **8**: 59-62.
- Thompson, J.D., Higgins, D.G. y Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**: 4673-4680.
- Turk, V. y Bode, W. (1991). The cystatins: protein inhibitors of cysteine proteinases. *FEBS Lett.*, **285**: 213-219.

- Verlet, L. (1967). Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, **159**: 98-103.
- Vernet, T., Tessier, D.C., Richardson, C., Laliberte, F., Khouri, H.E., Bell, A.W., Storer, A.C. y Thomas, D.Y. (1990). Secretion of functional papain precursor from insect cells. Requirement for N-glycosylation of the pro-region. *J. Biol. Chem.*, **265**: 16661-16666.
- Vernet, T., Khouri, H.E., Laflamme, P., Tessier, D.C., Musil, R., Gour-Salin, B.J., Storer, A.C. y Thomas, D.Y. (1991). Processing of the papain precursor. Purification of the zymogen and characterization of its mechanism of processing. *J. Biol. Chem.*, **266**: 21451-21457.
- Vernet, T., Berti, P.J., de Montigny, C., Musil, R., Tessier, D.C., Ménard, R., Magny, M.C., Storer, A.C. y Thomas, D.Y. (1995). Processing of the papain precursor. The ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. *J. Biol. Chem.*, **270**: 10838-10846.
- von Rague-Schleyer, P., Allinger, N.L., Clark, T.C., Gasteiger, J., Kollman, P.A. y Schaefer, H.F. (1998). Protein structure prediction in 1D, 2D and 3D. In *Encyclopedia of Computational Chemistry*, John Wiley, Sussex, 2242-2255. <http://www.embl-heidelberg.de/~rost/Papers/1998-encyclopedia/abstract.html>
- White, J.V., Stultz, C.M. y Smith, T.F. (1994). Protein classification by stochastic modeling and optimal filtering of amino-acid sequences. *Math. Biosci.*, **119**: 35-75.
- Winther, J.R. y Sorensen, P. (1991). Propeptide of carboxypeptidase Y provides a chaperone-like function as well as inhibition of the enzymatic activity. *Proc. Natl. Acad. Sci. USA*, **88**: 9330-9334.
- Winther, J.R., Sorensen, P. y Kielland-Brandt, M.C. (1994). Refolding of a carboxypeptidase Y folding intermediate *in vitro* by low-affinity binding of the proregion. *J. Biol. Chem.*, **269**: 22007-22013.
- Yamamoto, Y., Watabe, S., Kageyama, T. y Takahashi, S.Y. (1999). A novel inhibitor protein for *Bombyx* cysteine proteinase is homologous to propeptide regions of cysteine proteinases. *FEBS Lett.*, **448**: 257-260.
- Zvelebil, M.J., Barton, G.J., Taylor, W.R. y Sternberg, M.J. (1987). Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J. Mol. Biol.*, **195**: 957-961.